

ASPECTEN VAN DE  
'GEMIDDELDE VAN NORMALEN'-METHODE  
IN DE KLINISCHE CHEMIE

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN  
DE GENEESKUNDE AAN DE ERASMUS UNIVERSITEIT  
TE ROTTERDAM OP GEZAG VAN DE RECTOR MAGNI-  
FICUS PROF. DR. B. LEIJNSE EN VOLGENS BESLUIT  
VAN HET COLLEGE VAN DEKANEN. DE OPENBARE  
VERDEDIGING ZAL PLAATS VINDEN OP WOENSDAG  
6 OCTOBER 1976 DES NAMIDDAGS TE 4.15 UUR  
PRECIES

DOOR

JOHANNES ODO OLIVIER HOEKE

GEBOREN TE BANDOENG (INDONESIË)

DRUKKERIJ ELINKWIJK B.V. -UTRECHT

PROMOTOREN: PROF. DR. B. LEIJNSE  
DR. J. DE WAEL

CO-REFERENTEN: PROF. DR. J.B.J. SOONS  
R. VAN STRIK

## VOORWOORD.

Bij de voltooiing van dit proefschrift zou ik gaarne mijn erkentelijkheid betuigen aan allen, die op enigerlei wijze hebben bijgedragen aan het tót stand komen daarvan.

Dit geldt, uiteraard, in de eerste plaats mijn promotoren en daarnaast de beide co-referenten, die, ondanks drukke werkzaamheden, niet alleen tijd hebben vrijgemaakt om zich in mijn onderzoek te verdiepen, maar ook bereid bleken tot diepgaande discussies, die vaak meer omvatten dan inhoud en vorm van dit proefschrift.

Het was mij in de afgelopen periode mogelijk de benodigde tijd aan dit onderzoek te besteden, doordat drs. K.H. Peuker en Mej. C. Voorrips een aanzienlijk deel van mijn andere werkzaamheden voor hun rekening hebben genomen.

Prof.dr. B. Leijnse en dr. J. de Wael verleenden toestemming bepalingen-uitkomsten te verwerken afkomstig uit resp. de afdeling Klinische Chemie van het Academisch Ziekenhuis Rotterdam - Dijkzigt en het Laboratorium voor Klinische Biochemie van de Kliniek voor Inwendige Geneeskunde van het Academisch Ziekenhuis Utrecht. Dat bij de uitvoering van de bepalingen in beide laboratoria een groot aantal medewerkers betrokken is geweest behoeft geen betoog.

Bij het verzamelen van de Rotterdamse bepalingen-uitkomsten werd onmisbare hulp verleend door de dames L.Th. Freimann, A. van der Hulst-Casteleijn, H.A.G. Lagas-Smit en H.A. Roetering.

Voor de verwerking van de grote reeksen bepalingen-uitkomsten en voor een deel van de modelstudies kon gebruik worden gemaakt van computer-faciliteiten dank zij de enthousiast verleende steun van verscheidene medewerkers van de Centrale Computerafdeling (Hoofd: Ir. P.L. Schuerman) van het A.Z.U. en van de Computergroep (onder leiding van drs. H.N. Herbschleb en Ir. R. van Poelgeest) van de vakgroep Cardiologie

van de Rijksuniversiteit Utrecht.

Medewerkers van de bibliotheek waren, onder leiding van Mej. M.G. te Kolstê, behulpzaam bij het beschikbaar krijgen van de benodigde literatuur. De medewerkers van de reproductieafdeling (F.A. de Bruin en C.K.J. Lodder) zorgden voor vele copieën.

De figuren werden verzorgd door Mej. I. Jansen (illustratrice) en de Heer P. de Meier (fotograaf).

Het manuscript werd persklaar gemaakt door mijn vrouw, die daarmee echter niet haar belangrijkste bijdrage leverde aan het tot stand komen van mijn proefschrift.

Hen allen, maar ook velen die ik niet genoemd heb (het zij mij ver-  
geven) betuig ik mijn hartelijke dank.

