

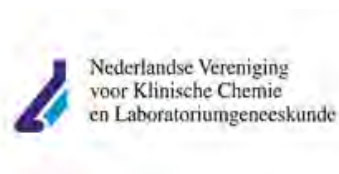
LEIDRAAD ANALYTISCHE KWALITEITSCONTROLE MET BEHULP VAN SIX SIGMA

INITIATIEF

Nederlandse Vereniging voor Klinische Chemie en Laboratoriumgeneeskunde – Vereniging Artsen Laboratoriumdiagnostiek (NVKC-VAL)

FINANCIERING

Ontwikkeling van de leidraad werd gefinancierd uit de Kwaliteitsgelden Medisch Specialisten (SKMS), projectnummer 14957950



COLOFON

LEIDRAAD ANALYTISCHE KWALITEITSCONTROLE MET BEHULP VAN SIX SIGMA

© 2015 Nederlandse Vereniging voor Klinische Chemie en Laboratoriumgeneeskunde –
Vereniging Artsen Laboratoriumdiagnostiek

website: www.nvkc.nl, www.labarts.nl

INHOUDSOPGAVE

Hoofdstuk 1	Samenstelling van de werkgroep	3
Hoofdstuk 2	Overzicht van de aanbevelingen	4
Hoofdstuk 3	Aanleiding	5
Hoofdstuk 4	Doelstelling	6
Hoofdstuk 5	Toepassingsgebied en beperkingen (scope)	8
Hoofdstuk 6	Algemene inleiding	10
6.1	Analytische kwaliteitscontrole	10
6.2	Het Six Sigma concept	11
6.3	Samenhang met andere documenten c.q. richtlijnen	13
Hoofdstuk 7	Berekening van de Sigma score	15
Hoofdstuk 8	Uitvoering van kwaliteitscontrole	24
8.1	Toepassen van kwaliteitscontroleregels op basis van berekende Sigma score	24
8.2	Vaststellen van maatregelen bij een bepaling met een Sigma score < 3	26
8.3	Vaststellen hoe de Six Sigma benadering kan worden toegepast bij meerdere apparaten	27
Hoofdstuk 9	Stappenplan voor het inrichten van de interne kwaliteitscontrole met behulp van Six Sigma	30
Hoofdstuk 10	Implementatie van de leidraad	32
Literatuurreferenties		33
Bijlagen		
I	Vaststellen van de toevallige fout of <i>random error</i>	38
II	Vaststellen van de systematische fout of <i>bias</i>	40
III	Vaststellen van de biologische variatie	47
IV	Vaststellen van het <i>total error budget</i>	50
Begrippenlijst en afkortingen		54
Noten		58

HOOFDSTUK 1 SAMENSTELLING VAN DE WERKGROEP

Dr. ir. N. de Jonge (voorzitter), *Klinisch Chemicus*, Ziekenhuis Bronovo, Den Haag

Dr. W.P.H.G. Verboeket-van de Venne (secretaris), *Senior Onderzoeker Klinische Chemie*,
Atrium-Orbis Medisch Centrum, Heerlen

Dr. N. Jonker, *Klinisch Chemicus*, Certe – Wilhelmina Ziekenhuis, Assen

Dr. J.E. Kootstra-Ros, *Klinisch Chemicus*, Universitair Medisch Centrum Groningen, Groningen

Dr. D. van Loon, *Klinisch Chemicus*, Sint Antonius Ziekenhuis, Nieuwegein

Dr. F.J. Loupatty, *Klinisch Chemicus*, Reinier de Graaf Groep Diagnostisch Centrum SSDZ,
Delft

Dr. drs. W.P. Oosterhuis, *Arts Klinische Chemie*, Atrium-Orbis Medisch Centrum, Heerlen

Dr. H.J. Vermeer, *Klinisch Chemicus*, Albert Schweitzer Ziekenhuis, Dordrecht

Dr. ir. R.W. Wulkan, *Klinisch Chemicus*, Maasstad Ziekenhuis, Rotterdam

HOOFDSTUK 2 OVERZICHT VAN DE AANBEVELINGEN

Toepassingsgebied en beperkingen (scope)

- Aanbeveling 1 ⇒ De analytische kwaliteitseisen worden vastgesteld in hiërarchisch aflopende volgorde, conform de Stockholm criteria [Kallner *et al.*, 1999; Kenny *et al.*, 1999]*:
1. het effect op de klinische uitkomst binnen een specifieke klinische vraagstelling
 2. het effect op de algemene klinische besluitvorming ingegeven door
 - a. de biologische variatie
 - b. klinische ervaring
 3. professionele aanbevelingen (gepubliceerd)
 - a. vanuit (inter)nationale beroepsverenigingen
 - b. expertise van regionale samenwerkingen
 4. een prestatiedoel ingegeven door een regelgevende instantie
 5. een prestatiedoel ingegeven door *State of the Art* technologie (in deze leidraad: zoals gedefinieerd door de SKML (Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek)).

* Indien beschikbaar en toepasbaar genieten eisen hoger in de hiërarchie de voorkeur boven eisen lager in deze rangschikking.

Berekening van de Sigma score

- Aanbeveling 2 ⇒ Voor de berekening van de Sigma score wordt gebruik gemaakt van de volgende formule : $(TE_a - bias)/CV_a$.
- Aanbeveling 3 ⇒ Sigma scores worden voor kwantitatieve methoden bepaald en eventueel bijgesteld als daar aanleiding toe is (bijvoorbeeld bij ingebruikname van nieuwe apparatuur of bij technische problemen).
- Aanbeveling 4 ⇒ Bij het ontwerpen van de interne kwaliteitscontrole wordt – met een zo hoog mogelijk kwaliteitscriterium – gestreefd naar een Sigma score > 3.
- Aanbeveling 5 ⇒ De CV van een bepaling wordt bepaald op het controleniveau dat het dichtst bij de klinische beslisgrens ligt. De laboratoriumspecialist stelt vast op welk niveau dit is.

Uitvoering van kwaliteitscontrole

- Aanbeveling 6 ⇒ Per Sigma score wordt een bewakingsregel of combinatie van bewakingsregels toegepast.
- Aanbeveling 7 ⇒ De CV van het virtuele apparaat wordt bepaald door de CV van alle gemeten controles van een lotnummer op alle apparaten te bepalen ten opzichte van de gemiddeld bepaalde waarde voor deze controles over alle apparaten.

HOOFDSTUK 3 AANLEIDING

Om in de laboratoriumdiagnostiek tot een betrouwbaar resultaat te komen is kwaliteitsbewaking noodzakelijk. Analytische kwaliteitscontrole is hierbij een belangrijk instrument, naast borging van de randvoorwaarden om tot betrouwbare resultaten te komen zoals deskundigheid van medewerkers en keuze en onderhoud van apparatuur.

Analytische kwaliteitscontrole wordt onderverdeeld in externe kwaliteitscontrole (*External Quality Assessment*, EQA) en interne kwaliteitscontrole (*Internal Quality Control*, IQC). EQA is gericht op vergelijking met en afstemming op definitieve methoden, referentiemethoden of consensuswaarden. Interne kwaliteitscontrole richt zich op juistheid en precisie, door middel van continue (dagelijkse) controle van het (totale) analyseproces. Hiertoe worden onder meer kwaliteitscontrolemonsters als surrogaat-patiëntenmonsters gemeten waarbij de resultaten worden getoetst aan de hand van criteria. In klinisch chemische laboratoria wordt hiervoor sinds de jaren '70-'80 van de vorige eeuw veelvuldig gebruik gemaakt van criteria en beoordelingsmethodieken die zijn beschreven door Levey en Jennings [1950], respectievelijk Westgard *et al.* [1981]. Deze wijze van kwaliteitscontrole is gebaseerd op een statistische benadering, zoals oorspronkelijk door Shewhart beschreven [1931]. Hierbij werd echter geen rekening gehouden met de verschillen in biologische en analytische variatie voor de verschillende bepalingen. Met behulp van de rond 1985 ontwikkelde Six Sigma methodiek kan de toepassing van kwaliteitscontroleregels verbeterd worden, om zowel tot optimaal resultaat (hoge foutdetectiekans en lage onterechte afkeuring) als tot optimaal gebruik van middelen te komen. Deze leidraad geeft aanbevelingen voor het ontwerpen van de interne kwaliteitsbewaking **indien gekozen wordt voor een methode gebaseerd op Six Sigma** en beperkt zich tot kwantitatieve bepalingen.

HOOFDSTUK 4 DOELSTELLING

Six Sigma is een kwaliteitsmanagementmethode die erop gericht is de kwaliteit van de resultaten van bedrijfskundige processen te verbeteren. Door oorzaken van defecten of fouten op te sporen en te verwijderen wordt de variatie in de processen gereduceerd. Deze methode kan ook toegepast worden om kwaliteitscontrolesystemen voor laboratoriumbepalingen te ontwerpen. Hierbij wordt naast de analytische nauwkeurigheid, ook rekening gehouden met de biologische variatie (binnen-patiënt en tussen-patiënt). Hoewel deze methode theoretisch voldoende is ontwikkeld, wordt deze in de praktijk echter in beperkte mate toegepast. Voor veel bepalingen worden nu te strenge interne kwaliteitscriteria gehanteerd, omdat geen rekening wordt gehouden met de biologische variatie van de te analyseren stof. Voor andere bepalingen echter is de kwaliteit – volgens deze criteria – juist onvoldoende. Striktere controle of aanpassing van de methodiek zijn in dit geval mogelijke acties. Toepassing van de Six Sigma benadering leidt tot een meer op maat gesneden analytische kwaliteitsbewaking. Mensen en middelen kunnen zo beter worden ingezet, waarmee ook financiële besparingen zijn te realiseren. Six Sigma is een pragmatische benadering om onderscheid te maken tussen analytische procedures die met minimale kwaliteitscontrole bewaakt kunnen worden en analytische procedures die intensieve kwaliteitsbewaking c.q. verbetering vergen.

Deze benadering geeft ook een praktische handreiking om keuzes te maken tussen bepalingen met verschillende analytische kwaliteit. Ontwikkelaars van bepalingen (i.c. diagnostische industrie) verkrijgen met deze benadering duidelijk inzicht of bepalingen geschikt zijn voor het beoogde doel.

In de voorliggende leidraad wordt – na een inleiding over interne kwaliteitscontrole (Hoofdstuk 6) en Six Sigma (Hoofdstuk 7) – een uitwerking gegeven van het theoretisch concept met betrekking tot de toepassing van Six Sigma in de interne kwaliteitsbewaking (Hoofdstuk 8). In hoofdstuk 9 staat een praktische samenvatting met een stappenplan, waarmee Six Sigma op pragmatische wijze in het laboratorium kan worden geïmplementeerd. Voor meer verdieping wordt u verwezen naar de bijlagen en voetnoten.

Doelstelling is het opstellen van een leidraad die – in praktische zin – een handleiding vormt voor (medische) laboratoria die ervoor kiezen om de Six Sigma benadering toe te passen bij het ontwerpen van de (interne) analytische kwaliteitscontrole.

HOOFDSTUK 5 TOEPASSINGSGEBIED EN BEPERKINGEN (SCOPE)

De analytische kwaliteit van een procedure waarmee een kwantitatief resultaat wordt gerapporteerd laat zich relatief eenvoudig beschrijven in termen van o.a. reproduceerbaarheid, repeteerbaarheid, totale fout, systematische fout en toevallige fout. In 1999 is onder auspiciën van de IUPAC Clinical Chemistry Section, de IFCC en de WHO een bijeenkomst gehouden met het primaire doel om een hiërarchisch raamwerk te ontwikkelen voor de benodigde analytische kwaliteitseisen binnen de laboratoriumgeneeskunde. Het resultaat van deze consensus, de zogenaamde Stockholm Criteria [Kallner *et al.*, 1999; Kenny *et al.*, 1999], staat hieronder weergegeven.

Aanbeveling 1:

De analytische kwaliteitseisen worden vastgesteld in hiërarchisch aflopende volgorde, conform de Stockholm criteria [Kallner *et al.*, 1999; Kenny *et al.*, 1999]*:

1. het effect op de klinische uitkomst binnen een specifieke klinische vraagstelling
2. het effect op de algemene klinische besluitvorming ingegeven door
 - a. de biologische variatie
 - b. klinische ervaring
3. professionele aanbevelingen (gepubliceerd)
 - a. vanuit (inter)nationale beroepsverenigingen
 - b. expertise van regionale samenwerkingen
4. een prestatiedoel ingegeven door een regelgevende instantie
5. een prestatiedoel ingegeven door *State of the Art* technologie (in deze leidraad: zoals gedefinieerd door de SKML (Stichting Kwaliteitsbewaking Medische Laboratoriumdiagnostiek)).

* Indien beschikbaar en toepasbaar genieten eisen hoger in de hiërarchie de voorkeur boven eisen lager in deze rangschikking.

In de praktijk is de Six Sigma benadering alleen geschikt voor kwantitatieve bepalingen, waarbij de *allowable Total Error* (TE_a) en de imprecisie (CV_a) van deze bepalingen bekend

zijn. De Six Sigma benadering vormt samen met bovenstaande hiërarchie een krachtige combinatie om de analytische kwaliteit te borgen. Deze combinatie vindt toepassing op verschillende gebieden waarbij kwaliteitseisen een rol spelen, zoals bij het selecteren van analytische bepalingen, het ontwerpen van nieuwe bepalingen, het valideren of verifiëren van methoden, en het inrichten van de externe en interne kwaliteitsbewaking. Deze leidraad heeft uitsluitend betrekking op de interne kwaliteitsbewaking.

Deze leidraad geeft aanbevelingen voor het ontwerpen van de interne kwaliteitsbewaking met behulp van de Six Sigma methode en beperkt zich tot kwantitatieve bepalingen. Deze leidraad is gebaseerd op modellen en aannames, waarover niet altijd consensus bestaat. Naar verwachting zullen nieuwe inzichten en concepten zich in de toekomst aandienen.

HOOFDSTUK 6 ALGEMENE INLEIDING

6.1 Analytische kwaliteitscontrole

Toevallige fouten en systematische fouten dienen vroegtijdig te worden opgemerkt zodat deze tijdig gecorrigeerd kunnen worden. De werkwijze om deze fouten te herkennen bestaat uit het gebruik van statistische kwaliteitscontrole (*Statistical Quality Control, SQC*) op twee specifieke manieren:

Interne kwaliteitscontrole (IQC)

De IQC omvat alle SQC methoden welke dagelijks/binnen een run worden toegepast, gebruikmakend van de analytische systemen en bijbehorende reagentia, standaarden en controlematerialen. De IQC is gericht op het beheersen van juistheid en precisie op de korte termijn, door middel van continue (dagelijkse) controle van het (totale) analyseproces met stabiel/gestabiliseerd controlemateriaal.

Interne kwaliteitscontrole binnen de medische laboratoria wordt uitgevoerd door middel van metingen in controlematerialen met als doel te beoordelen of de onderzoeksprocedures beheerst zijn. Afhankelijk van de resultaten van deze controlemetingen worden de resultaten verkregen met de onderzoeksprocedure vrijgegeven voor de diagnostiek. Onder de onderzoeksprocedure wordt verstaan het analysesysteem (apparatuur), de gebruikte reagentia en standaarden.

Externe kwaliteitscontrole (EQA)

De EQA omvat alle SQC methoden welke periodiek worden uitgevoerd waarbij gebruik gemaakt wordt van de diensten/programma's van een externe organisatie (referentielaboratorium, organisator van externe rondzendingen, industrie). De EQA is gericht op vergelijking met en afstemming op definitieve methoden, referentiemethoden of consensuswaarden op de lange termijn; en toepasbaar voor harmonisatie van uitslagen tussen laboratoria [Perich *et al.*, 2014]. In Nederland verzorgt de organisatie SKML externe rondzendingen voor alle medische laboratoria.

