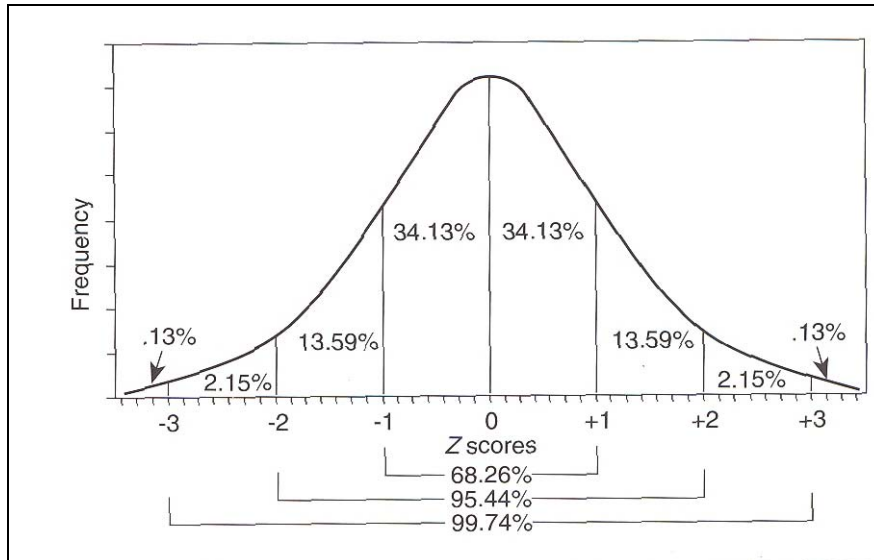


Kwaliteitscontrole door middel van Biologisch ijken

Patrick Jak (PMC.Jak@Vumc.nl) en Herman Groepenhoff (H.Groepenhoff@vumc.nl)
VU Medisch Centrum, Amsterdam.

Een belangrijk hulpmiddel bij biologisch ijken is het statistische gegeven dat in een stabiele laboratoriumomgeving herhaalde longfunctie metingen bij een gezonde proefpersoon (PP) normaal verdeeld zijn. [1] (Figuur 1.)



Figuur 1: Normale verdeling

Bij een normale verdeling geldt dat ongeveer:

1. 65 % van de metingen binnen ± 1 standaarddeviatie (SD) van het gemiddelde (M) valt
2. 95 % van de metingen binnen ± 2 SD van M valt
3. 99 % van de metingen binnen ± 3 SD van M valt

Westgard regels

J.O. Westgard heeft uitgebreid onderzoek gedaan naar kwaliteitscontrole binnen het klinisch chemisch laboratorium. Dit heeft ondermeer geresulteerd in “Westgard” regels [2] welke het mogelijk maken door middel van biologische ijken waarden de kwaliteit van de meetuitslagen te beoordelen.

- A. Eén meting buiten de ± 2 SD grenzen valt resulteert in een **“WARNING”** situatie
- B. Eén meting buiten de ± 3 SD grenzen valt resulteert in een **“OUT OF CONTROL”** situatie
- C. Twee op één volgende metingen buiten de ± 2 SD grenzen resulteert eveneens in een **“OUT OF CONTROL”** situatie
- D. Vier op elkaar volgende metingen buiten de ± 1 SD grens aan dezelfde kant van M resulteert eveneens in een **“OUT OF CONTROL”** situatie
- E. Tien op elkaar volgende metingen aan dezelfde kant van M vallen resulteert eveneens in een **“OUT OF CONTROL”** situatie