## INHOUD

1.	INLEIDING	9
2.	LITERATUURONDERZOEK BETREFFENDE KWALITEITSCONTROLE	
2.1.	Inleiding	11
2.2.	Fouten in het klinisch chemisch laboratorium	12
	2.2.1. Eerste categorie	12
	2.2.2. Tweede categorie	15
	2.2.3. Derde categorie	16
2.3.	Methoden van kwaliteitscontrole	16
2.4.	De kwaliteitscontrole door middel van resultaten van het aangevraagde onderzoek	18
3.	LITERATUURONDERZOEK BETREFFENDE 'NORMALE WAARDEN'	
3.1.	Inleiding	31
3.2.	Problemen, die samenhangen met de termen 'normaal' en 'normale waarden'	32
3.3.	Problemen, die het begrip 'normale waarden' betreffen	32
3.4.	Problemen, die het vaststellen en het gebruik van 'normale waarden' betreffen	34
3.5.	De keuze van de statistische methode	41
4.	ONDERZOEK NAAR ENKELE EIGENSCHAPPEN VAN DE 'GEMIDDEIDE VAN NORMALEN'-METHODE	
4.1.	Inleiding	47
4.2.	Modelproeven	48
	4.2.1. Berekeningen uitgaande van de standaardnormale verdeling	48
	4.2.2. Steekproeven uit een tweezijdig afgekapte normale verdeling	49
	4.2.3. Steekproeven, die zijn samengesteld uit elementen van twee verschillende populaties	50

4.3.	Oriënterende toepassing op reeksen bepalings- uitkomsten	52
4.4.	Conclusie	54
5.	ONDERZOEK VAN ENKELE METHODEN VOOR DE SCHATTING VAN PARAMETERS	
5.1.	Inleiding	57
5.2.	Modelproeven	58
	5.2.1. Toepassing op modellen, die gebaseerd zijn op de standaardnormale verdeling	58
	5.2.2. Toepassing op grote steekproeven	61
5.3.	Toepassing op reeksen bepalingsuitkomsten	63
5.4.	De 'Best passende parabool'-methode	63
5.5.	Discussie en conclusie	65
6.	DE 'GEMIDDELDE VAN NORMALEN'-METHODE IN DE PRAKTIJK	
6.1.	Inleiding	68
6.2.	Verwerking van het materiaal en bespreking van de resultaten	69
6.3.	Slotbeschouwing	71
	Samenvatting	74
	Summary	77
	Literatuurlijst	80
	Appendix: Tabellen en Figuren	

## HOOFDSTUK 1.

### INLEIDING.

Zoals wel bekend is aan diegenen, die in de afgelopen jaren op enigerlei wijze betrokken zijn geweest bij de gang van zaken in de wereld van de klinische chemie, verkeerde een aantal laboratoria ongeveer tien jaar geleden in een situatie die tamelijk uitzichtloos leek. De hoeveelheid werk was sterk toegenomen (een verdubbeling in vijf jaar tijd), terwijl de mogelijkheden tot personele en/of ruimtelijke uitbreiding meestal zeer beperkt waren. Door deze ongunstige combinatie van factoren gedwongen was een begin gemaakt met de automatisering, of beter gezegd mechanisering, van een deel van het laboratoriumwerk. Het succes van deze maatregelen hield in dat steeds grotere aantallen aanvragen en dus ook uitslagen verwerkt moesten worden, waardoor de administratieve begeleiding van het aangevraagde onderzoek tot flessenhals dreigde te worden.

Deze dreiging was de achtergrond van ons streven naar invoering van de elektronische informatieverwerking in het laboratorium. Het ligt dan ook voor de hand dat wij bij het opstellen van het realisatieschema een hoge prioriteit hebben toegekend aan de on-line data-acquisitie en de administratieve begeleiding van het aangevraagde onderzoek. Het zal insiders niet verbazen dat het realiseren hiervan belangrijk meer tijd kostte dan was voorzien.

Tijdens de implementatie-periode van dit systeem werd het plan opgevat onderzoek te verrichten naar mogelijkheden de bruikbaarheid van klinisch-chemische uitkomsten te vergroten door verdergaande toepassing van de elektronische informatieverwerking. Hierbij werd gedacht aan:

- 1) statistische kwaliteitscontrole
- 2) het vaststellen van grenzen van 'normale gebieden'

- 3) het verzamelen van gegevens over de frekwentieverdelingen van pathologische uitkomsten
- 4) het analyseren van de patronen van de uitkomsten van groepen bepalingen.

Het in dit proefschrift beschreven onderzoek heeft betrekking op de eerste twee punten.

De doelstelling is in de eerste plaats inzicht te verkrijgen in de achtergronden van methoden voor kwaliteitscontrole, die gebruik maken van uitkomsten van aangevraagde bepalingen.