Een belangrijke voorwaarde om de juistheid te kunnen vastleggen is calibratie van de analytische procedure. Van de gebruikte standaarden dient de toegekende waarde en de commuteerbaarheid bekend te zijn en indien mogelijk herleidbaar naar een primaire referentiemethode. Een andere voorwaarde is de commuteerbaarheid van het externe controlemateriaal. Voor verdere beschouwing van de externe kwaliteitsbewaking, zie Jansen *et al.*, 2013.

6.2 Het Six Sigma concept

Algemeen

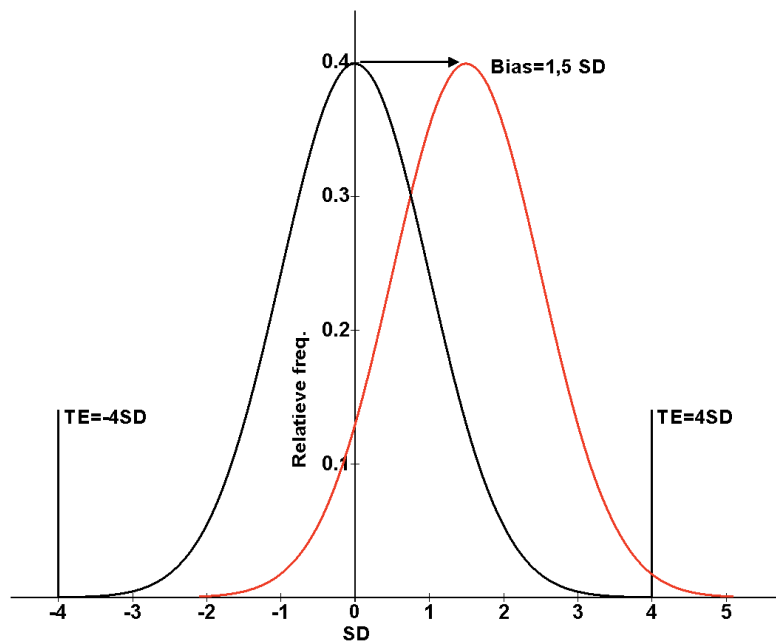
Six Sigma is een managementstrategie die gebaseerd is op een concept dat in 1986 bij Motorola door Bill Smith is ontwikkeld. Dit concept was gericht op minimalisering van de variabiliteit in productie door middel van een gestandaardiseerde methodiek voor het meten van defecten. In 2008 is dit concept geïntegreerd in de *lean* productie methodologie. Het doel van deze methodologie, inmiddels bekend als *Lean Six Sigma*, is optimale waardecreatie voor de klant en minimale verspilling [Motorola Solutions].

Nader beschouwd is Six Sigma toepasbaar als methode, managementstrategie of meetinstrument. Kern van de methode is de DMAIC-verbetercyclus (*Define, Measure, Analyze, Improve, Control*). Als managementstrategie is Six Sigma een gestructureerde methode om procesverbetering binnen organisaties te organiseren en vast te houden. Als meetinstrument is de Sigma score (*Sigma metric*) een maat voor de kwaliteit van een proces, product of dienst. Een proces of product met een geringe variatie ten opzichte van de gedefinieerde tolerantie heeft een betere kwaliteit dan een proces of product met een grotere variatie. De *Sigma metric* is gebaseerd op het aantal standaarddeviaties proces- of productvariatie dat binnen de tolerantiegrenzen past. Een twee Sigma product heeft 4,55 defecten per 100, een zes Sigma product heeft 0,002 defecten per miljoen gebeurtenissen (*Defects Per Million Opportunities, DPMO*). Een proces of product dient bovendien robuust genoeg te zijn om een bepaalde mate van verslechtering (bijvoorbeeld door invloeden van buiten af) op te kunnen vangen. In Tabel 1 staat voor een aantal Sigma scores het corresponderende aantal DPMO.

Tabel 1: Sigma score en corresponderende foutscore

Sigma score	Fouten per miljoen gebeurtenissen (DPMO) (2-zijdig), proces zonder <i>bias</i>	DPMO bij systematische fout (<i>bias</i>) van 1,5 SD (2-zijdig)
1	317.310	697.700
2	45.500	308.770
3	2.700	66.810
4	63	6.210
5	0,57	233
6	0,002	3,4

Figuur 1: Overschrijding van de $\pm 4 * SD$ limiet (4 Sigma proces) zonder *bias*: DPMO = 63; overschrijding met positieve *bias* van 1,5 SD: DPMO= 6.210.



Toepassing van Six Sigma in het klinisch-chemisch laboratorium

Six Sigma heeft ook ingang gevonden binnen de klinisch-chemische laboratoria [Nevalainen *et al.*, 2000; Kazmierczak, 2003; Gras en Philippe, 2007; Stankovic en Romeo, 2007; Schoenmakers *et al.*, 2011; Carlson *et al.*, 2012]. Schoenmakers *et al.* [2011] beschreven een praktische werkwijze voor de selectie van Westgardregels, gebaseerd op de analytische

kwaliteit van een bepaling uitgedrukt in de Sigma score. Deze benadering wordt in deze leidraad ook gevolgd.

6.3 Samenhang met andere documenten c.q. richtlijnen

In een aantal landen is de analytische kwaliteitsbewaking aan regelgeving onderhevig. In de Verenigde Staten is hiervoor CLIA (Clinical Laboratory Improvement Amendments) van toepassing [CLIA], in Duitsland RiliBÄK (Richtlinie der Bundesärztekammer)[RiliBÄK]. Deze regelgeving is vooral gericht op het voorkómen van ondermaatse bepalingen (resultaten) via externe kwaliteitscontrole. In Nederland is dergelijke regelgeving niet verplicht.

De norm NEN-EN-ISO 15189 (Medische laboratoria – Bijzondere eisen voor kwaliteit en competentie) waartegen medische laboratoria geaccrediteerd worden, stelt slechts in algemene termen eisen aan het systeem van kwaliteitsbewaking:

5.6.2 Kwaliteitscontrole

Het laboratorium moet kwaliteitscontroleprocedures ontwerpen die verifiëren dat de beoogde kwaliteit van resultaten wordt verkregen.

Six Sigma geeft op kwantitatieve wijze aan of een bepaling voldoet aan de gestelde kwaliteitseisen.

5.6.2.2 Materialen voor kwaliteitscontrole

Het laboratorium moet voor kwaliteitscontrole materialen gebruiken die op het onderzoekssysteem reageren op een manier die patiëntenmonsters zo dicht mogelijk benadert.

Materialen voor kwaliteitscontrole moeten periodiek worden onderzocht met een frequentie die is gebaseerd op de stabiliteit van de procedure en op het risico van schade voor de patiënt door een onjuist resultaat.

OPMERKING 1 Het laboratorium moet indien mogelijk concentraties van controlematerialen kiezen precies op of rond medische beslissingswaarden hetgeen de geldigheid van de genomen beslissingen bewerkstelligt.

OPMERKING 2 Gebruik van controlematerialen afkomstig van onafhankelijke derden behoort te worden overwogen, hetzij in plaats van of naast controlematerialen die door de fabrikant van het reagens of instrument worden geleverd.

5.6.2.3 Gegevens betreffende kwaliteitscontrole

Het laboratorium moet beschikken over een procedure om vrijgave van patiëntenresultaten in geval van een afwijking in de kwaliteitscontrole te voorkomen.

Als de regels voor kwaliteitscontrole zijn overtreden en er aanwijzingen bestaan dat onderzoeksresultaten waarschijnlijk medisch significante fouten bevatten, moeten de resultaten worden afgekeurd en relevante patiëntenmonsters opnieuw worden onderzocht nadat de foutieve omstandigheid is hersteld en is geverifieerd dat de prestatie binnen de specificatie plaatsvindt. Het laboratorium moet ook de resultaten van patiëntenmonsters evalueren die zijn onderzocht na de laatste succesvolle uitvoering van kwaliteitscontrole.

Gegevens betreffende kwaliteitscontrole moeten met regelmatige intervallen opnieuw worden beoordeeld om trends in de uitvoering van onderzoek te ontdekken die een indicatie kunnen zijn voor problemen in het onderzoekssysteem. Als dergelijke trends worden genoteerd, moeten preventieve maatregelen worden getroffen en geregistreerd.

Deze leidraad biedt een strategie om de gestelde eis (ISO) te vertalen in concrete criteria.

OPMERKING Om continu de prestatie van het onderzoekssysteem te monitoren behoren waar mogelijk statistische en niet-statistische technieken voor procesbeheersing te worden toegepast.

Six Sigma geeft hier een antwoord op.

Samenvattend kan gesteld worden dat de norm NEN-EN-ISO 15189 kwaliteitsbewaking vereist. De huidige leidraad geeft een praktische handreiking ten aanzien van de interne kwaliteitsbewaking.

HOOFDSTUK 7 BEREKENING VAN DE SIGMA SCORE

De Sigma score voor de onderzoeksprocedure kan worden afgeleid uit de maximaal toegestane systematische afwijking in de meetwaarde en de analytische variatie CV_a [Westgard, 2000; Nevalainen *et al.*, 2000; Gras en Philippe, 2007; Schoenmakers *et al.*, 2011]. In het algemeen geldt dat de totale afwijking van een enkelvoudig kwaliteitscontrole resultaat een combinatie is van de *bias* (systematische fout) en de toevallige fout van dat resultaat:

$$\text{Totale meetfout (TE, total error)} = \text{bias} + \text{toevallige fout} \quad (\text{vergelijking 1})$$

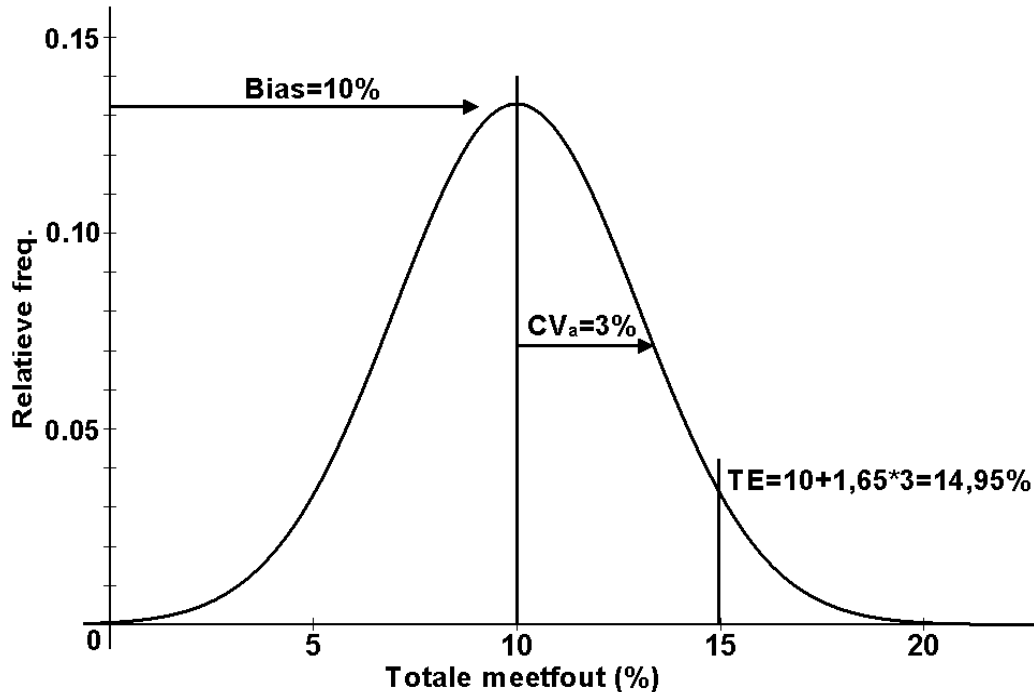
Beschouwt men een groot aantal resultaten, dan kan men statistisch de waarschijnlijkheid aangeven van een bepaalde afwijking van de kwaliteitscontrole resultaten:

$$TE = \text{bias} + z * CV_a \quad (\text{vergelijking 2})$$

TE geeft een grenswaarde aan waarbij een vastgesteld percentage resultaten deze grens overschrijdt. De z-waarde bepaalt dit percentage. Dit geldt strikt genomen alleen voor resultaten die normaal verdeeld zijn. Vanwege praktische overwegingen wordt uitgegaan van een normale verdeling van testresultaten.

Voorbeeld: met een CV_a van 3%, een bias van 10% en een z-waarde van 1,65 is de berekende TE 14,95%, 5% van de resultaten zal de TE-grens overschrijden (zie Figuur 2)(in dit geval eenzijdig; een tweezijdige limiet met corresponderende z-waarde is echter ook mogelijk).

Figuur 2: Berekening van de totale meetfout op basis van *bias* en analytische variatie.



In bovenstaand voorbeeld wordt aangegeven, hoe de totale meetfout wordt bepaald door de systematische fout, de analytische variatie en door de keuze van de z-waarde. Men kan het ook omkeren: definieert men nu een vaste grenswaarde voor TE (ofwel men definieert de limiet TE_a ofwel de *Total Error allowable*), dan stelt dit een grens aan de maximale waarden van de *bias* en de analytische variatie:

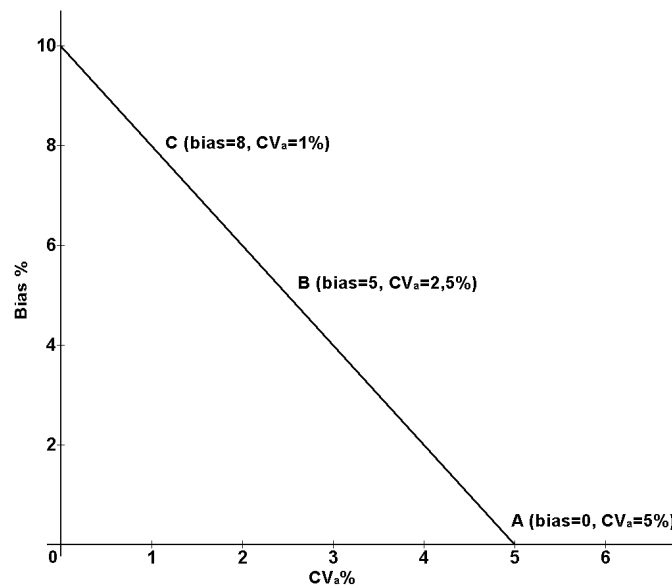
$$TE_a = bias + z * CV_a \quad (\text{vergelijking 2})$$

Deze vergelijking is om te schrijven tot:

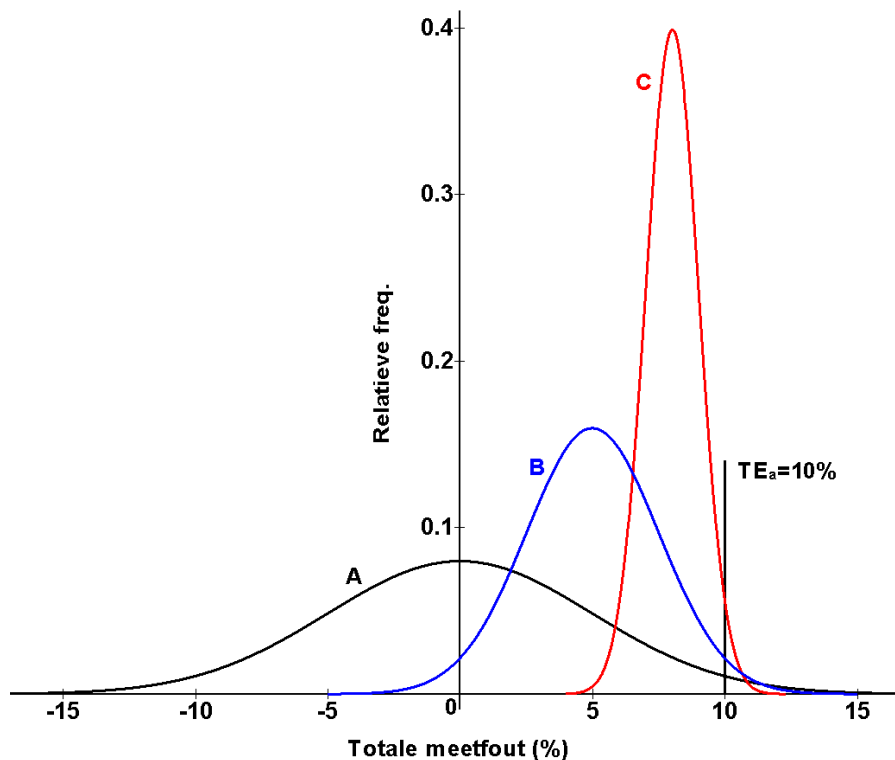
$$bias = TE_a - z * CV_a \quad (\text{vergelijking 3})$$

In dit model bestaat er een lineaire relatie tussen *bias* en CV_a , waarbij TE_a een constante waarde heeft. Bij een hypothetische waarde $CV_a = 0$ geldt $TE_a = bias$. De richtingscoëfficiënt van deze lijn is gelijk aan $-z$.