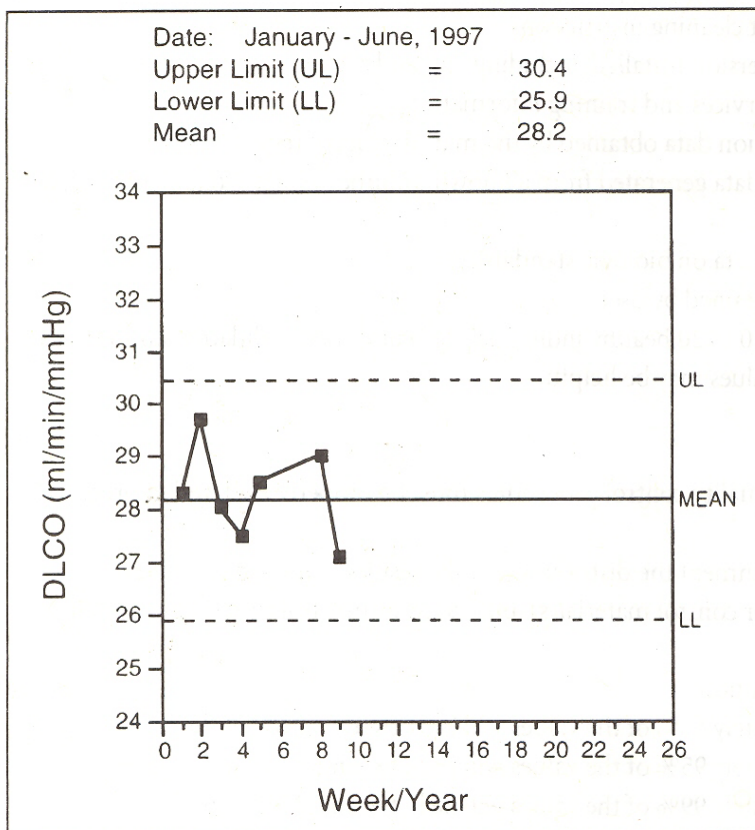
Bij een “**WARNING**” situatie moet de meting herhaald worden (wanneer mogelijk door een andere PP). Bij een “**OUT OF CONTROL**” situatie wijken de waarden zodanig af dat de meetopstelling gecontroleerd moet worden (longfunctie assistent, technicus ziekenhuis en/of firma).

Hoe te beginnen met de biologische ijk?

Allereerst moet er iemand worden geselecteerd die de bioijk procedure gaat uitvoeren. Dit moet vanzelfsprekend een gezonde PP zijn met een goede “longfunctie” techniek die voor een langere periode beschikbaar is om herhaaldelijk (wekelijks) de gewenste testen uit te voeren. Deze persoon moet dan binnen korte tijd (< 2 weken) gegevens verzamelen van minimaal 6 herhaalde metingen van waaruit een betrouwbaar gemiddelde en SD bepaald kan worden. Hoe meer metingen hoe betrouwbaarder het gemiddelde! Waarna de bioijk grafiek (Figuur 2.) opgemaakt kan worden met het gemiddelde en de $\pm 2SD$ grenzen van de herhaalde longfunctie meting op de Y-as en de tijd in dagen of weken op de X-as. Vervolgens kunnen dan de wekelijks gemeten waarden in de grafiek worden geplaatst. Waardoor de beoordeling volgens “Westgard” mogelijk wordt.

Eén bioijk, één meetinstrument en één longfunctie meting

Boven beschreven theorie is makkelijk toepasbaar wanneer er maar één meetinstrument voor één soort longfunctie meting met één PP als bioijk. In Figuur 2. is deze situatie uitgewerkt voor één diffusie meetopstelling met één PP.



Figuur 2: Bio-ijk grafiek

(UL “upper limit” en LL “lower limit” staan voor de $\pm 2SD$ grenzen)

De praktijk: meerdere proefpersonen, meerdere meetinstrumenten voor één longfunctie meting

In de klinische longfunctie praktijk komt het echter maar zeer zelden voor dat er maar één meetinstrument is voor één soort longfunctie meting. De meeste afdelingen hebben meerdere meetinstrumenten voor één zelfde longfunctie meting. Wat op zich nog niet een groot probleem zou opleveren mits de vooraf gemeten herhaalde metingen, exact hetzelfde gemiddelde en SD zouden opleveren voor de verschillende meetopstellingen. In dit zéér hypothetische voorbeeld kun je, omdat er maar één gemiddelde en één en dezelfde $\pm 2SD$ grenzen zijn, meerdere puntjes en/of lijntjes (voor de verschillende meetopstellingen) in één grafiek plaatsen. (*de praktijk is jammer genoeg niet zo ideaal, probeer het maar eens uit!*). Verder is het raadzaam om meerdere proefpersonen als bioijk aan te wijzen. De belangrijkste drie redenen hiervoor zijn:

1. Verbetering continuïteit. De PP kan onverwacht niet beschikbaar zijn door bv. ziekte, vakantie of andere werkplek en kan nu makkelijk vervangen worden.
2. Verbreding meetrange. Zeker wanneer verschillende personen van verschillende leeftijd, lengte en geslacht geïnccludeerd worden.
3. Minder tijdbelasting per proefpersoon. Niet meer wekelijks aan de beurt (maar 1/n weken, waarbij n = aantal PP), maar eens per zoveel weken, afhankelijk van het aantal proefpersonen.