## HOOFDSTUK 2.

### LITERATUURONDERZOEK BETREFFENDE KWALITEITSCONTROLE.

#### 2.1. INLEIDING.

In klinisch-chemische kring wordt, vooral de laatste jaren, veel belang gehecht aan de kwaliteitscontrole in het laboratorium. Dit blijkt niet alleen uit aan dit onderwerp gewijde publicaties en voordrachten op klinisch-chemische bijeenkomsten, maar tevens uit diverse activiteiten in het kader van regionale of nationale kwaliteitscontrole (Tonks, 1963; Copeland e.a., 1967; Stamm & Büttner, 1969; Howarth e.a., 1973; Whitehead e.a., 1973; Statuten van de Stichting Kwaliteitsbewaking klinisch-chemische laboratoria, 1973).

Ter vermindering van misverstanden dient te worden opgemerkt, dat in dit verband de termen 'kwaliteitscontrole', 'Qualitätskontrolle', 'contrôle de qualité' en 'quality control' soms worden gebruikt in een betekenis, die ligt tussen de algemeen gangbare betekenis van de nederlandse, duitse en franse termen en die van de engelse. De term 'kwaliteitscontrole', en dit geldt ook voor de duitse en de franse term, duidt slechts aan: 'het nagaan of het produkt aan bepaalde kwaliteitseisen voldoet'. De term 'quality control' heeft daarentegen een ruimere betekenis, die in het nederlands weergegeven wordt door de term 'kwaliteitsbeheersing'. Zo moet volgens Copeland (1968) aan de term 'quality control' voor het gebied van de klinische chemie eigenlijk een zeer ruime betekenis worden toegekend:

"Quality control in its complete sense includes all efforts to produce precise and accurate measurement for medical diagnosis and therapy. Therefore, the quality of the training of the laboratory personnel, the quality of the laboratory instruments,

reagent chemicals, primary and secondary standards are important parts of total quality control. In addition, the laboratory space itself, lighting, the arrangement of work benches and other physical working arrangements will affect the quality of measurements. The morale and teamwork of the laboratory personnel is another important factor. From the glassware washer to the recording clerk, every person in the laboratory is important in the complex system which begins with the physician's request for a chemical measurement and ends with the recording of the result in the patient's record."

In publicaties over kwaliteitscontrole in de klinische chemie wordt hieronder echter, ook door angelsaksische auteurs, meestal verstaan de kwaliteitscontrole in de nederlandse betekenis van het woord of wel het geheel van maatregelen, die, binnen het kader van de normale laboratoriumwerkzaamheden, kunnen worden genomen om te bevorderen, dat de uitslagen aan de gestelde kwaliteitseisen blijven voldoen. De kwaliteitseisen houden in het algemeen gesproken in, dat de uitslagen zo min mogelijk zijn beïnvloed door fouten en dat, voorzover fouten niet vermeden konden worden, de onzekerheid over hun grootte zoveel mogelijk wordt beperkt.

## 2.2. FOUTEN IN HET KLINISCH CHEMISCH LABORATORIUM.

De fouten waarmee in het laboratorium rekening dient te worden gehouden zijn op uiteenlopende manieren in categorieën te verdelen. Whitby c.s. (1967) bespreken hiervan enkele voorbeelden. Veel gebruikt wordt de van C.F.Gauss afkomstige indeling, waarbij drie categorieën worden onderscheiden

- 1) toevallige fouten
- 2) systematische fouten
- 3) grove fouten

### 2.2.1 Eerste categorie.

In Nederland wordt de eerste categorie tegenwoordig bij voorkeur aangeduid met de term 'fluctuaties'. Het optreden van fluctuaties komt tot uiting in de variabiliteit van de resultaten wanneer bijvoorbeeld aan één object de meting van een, gedurende het experiment als constant te beschouwen, fysische grootte een groot aantal

malen wordt verricht. Meestal is er sprake van een zekere groepering van de resultaten rond een bepaalde waarde, waarbij kleine afwijkingen ten opzichte van deze waarde vaker voorkomen dan grote. Vrijwel algemeen wordt aangenomen, dat de frekwentieverdeling van dergelijke meetresultaten bij een voldoende groot aantal metingen nadert tot een verdeling volgens Gauss. De waarde waartoe het gemiddelde van de meetwaarden zou naderen, indien het aantal metingen zeer sterk zou worden opgevoerd, wordt aangeduid met de term 'meetverwachting'. Het gemiddelde van een korte reeks metingen kan als een schatting van de meetverwachting worden beschouwd.