Figuur 3: Lineaire relatie tussen *bias* en analytische variatie. TE_a is vastgesteld op 10%, de z -waarde op 2 (dit zijn arbitrair gekozen waarden). Bij deze z -waarde valt 2,3% van de meetwaarden buiten de limiet van 10% maximale afwijking. Drie combinaties A, B en C voldoen aan de voorwaarden.



Figuur 4: Drie combinaties van analytische variatie en *bias* met een Sigma score van 2.



TE_a definieert de tolerantiegrens waar in dit model een vastgesteld percentage van de resultaten buiten valt. Dit percentage wordt bepaald door de waarde van z . Stel men heeft een beginsituatie met $CV_a = 5\%$, $bias = 0$ en $z = 2$, met $TE_a = 10\%$; 2,3% van de resultaten overschrijdt de grenswaarde TE_a (eenzijdig). De drie curves in Figuur 4 corresponderen met punten A, B en C in Figuur 3. In elk van deze combinaties van *bias* en CV_a is het percentage resultaten dat de tolerantiegrens TE_a overschrijdt identiek en gelijk aan 2,3%. Dit illustreert dat een hogere *bias* gecompenseerd wordt door een lagere analytische variatie; omgekeerd laat een lagere analytische variatie een hogere *bias* toe.

Lineair model versus niet lineaire modellen

Het hier beschreven model is lineair: bij normaal verdeelde uitkomsten bestaat er een bovenbeschreven lineair verband tussen *bias* en imprecisie. Dit model geldt alleen, als de spreiding van de normaalverdeling uitsluitend wordt bepaald door de analytische variatie, zoals het geval is bij de uitkomsten van de kwaliteitscontrole resultaten.

Er zijn echter ook andere modellen ontwikkeld, waarbij de verdeling bepaald wordt door zowel biologische als analytische variatie van resultaten. Referentiewaarden of kritische verschillen worden immers ook bepaald door beide vormen van variatie. Kwaliteitscriteria

zijn afgeleid van referentiewaarden en kritische verschillen, en daardoor gebaseerd op zowel biologische als analytische variatie. Volgens deze “gecombineerde” modellen bestaat er echter geen lineaire relatie tussen *bias* en analytische variatie, maar wordt deze relatie beschreven door een curve [Gowans *et al.*, 1988; Larsen *et al.*, 1991; Oosterhuis, 2011].

Het lineaire model is robuust en wordt het meest toegepast. Het is echter minder geschikt in andere situaties, bijvoorbeeld waar de kwaliteit van meerdere apparaten wordt bewaakt. Bovendien wordt de maximaal toegestane *bias* vaak overschat (zie verder Hoofdstuk 8.3 en Bijlage II).

Sigma score

In vergelijking 3 ($TE_a = bias + z * CV_a$) is de richtingscoëfficiënt z gelijk aan de Sigma score. Deze vergelijking kan ook geschreven worden als:

Aanbeveling 2:

Voor de berekening van de Sigma score wordt gebruik gemaakt van de volgende formule:

$$(TE_a - bias) / CV_a \quad \text{(vergelijking 4)}$$

Aanbeveling 3:

Sigma scores worden voor kwantitatieve methoden bepaald en eventueel bijgesteld als daar aanleiding toe is (bijvoorbeeld bij ingebruikname van nieuwe apparatuur of bij technische problemen).

Dit komt overeen met het aantal maal dat de analytische variatie "past" in de tolerantie TE_a , na correctie voor de *bias*. In Figuur 4 is dit steeds twee maal.

De Sigma score van een bepaling is een kwantitatieve maat voor de kwaliteit van een bepaling. Deze score vertegenwoordigt ook een kwaliteitsdoel voor de betreffende bepaling, en dient daarmee als basis om een programma van interne kwaliteitscontrole te ontwerpen. De interpretatie van deze score is als volgt:

< 3 onvoldoende kwaliteit, niet met conventionele (*multi-rules*) SQC te beheersen

- 3-4 geschikt voor beoogd doel, echter frequente controle is noodzakelijk
- 4-6 geschikt voor beoogd doel, d.w.z. met eenvoudige SQC regels te beheersen
- > 6 uitstekende kwaliteit

Aanbeveling 4:

Bij het ontwerpen van de interne kwaliteitscontrole wordt – met een zo hoog mogelijk kwaliteitscriterium – gestreefd naar een Sigma score > 3.

Waarom dient de Sigma score minimaal 3 zijn?

Als een test voldoet aan het kwaliteitscriterium $CV_a = 0,5 * CV_i$, zal de Sigma score toch vaak lager dan 3 zijn (zie berekening Noot 3). Er is additioneel een bepaalde marge noodzakelijk voor de borging van deze kwaliteit. Deze marge wordt bepaald door de sensitiviteit van de kwaliteitscontrole voor een bepaalde afwijking. Dit wordt mede veroorzaakt door het lage aantal controlemetingen (soms één waarde). Daarmee is vaak al lastig een afwijking van 1 of 1,5 SD vast te stellen. Dit komt tot uitdrukking in Tabel 2; bij een Sigma score van 4 wordt nog net voldaan aan de uitgangspunten van de kwaliteitsregels (P_{ed} : sensitiviteit voor detectie van afwijkingen minstens 90%, P_{fr} : fout-positieven lager dan 5%). De kans om echte fouten te ontdekken wordt bij een Sigma score van 3 of lager kleiner dan 0,36 bij een P_{fr} van 0,02, waarbij al 3 controlewaarnemingen zijn betrokken. De gebruikte combinatie van bewakingsregels van Westgard ($1_{3S}/2_{2S}/R_{4S}/4_{1S}$) is bij een Sigma score van 3 gericht op de herkenning van een systematische bias van 1 SD.

Berekenen van de Sigma score

Voor het berekenen van de Sigma score voor een analytisch proces wordt allereerst de toegestane fout (TE_a) in het resultaat vastgesteld. Het eerste Stockholm criterium, effect op klinische uitkomst, is beperkt toepasbaar. Veel vaker wordt de TE_a gebaseerd op gegevens omtrent biologische variatie (zie Bijlage IV) of op basis van EQA resultaten.

Bepalen van de analytische variatie

In Bijlage I wordt ingegaan op het vaststellen van de toevallige fout. Praktisch gezien wordt de CV van een bepaling bepaald op het controleniveau dat het dichtst bij de klinische beslisgrens ligt (keuze laboratoriumspecialist).

Aanbeveling 5:

De CV van een bepaling wordt bepaald op het controleniveau dat het dichtst bij de klinische beslisgrens ligt. De laboratoriumspecialist stelt vast op welk niveau dit is.

Wanneer er geen grote veranderingen zijn in het concentratieniveau van de controles en er niet op een ander type controle wordt overgestapt, kunnen de CV_a waarden gebaseerd worden op de uitslagen van een eerder lotnummer van de controles. Wanneer er wel een verandering in de controles plaatsvindt, kan minimaal 20 keer de controle bepaald worden en wordt de hieruit resulterende CV gebruikt. Bij ingebruikname van een nieuwe onderzoeksprocedure kunnen hiervoor bijvoorbeeld data uit het EP15 of EP5 protocol gebruikt worden.

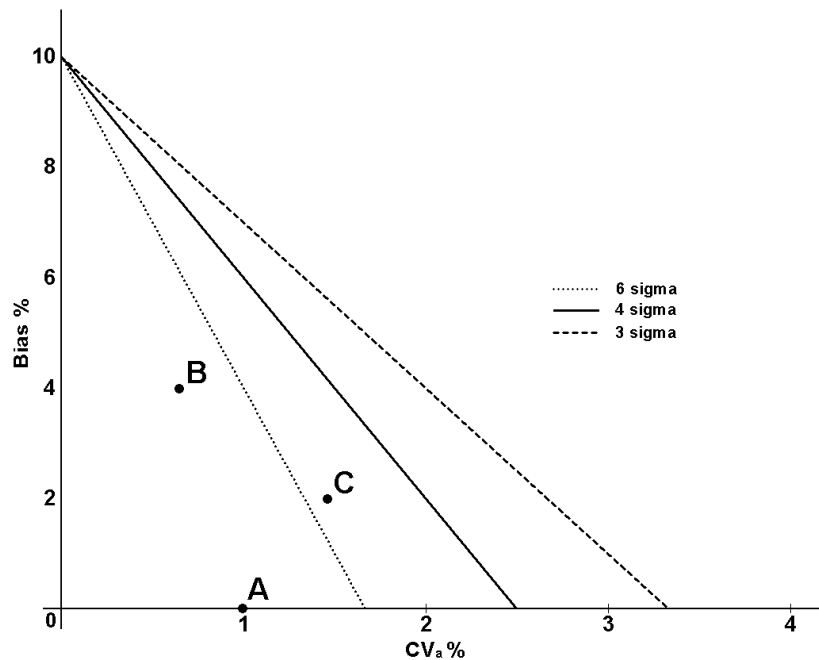
Vaststellen van de bias

In Bijlage II wordt ingegaan op het vaststellen van de *bias*. Er zijn verschillende situaties te onderscheiden, waarbij op verschillende manieren rekening wordt gehouden met de *bias*. We onderscheiden drie situaties:

- 1) met behulp van *Sigma Metrics* de interne kwaliteitscontrole instellen. Bij initieel ontwerp van de IQC speelt de *bias* geen rol als er sprake is van één apparaat en kan deze – voor de berekening van de Sigma score – op nul gesteld worden. Het instellen van de IQC is namelijk gericht op het handhaven van een (uitgangs)niveau. Voor het instellen van de IQC is de ligging van dat niveau niet van belang. Het uitgangsniveau wordt gecontroleerd door toepassing van EQC. Met IQC wordt het uitgangsniveau binnen bepaalde bandbreedte gehandhaafd. Dit niveau zou geen *bias* moeten hebben. Als er toch een *bias* is ten opzichte van een standaard, dan wordt deze met andere middelen vastgesteld (EQC) en zo nodig gecorrigeerd. Bewaking van dit gecorrigeerde niveau zal met dezelfde IQC, gebaseerd op dezelfde Sigma score kunnen gebeuren. Bij meerdere apparaten speelt de *bias* – als systematisch verschil tussen apparaten – wel een rol (zie Paragraaf 8.3).
- 2) toepassing van *Sigma Metrics* en externe kwaliteitscontrole; hierbij speelt de *bias* wel een rol¹, waarbij ook wordt ingegaan op de Sigma score volgens MUSE.

3) TE en Sigma score bij het valideren van een bepaling, waarbij de *bias* geen rol speelt². Ten overvloede: situatie 2 en 3 vallen buiten het bestek van deze leidraad.

Figuur 5: Combinaties van *bias* en analytische variatie voor drie verschillende Sigma scores.



In dit voorbeeld is TE_a gesteld op 10%. Merk op dat bij CV_a = 0 geldt *bias* = TE_a.

Punt A geeft een bepaling weer zonder *bias* en met een analytische variatie van 1%. Berekening van de Sigma score geeft een waarde van $(10 - 0)/1 = 10$.

Punt B geeft een bepaling weer met een *bias* van 4% en een analytische variatie van 0,6%. De Sigma score $(10 - 4)/0,6$ is hier eveneens 10.

Punt C geeft een bepaling weer met een *bias* van 2% en een analytische variatie van 1,5%. De Sigma score is $(10 - 2)/1,5 = 5,3$.

Berekenen van de Sigma score op basis van de biologische en analytische variatie

Zoals beschreven in Bijlage IV wordt TE_a meestal berekend op basis van de biologische variatie. Bij het berekenen van de Sigma score voor een analytisch proces wordt allereerst de tolerantie (TE_a)³ in het resultaat vastgesteld volgens de Stockholm criteria. Aan Stockholm criterium 1 (het effect op de klinische uitkomst) voldoen slechts weinig bepalingen. Er zijn enkele uitzonderingen, waaronder de kwaliteitsnorm voor troponine [Sandoval en Apple,

2014]. Voor HbA1c zijn kwaliteitscriteria theoretisch uitgewerkt, maar deze hebben geen brede ingang gevonden [Jørgensen *et al.*, 2002].

In klinische richtlijnen worden beslisgrenzen vermeld, die zich niet eenvoudig laten herleiden tot een kwaliteitsnorm. Vaker wordt de tolerantie afgeleid van de biologische variatie. Voor ongeveer 400 veel gebruikte bepalingen is de biologische variatie gepubliceerd. (Ricós; TE_a). Enige voorzichtigheid is hier geboden. Bij het vaststellen van de biologische variatie dienen allerlei bronnen van variatie te worden uitgesloten (denk aan medicatie, alcohol). Andere bronnen van variatie zijn moeilijk te beheersen of zijn minder eenduidig (binnen dag variatie, seizoensvariatie). Daarnaast kan de tolerantie worden afgeleid van de *State of the Art* (TE_{SA}) op basis van de EQA resultaten.

TE_a wordt op basis van de biologische variaties berekend met de volgende vergelijking (zie bijlage IV):

$$TE_a = 0,25 * (CV_i^2 + CV_g^2)^{0,5} + 1,65 * 0,5 * CV_i \quad (\text{vergelijking 5})$$

met CV_i = *within-subject* of intra-individuele variatie, CV_g = *between-subject* of inter-individuele variatie

Voorbeeld: Voor creatine fosfokinase (CK) geldt: $CV_i = 22,8\%$, $CV_g = 40\%$ (volgens Ricós tabel). TE_a wordt als volgt berekend:

$$\begin{aligned} TE_a &= 0,25 * (22,8^2 + 40^2)^{0,5} + 1,65 * 0,5 * 22,8 \\ &= 30,3\% \end{aligned}$$

Voor veel bepalingen zijn de waarden voor CV_i , CV_g en TE_a in de Ricós tabel opgenomen.

HOOFDSTUK 8 UITVOERING VAN KWALITEITSCONTROLE

8.1 Toepassen van kwaliteitscontroleregels op basis van berekende Sigma score

De Westgardregels⁴ vormen een combinatie van bewakingsregels die zodanig zijn opgesteld, dat voldaan wordt aan twee voorwaarden [Westgard en Burnett, 1990]:

- 1) er zijn minder dan 5% foutpositieve resultaten van interne kwaliteitscontrole; (*probability for false rejection* $P_{fr} < 0.05$);
- 2) de sensitiviteit (of *power*) voor het aantonen van een significante fout (zowel van *bias* als van toevallige fout/imprecisie) is bij voorkeur 90% (*probability for error detection* $P_{ed} > 0.90$).

Aanbeveling 6:

Per Sigma score wordt een bewakingsregel of combinatie van bewakingsregels toegepast.

Tabel 2 geeft een overzicht van mogelijke bewakingsregels per Sigma score. De waarde N geeft het aantal controles weer binnen een serie metingen. De huidige processen zijn continue processen. De onderliggende statistiek betreft niet het aantal patiëntenuitslagen in de berekening. In het algemeen wordt de waarde N per 24 uur beschouwd.

Tabel 2: Overzicht Sigma score en mogelijke controleregels [Schoenmakers *et al.*, 2011].