Dus naast het “probleem” van meerde meetinstrumenten is er nu ook het “probleem” van de meerdere proefpersonen (van verschillende lengte, leeftijd en geslacht) bij gekomen. De verschillende meetwaarden van deze verschillende personen krijg je met geen mogelijkheid meer in één grafiek en dat is wel het doel. Je wilt een overzichtelijke grafiek waarbij het mogelijk is de uitslagen, gebruikmakend van de “Westgard” regels, te beoordelen.

De oplossing:

Na de start, zoals eerder beschreven met het creëren van een database voor de verschillende proefpersonen voor de verschillende meetopstellingen, moet allereerst worden getoetst of er sprake is van een significant verschil tussen de verschillende meetopstellingen. Bij twee testopstellingen kan dit redelijk eenvoudig dmv. de gepaarde T-test (Excel). Bij drie of meer opstellingen moet je de herhaalde metingen variantie analyse (ANOVA) gebruiken. Dit kan met behulp van elk statistisch software pakket (bv. Graphpad Prism of SPSS) meestal wel beschikbaar zijn het ziekenhuis (zo niet email sturen, zeker niet laten afschrikken!!). Wanneer er een significant verschil gevonden wordt dan moet, zeker wanneer dit ook een klinisch relevant verschil is, dit eerst worden verholpen. Een (heel) klein verschil kan eventueel geaccepteerd worden als zijnde niet klinisch relevant. Deze beslissing moet je echter zelf nemen. De metingen worden dan beschouwd als komende vanuit één populatie. Waarna het overall gemiddelde en SD van al de meetopstellingen per proefpersoon kunnen worden berekend (Excel). Waarmee de herhaalde wekelijkse bioijk uitslagen kunnen worden gestandaardiseerd. Standaardisatie houd in dat de meetuitslagen worden uitgedrukt in aantal SD afwijkingen van het vooraf bepaalde overall gemiddelde.

$$\text{meetuitslag}_{\text{gestandaardiseerd}} = (\text{meetuitslag} - \text{overall gemiddelde}) / \text{SD}$$

waardoor de meetuitslagen van de verschillende proefpersonen vergelijkbaar worden en dus in één grafiek geplaatst kunnen worden.

Reken voorbeeld:

- 3 diffusie meet opstellingen (tlco1, tlco2 en tlco3)
- 2 proefpersonen (PP 1 en PP 2)

Tabel 1: Fictieve diffusie uitslagen van drie meetssystemen van proefpersonen 1 en 2.

PP	tlco1	tlco2	tlco3
1	13,1	13,3	13,1
1	13,7	13,6	13,2
1	13,2	13,6	13,1
1	13,4	13,6	13,1
1	13,8	13,5	13,8
1	13,5	13,5	13,7
1	13,1	13	13,3
1	13,8	13,8	13,3
1	13,4	13,2	13,4
1	13,6	13,8	13
2	6,2	6,4	6,3
2	6,5	6,3	6,2
2	6,5	6,8	6,5
2	6,1	6,8	6,5
2	7,1	6,7	6,6
2	6,5	6,6	6,7
2	6,7	6,3	6,1
2	6,7	6,1	6,9
2	6,3	6,8	6,9
2	6,3	6,7	6,7

Tabel 1. laat de 10 herhaalde diffusie metingen van beide proefpersonen zien. PP 1 heeft een gemiddeld hogere waarde dan PP 2 (13.41 ± 0.268 en 6.53 ± 0.268 , respectievelijk) (Tabel 2.) Aan de voorwaarde om twee “verschillende” proefpersonen als bioijk te includeren is dus prima voldaan. Allereerst willen we toetsen of er een significant verschil in uitkomsten tussen de meetopstellingen aanwezig is. Daar we meer dan twee diffusieopstellingen willen vergelijken, kan dit niet meer met de gepaarde T-test maar, moet dit via een herhaalde metingen ANOVA getoetst worden. In ons voorbeeld wordt een *p*-waarde van 0.51 gevonden (tabel 3.) wat duidt op géén significant verschil (*gelukkig!*, *maakt het voorbeeld wel zo makkelijk*). De drie verschillende opstellingen kunnen dus als niet verschillend beschouwd worden en dat is wat je wilt, starten met niet verschillende uitkomsten op de verschillende systemen.

Tabel 2: Gemiddelde diffusiewaarden proefpersoon 1 en 2.