Als maat voor de grootte van de fluctuaties kunnen worden gebruikt: de spreidingsbreedte, de standaardafwijking en de variatiecoëfficiënt. Voor uitspraken van kwalitatieve aard worden gebruikt de termen 'precisie', 'reproduceerbaarheid' en 'dupliceerbaarheid'. Voor de laatste drie geldt, dat ze beter worden naarmate de fluctuaties kleiner worden. Hoewel de grootte van de fluctuaties van meetwaarden kan worden beperkt door zorgvuldig werken en gebruik van goede hulpmiddelen, is het principieel onmogelijk deze geheel te elimineren. De precisie van een resultaat kan wel worden verbeterd door niet de uitkomst van een enkelvoudige meting als resultaat te beschouwen, maar het gemiddelde van de uitkomsten van een reeks onafhankelijke metingen.

Ook klinisch-chemische bepalingen, die meestal bestaan uit een reeks voorbereidende bewerkingen, met daarop aansluitend één of meer metingen en tenslotte een berekening van de concentratie, zijn onderhevig aan fluctuaties. Dit zal tot uiting komen in de variabiliteit van de uitkomsten wanneer de concentratie van een bestanddeel van een, als homogeen te beschouwen, materiaal in een groot aantal monsters van dat materiaal wordt bepaald met dezelfde bepalingmethode en onder vergelijkbare omstandigheden. In het algemeen wordt aangenomen, dat de frekwentieverdeling van de bepalinguitkomsten ook in dit geval tot een Gauss-verdeling nadert als het aantal bepalingen voldoende groot is (bijv. Caraway, 1971). Het blijkt echter dat experimentele gegevens hierover tamelijk schaars zijn. Door Richterich en Colombo (1962) werd de frekwentieverdeling bepaald van 100 uitkomsten van een glucosebepaling met een glucose-oxydase-methode en Büttner (1967) deed dit van 241 uitkomsten van een eiwitbepaling met een biureetmethode. Deze beide verdelingen lijken

in overeenstemming met bovengenoemde veronderstelling. Door Campbell en Owen (1969) en Büttner e.a. (1970) is er op gewezen, dat het aannemen van een verdeling volgens Gauss niet altijd juist is. Van de door Büttner bepaalde cumulatieve frekwentieverdelingen <sup>1)</sup> van 474 uitkomsten van ureumbepalingen (diacetylmonoxim-methode) en van 769 glucosebepalingen (kaliumferricyanide-micromethode) bleek die van de ureumuitkomsten wel een Gauss-verdeling te vertonen, maar die van de glucoseuitkomsten duidelijk niet (fig 2-1). Helaas vermeldt Büttner geen details over de wijze waarop de reeksen uitkomsten zijn verkregen, zodat elke poging tot verklaring van zijn bevindingen een speculatief karakter zou hebben. Daarom wil ik er slechts op wijzen, dat een reeks van 769 bepalingen, bij de gebruikelijke uitvoering (60 monsters per uur), zelfs zonder tussengevoegde controlemonsters, meer dan 12 uren vergt.

Owen en Campbell (1968-B) hebben de frekwentieverdeling nagegaan van de met een vijfkanals Autoanalyzer verkregen bepalinguitkomsten voor  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Cl}^-$ ,  $\text{HCO}_3^-$  en ureum. Verdeeld over een periode van zes maanden werden 420 monsters van een poolserum onderzocht, die in plastic buizen bij  $-10^{\circ}\text{C}$  werden bewaard. Voor elke bepalingsoort werden, zowel rechtstreeks als na voorafgaande logaritmische transformatie van de uitkomsten, het gemiddelde, de standaardafwijking en ook de grootheden  $g_1$  (als maat voor de scheefheid) en  $g_2$  (als maat voor de kurtose, of slankheid) berekend. De resultaten zijn samengevat in tabel 2-1. Het blijkt, dat de verdeling van de logaritmisch getransformeerde uitkomsten kleinere afwijkingen ten opzichte van een verdeling volgens Gauss vertoont dan de verdeling van de uitkomsten zelf, (de meetverwachtingen van  $g_1$  en  $g_2$  zijn voor een Gauss-verdeling gelijk aan nul).

In sommige gevallen is op grond van de bij de berekening gebruikte formule al niet te verwachten, dat de gemeten grootheid en de daaruit berekende concentratie beide een verdeling volgens Gauss kunnen bezitten. Dit geldt, in theorie, ook voor de in de klinische chemie gebruikelijke colorimetrische bepalingen. De grootheid, die hierbij wordt gemeten, is meestal een spanning (V), die evenredig is met de transmissie. De concentratie kan hieruit, indien althans de wet van

<sup>1)</sup> Voor een verdeling volgens Gauss liggen de punten van de cumulatieve frekwentieverdeling, uitgezet op z.g. waarschijnlijkheidspapier, op een rechte lijn.