Sigma	Regel	P _{fr}	P _{ed}	N	Niveaus
3,0	1 _{3S} /2 _{2S} /R _{4S} /4 _{1S}	0.02	0.36	2	3
3,2	1 _{3S} /2 _{2S} /R _{4S} /4 _{1S}	0.03	0.48	2	3
3,4	1 _{3S} /2 _{2S} /R _{4S} /4 _{1S}	0.03	0.65	2	2
3,6	1 _{3S} /2 _{2S} /R _{4S} /4 _{1S}	0.03	0.79	2	2
3,8	1 _{3S} /2 _{2S} /R _{4S} /4 _{1S}	0.03	0.86	2	2
4,0	1 _{3S} /2 _{2S} /R _{4S} /4 _{1S}	0.03	0.91	2	2
4,2	1 _{2,5S}	0.04	0.92	1	2
4,4	1 _{2,5S}	0.04	0.96	1	2
4,6	1 _{3S}	0.01	0.92	1	2
4,8	1 _{2,5S}	0.03	0.93	1	2
5,0	1 _{2,5S}	0.03	0.96	1	2
5,2	1 _{3S}	<0.01	0.91	1	2
5,4	1 _{3S}	<0.01	0.94	1	2
5,6	1 _{3S}	<0.01	0.97	1	2
5,8	1 _{3,5S}	<0.01	>0.97	1	2
6,0	1 _{3,5S}	<0.01	>0.97	1	2

Sigma scores hoger dan 6 zijn ideale processen en kunnen bewaakt worden met een enkelvoudige bewakingsregel welke voldoet aan de regel 1_{4,35S}. De tolerantie limiet is hier 4,35 SD (6 – 1,65); in de praktijk is de 1_{4S} regel geschikt. Dit is gelijk aan de zogenaamde kritische *bias* van het systeem. De kritische *bias* is de *bias* waarbij 5% van de resultaten de tolerantielimiet overschrijdt.

Bij een Sigma score tussen 3 en 6 kan de IQC ontworpen worden zoals beschreven in Tabel 2, waarbij de bron van de TE vastgelegd wordt (op basis van biologische variatie [Ricós *et al.*, 1999] of TE_{SA}) en P_{ed}, P_{fr}, de frequentie en het aantal controleniveaus zijn aangegeven. Voor processen met een Sigma score < 3 kan geen adequate foutdetectie geborgd worden (zie Paragraaf 8.2).

Er is ook commerciële software beschikbaar die het mogelijk maakt om – na invoer van de mate van de tolerantie in het resultaat (TE_a of TE_{SA}) en de CV_a – de bewakingsregel of set van bewakingsregels te kiezen die voldoet aan de genoemde voorwaarden.

8.2 Vaststellen van maatregelen bij een bepaling met een Sigma score < 3

Het ontwerpen van de IQC is gericht op het efficiënt organiseren van de bewaking van een analytisch proces met als uitgangspunten een $P_{fr} < 0.05$ en een $P_{ed} > 0.90$. Een Sigma score < 3 betekent dat de prestatie van het analytisch systeem voor het doel waarvoor het wordt ingezet als onvoldoende wordt beoordeeld.

Bij gebruik van de biologische variatie als bron voor de TE_a kan de TE_a verruimd worden door te kiezen voor *minimum performance* (zie bijlage IV). Mocht dit niet toereikend zijn om de Sigma score > 3 te krijgen, dan kan als bron voor de TE overgegaan worden op de TE_{SA} , gebaseerd op normen zoals SKML, RiliBÄK of CLIA. De $TE_{SA} = k * CV_{SA}$, met k gelijk aan 3 (zie ook de paragraaf over selectie *bias* in bijlage III).

Als de CV_{SA} is vastgesteld op basis van resultaten van externe kwaliteitscontrole, dan zal doorgaans 90% van de deelnemers binnen deze norm vallen. Hieruit volgt dat voor deze groep de Sigma score groter is dan 3. Voor deze bepalingen is het belangrijk om te streven naar het verbeteren van de analytische procedure.

Een andere (theoretische) mogelijkheid is het verbeteren van IQC procedures zodat de sensitiviteit voor afwijkingen (P_{ed}) toeneemt. Zoals eerder vermeld hebben IQC procedures met kwaliteitscontrolemonsters inherent een lage sensitiviteit, door het lage aantal waarnemingen. Een alternatieve methode is die van het patiëntgemiddelde of -mediaan, waarbij al dan niet na filtering uit deze gegevens een kwaliteitsparameter wordt berekend. Door het hoge aantal waarnemingen kan de sensitiviteit voor afwijkingen hoog zijn. Daardoor zou een Sigma score lager dan drie ook acceptabel kunnen worden. Binnen het kader van deze leidraad zal op deze methode echter niet verder worden ingegaan.

8.3 Vaststellen hoe de Six Sigma benadering kan worden toegepast bij meerdere apparaten

De tot dusverre gepresenteerde Six Sigma methodologie is gebaseerd op één apparaat en de gemeten CV_a op dat apparaat. In de praktijk worden in de meeste laboratoria echter de routine parameters op meerdere apparaten geïnstalleerd, om continue productie te kunnen garanderen. Wanneer een bepaling op meerdere apparaten gemeten wordt, wordt een extra onzekerheid geïntroduceerd: apparaten kunnen onderling verschillen in juistheid en precisie.

Het is belangrijk wat de relatie van de apparaten tot elkaar is. In het geval dat het vervolgen van een patiënt niet per se op één apparaat gebeurt – zoals in een groot laboratorium waar monsters willekeurig verdeeld worden over meerdere apparaten met dezelfde methodiek – kunnen de verschillende apparaten beschouwd worden als één geheel, dat in principe aan dezelfde eisen voldoet als ware het één enkel apparaat. Dit concept noemen we het *virtuele apparaat*. Bij het virtuele apparaat wordt de gemiddelde waarde van het controlemateriaal als doelwaarde gehanteerd: de *bias* van het virtuele apparaat is 0. De CV van het virtuele apparaat ($CV_{a,va}$) wordt berekend over de controlemonsters van alle apparaten. Een eis aan het gebruik van een virtueel apparaat is wel dat de verschillende apparaten dezelfde prestaties dienen te leveren. De CV dient vergelijkbaar te zijn, en een onderlinge *bias* dient beperkt te zijn of door middel van calibratie verminderd te worden. Voor de *bias* zijn zowel arbitraire grenzen gesuggereerd (bijvoorbeeld 10%), als grenzen gebaseerd op de biologische variatie ($1/3 * CV_i$)[Calleja, 2008]. Voor de CV ontbreken deze eisen. Het is aan de laboratoriumspecialist om de juiste inschatting te maken of meerdere apparaten voldoende vergelijkbare karakteristieken hebben om als gelijk beschouwd te kunnen worden, en in het concept van het virtuele apparaat passen.

Aanbeveling 7:

De CV van het virtuele apparaat wordt bepaald door de CV van alle gemeten controles van een lotnummer op alle apparaten te bepalen ten opzichte van de gemiddeld bepaalde waarde voor deze controles over alle apparaten.

Door voor het vaststellen van de $CV_{a,va}$ gebruik te maken van controlemonsters, die op alle apparaten even frequent gemeten worden, hoeft er geen rekening gehouden te worden met verschillen in het absolute aantal patiëntenmonsters dat op de verschillende apparaten gemeten wordt.

Met deze $CV_{a,va}$ kan een Sigma score berekend worden voor het virtuele apparaat. Logischerwijs is de $CV_{a,va}$ groter of gelijk aan de CV's van de afzonderlijke apparaten. Dit heeft tot gevolg dat het virtuele apparaat dan ook een lagere Sigma score heeft dan de Sigma score van de afzonderlijke apparaten en daarmee strikter bewaakt wordt. Groot voordeel van deze werkwijze is dat er per bepaling voor alle apparaten dezelfde doelwaarde en $CV_{a,va}$ geldt en dat een bepaling steeds met dezelfde regels gecontroleerd wordt.

Als alternatief voor het concept van het virtuele apparaat kan ervoor gekozen worden om de Sigma score van ieder apparaat individueel te berekenen. Controleregels worden dan per apparaat ingesteld, gebaseerd op de CV_a van het betreffende apparaat. Overigens laat niet elk LIS/QC beheersysteem toe, dat apparaten apart worden bewaakt.

Wanneer een bepaling op meerdere apparaten staat, dient voorkomen te worden dat een verschil tussen twee testuitslagen van twee verschillende apparaten geïnterpreteerd wordt als een significante verandering, terwijl dit veroorzaakt wordt door *bias* tussen de twee apparaten: de *bias* tussen apparaten is aan een maximum gebonden.

Het eerder beschreven lineaire model is minder geschikt om kwaliteitseisen mee af te leiden omdat het een overschatting van mogelijke *bias* geeft. Daarom wordt in onderstaande tekst voor het vaststellen van de maximaal toegestane *bias* van het criterium voor kritisch verschil uitgegaan. Dit is gebaseerd op het gecombineerd model [Larsen *et al.*, 1991]. Onderbouwing is terug te vinden in Bijlage IV.

a) $CV_a \leq 0,2 * CV_i$

Stöckl *et al.* [1995] hebben uit de berekening voor het kritisch verschil afgeleid dat de maximaal toegestane *bias* bij een verwaarloosbare CV_a ($CV_a \ll CV_i$) gelijk is aan $0,33 * CV_i$.

Voor praktische toepassing dient duidelijk te zijn wanneer deze regel toegepast kan worden, dus wanneer $CV_a \ll CV_i$ geldt. Bij $CV_a = 0,2 * CV_i$ geldt dat de maximale toelaatbare *bias* 17% afwijkt van $0,33 * CV_i$. Zie voor de afleiding Tabel 3: als $CV_a / CV_i = 0$ dan is de maximale *bias* 3,27 (afgerond 3,3). Als $CV_a / CV_i = 0,2$, dan is de maximale *bias* 2,72. Het verschil tussen 3,27 en 2,72 is 17%. Naar mening van de werkgroep is dit nog een acceptabele afwijking. Daarmee geldt: als $CV_a \leq 0,2 * CV_i$ is de maximaal toegestane *bias* $0,33 * CV_i$.

b) $CV_a / CV_i > 0,2$

Wordt CV_a te groot ten opzichte van CV_i , dan geldt de bovenstaande benadering niet meer en wordt de maximale *bias* berekend op basis van het gecombineerde model [Larsen *et al.*, 1991]. Voor verschillende verhoudingen van CV_a en CV_i wordt de berekende maximale *bias* weergegeven in Tabel 3. In het model wordt uitgegaan van een maximale CV_a van $0,5 * CV_i$. Voor bepalingen van mindere kwaliteit kan de norm verlegd worden naar $0,75 * CV_i$.

Tabel 3: Berekening van de limiet van de *bias* op basis van de verhouding analytische variatie en biologische variatie (volgens Larsen *et al.*, 1991). Hierbij is onderscheid gemaakt tussen *desirable* en *minimum performance* (zie Bijlage IV).

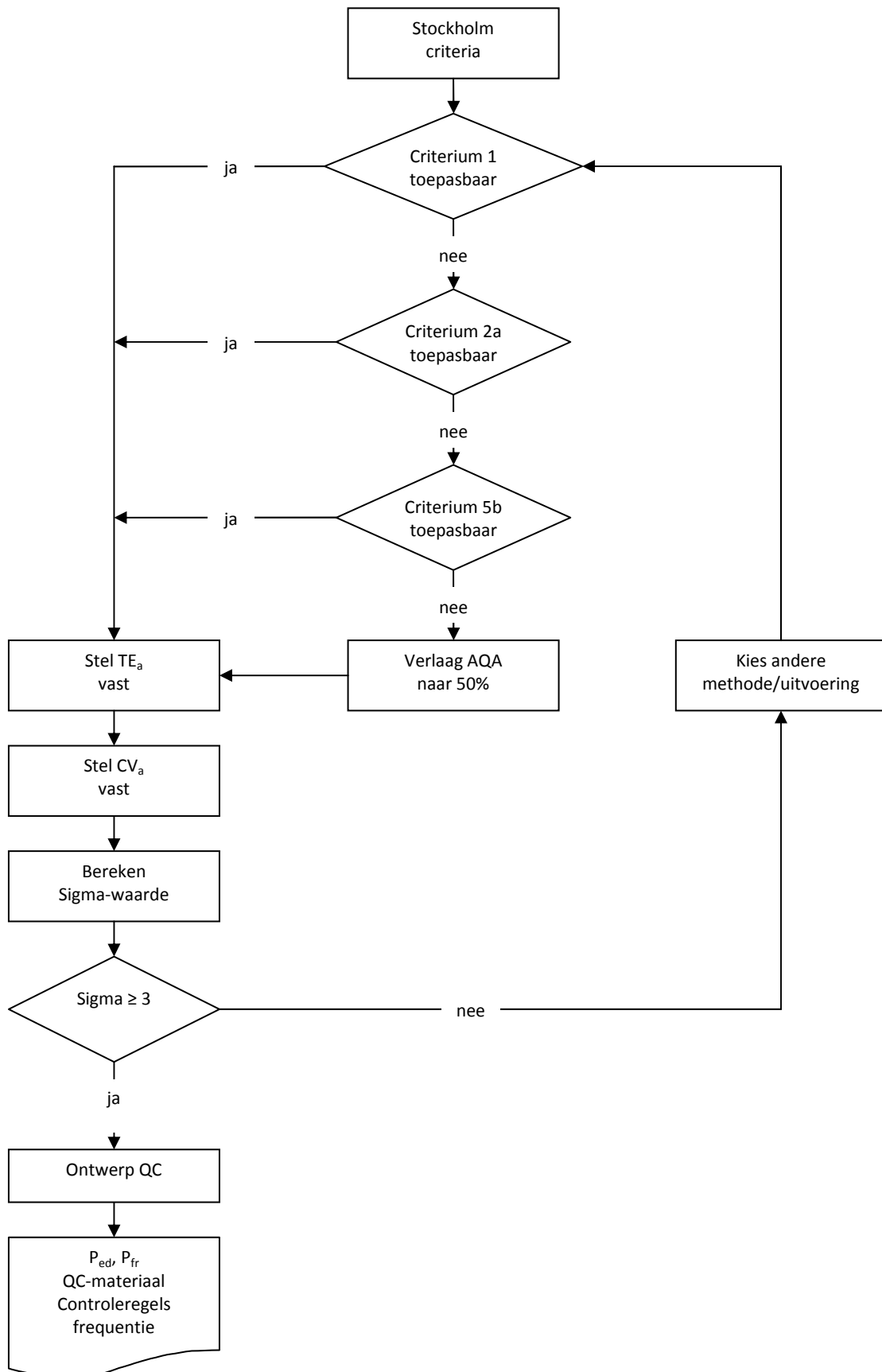
CV_a / CV_i	<i>Desirable bias / CV_i (%)</i>	<i>Minimum bias / CV_i (%)</i>
0,00	32,70	69,30
0,05	32,40	69,00
0,10	31,30	67,90
0,15	29,60	66,20
0,20	27,20	63,80
0,25	24,20	60,80
0,30	20,50	57,10
0,35	16,20	52,80
0,40	11,40	47,90
0,45	5,90	42,50
0,50	0,00	36,60
0,55		30,10
0,60		23,20
0,65		15,90
0,70		8,10
0,75		0,00

HOOFDSTUK 9 STAPPENPLAN VOOR HET INRICHTEN VAN DE INTERNE KWALITEITSCONTROLE MET BEHULP VAN SIX SIGMA

- 1) Stel de toegestane afwijking (TE_a) van een bepaling vast.
Streef bij de keuze van de TE_a naar het volgen van de Stockholm criteria (Hoofdstuk 5, pagina 8). Veelal zal de TE_a gebaseerd zijn op de biologische variatie (Bijlage IV).
- 2) Stel de analytische imprecisie van de bepaling vast.
Doe dit voor ieder apparaat waarop de bepaling uitgevoerd wordt (Hoofdstuk 6).
- 3) Wanneer de bepaling op meerdere – gelijkwaardige – apparaten uitgevoerd wordt, kan ervoor gekozen worden om gebruik te maken van het concept van het virtuele apparaat. Stel hiervoor de totale imprecisie van de apparaten samen ($CV_{a,va}$) vast (Paragraaf 8.3).
- 4) Bereken Sigma, stel *bias* op 0.
Gebruik CV_a (individueel apparaat) of $CV_{a,va}$ (virtueel apparaat).
 $\text{Sigma} > 3 \Rightarrow$ stel controleregels in volgens Tabel 2 (Paragraaf 8.1).
 $\text{Sigma} < 3 \Rightarrow$ stel zo nodig gekozen TE_a bij (Paragraaf 8.2) en bereken Sigma opnieuw.
- 5) Indien de bepaling wordt uitgevoerd op meerdere apparaten: bereken de maximaal toegestane *bias* tussen de apparaten (Tabel 3).