N	30	N	30
Mean	13,42	Mean	6,53
Median	13,40	Median	6,50
S td. Deviation	,268	Std. Deviation	,266
TLCO PP1		TLCO PP2	

Tabel 3: ANOVA uitslag. (SS = totale variantie, df = vrijheidsgraden, MS = gemiddelde variantie, F = MS tussen meetsystemen / MS error)

	SS	df	MS	F	P
<i>Tussen meetsystemen</i>	0,1003	2	0,050	0,69	0,51
<i>Tussen personen</i>	712,1	1	712,1		
<i>Error</i>	4,04	56	0,072		

Nu moet voor iedere proefpersoon afzonderlijk het overall gemiddelde en SD berekend worden. Het gemiddelde van de 30 metingen van PP1 is, zoals we al eerder hadden gezien in tabel 2, 13.42 (het kleine verschil met de mediaan van 13.4 duidt op een mooi symmetrische verdeling van de gemeten waarden wat eigenlijk een voorwaarde is voor de gebruikte statistische testen) met een SD van 0.268. PP2 laat een gemiddelde van 6.53 met een SD van even eens een SD 0.266 zien.

Waarna de wekelijkse controle door PP1 en PP2 kan starten.

In week 1 is eerst en alleen PP1 aan de beurt en voert (binnen één of twee dagen) op alle drie de diffusie opstellingen een diffusie meting uit met de volgende meetuitslagen:

1. tlco1 = 13.8
2. tlco2 = 13.1
3. tlco3 = 13.5

De meetuitslagen moeten nu worden gestandaardiseerd mbv het vooraf bepaalde overall gemiddelde en SD.

$$\text{tlco}_{\text{gestandaardiseerd}} = \text{tlco}_{\text{gemeten}} - \text{overall gemiddelde} / \text{SD}$$

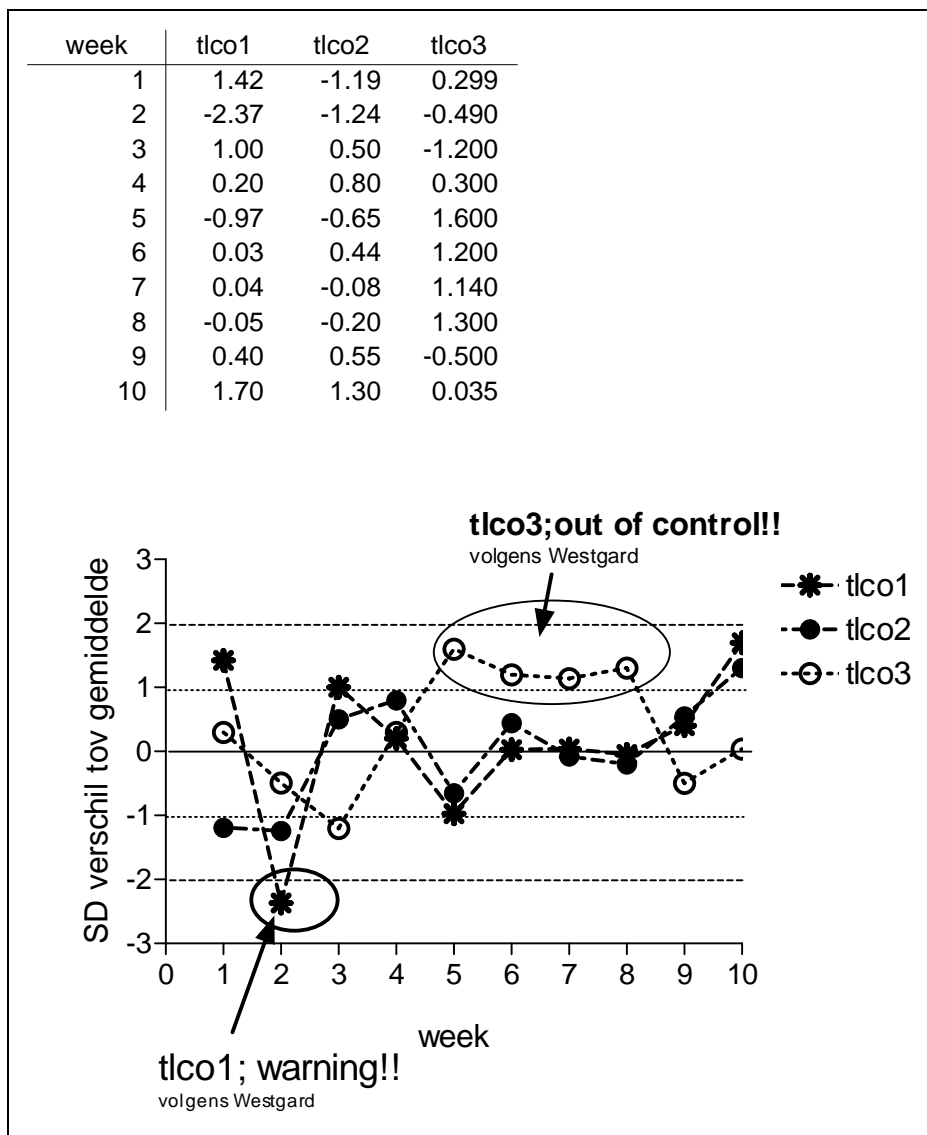
wat resulteert dit in de volgende getallen:

1. $\text{tlco}_{\text{gestandaardiseerd}1} = (13.8 - 13.42) / 0.268 = + 1.42$ de gemeten TLCO ligt dus 1.42 x SD boven het vooraf gemeten gemiddelde van PP1. (Is dit een waarde die je statistisch gezien wel of niet verwacht? Ja, immers 95% van de waarden zal zich bevinden tussen $\pm 2\text{SD}$)
2. $\text{tlco}_{\text{gestandaardiseerd}2} = (13.1 - 13.42) / 0.268 = - 1.19$
3. $\text{tlco}_{\text{gestandaardiseerd}3} = (13.5 - 13.42) / 0.268 = + 0.30$

In week 2 is PP2 aan de beurt en deze meet bij zichzelf de volgende 3 waarden:

1. tlco1 = 5.9; $\text{tlco}_{\text{gestandaardiseerd}1} = (5.9 - 6.53) / 0.266 = - 2.37$
2. tlco2 = 6.2; $\text{tlco}_{\text{gestandaardiseerd}2} = (6.2 - 6.53) / 0.266 = - 1.24$
3. tlco3 = 6.4; $\text{tlco}_{\text{gestandaardiseerd}3} = (6.4 - 6.53) / 0.266 = - 0.49$

Om de week is nu PP 1 of PP 2 aan de beurt. De gestandaardiseerde waarden van de drie diffusie meetopstellingen van beide proefpersonen nu wekelijks grafisch weergeven in één grafiek, waardoor het mogelijk is wordt om de waarden van meerdere proefpersonen volgens de Westgard regels te beoordelen (Figuur 3.).



Figuur 3. Biologische ijk grafiek over 10 weken van twee proefpersonen op drie diffusie systemen (tlco1,tlco2 en tlco3).

Figuur 3 laat nu één grafiek zien met daarin opgenomen de gestandaardiseerde data van de twee proefpersonen gemeten op drie meetopstellingen voor één longfunctie test. Voor iedere andere longfunctietest is natuurlijk wel een aparte grafiek nodig.

Klinisch relevant?

Bedenk wel dat de verschillen in grafiek niets zeggen over absolute verschillen in meetuitslagen tussen de meetopstellingen. Wil je beoordelen of een significant verschil ook een klinisch relevant verschil is dan moet je terug naar je database met de ruwe data. Van waar uit je dan de overall gemiddelden van de verschillende meetopstellingen met elkaar kunt vergelijken.

Tabel 4. Fictieve diffusie data van de eerste 20 weken op drie diffusie systemen.

bio-ijk	tlco1	tlco2	tlco3
1 (wk1)	13,1	13,3	13,1
1 (wk3)	13,7	13,6	13,2
1 (wk5)	13,2	13,6	13,1
1 (wk7)	13,4	13,6	13,1
1 (wk9)	13,8	13,5	13,8
1(wk11)	13,5	13,5	13,7
1(wk13)	13,1	13	13,3
1(wk15)	13,8	13,8	13,3
1(wk17)	13,4	13,2	13,4
1(wk19)	13,6	13,8	13
2 (wk2)	6,2	6,4	6,3
2 (wk4)	6,5	6,3	6,2
2 (wk6)	6,5	6,8	6,5
2 (wk8)	6,1	6,8	6,5
2(wk10)	7,1	6,7	6,6
2(wk12)	6,5	6,6	6,7
2(wk14)	6,7	6,3	6,1
2(wk16)	6,7	6,1	6,9
2(wk18)	6,3	6,8	6,9
2(wk20)	6,3	6,7	6,7

Hiervoor gaan we naar de dataset met de waarden van de eerste 20 weken (Tabel 4.) Tabel 5 laat zien dat de absolute gemiddelden erg dicht bij elkaar liggen, géén klinisch relevant verschil lijkt mij. Wanneer je wilt toetsen of er een significant verschil is dan gaat dit volgens dezelfde procedure als eerder genoemd via een gepaarde T-test bij twee opstellingen of anders de ANOVA.