Lambert geldt, worden berekend met de formule  $C = -k_2 \log k_1 V$ . Indien  $V$  een Gauss-verdeling bezit zal  $C$  een log-Gauss verdeling hebben.

### 2.2.2. De tweede categorie.

Tot de categorie van de systematische fouten worden fouten gerekend, die bestaan uit een verschuiving van alle betreffende waarnemingsresultaten in een zelfde richting.

Aangezien het onmogelijk is het optreden van fluctuaties bij onze waarneming uit te sluiten, moeten we bepalingsuitkomsten, die uitgaande van een homogeen materiaal zijn verkregen, beschouwen als een steekproef uit de populatie van alle bepalingsuitkomsten, die uitgaande van dat materiaal met dezelfde werkwijze verkregen zouden kunnen worden. Bij afwezigheid van andere fouten dan de onvermijdelijke fluctuaties zou de meetverwachting  $\mu_0$  gelijk zijn aan de gezochte waarde  $x_g$ . Een steekproefgemiddelde ( $\bar{x}_0$ ) is dan niet alleen als een schatter van  $\mu_0$ , maar ook als een schatter van  $x_g$  te beschouwen. Bij aanwezigheid van een systematische fout is een steekproefgemiddelde ( $\bar{x}_f$ ) een schatter van het gemiddelde  $\mu_f$  van de populatie van de door de systematische fout gemodificeerde uitkomsten. De grootte van een systematische fout zou in theorie het beste weergegeven kunnen worden door de afwijking van  $\mu_f$  ten opzichte van  $\mu_0$ . In de praktijk stuit dit echter op ten dele onoverkomelijke moeilijkheden. Zo is het in het algemeen niet mogelijk de populatiegemiddelden te bepalen, zodat genoeg zal moeten worden genomen met een schatting daarvan door middel van steekproefgemiddelden. Maar ook dan blijft nog het probleem, dat het zeer moeilijk is een betrouwbare waarde voor  $\bar{x}_0$  te verkrijgen omdat de aanwezigheid van een systematische fout niet altijd met zekerheid uit te sluiten is. Van sommige bepalingsmethoden is bekend dat hun systematische fout slechts gering is. Vaak zijn deze methoden echter te bewerkelijk voor het verwerken van grote series monsters en voor toepassing in de dagelijkse kwaliteitscontrole. Wel kunnen ze gebruikt worden als referentiemethode, bijvoorbeeld bij het testen van een nieuwe bepalingsmethode.

Naast de in de bepalingsmethode gelegen bronnen van systematische fouten kunnen als zodanig optreden afwijkingen in de gebruikte hulpmiddelen en onvolkomenheden in de uitvoering. De dagelijkse kwali-

teitscontrole zal zelfs voornamelijk op dergelijke ontsporingen gericht moeten zijn.

### 2.2.3. De derde categorie.

Tegen de gebruikelijke aanduiding van de derde categorie (grove fouten) is aan te voeren dat het niet altijd om grove fouten gaat. In de angelsaksische literatuur worden wel de termen 'non chemical errors' en 'clerical errors' gebruikt, maar ook deze aanduidingen omvatten niet alle fouten die naast de fluctuaties en systematische fouten een rol kunnen spelen. Daarom zou ik voor deze categorie de term 'incidentele fouten' willen gebruiken en daaronder fouten willen verstaan die slechts bij een (meestal gering) deel van de resultaten een rol spelen en waarvan het tijdstip van optreden, de grootte en de richting niet voorspelbaar zijn.

Als oorzaken kunnen worden genoemd:

- 1) onopgemerkte afwijkingen van de gebruikelijke procedure bij de monstervoorbereiding of bij de meting
- 2) kortdurende, niet opgemerkte storingen van meetinstrumenten
- 3) vergissingen bij het vastleggen of de verdere verwerking van de meetresultaten en andere op de betreffende monsters betrekking hebbende gegevens.

Door Mc Swiney en Woodrow (1969) en ook door Grannis e.a. (1971 en 1972) wordt de frekwentie van deze categorie fouten op ongeveer 3% geschat. Laatstgenoemde auteurs geven de volgende onderverdeling:

1) monsterverwisseling	0,89%
2) afleesfout	0,66%
3) rekenfout	0,60%
4) verkeerde reagentia	0,75%
5) andere oorzaken	0,56%