Het stappenplan is grafisch weergegeven in Figuur 6.

Figuur 6: Ontwerp IQC met behulp van Six Sigma.



HOOFDSTUK 10 IMPLEMENTATIE VAN DE LEIDRAAD

De vertaling van conceptuele inzichten naar breed geaccepteerde praktische werkwijzen kost tijd. Eenmaal in praktijk gebracht, blijkt het ook niet eenvoudig om bestaande werkwijzen aan te passen aan nieuwe inzichten. Dat geldt ook voor analytische kwaliteitsbewaking: de door Shewhart in 1931 beschreven statistische procescontrole vond pas na de publicatie van Levey en Jennings [1950] ingang in klinische laboratoria. De in 1981 beschreven *multi-rules* (beter bekend als Westgardregels [Westgard *et al.*, 1981]) zijn nog altijd algemeen gangbaar, hoewel Cotlove *et al.* reeds in 1970 beschreven dat de tolerantiegrenzen per bepaling afhankelijk zijn van biologische en analytische variatie.

Elk laboratorium is vrij in zijn keuze om te beslissen of, en in welke volgorde, bepalingen in aanmerking komen voor de in deze leidraad beschreven Six Sigma benadering. Indien men de Six Sigma benadering wil toepassen, ligt het voor de hand om te beginnen met hoogvolume chemometrische bepalingen. Niet alleen zijn hiervoor de input parameters eenvoudig vast te stellen (analytische variatie en TE_a). De kwaliteitscontroleregels zijn ook relatief eenvoudig vast te stellen. Het effect van de implementatie is bovendien snel zichtbaar.

De werkgroep is er van overtuigd dat de leidraad bijdraagt aan een meer rationeel onderbouwde analytische kwaliteitsbewaking en een doelmatiger inzet van mensen en middelen. De bereidheid van laboratoria om ervaringen en *best practices* te delen zal zeker bijdragen aan implementatie van deze leidraad.

LITERATUURREFERENTIES

Barwick V. Evaluating measurement uncertainty in clinical chemistry. UK National Measurement System 2012.

http://www.nmschembio.org.uk/dm_documents/Clinical_worked_examples_report_Final_f26JC.pdf

Burnett D, Ceriotti F, Cooper G, Parvin C, Plebani M, Westgard J. Collective opinion paper on findings of the 2009 convocation of experts on quality control. Clin Chem Lab Med 2010; 48: 41-52.

Calleja J. Parallel processing and maintaining adequate alignment between instruments and methods. Clin Biochem Rev 2008; 29: S71-S77.

Carlson RO, Amirahmadi F, Hernandez JS. A primer on the cost of quality for improvement of laboratory and pathology specimen processes. Am J Clin Pathol 2012; 138: 347-354.

Carobene A, Braga F, Røraas T, Sandberg S, Bartlett WA. A systematic review of data on biological variation for alanine aminotransferase, aspartate aminotransferase and γ -glutamyl transferase. Clin Chem Lab Med 2013; 51: 1997-2007.

Clinical Laboratory Improvement Amendments (CLIA).

<http://www.cms.gov/Regulations-and-Guidance/Legislation/CLIA/index.html>,
geraadpleegd op 27-03-2014.

Cotlove E, Harris EK, Williams GZ. Biological and analytical components of variation in long-term studies of serum constituents in normal subjects. III. Physiological and medical implications. Clin Chem 1970; 16: 1028-1032.

Fraser CG, Harris EK. Generation and application of data on biological variation in clinical chemistry. Crit Rev Clin Lab Sci 1989; 27: 409-437.

Fraser CG, Hyltoft Peterson P, Larsen ML. Setting analytical goals for random analytical error in specific clinical monitoring situations. Clin Chem 1990; 36: 1625-1628.

Fraser CG, Petersen PH. Quality goals in external quality assessment are best based on biology. Scan J Clin Lab Invest Suppl. 1993; 212: 8-9.

Fraser CG, Hyltoft Petersen P. Analytical performance characteristics should be judged against objective quality specifications. Clin Chem 1999; 45: 321-323.

- Gowans, EM, Petersen PH, Blaabjerg O, Hørder M. Analytical goals for the acceptance of common reference intervals for laboratories throughout a geographical area. *Scand J Clin Lab Invest* 1988; 48: 757-764.
- Gras JM, Philippe M. Application of the Six Sigma concept in clinical laboratories: a review. *Clin Chem Lab Med* 2007; 45:789-796.
- Guide to the expression of uncertainty in measurement (GUM). JCGM 100: 2008 – GUM 1995 with minor corrections.
http://www.bipm.org/utils/common/documents/jcgm/JCGM_100_2008_E.pdf,
geraadpleegd op 09-05-2014.
- Haesselbarth W. Accounting for bias in measurement uncertainty estimation. *Accred Qual Assur* 2004; 9: 509-514.
- ISO 15189: 2012. Medical laboratories – Requirements for quality and competence. ISO (International Organization for Standardization), Genève, Zwitserland.
- ISO 3534-1: 2006. Statistics – Vocabulary and symbols – Part 1: General statistical terms and terms used in probability. ISO (International Organization for Standardization), Genève, Zwitserland.
- Jansen R, Jassam N, Thomas A, Perich C, Fernandez-Calle P, Faria AP, Correia H, Barth JH, Weykamp C, Cobbaert C, Thelen M, Ricós C. A category 1 EQA scheme for comparison of laboratory performance and method performance: An international pilot study in the framework of the Calibration 2000 project. *Clin Chim Acta* 2013, nov. 14 (Epub ahead of print).
- Jørgensen LGM, Brandslund I, Stahl M, Hyltoft Petersen P, Iversen S, Klitgaard N, Olivarius N de Fine. Upper reference limit, analytical quality specifications and clinical use of haemoglobin A1C. *Scand J Clin Lab Invest* 2002; 62: 609-622.
- Kallner A, McQueen M, Heuck C. The Stockholm consensus conference on quality specifications in laboratory medicine, 25-26 April 1999. *Scand J Clin Lab Invest* 1999; 59: 475-476.
- Kazmierczak SC. Laboratory quality control: using patient data to assess analytical performance. *Clin Chem Lab Med* 2003; 41: 617-627.
- Kenny D, Fraser CG, Hyltoft Petersen P, Kallner A. Consensus agreement. *Scand J Clin Lab Invest* 1999; 59: 585.

Klee GG. Tolerance limits for short-term analytical bias and analytical imprecision derived from clinical assay specificity. *Clin Chem* 1993; 39: 1514-1518.

Larsen ML, Fraser CG, Petersen PH. A comparison of analytical goals for haemoglobin A_{1c} assays derived using different strategies. *Ann Clin Biochem* 1991; 28: 272-278.

Levey S, Jennings ER. The use of control charts in the clinical laboratory. *Am J Clin Pathol* 1950; 20: 1059-1066.

Linko S, Örnemark U, Kessel R, Taylor PDP. Evaluation of uncertainty of measurement in routine clinical chemistry – Applications to determination of the substance concentration of calcium and glucose in serum. *Clin Chem Lab Med* 2002; 40: 391–398.

Motorola Solutions.

<http://www.motorola.com/Business/US-EN/Training+Home/Lean+Six+Sigma>,

geraadpleegd op 27-03-2014.

National Pathology Accreditation Advisory Council (NPAAC) – Requirements for the estimation of measurement uncertainty, 2007 (available at www.health.gov.au/npaac).

Nevalainen D, Berte L, Kraft C, Leigh E, Picaso L, Morgan T. Evaluating laboratory performance on quality indicators with the six sigma scale. *Arch Pathol Lab Med* 2000; 124: 516-519.

Oosterhuis WP. Gross overestimation of total allowable error based on biological variation. *Clin Chem* 2011; 57: 1334-1336.

Perich C, Ricós C, Alvarez V, Biosca C, Boned B, Cava F, Doménech MV, Fernández-Calle P, Fernández-Fernández P, García-Lario JV, Minchinela J, Simón M, Jansen R. External quality assurance programs as a tool for verifying standardization of measurement procedures: Pilot collaboration in Europe. *Clin Chim Acta* 2014; 432: 82-89.

Ricós C, Alvarez V, Cava F, García-Lario JV, Hernández A, Jiménez CV, Minchinela J, Perich C, Simón M. Current databases on biological variation: pros, cons and progress. *Scand J Clin Lab Invest* 1999; 59: 491-500. Update in: Minchinela J, Ricós C, Perich C, Fernández-Calle P, Alvarez V, Domenech M, Simón M, Biosca C, Boned B, Cava F, García-Lario JV, Fernández-Fernández P. Biological variation database, and quality specifications for imprecision, bias and total error (desirable and minimum). The 2014 update.

<http://www.westgard.com/biodatabase-2014-update.htm>

Richtlinie der Bundesärztekammer (RiliBÄK).

<http://www.bundesaerztekammer.de/page.asp?his=0.7.45>, geraadpleegd op 27-03-2014

Røraas T, Petersen PH, Sandberg S. Confidence intervals and power calculations for within-person biological variation: effect of analytical imprecision, number of replicates, number of samples, and number of individuals. *Clin Chem* 2012; 58: 1306-1313. Erratum in *Clin Chem* 2012; 58: 1724.

Rynning M, Wentzel-Larsen T, Bolann BJ. A model for an uncertainty budget for preanalytical variables in clinical chemistry analyses. *Clin Chem* 2007; 53: 1343-1348.

Sandoval Y, Apple FS. The global need to define normality: the 99th percentile value of cardiac troponin. *Clin Chem* 2014; 60: 455-462.

Shewhart WA. Economic control of quality of manufactured product. Van Nostrand, New York, NY, 1931.

Schoenmakers CHH, Naus AJM, Vermeer HJ, Van Loon D, Steen G. Practical application of Sigma Metrics QC procedures in clinical chemistry. *Clin Chem Lab Med* 2011; 49: 1837-1843.

Solberg HE. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section, Expert Panel on Theory of Reference Values, and International Committee for Standardization in Haematology (ICSH), Standing Committee on Reference Values, Approved Recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *J Clin Chem Clin Biochem* 1987; 25: 337-342.

Stankovic AK, Romeo P. The role of in vitro diagnostic companies in reducing laboratory error. *Clin Chem Lab Med* 2007; 45:781-788.

Stöckl D, Baadenhuijsen H, Fraser CG, Libeer J-C, Petersen PH, Ricós 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: 157-169.

Stöckl D, Thienpont LM. About the z-multiplier in total error calculations. *Clin Chem Lab Med* 2008; 46: 1648-1649.

Westgard JO, Carey RN, Wold S. Criteria for judging precision and accuracy in method development and evaluation. *Clin Chem* 1974; 20: 825-833. Corrections in *Clin Chem* 1975; 21: 157.

Westgard JO, Barry PL, Hunt MR, Groth T. A multi-rule Shewhart chart for quality control in clinical chemistry. *Clin Chem* 1981; 27: 493-501.

Westgard JO, Burnett RW. Precision requirements for cost-effective operation of analytical processes. Clin Chem 1990; 36: 1629-1632.

Westgard JO. Six Sigma Quality Design & Control: Desirable Precision and Requisite QC for Laboratory Measurement Processes. Madison, WI: Westgard QC; 2000.

White GH, Farrance I (on behalf of the AACB Uncertainty of Measurement Working Group). Uncertainty of measurement in quantitative medical testing – A laboratory implementation guide. Clin Biochem Rev 2004; 25, Suppl. (ii): S1-S24.

BIJLAGE I VASTSTELLEN VAN DE TOEVALLIGE FOUT OF *RANDOM ERROR*

Fouten in analytische metingen kunnen worden onderverdeeld in twee categorieën: toevallige fouten (*random errors*) en systematische fouten (*systematic errors*). Toevallige fouten worden veroorzaakt door onvoorspelbare fluctuaties in de meetsystemen van meetapparatuur. Dergelijke fouten zijn inherent aan meetsystemen. De definitie van de toevallige fout of *random error* geeft vaak aanleiding tot verwarring of begripsproblemen. Het is daarom van belang te benadrukken dat ook binnen de klinische chemie de statistische definitie van de *random error* wordt gehanteerd. Er wordt dus van uitgegaan dat de meetuitkomsten van analytische laboratoriumbepalingen op correcte wijze zijn uitgevoerd en verkregen. Afwijkingen in de meetresultaten die het gevolg zijn van onjuiste calibratie, falende kwaliteitscontrole of triviale fouten als verwisseling van reagens worden niet onder de toevallige fouten gerekend.

Belangrijk is dat toevallige fouten zowel een positieve als negatieve dispersie rondom het rekenkundig gemiddelde vertonen (vergelijking 6). De sommatie van alle toevallige fouten tendeert naar nul. In de klinische chemie wordt het gemiddelde vaak als de juiste waarde of *target value* beschouwd. Onnodig hierbij te vermelden dat de ware waarde principieel niet kan worden gemeten (zie Bijlage II).

Het concept van de toevallige fout is direct gerelateerd aan de imprecisie. Hoe lager de imprecisie van een bepaling des te kleiner de variatie (= standaarddeviatie) van de fluctuaties in de meetuitkomst. De imprecisie wordt gewoonlijk gekwantificeerd door berekening van de standaarddeviatie (SD) van een dataset bestaande uit herhalingsmetingen (vergelijking 7). Aangezien de SD doorgaans groter wordt bij toenemende concentratie is het in de klinische chemie inzichtelijker om de variatiecoëfficiënt (CV, *Coefficient of Variation*) te berekenen waarbij de SD uitgedrukt wordt als percentage van de gemiddelde concentratie uit een herhalingsexperiment (vergelijking 8).

$$x_{\text{gem}} = \sum x_i / N \quad (\text{vergelijking 6})$$

$$SD = \sqrt{(\sum (x_i - x_{\text{gem}})^2) / N - 1} \quad (\text{vergelijking 7})$$

$$CV = 100 * s / x_{\text{gem}} \quad (\text{vergelijking 8})$$

x_i = individuele meting

x_{gem} = rekenkundig gemiddelde van serie herhalingsmetingen

N = totaal aantal bepalingen

Opmerking: in vergelijking 7 bestaat de noemer uit de term $N-1$ aangezien het een steekproef betreft en er derhalve sprake is van een steekproefstandaarddeviatie s . De standaarddeviatie van een populatie wordt immers gegeven door σ die per definitie bestaat uit de verzameling van alle mogelijke meetuitkomsten (dus $N \rightarrow \infty$).

Wat verstaan we onder een systematische fout?

Een systematische fout, ook wel *bias* genoemd, veroorzaakt een vertekening (dat wil zeggen, een afwijking ten opzichte van het juiste resultaat) welke niet te wijten is aan toevallige effecten. *Bias* wordt gedefinieerd als ‘*the difference between the expectation of the test results and an accepted reference value*’ [ISO 3534-1].

Waar de toevallige fout (imprecisie) het gemiddelde van de waarnemingen, bij voldoende steekproefgrootte, niet verandert, verandert de systematische fout het gemiddelde wel. De afwijking is reproduceerbaar. De aard van de fout is hierbij niet van belang: een afwijking zou systematisch kunnen zijn (bijvoorbeeld een afwijking die in een serie afnamebuizen voorkomt). Als deze fout echter enkelvoudig (niet herhaalbaar) voorkomt, is het een toevallige fout.