Tabel 5. Overall gemiddelde diffusiewaarde van de drie opstellingen.

	TLCO 1	TLCO 2	TLCO 3
N	20	20	20
Mean	9,98	10,02	9,92
Std. Deviation	3,586	3,569	3,478

Variantie coëfficiënt (CV)

Uit de ruwe data is ook de variatie coëfficiënt (CV) te berekenen. De CV is een belangrijke maat voor de precisie van de metingen. De CV is de SD gedeeld door het gemiddelde. Het misschien verleidelijk om de CV ook uit de overall data output van tabel 5. te halen maar dat is niet correct!! Waarom niet ?

In de SD van de overall data zit ook de variantie tussen de twee proefpersonen en deze was erg groot (Tabel 2 en 3). Voor de berekening van de CV moet het gemiddelden en SD van de afzonderlijke proefpersonen gebruikt worden.(Tabel. 6) Waarna het vaststellen van de CV voor de verschillende TLCO opstellingen een fluitje van een cent is geworden.

Tabel 6: Gemiddelde diffusie waarden per opstelling per proefpersoon.

	TLCO 1	TLCO 2	TLCO 3	PP1
N	10	10	10	
Mean	13,46	13,49	13,30	
Std. Deviation	,2675	,2558	,2667	

	TLCO 1	TLCO 2	TLCO 3	PP2
N	10	10	10	
Mean	6,49	6,55	6,54	
Std. Deviation	,2923	,2550	,2757	

PP1

1. CV Tlco1 = $.2675 / 13.46 = \mathbf{0.020}$
2. CV Tlco2 = $.2558 / 13.49 = \mathbf{0.019}$
3. CV Tlco3 = $.2667 / 13.30 = \mathbf{0.020}$

PP2

1. CV Tlco1 = $.2923 / 6.49 = \mathbf{0.045}$
2. CV Tlco2 = $.2550 / 6.55 = \mathbf{0.038}$
3. CV Tlco3 = $.2757 / 6.54 = \mathbf{0.042}$

Hier dan weer de gemiddelde CV per opstelling uitrekenen:

1. CVTlco1 gemiddeld = $(0.020 + 0.045) / 2 = \mathbf{0.03}$
2. CVTlco2 gemiddeld = $(0.019 + 0.038) / 2 = \mathbf{0.03}$
3. CVTlco3 gemiddeld = $(0.020 + 0.042) / 2 = \mathbf{0.03}$

Kortom géén verschil in CV tussen de verschillende diffusie opstellingen.

(nogmaals de dataset bevat fictieve waarden, voor diffusie metingen is een CV van 0.05 heel netjes[3]) In de praktijk zou het heel zinvol zijn de CV waarden van verschillende longfunctie afdelingen voor de verschillende longfunctie testen met elkaar te vergelijken en met de waarden aanbevolen door de ERS.[3]

Belangrijk!

Het getallen voorbeeld was gebaseerd op twee proefpersonen en drie diffusie opstellingen, maar is natuurlijk van toepassing op iedere andere longfunctiemeting, waarbij het aantal systemen en aantal proefpersonen kan variëren. Probeer het maar!

(Vragen en op- of aanmerkingen graag naar H.Groepenhoff@VUMC.NL)

Referenties

- (1) J.Wanger, R.O.Crapo, C.G.Irvin: Pulmonary Function Laboratory Management and Procedure Manual; A project of the American Thoracic Society; Chapter 5: Quality Control.
- (2) Westgard JO, Groth T, Aronsson T, Falk H, de Verdier CH: Performance characteristics of rules for internal quality control: probabilities for false rejection and error detection. Clin Chem 1977;23:1857-1867.
- (3) Quanjer PH, Tammeling GJ, Cotes JE, Pedersen OF, Peslin R, Yernault JC: [Lung volumes and forced ventilatory flows. Work Group on Standardization of Respiratory Function Tests. European Community for Coal and Steel. Official position of the European Respiratory Society]. Rev Mal Respir 1994;11 Suppl 3:5-40.