### 2.3. METHODEN VAN KALITEITSCONTROLE.

Informatie over de grootte van de fluctuaties op de bepalingsuitkomsten zou in principe kunnen worden verkregen door het aangevraagde onderzoek in meervoud en onafhankelijk van elkaar uit te voeren. Een bijkomend voordeel zou dan nog zijn, dat de gemiddelden van de bijbehorende uitkomsten betere schatters zijn van de meetverwachtingen dan de uitkomsten van enkelvoudige bepalingen. In de praktijk

blijkt het in meervoud analyseren vaak moeilijk als algemene regel door te voeren. Door Whitby c.s. (1967) wordt het in duplo uitvoeren van alle aangevraagde bepalingen zelfs niet aanbevelenswaardig geacht, vooral niet omdat in vele laboratoria de druk van het routinematige werk de kwaliteit reeds in ongunstige zin beïnvloedt. Het in duplo uitvoeren van een beperkt aantal bepalingen is naar hun mening voldoende om een indruk te krijgen van de invloed van de fluctuaties. Hierbij dient te worden aangetekend, dat de keuze en de verwerking van de duplo's eigenlijk 'blind' zou moeten gebeuren. Dat wil zeggen dat degene, die de bepalingen verricht, de duplo-monsters niet als zodanig kent of kan herkennen. Meestal blijft het in meervoud analyseren beperkt tot bijzondere gevallen, zoals bijvoorbeeld de voorbereiding van het invoeren van een nieuwe bepalingmethode.

Tot een dergelijke voorbereiding behoort ook een onderzoek naar de eventuele systematische fout. Hiervoor zou men onder meer de beschikking moeten hebben over ijkstandaarden en controleoplossingen, die aan hoge eisen voldoen en ook over een bepalingmethode, die geschikt is als referentiemethode (Radin, 1967; Young & Mears, 1968; Zief & Michelotti, 1971; Barnard e.a., 1971; Stamm, 1974). Hoewel er in de afgelopen jaren veel werk is verricht, zijn er op het gebied van de klinische chemie nog maar weinig mogelijkheden om zo'n onderzoek te realiseren (Stamm, 1974).

Zoals eerder werd gesteld, is de dagelijkse kwaliteitscontrole voornamelijk gericht op het aan het licht brengen van eventuele ontsporingen, die gepaard gaan met een (verandering van de) systematische fout en daardoor de vergelijkbaarheid van de door het laboratorium geleverde uitkomsten in ongunstige zin beïnvloeden.

In principe zijn hiervoor de volgende methoden beschikbaar:

- 1) Monsters uit een bepalingsserie ook analyseren m.b.v. een andere, onafhankelijke methode.
- 2) In een bepalingsserie monsters opnemen, waarvan een zo goed mogelijke schatting van de meetverwachting bekend is.
- 3) Statistische verwerking van de uitkomsten van het aangevraagde onderzoek.

De eerstgenoemde methode wordt weinig toegepast in de dagelijkse kwaliteitscontrole.

De tweede methode is daarentegen de meest gebruikelijke. Levey &

Jennings (1950) introduceerden de combinatie van deze methode met de toepassing van controlekaarten zoals die bij industriële kwaliteitscontrole reeds werden gebruikt. Volgens de door deze auteurs voorgestelde procedure worden, als regel twee maal per week, twee monsters van een controleserum in de bepalingsseries meegenomen. Als controlegrootheden dienen het gemiddelde en het verschil van de twee uitkomsten. Schattingen van de meetverwachtingen voor de concentraties van de verschillende bestanddelen worden verkregen door het gemiddelde te berekenen van de eerste twintig paren uitkomsten, die in ongeveer een maand werden verzameld. De controlegrootheden worden uitgezet in grafieken, waarin tevoren de schatting van de meetverwachting van de concentratie en de tolerantiegrenzen voor het gemiddelde en voor het verschil van de duplo's worden aangegeven. De grenzen liggen resp. op een afstand van  $\pm 1,88 \times \bar{R}$  van het gemiddelde en op  $0$  en  $3,27 \times \bar{R}$ , waarbij  $\bar{R}$  het gemiddelde is van de verschillen van de eerste twintig paren uitkomsten. De kans op overschrijding van de grenzen bedraagt volgens de auteurs voor het gemiddelde en voor de spreidingsbreedte ongeveer  $0,003$ <sup>1</sup>). Modificaties, die vaak door gebruikers van deze methode worden aangebracht, zijn:

- 1) Het opvoeren van de frekwentie, waarmee controlemonsters worden meegenomen.
- 2) Zodanige keuze van de tolerantiegrenzen, dat de overschrijdingskans ongeveer  $0,05$  bedraagt.

Op de derde methode zal in de volgende paragraaf worden ingegaan.