Een systematische fout kan verschillende vormen aannemen: additief als een afwijking van de werkelijke meetwaarde X van monster i met een constante waarde c als $(X_i + c)$, proportioneel met een factor a als aX_i , of in een andere mathematische vorm. In tegenstelling met de imprecisie kan voor de *bias* wel gecorrigeerd worden. De *bias* hoeft niet constant in de tijd te zijn, maar kan bijvoorbeeld veranderen na calibratie of nieuw reagens.

Waarom is de systematische fout van belang?

Naast het feit dat systematische fouten leiden tot onjuiste resultaten, eist de ISO 15189 bovendien dat het laboratorium de meetonzekerheid van bepalingen vaststelt:

Het laboratorium moet de meetonzekerheid vaststellen voor elke meetprocedure in de onderzoeksfase die wordt toegepast om in monsters van patiënten gemeten kwantitatieve waarden te rapporteren. Het laboratorium moet de prestatie-eisen definiëren voor de meetonzekerheid van elke meetprocedure en de geschatte waarden van meetonzekerheid beoordelen [ISO 15189, 5.5.1.4.].

Deze meetonzekerheid wordt bepaald door zowel *bias* als imprecisie. Daarnaast wordt de Sigma score berekend op basis van *bias* en imprecisie.

Sigma score en bias

Hoewel de *bias* deel uitmaakt van de vergelijking voor berekening van de Sigma score, wordt de *bias* niet meegenomen in de berekening van de Sigma score voor het ontwerpen van de IQC. Hierbij is alleen de imprecisie van belang. Hiervoor zijn verschillende redenen.

- de *bias* zal vaak veranderen. De interne kwaliteitscontrole kan beter niet aangepast worden op basis van veranderingen die op korte termijn optreden.

- er kan sprake zijn van een methodespecifieke *bias*, bijvoorbeeld op basis van externe kwaliteitscontrole resultaten. Men kan deze *bias* op verschillende gronden accepteren. Het is dan niet logisch om de interne kwaliteitscontrole te ontwerpen op basis van deze *bias*. Intensieve kwaliteitscontrole zal deze *bias* immers niet veranderen.

- de *bias* op basis van consensuswaarden van externe kwaliteitscontrole kan onbetrouwbaar zijn, onder meer omdat deelnemers met correctiefactoren kunnen werken. De IQC is er natuurlijk wel op gericht *bias* te detecteren en te minimaliseren. Onderstaand worden verschillende aspecten van de *bias* behandeld, die van belang kunnen zijn voor het begrip en toepassing van de Six Sigma methode.

Tijdsfactor van de systematische fout

Het onderscheid tussen imprecisie en *bias* is niet altijd eenvoudig [Klee, 1993]. Stel dat er verschillende discrete series bepalingen worden uitgevoerd. Elke serie heeft een kleine systematische afwijking, soms positief, soms negatief. Worden deze series samen genomen, dan kan de netto *bias* gelijk zijn aan nul, maar draagt de *bias* van elke serie wel bij tot de totale imprecisie. Met andere woorden: wordt over een langer tijdsinterval gemeten, dan presenteert *bias* zich als imprecisie, en draagt bij aan de totale meetonzekerheid:

$$SD_{\text{tot}} = \sqrt{\{(n-1)/n\}SD_a^2 + C^2} \quad \text{(vergelijking 9)}$$

SD_{tot} = totale variantie

SD_a = analytische variantie

C = *bias*

In de praktijk kan de *bias* gedefinieerd worden als een afwijking (shift) van de testwaarden in het beschouwde tijdsinterval ('*over the time interval considered*') [Klee, 1993]. Bij

geautomatiseerde apparaten is er geen sprake meer van bepalingsseries, en is het minder duidelijk over welke periode de *bias* dient te worden gedefinieerd of vastgesteld.

Wat is de “werkelijke waarde”?

Bias wordt gedefinieerd als de reproduceerbare afwijking van de meetwaarde tot de werkelijke waarde. Het is daarom van belang de (schatting van) de werkelijke waarde te kennen. In het ideale geval is er sprake van een goed gedefinieerde te meten stof (natrium, kalium) waarvoor een referentie bepalingmethode bestaat en gecertificeerd en commuteerbaar referentiemateriaal. In de klinische chemie is hier in veel gevallen geen sprake van. Vaak kan de te meten stof niet voldoende gedefinieerd worden. Er kan sprake zijn van verschillende moleculaire vormen, die in verschillende mate worden meebepaald (bijvoorbeeld HCG (vrij, beta, nicked), vitamine D (D2, D3)). Ook kan een referentiemethode ontbreken (PSA). Standaardisatie wordt daarmee erg lastig, zodat men uitkomt op methodespecifieke standaarden en consensuswaarden.

Guide to the expression of uncertainty in measurement (GUM): bottom-up

In 1993 verscheen de eerste versie van een standaard voor het rapporteren van de meetfout [GUM]. De hierin aanbevolen methode is *bottom-up*. Hiermee wordt bedoeld dat elk onderdeel van een meting dat analytische variatie kan veroorzaken, apart wordt beoordeeld. Van elk onderdeel wordt de bijdrage aan de uiteindelijke meetfout vastgesteld. Deze verschillende bijdragen worden gecombineerd. Dit geeft een schatting van de totale meetfout. Deze methode werd in de klinische chemie als voorbeeld uitgewerkt voor enkele bepalingen [Linko *et al.*, 2002]. Voorbeelden van onderdelen van de meetprocedure die bijdragen aan de meetfout zijn bijvoorbeeld de onnauwkeurigheid in de bepaling van blanco, calibrator en monster, en in de concentratie van calibrator in de calibratievloeistof.

Een belangrijke stap in de GUM methode is het vaststellen van een eventuele *bias*, en om daarvoor te corrigeren als deze significant verschilt van nul [Haesselbarth, 2004]:

3.2.3 Systematic error, like random error, cannot be eliminated but it too can often be reduced.

If a systematic error arises from a recognized effect of an influence quantity on a measurement result, hereafter termed a systematic effect, the effect can be quantified and, if it is significant in size relative to the required accuracy of the measurement, a correction (B.2.23) or correction

factor (B.2.24) can be applied to compensate for the effect. It is assumed that, after correction, the expectation or expected value of the error arising from a systematic effect is zero.

Volgens GUM kan er dus geen sprake zijn van *bias*, omdat daarvoor gecorrigeerd dient te worden. De onzekerheid van de correctie voor *bias* is hier echter wel onderdeel van de berekening van de totale meetfout, net als andere vormen van imprecisie.

Bias in de klinische chemie

In de praktijk kan het voorkomen dat er sprake is van significante *bias*, maar dat de gegevens niet volstaan voor het afleiden van een goede correctie. Het kan te ingewikkeld en te duur zijn om de *bias* nauwkeurig vast te stellen. In de klinische chemie is dit vaak het geval. Via interne of externe kwaliteitscontrole blijkt er dan sprake te zijn van enige *bias*, maar de gegevens zijn vaak niet voldoende om daarvoor te corrigeren. Een pragmatisch alternatief is dan om deze afwijking in rekening te brengen in de totale meetonzekerheid. GUM geeft weliswaar enige ruimte voor deze benadering, maar lijkt deze methode echter te ontmoedigen⁵. Het verhogen van de meetonzekerheid als methode om te corrigeren voor belangrijke *bias* lijkt echter beter dan het toepassen van een twijfelachtige correctie of, erger nog, het negeren van deze afwijking.

Momenteel zijn er verschillende benaderingen om rekening te houden met ongecorrigeerde *bias*; er is echter nog geen algemeen aanvaarde procedure [Haesselbarth, 2004]. Merk op dat in het lineaire model (zie hoofdstuk 7) dat door Westgard wordt gevolgd, de totale meetfout berekend wordt door de *bias* op te tellen bij de imprecisie [Westgard *et al.*, 1974]. In deze berekening wordt geen rekening gehouden met de onzekerheid van de *bias*.

Klinische chemie: top-down

In de klinische chemie is de *bottom-up* benadering minder goed toepasbaar dan de *top-down* benadering. Bij de *bottom-up* benadering worden ruwe data (zoals bijvoorbeeld gemeten extinctie) gebruikt om de totale meetonzekerheid vast te stellen. Bij de *top-down* benadering worden meetgegevens gebruikt (bepalingsresultaten, controles). *Bias* en imprecisie kunnen bijvoorbeeld worden vastgesteld door de correlatie met en afwijking ten opzichte van controlemonsters. Dit levert volgens GUM echter een beperkte schatting van de totale onzekerheid van de meting, omdat verschillende oorzaken van variatie niet worden meegenomen maar constant blijven (apparaatfactoren, calibratiefactoren, reagensfactoren).

In een recent rapport is deze benadering geëvalueerd [Barwick, 2012]. Hierin wordt onderbouwd, dat de *top-down* benadering volgens huidige richtlijnen wel kan worden toegepast in de klinische chemie:

The requirements in ISO/IEC 17025 make it clear that an uncertainty estimate can use existing method performance data, for example from method validation studies. There are a number of papers and guides which describe the use of method performance data in the evaluation of measurement uncertainty for clinical measurement procedures [...]. The 'top-down' approach is discussed in a number of documents including the Eurachem/CITAC Guide [...], the Eurolab report on alternative approaches to uncertainty evaluation [...] and ISO 21748 [...]. The latter focuses on the use of data from interlaboratory studies for method validation.

Ten opzichte van welk ijkpunt wordt de systematische afwijking vastgesteld?

Er zijn verschillende mogelijkheden om de *bias* vast te stellen [Barwick, 2012]. De *bias* geeft de afwijking aan tussen de gemeten waarde en de werkelijke waarde. Het wordt experimenteel geschat door het verschil te bepalen tussen het gemiddelde van een serie bepalingen en een geschikte referentiewaarde. De meest gebruikelijke benaderingen voor het bepalen van *bias* zijn:

Schatting van de *bias* op basis van gecertificeerd referentiemateriaal

Gecertificeerd referentiemateriaal geeft een traceerbare referentie(ijk)waarde. Deze waarde dient bij voorkeur vastgesteld te worden met een beschikbare referentiemethode. Als dit materiaal en een referentiemethode beschikbaar zijn, is dit de eerste keus om *bias* vast te stellen. Het aantal gecertificeerde referentiematerialen die voldoen aan internationaal vastgestelde normen is echter beperkt, vergeleken met de grote verscheidenheid aan monster/test combinaties die in het laboratorium voorkomen. De SKML werkt zoveel mogelijk met commuteerbare monsters (EQA categorie 1). Het vaststellen en de correctie van eventuele *bias* zal leiden tot een bijgesteld meetniveau. Dit heeft echter geen invloed op de berekening van de Sigma score, en daarmee geen invloed op het ontwerp van de IQC.

Schatting van de *bias* op basis van de bepaling van "*gespikete*" monsters (eventueel met toevoegingen)

In dit geval wordt referentiemateriaal bereid door een bekende hoeveelheid van de te analyseren stof aan een geschikte matrix toe te voegen. De concentratie van de te analyseren stof wordt berekend en geeft de referentiewaarde ten behoeve van de bepaling van de *bias*. Toegevoegd materiaal gedraagt zich echter niet altijd op dezelfde wijze als oorspronkelijk materiaal.

Schatting van de *bias* op basis van externe kwaliteitscontrole

In het geval van goed gedefinieerde methoden is het mogelijk om de *bias* te bepalen via externe kwaliteitscontrole gegevens. Hiervoor dient aan bepaalde voorwaarden te zijn voldaan. De toegekende waarde van het referentiemateriaal dient hierbij als referentiewaarde. Echter, wanneer de toegekende waarde wordt vastgesteld op basis van een consensus, is er geen sprake van een traceerbaar resultaat. De toegekende waarde zal vaak methodeafhankelijk zijn, waarmee de schatting van de *bias* afhankelijk wordt van de uitkomsten van laboratoria die dezelfde methode gebruiken. Deelnemende laboratoria kunnen correctiefactoren toepassen, wat invloed kan hebben op de consensuswaarde. Ondanks deze beperkingen kan dit de enige beschikbare methodiek zijn om de *bias* te schatten [Linko *et al.*, 2002; National Pathology Accreditation Advisory Council (NPAAC), 2007; White en Farrance, 2004; Rynning *et al.*, 2007].

Door Schoenmakers *et al.* [2011] wordt ook voorgesteld de systematische fout vast te stellen ten opzichte van de referentiemethode, of – als deze niet beschikbaar is – de consensuswaarde. Dit uitgangspunt geldt overigens alleen, als de uitkomst van de referentiemethode (of van de consensusmethode) overeenkomt met de streefwaarde van het laboratorium, en correspondeert met de referentiewaarden van het laboratorium. Dit is bijvoorbeeld het geval bij bepalingen waarbij geharmoniseerde referentiewaarden gelden. Dit is niet het geval als het laboratorium een andere streefwaarde hanteert, met daarbij horende – vaak laboratoriumspecifieke – referentiewaarden.

Schatting van de *bias* op basis van patiëntgemiddelden

Heeft een laboratorium een bepalingsmethode in gebruik genomen, dan streeft men ernaar de methode stabiel te houden. Dit betekent, dat het niveau zoveel mogelijk constant wordt gehouden ten opzichte van het uitgangsniveau, zodat de referentiewaarden hun geldigheid

behouden en de uitkomsten van patiënten in de tijd vergelijkbaar blijven. Klee [1993] stelde daarom, dat in de praktijk *bias* gedefinieerd moet worden als een afwijking (*shift*) van de testwaarden [...] ten opzichte van een ijkpunt van de bepalingmethode dat gold ten tijde van het vaststellen van de referentiewaarden. Patiëntgemiddelden zijn uitvoerig onderzocht en beschreven om de stabiliteit van bepalingen te controleren (AON *average of normals*)[Kazmierczak, 2003]. Het voordeel van kwaliteitscontrole op basis van patiëntenresultaten kan zijn, dat de sensitiviteit voor detectie van afwijkingen wordt verhoogd. Bij gebruik van kwaliteitscontrolemonsters is het aantal waarnemingen immers erg laag, waardoor de sensitiviteit ook beperkt is. Een hogere sensitiviteit laat een lagere Sigma score toe. Juist voor bepalingen met een lage biologische variatie (en daardoor vaak een lage Sigma score) zou toepassing van kwaliteitscontrole op basis van patiëntenresultaten een toegevoegde waarde kunnen hebben. Deze methode valt buiten het kader van deze leidraad en zal daarom niet verder worden beschreven.

***Bias* tussen verschillende apparaten**

Wanneer dezelfde bepaling op twee of meer apparaten binnen één laboratorium wordt uitgevoerd, zullen er onherroepelijk niveauverschillen optreden tussen de apparaten. Er is daarmee dus sprake van *bias*. In het hypothetische geval dat de exacte *bias* van elk apparaat ten opzichte van de werkelijke waarde bekend zou zijn, zou dit in rekening gebracht kunnen worden bij de berekening van de Sigma score. In de praktijk is dit om bovengenoemde redenen niet mogelijk.

Er is in deze leidraad een pragmatische benadering gevolgd: die van het *virtuele apparaat* waarbij de resultaten van de IQC van meerdere apparaten worden samengevoegd. In feite zal *bias* tussen apparaten een bijdrage leveren aan de totale imprecisie, en daarmee de Sigma score beïnvloeden.

Een andere mogelijkheid is om de verschillende apparaten apart te bewaken en eventueel zelfs aparte Sigma scores te hanteren op basis van de respectievelijke imprecisies. Hierbij wordt de *bias* in feite genegeerd.

In alle gevallen wordt aanbevolen de *bias* binnen bepaalde grenzen te houden. Deze grenzen zijn gebaseerd op berekening van het kritische verschil (zie ook de hoofdtekst van deze leidraad).

Voor de keuze van de TE_a wordt bij de Six Sigma methodologie veelvuldig gebruikt gemaakt van de biologische variatie van een bepaling binnen een gezond persoon of een groep gezonde personen.

Het werk van Ricós *et al.* [1999] is dan ook baanbrekend en essentieel voor de toepassing van Six Sigma methodieken in de laboratoriumgeneeskunde, omdat zij door middel van uitgebreid literatuuronderzoek de biologische variatie heeft ingeschat voor een groot aantal veel voorkomende laboratoriumbepalingen, en deze data beschikbaar heeft gesteld aan het brede publiek. Het is echter belangrijk om te begrijpen wat de beperkingen zijn in de data zoals die uit het werk van Ricós *et al.* zijn voortgekomen en gepubliceerd op de website van Westgard. Om te beginnen gaat het werk van Ricós om een uitgebreide literatuurstudie, waarbij gemiddelden zijn genomen van studies die op verschillende wijze zijn georganiseerd en uitgevoerd, voornamelijk omdat ze ook verschillende doelen hadden. Bij gebruik van de data dient ondermeer met de volgende aspecten rekening gehouden te worden:

Selectie bias

Biologische variatie wordt vaak bepaald in een populatie van jonge gezonde personen, bijvoorbeeld studenten. Biologische variatie kan echter – net als referentiewaarden – verschillen per leeftijdscategorie of per geslacht. [Carobene *et al.*, 2013]. Wat betreft de selectie van de testpopulatie zijn er dus een aantal verschillende invalshoeken. De eerste is om de biologische variatie te bepalen van de populatie die de minste variatie vertoont. Dit leidt tot de strengste regels, en een controlesystematiek die geschikt is voor alle doelgroepen, maar wellicht te veel tijd en geld in controle en calibratie steekt. Een alternatief is om de biologische variatie te bepalen in een groep die een correcte afspiegeling is van de gehele populatie, of van de deelpopulatie die de hoogste incidentie vertoont van de pathologie die men met de bepaling probeert te diagnosticeren. Zo lang zowel met het bepalen van de biologische variatie als met de inrichting van de interne kwaliteitscontrole rekening wordt gehouden met het doel van de bepaling, kan iedere aanpak valide zijn.

Tijdsintervallen

Wanneer duidelijk is binnen welke tijd een bepaling herhaald wordt om een significante afwijking aan te tonen, dient in principe de biologische variatie ook over die tijdsperiode bepaald te worden. Voor HbA1c betreffende diabetici die vier keer per jaar voor controle komen, wordt de biologische variatie bepaald over intervallen van drie maanden. Als het natrium betreft van IC patiënten die dagelijks geprikt worden, wordt de biologische variatie ook met intervallen van één dag bepaald. Carobene *et al.* [2013] hebben voor een aantal voorbeelden een overzicht gemaakt van de verschillende resultaten die kunnen volgen uit verschillende tijdsintervallen. Hieruit blijkt dat er significante verschillen kunnen optreden wanneer verschillende tijdsintervallen worden gebruikt.

Samengevat is het belangrijk om te weten of de biologische variatie die gebruikt wordt voor het bepalen van de interne kwaliteitscontrole, toepasbaar is op de specifieke klinische vraagstelling. Daarom kan het voorkomen dat het beter is om zelf de biologische variatie te bepalen in plaats van waarden uit de literatuur te gebruiken. Er zijn tot dusver echter geen eenduidige protocollen die stapsgewijs uitleggen hoe de biologische variatie van een bepaalde entiteit bepaald wordt. Dit komt hoofdzakelijk doordat het aantal proefpersonen en datapunten dat nodig is om statistisch berouwbaar data te produceren, afhankelijk is van de verhouding tussen de analytische en biologische variatiecoëfficiënt [Røraas *et al.*, 2012]. Echter, Fraser en Harris [1989] stelden dat betrouwbare data voor biologische variatie afgeleid kunnen worden uit een vrij kleine groep gezonde personen, met relatief weinig datapunten. Door Røraas *et al.* [2012] is met behulp van power berekeningen uitgewerkt dat voor alle klinisch chemische parameters in principe hetzelfde protocol geschikt zou zijn.

Uit de berekeningen van Røraas *et al.* [2012] blijkt dat voor iedere bepaling waarbij de CV_a kleiner is dan de CV_i , een protocol volstaat waarbij 20 individuen op 4 tijdspunten geprikt worden, waarna al deze monsters in duplo bepaald worden. Notoir zwakke bepalingen als totaal eiwit en natrium voldoen aan deze eis. Wanneer dit protocol gevolgd wordt kan met een power van 1.0 de CV_i bepaald worden met een betrouwbaarheidsinterval van minimaal 95%. Om *between-run* variatie te voorkomen worden alle monsters na afname ingevroren en in één run gemeten. Statistische analyse vindt plaats door middel van een *nested* ANOVA. De methode hiervoor staat beschreven in appendix 1 van het artikel van Fraser en Harris [1989], inclusief het vinden van uitbijters en een test op de homogeniteit van de data. Er zijn

echter voldoende software pakketten op de markt waarmee dit geautomatiseerd uitgevoerd kan worden.

Het concept van de totale meetfout (TE)

De TE geeft de totale meetfout weer. Westgard *et al.* [1974] hebben dit concept geïntroduceerd met de volgende argumenten:

The physician thinks rather in terms of the total analytical error, which includes both random and systematic components. From his point of view, all types of analytic error are acceptable as long as the total analytic error is less than a specified amount.

Er is een bepaalde waarschijnlijkheid dat een meetresultaat de TE limiet overschrijdt. Meestal is deze kans vastgesteld op 5%. De uitdrukking die algemeen gebruikt wordt voor de berekening van de TE is [Westgard *et al.*, 1974]:

$$TE = bias + z * CV_a \quad \text{(vergelijking 2)}$$

De z-waarde wordt in het algemeen vastgesteld op 1,65, waarbij 95% van de testresultaten binnen de TE-limiet vallen (eenzijdige limiet)⁶.

Toegestane fout (TE_a)

Eenzijds kan de TE worden berekend uit de *bias* en imprecisie (indien deze bekend zijn). Anderzijds kunnen de maximaal toegestane *bias* en imprecisie worden afgeleid, wanneer een limiet voor TE vooraf is gedefinieerd (TE_a). Volgens de Stockholm criteria worden indien mogelijk tolerantiegrenzen rond medische beslisgrenzen vastgesteld. Bij deze concentraties is de juistheid en precisie van een bepalingsmethode het meest kritisch.

TE_a gebaseerd op biologische variatie

Zoals eerder vermeld, zou de klinische noodzaak de kwaliteitsspecificaties dienen te bepalen. Dit is echter vaak lastig. Er kan sprake zijn van meerdere klinische beslisgrenzen, of deze beslisgrenzen kunnen niet precies gedefinieerd worden. Volgens de Stockholm consensus is het een alternatief om de kwaliteitsspecificaties af te leiden uit de biologische variatie.

Fraser en Hyltoft Petersen [1999] definieerden drie kwaliteitsniveaus voor zowel imprecisie als *bias*: *minimum performance*, *desirable performance* en *optimum performance*:

	imprecisie	<i>bias</i>	
<i>minimum</i>	$< 0,75 * CV_i$	$< 0,375 * (CV_i^2 + CV_g^2)^{0,5}$	(vergelijking 10)
<i>desirable</i>	$< 0,50 * CV_i$	$< 0,25 * (CV_i^2 + CV_g^2)^{0,5}$	(vergelijking 11)
<i>optimum</i>	$< 0,25 * CV_i$	$< 0,125 * (CV_i^2 + CV_g^2)^{0,5}$	(vergelijking 12)

met CV_i = *within-subject* of intra-individuele variatie, CV_g = *between-subject* of inter-individuele variatie

Bias en imprecisie kunnen vervolgens worden gesubstitueerd in vergelijking 2⁷:

$$TE_a = 0,25 * (CV_i^2 + CV_g^2)^{0,5} + 1,65 * 0,5 * CV_i \quad (\text{vergelijking 5})$$

Deze vergelijking geldt voor *monitoring* (vervolgen van patiënten). Aangezien de meeste bepalingen gebruikt worden om patiënten te vervolgen, wordt deze vergelijking het meest toegepast. Deze is ook algemeen geaccepteerd, en wordt o.a. gebruikt voor het ontwerpen van kwaliteitscontrole programma's [Burnett 2010; Schoenmakers *et al.*, 2011]. Voor veel bepalingen zijn de TE_a -waarden vastgesteld [Ricós *et al.*, 1999].

Voor de TE_a gelden bij *monitoring* (vervolgen van patiënten) strengere grenzen dan bij het stellen van een diagnose (dan geldt voor de maximale imprecisie: $0,5 * (CV_i^2 + CV_g^2)^{0,5}$, Gowans *et al.* [1988]). In Tabel 4 zijn voor een aantal bepalingen TE_a en Sigma scores weergegeven. De waarden CV_i en CV_g zijn uit de Ricós tabel gehaald; de CV_a waarden zijn hypothetisch en dienen als voorbeeld.

Tabel 4: Voorbeeldberekeningen van TE_a en Sigma scores voor verschillende bepalingen volgens het standaard model [zie Ricós *et al.*, 1999]. TE_a is berekend volgens vergelijking 5 ($TE_a = 0,25 * (CV_i^2 + CV_g^2)^{0,5} + 1,65 * 0,5 * CV_i$); Sigma is berekend volgens vergelijking 4 (Sigma score = $(TE_a - bias)/CV_a$) waarbij de *bias* buiten beschouwing wordt gelaten.

	CV_i	CV_g	TE_a	CV_a	Sigma
Natrium	0,60	0,70	0,73	1,06	0,68
Calcium	2,10	2,50	2,55	2,00	1,27
Transferrine	3,00	4,30	3,79	1,28	2,95
Totaal eiwit	2,75	4,70	3,63	0,91	3,99
LDH	8,60	14,70	11,35	2,17	5,23
Kalium	4,60	5,60	5,61	1,07	5,24
AST	12,30	23,10	16,69	2,91	5,74
Haptoglobine	20,40	36,40	27,26	2,04	13,36
CK	22,80	40,00	30,32	1,17	25,91
ALT	19,40	41,60	27,48	1,06	25,92

Alternatieve modellen voor de TE_a berekening

Er zijn alternatieve modellen, waarbij het gebrek aan theoretische onderbouwing voor de TE_a berekening niet geldt [Gowans *et al.*, 1988; Oosterhuis, 2011]. Het model van Gowans *et al.* stelt kwaliteitseisen op, gebaseerd op de handhaving van de geldigheid van referentiewaarden. In dit model worden *bias* and imprecisie gecombineerd. De invloed van de toename van *bias* en imprecisie op het aantal fout-positieven wordt hierin berekend. Wanneer *bias*, imprecisie, of een combinatie van deze twee toeneemt, zullen meer meetwaarden buiten de referentielimieten vallen. In plaats van de gebruikelijke 2,5% testresultaten buiten een referentielimiet (tweezijdig bij 1,96 SD), wordt in dit model een maximum van 4,6% buiten de eerder vastgestelde referentiewaarden aangenomen als acceptabel. Dit maximum van 4,6% is gebaseerd op de IFCC richtlijn over referentiewaarden [Solberg, 1987]. Wanneer deze richtlijn wordt gevolgd, worden de referentiewaarden berekend op basis van de resultaten van een groep van 120 personen. Er is daarbij sprake van een bepaalde onzekerheid van de referentiewaarden. Zowel de onderste als de bovenste referentiewaarde heeft een 95% betrouwbaarheidsinterval, met een binnenste en buitenste grens. Deze onzekerheid is zodanig, dat maximaal 4,6% (eenzijdig) van de resultaten buiten de binnenste grens van het betrouwbaarheidsinterval van de referentiewaarden (bij 1,96 SD) kunnen vallen.

Volgens dit model kan een curve worden berekend die de combinaties van maximum toelaatbare *bias* en imprecisie weergeeft. De maximale *bias* en imprecisie aan begin en einde van de curve zijn:

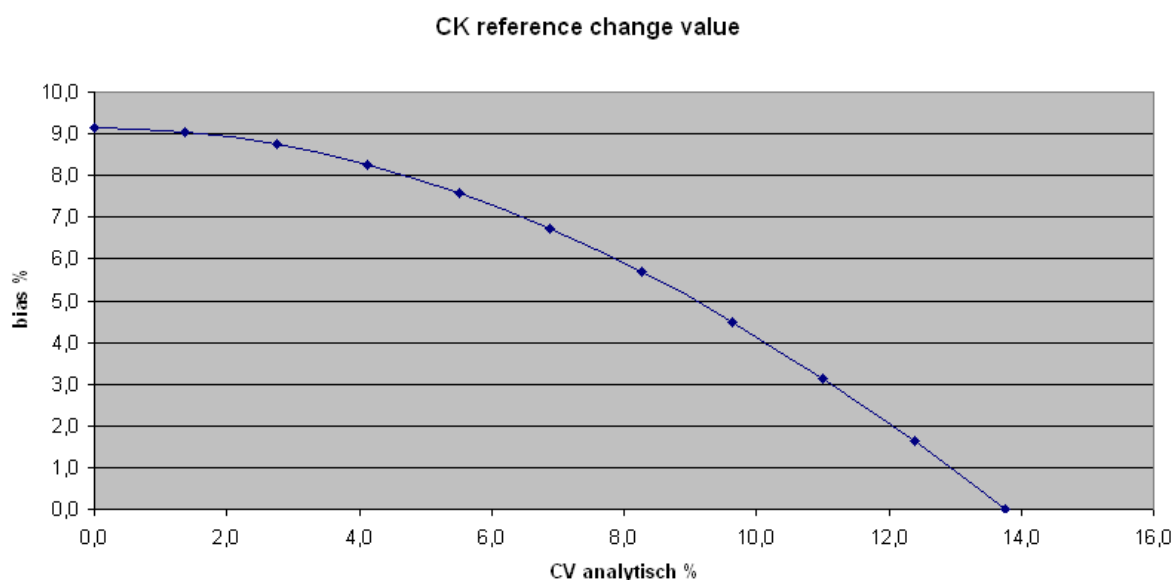
Maximale *bias* (met imprecisie = 0): $0,275 * (CV_b)$ (vergelijking 13)

Maximale imprecisie (met *bias* =0): 0,59 (CV_b) (vergelijking 14)

met CV_b = totale biologische variatie (in dit model werd geen onderscheid gemaakt tussen inter- en intra-individuele variatie)⁸.

Dit model kan worden aangepast aan de striktere norm voor *monitoring*, door uit te gaan van *Reference Change Values* (RCV)[Fraser *et al.*, 1990; Larsen *et al.*, 1991; Stöckl *et al.*, 1995]. In Figuur 7 is dit uitgewerkt voor creatine fosfokinase (CK).

Figuur 7: De curve beschrijft combinaties van *bias* en CV_a, waarbij voldaan wordt aan de voorwaarde dat de grens van de RCV op P95,4% ligt (i.p.v. P97,5).



BEGRIPPENLIJST EN AFKORTINGEN

In de leidraad worden de begrippen ‘bepaling’ (synoniem: test, analyse) en ‘apparaat’ (synoniem: toestel, instrument, *analyser*) gehanteerd. In ISO 15189 worden deze begrippen samen aangeduid als ‘onderzoeksprocedure’. Vanwege de aansluiting bij wetenschappelijke literatuur is ervoor gekozen om in sommige gevallen gebruik te maken van Engelstalige begrippen.