#### 2.4. DE KWALITEITSCONTROLE DOOR MIDDEL VAN RESULTATEN VAN HET AANGEVRAAGDE ONDERZOEK.

Ongeveer twintig jaar geleden introduceerden Waid en Hoffmann (1955) een methode voor kwaliteitscontrole, die gebruik maakte van de resultaten, die bij uitvoering van het aangevraagde onderzoek werden ver-

<sup>1</sup>) Deze overschrijdingkans geldt eigenlijk alleen indien het verdelingstype voldoende overeenstemming vertoont met een verdeling volgens Gauss. Dit is voor de spreidingsbreedte van duplo's zeker niet het geval. Indien rekening wordt gehouden met de scheefheid van de verdeling van de spreidingsbreedte, wordt voor de bovengrens daarvan een (éénzijdige) overschrijdingskans van ongeveer  $0,01$  gevonden (Beyer, 1968).



kregen. Deze methode berust op de veronderstelling, dat de gemiddelden van de dagelijks geproduceerde series bepalingen binnen betrekkelijk nauwe grenzen zullen vallen, zolang de bepaling niet ontspoord is. Toepassing van de methode komt neer op het dagelijks berekenen en grafisch uitzetten van het rekenkundig gemiddelde van elke serie bepalingen. Zoals voor de hand ligt, wordt deze methode aangeduid als 'daggemiddelde-methode'.

Enkele jaren later beschreven Hoffmann en Waid (1963) een tweede methode, die zij 'number plus-methode' noemden. Deze berust op de veronderstelling dat voor een gegeven bepaling de verhouding tussen het aantal van de uitkomsten, die groter zijn dan de modale waarde en het totale aantal uitkomsten nagenoeg constant is, zolang er geen sprake is van een afwijking. De uitvoering bestaat uit het tellen van het aantal uitkomsten, dat in een reeks van vijftig achtereenvolgende uitkomsten voldoet aan het criterium.

Een enigszins gemodificeerde 'daggemiddelde-methode' werd beschreven door Hoffmann en Waid (1965) onder de naam 'gemiddelde van normalen-methode' (GvN-methode). De modificatie bestaat daaruit, dat voor de berekening van het gemiddelde alleen die resultaten worden gebruikt, die binnen de grenzen van het normale gebied vallen. Voor de toepassing van deze methode dienen dus de grenzen van het normale gebied bekend te zijn. De auteurs geven aan dat die door middel van een computer of wel op grafische wijze (Hoffmann 1963) te bepalen zijn uitgaande van een frekwentieverdeling van enkele honderden ongeselecteerde bepalingen. Het midden van het normale gebied dient volgens de auteurs samen te vallen met een redelijke schatting van de modale waarde van de frekwentieverdeling. Een schatting van de standaardafwijking van de dagelijks te berekenen gemiddelden van de normale uitkomsten wordt verkregen door de bovengenoemde schatting van de standaardafwijking van de normale uitkomsten te delen door de wortel van het aantal normale uitkomsten, dat voor de berekening van het gemiddelde gebruikt is. Door de verkregen waarde met 1.96 te vermenigvuldigen en het resultaat bij het midden van het normale gebied op te tellen respectievelijk daarvan af te trekken kunnen de boven- en ondergrens van het tolerantiegebied van de gemiddelden worden bepaald.

Kilgariff en Owen (1968) onderzochten de gevoeligheid van de 'gemiddelde van normalen-methode' voor de bepalingen van natrium, kalium, chloride,  $\text{CO}_2$ , calcium, magnesium, ureum en urinezuur. De grenzen van het normale gebied werden echter niet op de door Hoffmann en Waïd aangegeven wijze bepaald, maar uitgaande van de bepalingssuitkomsten van ongeveer honderd bloedmonsters afkomstig van gezonde personen. Van de binnen deze grenzen vallende bepalingssuitkomsten van patienten werd het gemiddelde en de standaardafwijking berekend. Deze waarden werden gebruikt voor de berekening van de tolerantiegrenzen voor een gemiddelde van 10 bepalingssuitkomsten. De series patientenuitkomsten werden daarna voor zover mogelijk verdeeld in blokken van tien uitkomsten, waarvan de gemiddelden werden berekend en vervolgens werden vergeleken met de tolerantiegrenzen. Voor het onderzoek naar de gevoeligheid van de methode werden de patientenuitkomsten zodanig gemodificeerd (zie tabel 2-II), dat het optreden van een systematische fout werd gesimuleerd. Op de gemodificeerde uitkomsten werd nu weer dezelfde procedure toegepast. De resultaten waren eigenlijk alleen voor de natrium- en de chlorideuitkomsten duidelijk verschillend (ongeveer 50% afgekeurde gem.) van de resultaten verkregen met de ongemodificeerde uitkomsten (5-10% afgekeurde gem.). De auteurs concludeerden hieruit dat de 'gemiddelde van normalen-methode' te ongevoelig is en dat de voorkeur moet worden gegeven aan het meenemen van een aantal controlemonsters in iedere bepalingsserie.