Term	Definitie
AQA, <i>Analytical Quality Assurance</i>	Analytische kwaliteitsborging. Hierbij wordt gebruik gemaakt van grafieken om het niveau van zekerheid aan te geven met betrekking tot de detectie van een kritische fout. Bijvoorbeeld: AQA 90% betekent dat er ten minste een 90% kans is op het opsporen van een kritische systematische fout, binnen de toegestane grenzen voor imprecisie en onnauwkeurigheid bij de vastgestelde bewakingsregels en het aantal controlemetingen.
Commuteerbaarheid	Mate van overeenstemming tussen meetresultaten voor een vastgestelde concentratie referentiemateriaal, verkregen volgens twee meetmethoden. Het betreffende materiaal is meestal een calibrator.
Controleregel	Een beslissend criterium voor het interpreteren van controlegegevens. Vaak wordt hierbij gebruik gemaakt van zg. <i>control charts</i> , een grafische methode waarbij beoordeeld wordt of een proces wel of niet voldoet aan de statistische kwaliteitscontrole. In laboratoria wordt meestal een Levey-Jennings grafiek (gehanteerd, waarbij de resultaten van de controlemetingen in de tijd worden weergegeven. Een voorbeeld van een controleregel is

	een overschrijding van een vooraf omschreven afwijking (bijvoorbeeld 2 of 3 SD).
<i>CV, Coefficient of Variation</i>	Variatiecoëfficiënt: de relatieve standaarddeviatie, dat wil zeggen de standaarddeviatie uitgedrukt als percentage van het gemiddelde (\bar{x}).
	$CV (\%) = (SD / \bar{x}) * 100$
CV_a	Analytische variatie.
CV_b	Biologische variatie. Hierbij wordt geen onderscheid gemaakt tussen intra-individuele en inter-individuele variatie.
CV_i	Intra-individuele of binnen-persoons biologische variatie (ook wel aangeduid als <i>within-subject</i> variatie of CV_w).
CV_g	Inter-individuele of tussen-persoons biologische variatie (ook wel aangeduid als <i>between-subject</i> variatie).
<i>EQA, External Quality Assessment</i>	Externe kwaliteitscontrole.
Gemiddelde (\bar{x})	Het rekenkundig gemiddelde van een reeks waarden.
<i>IQC, Internal Quality Control</i>	Interne kwaliteitscontrole.
Meetfout	Het verschil tussen een meetresultaat en de werkelijke waarde van de te meten grootte. De relatieve meetfout wordt berekend door de meetfout te delen door de streefwaarde.
Nauwkeurigheid, <i>accuracy</i>	De mate van overeenstemming tussen het resultaat van een meting en de werkelijke waarde van de te meten grootte ^{9, 10} .
P_{ed} , <i>probability for error detection</i>	Een prestatiekenmerk van een QC procedure die beschrijft hoe vaak een analysereeks wordt verworpen op basis van fouten die niet gerelateerd zijn aan de onnauwkeurigheid van de meting. Idealiter is de P_{ed} 1.00 voor klinisch relevante fouten.

	In de praktijk streven we naar een $P_{ed} > 0.90$ bij het selecteren en ontwerpen van QC procedures.
P_{fr} , probability for false rejection	Een prestatiekenmerk van een QC procedure die beschrijft hoe vaak een analysereeks wordt verworpen indien er geen fouten optreden, met uitzondering van de onnauwkeurigheid van de meting. Idealiter is de $P_{fr} 0.00$; in de praktijk streven we naar een $P_{fr} < 0.05$.
Precisie, <i>precision</i>	De mate van overeenstemming tussen onafhankelijke testresultaten verkregen onder vastgestelde voorwaarden ^{11, 12, 13} .
QC, <i>Quality Control</i>	Kwaliteitscontrole: een algemene term die verwijst naar bewaking en beoordeling van laboratoriumprocessen om zodoende problemen te identificeren en kwaliteit te handhaven.
<i>Reference Change Value</i>	Een bepaalde onzekerheid die het verschil weergeeft dat moet zijn waargenomen alvorens men een verandering van het meetresultaat in een patiënt als klinisch relevant beschouwd. Gedefinieerd door Fraser <i>et al.</i> [1990] als functie van de analytische variatie en de binnen-persoons biologische variatie.
Repeteerbaarheid, <i>repeatability</i>	Precisie onder herhaalbare omstandigheden; waarbij onafhankelijke testresultaten worden verkregen met dezelfde methode bij identiek analysemateriaal in hetzelfde laboratorium, uitgevoerd door dezelfde persoon, met dezelfde apparatuur in een kort tijdsbestek.
Reproduceerbaarheid, <i>reproducibility</i>	Precisie onder reproduceerbare omstandigheden; waarbij de testresultaten verkregen worden met dezelfde methode bij identiek analysemateriaal in verschillende laboratoria, uitgevoerd door

	verschillende medewerkers, met verschillende apparatuur.
Sigma metric	Een numerieke waarde die de prestatie van een bepaling karakteriseert. Met andere woorden: hoe vaak past de standaarddeviatie of de Sigma score in de tolerantiegrens of kwaliteitseis van een bepaling?
SQC, <i>Statistical Quality Control</i>	Statistische kwaliteitscontrole: die aspecten van de kwaliteitscontrole waarbij statistiek wordt toegepast, in tegenstelling tot het brede scala van kwaliteitscontrole die uit veel meer procedures bestaat (preventief onderhoud, apparaatcontrole, validatie van bepalingen). Statistische kwaliteitscontrole wordt vaak gebruikt om routinematig de kwaliteit van een methode te bewaken en een verandering in kwaliteit te signaleren.
Standaarddeviatie (SD)	Een maat voor de spreiding van de meetwaarden rond hun gemiddelde, uitgaande van een normale verdeling.
Systematische meetfout, <i>bias</i>	Gemiddelde van een oneindig aantal herhaalde metingen van een te meten stof, verminderd met de werkelijke waarde van de te meten grootte.
TE, <i>Total Error</i>	Totale meetfout: het netto- of gecombineerde effect van toevallige en systematische meetfouten.
TE _a , <i>allowable Total Error</i>	Tolerantiegrens: een analytische kwaliteitseis die aangeeft welke toevallige fout (<i>imprecision</i>) en welke systematische fout (<i>inaccuracy</i>) nog aanvaardbaar is in een enkele meting of een resultaat van een meting.
Toevallige meetfout, <i>random error</i>	Het verschil tussen een meetresultaat en het gemiddelde van een oneindig aantal herhaalde metingen van een te meten stof.

NOTEN

¹ **Bias – externe kwaliteitscontrole**

Bias en imprecisie zijn beide van belang bij externe kwaliteitscontrole. Met de Sigma score wordt een verband gelegd met de klinische relevantie van gevonden afwijkingen. In de MUSE-rapportage van de SKML wordt steeds een afzonderlijke testuitkomst met een Sigma score uitgedrukt. Deze score is gebaseerd op de conventionele *total error* berekening uitgaande van de biologische variatie. De Sigma score wordt berekend op basis van zowel *bias* als imprecisie. Heeft de bepaling in het algemeen een lage Sigma score, dan wordt de score berekend op basis van de *State of the Art*. Bij MUSE wordt met een Sigma score aangegeven, hoe een testuitslag zich verhoudt tot de klinisch noodzakelijke nauwkeurigheid. Het is niet aan te bevelen om de methode van interne kwaliteitscontrole te baseren op de uitkomsten van externe kwaliteitscontrole. Van belang is hier of de richtwaarde van het externe programma overeen met die van het laboratorium (en eigen referentiewaarden). Externe kwaliteitscontrole zal eerder leiden tot aanpassing van de bepalingsmethode of aanpassing (herijking) van de methode, maar niet tot aanpassing van de interne kwaliteitscontroleregels. De IQC wordt bij voorkeur ontworpen op basis van de imprecisie. De *bias* speelt geen rol. Aangezien de Sigma score in de SKML EQA mede gebaseerd is op de *bias*, leent deze score zich volgens de werkgroep niet, om de IQC mee te ontwerpen.

² **Bias – valideren van een bepaling**

Dit geldt wanneer men bijvoorbeeld wil beoordelen of een bepaling goed genoeg is voor de klinische toepassing waarvoor deze bedoeld is. Deze situatie doet zich voor bij introductie (bij validatie) van een bepaling. In dit geval dienen we de onnauwkeurigheid van de bepaling te meten. Dit gebeurt doorgaans met kwaliteitscontrole monsters en gedurende een langere periode, zodat variaties door reagens wisselingen en onderhoud van apparatuur ook in rekening worden gebracht. *Bias* wordt in dit geval echter niet opgenomen in de berekening. Het laboratorium kan ook eigen referentiewaarden vaststellen, waarmee de (methodeafhankelijke) *bias* op een andere manier wordt ingecalculeerd. *Bias* zal doorgaans geen invloed op de afweging hebben of een bepaling goed genoeg is om in de praktijk toegepast te worden; de imprecisie is daarvoor juist wel belangrijk. De Sigma score is: $(1,65 * 0,5 * CV_i) / CV_a$. Merk op dat volgens de Stockholm-consensus een bepaling aan het

volgende kwaliteitscriterium dient te voldoen: *desirable*: imprecisie $< 0,5 * CV_i$. Als een bepaling hieraan precies voldoet, geldt: $CV_a = 0,5 * CV_i$. De Sigma score is in dat geval: $(1,65 * 0,5 * CV_i) / (0,5 * CV_i) = \underline{1,65}$. Als de Sigma score groter dan drie is, kan de bepaling met IQC worden bewaakt. Als de score lager dan drie is, is de kwaliteit niet zonder meer voldoende om met IQC te borgen. Deze lagere score kan echter overeenkomen met de *State of the Art*.

³ TE_a (*allowable Total Error*)

$$TE_a = 0,25 * (CV_i^2 + CV_g^2)^{0,5} + 1,65 * 0,5 * CV_i \quad (\text{vergelijking 5})$$

Stel $CV_i^2 \approx CV_g^2$

en $bias = 0$

$$\begin{aligned} TE_a &= 0,25 * (2 * CV_i^2)^{0,5} + 1,65 * 0,5 * CV_i \\ &= 0,35 * CV_i + 0,825 * CV_i \\ &= 1,18 * CV_i \end{aligned}$$

$$\text{Sigma score} = TE_a / CV_a = 1,18 * CV_i / 0,5 * CV_i = 2,36$$

Meestal zal CV_g groter zijn dan CV_i . Als de verhouding CV_i / CV_g niet gelijk is aan 1, maar kleiner, dan blijft de Sigma score echter kleiner dan 3 totdat $CV_g > 6 * CV_i$ (bij $CV_g = 6 * CV_i$ is de Sigma score 2,98).

⁴ Westgardregels

In de oorspronkelijke publicatie van Westgard *et al.* [1981] werd de combinatie van bewakingsregels toegepast, als aan het ingangscriterium was voldaan, namelijk overtreding van de 1_{2s} regel. In latere versies van de bewakingsregels is deze voorwaardelijke regel vervallen, en worden powerfuncties berekend op basis van een combinatie van bewakingsregels zonder voorwaardelijk criterium. De meest gebruikte Westgardregels zijn:

$1_{2s}, 1_{3s}$ de tolerantie limiet wordt overschreden bij ± 2 (of 3) SD

2_{2s} overschrijding door 2 achtereenvolgende controle metingen plus 2 SD of min 2 SD

R_{4s} overschrijding van 1 controle meting plus 2 SD en een naastliggende min 2 SD

4_{1s} 4 achtereenvolgende controle metingen zijn meer dan plus 1 SD of min 1 SD

⁵ GUM – meetonzekerheid

Occasionally, one may find that a known correction b for a systematic effect has not been applied to the reported result of a measurement, but instead an attempt is made to take the effect into account by enlarging the “uncertainty” assigned to the result. This should be avoided; only in very special circumstances should corrections for known significant systematic effects not be applied to the result of a measurement [...]. Evaluating the uncertainty of a measurement result should not be confused with assigning a safety limit to some quantity.

⁶ z-waarde

Het is aangetoond dat de z -waarde afhankelijk is van de verhouding van *bias* en imprecisie. Is de *bias* groot ten opzichte van de imprecisie, dan geldt de z -waarde van 1,65. Deze waarde is 1,96 (tweezijdige limiet, 95%) wanneer de *bias* te verwaarlozen is vergeleken met de imprecisie. Strikt genomen heeft z in tussenliggende gevallen een waarde tussen 1,65 en 1,95 [Stöckl en Thienpont, 2008].

Wanneer de werkelijke meetwaarde van glucose 9 mmol/L is, met een *bias* van 6% en een imprecisie van 2%, dan is de berekening van de TE:

$$TE = 0,06 * 9 + 1,65 * (0,02 * 9) = 0,54 + 0,30 = 0,84 \text{ mmol/L.}$$

In dit geval is de *bias* positief en aanzienlijk groter dan de imprecisie. De z -waarde is dan 1,65, omdat alleen de bovengrens van $9 + 0,84 = 9,84$ mmol/L relevant is (eenzijdig). Het is daarmee 95% zeker dat de werkelijke concentratie glucose de grens van 9,84 mmol/L niet zal overschrijden. De andere grens is $9 + 0,54 - 0,30 = 9,24$ mmol/L. Er is 5% kans dat deze grens wordt overschreden, want dat zou de uitslag alleen dichterbij de werkelijke waarde brengen.

Als daarentegen de *bias* laag is, dient de z -waarde van 1,96 toegepast te worden. In ons voorbeeld, met *bias* gelijk aan nul, zou de werkelijke glucosewaarde met 95% zekerheid vallen binnen de grenzen $9 \pm (1,96 * 0,02 * 9) = 9 \pm 0,35$ mmol/L.

Het voorbeeld is misschien niet realistisch: in de praktijk zal een *bias*, die aanzienlijk groter is dan de imprecisie van een bepaling, in het algemeen onacceptabel zijn. In het algemeen wordt om pragmatische redenen de z-waarde van 1,65 aangehouden.

7

In de studie van Fraser en Petersen [1993] werd voorgesteld om de maximale grens voor de combinatie van imprecisie en *bias* op te stellen, voor toepassing in externe kwaliteitscontrole programma's. Een theoretische onderbouwing werd echter niet gegeven:

There are many data on within-subject (C_i) and between-subject (CV_g) biological variation which allow generation of quality goals for EQAS as:

allowable imprecision $< \frac{1}{2} CV_i$

allowable inaccuracy $< \frac{1}{4} (CV_i^2 + CV_g^2)^{\frac{1}{2}}$

These specifications should be fulfilled separately and EQAS appropriately designed. However, when only a single determination of each survey material is used or allowed, the 95% acceptance range (for total error) for each laboratory from the target value is:

$\pm [1.65(\frac{1}{2} CV_i) + \frac{1}{4} (CV_i^2 + CV_g^2)^{\frac{1}{2}}]$

8

Merk op dat deze resultaten in redelijke overeenstemming zijn met de Stockholm-consensus, met een maximum *bias* van $0,25 * (CV_i^2 + CV_g^2)^{0,5}$

Wat betreft de imprecisie, vergelijk $0,59 * (CV_b)$ with $0,5 * CV_i$; het model van Gowans *et al.* maakt echter geen onderscheid tussen monitoring en diagnose: alleen de totale biologische variatie wordt vermeld. Voor diagnostiek (met referentiewaarden gebaseerd op CV_i and CV_g) wordt TE_a :

$$TE_a = 0,275 * (CV_i^2 + CV_g^2)^{0,5} + 1,65 * 0,59 * (CV_i^2 + CV_g^2)^{0,5}$$

Bij monitoren wordt alleen CV_i in rekening gebracht:

$$TE_a = 0,275 * (CV_i) + 1,65 * 0,59 * (CV_i)$$

Er zijn bepalingen waarvoor de biologische variatie laag is, en de analytische variatie in verhouding hoog (bijvoorbeeld: natrium, chloor). Voor deze bepalingen geldt, dat de TE_a laag is, waarmee de eisen die worden gesteld aan de analytische kwaliteit soms onrealistisch hoog worden. Het model van Gowans gaat uit van referentiewaarden, die alleen gebaseerd zijn op biologische variatie. In werkelijkheid worden de referentiewaarden echter mede bepaald door analytische variatie. Door het model hiervoor aan te passen, worden er realistische waarden voor TE_a verkregen [Oosterhuis, in bewerking].

⁹ Nauwkeurigheid

“Nauwkeurigheid” is een kwalitatief begrip.

10

De term “precisie” mag niet worden gebruikt voor “nauwkeurigheid”.

11

De precisie hangt alleen af van de verdeling van toevallige fouten en staat niet in relatie tot de werkelijke waarde of de gespecificeerde waarde.

12

De mate van precisie wordt meestal uitgedrukt in termen van imprecisie en berekend als standaardafwijking van de testresultaten. Minder precisie betekent een grotere standaarddeviatie.

13

“Onafhankelijke testresultaten” betekent dat de resultaten verkregen zijn op een wijze die niet beïnvloed is door eerdere resultaten van dezelfde of een vergelijkbare bepaling.