Amador e.a. (1968) hebben de gevoeligheid van de door Waïd en Hoffmann (1955) en de beide door Hoffmann en Waïd (1963 en 1965) beschreven methoden vergeleken met die van de methode van Levey en Jennings (1950), die gebruik maakt van in de series opgenomen monsters van een controleserum.

De uitkomsten van de bepaling van glucose en van de GOT-activiteit in een poolserum, verricht op twintig achtereenvolgende dagen, werden uitgezet in een grafiek volgens Levey en Jennings. Als grenzen werden gekozen het gemiddelde  $\pm 2$  standaardafwijkingen, zoals berekend voor de twintig uitkomsten. Daarna werden verschillende series gemodificeerde uitkomsten berekend door elke uitkomst achtereenvolgens met 2.5%, 5%, 10% en 20% te verhogen of te verlagen. De gemodificeerde uitkomsten werden ook grafisch uitgezet, waarbij dezelfde grenzen

werden gebruikt als op de grafieken van de oorspronkelijke uitkomsten. Uit deze grafieken was de laagste modificatie te vinden, waarbij tenminste de helft van de punten buiten één van de grenzen lag en bijna alle resterende punten aan dezelfde kant van het gemiddelde lagen. Voor de glucosebepaling was dit  $\pm 5\%$  en voor de GOT-bepaling  $\pm 10\%$  (zie ook tabel 2-III). Als materiaal voor het testen van de drie andere methoden werden 1000 nuchtere bloedsuiker- en 1000 SGOT-uitkomsten uit dezelfde periode van twintig dagen gebruikt. Voor de toepassing van de eerste twee methoden werden beide series verdeeld in blokken van 50 uitkomsten, waarna deze volgens de voor die methoden aangegeven procedures werden verwerkt. Voor de toepassing van de GvN-methode werden uit beide series ongeveer 40 blokken van 10 'normale' uitkomsten gevormd, die daarna op de aangegeven wijze werden verwerkt. Voor de vergelijking van de gevoeligheid van de methoden werden de bloedsuikeruitkomsten nu gemodificeerd met  $+5\%$  en  $-5\%$  en de SGOT-uitkomsten met  $+10\%$  en  $-10\%$  en daarna weer verwerkt. De resultaten zijn samengevat in tabel 2-III. De auteurs concludeerden daaruit dat de op het gebruik van een referentiemonster gebaseerde methode van Levey & Jennings een veel gevoeliger middel is voor het aantonen van systematische fouten, dan de drie andere methoden, waarvan de gevoeligheid als onvoldoende werd gekwalificeerd.

Met betrekking tot de GvN-methode dient te worden opgemerkt, dat het aantal (10) uitkomsten dat voor de berekening van de gemiddelden werd gebruikt, hoewel niet in strijd met de aanwijzingen van Hoffmann en Waid, aan de lage kant is.

Owen en Campbell (1968-A) hebben een onderzoek verricht, waarbij de uitkomsten van de natrium-, kalium-, chloride-,  $\text{CO}_2$ - en ureumbepalingen gedurende een periode van drie maanden volgens verschillende procedures werden verwerkt. De uitkomsten werden per bepalingssoort gegroepeerd als dagserie en daarnaast in groepen van 25 en van 50 achtereenvolgende uitkomsten. Per groep werden berekend het gemiddelde van alle uitkomsten en een aantal gemiddelden van de uitkomsten die binnen op verschillende wijze vastgestelde grenzenparen lagen. Deze grenzenparen waren:

- 1) het gemiddelde  $\pm 2$  maal de standaardafwijking
- 2) het gemiddelde  $\pm 3$  maal de standaardafwijking

3) het 10<sup>e</sup> en het 90<sup>e</sup> percentiel van de groep

4) het 25<sup>e</sup> en het 75<sup>e</sup> percentiel van de groep

Tenslotte werd ook het percentage berekend van de uitkomsten, die binnen de onder 1) en 2) genoemde grenzenparen vielen.

Voor het onderzoek naar de gevoeligheid van de procedures werden de uitkomsten gemodificeerd door ze met 2%, 5% en 10% te verhogen of te verlagen, waarna ze weer volgens dezelfde procedure werden verwerkt. Bij de beoordeling en de onderlinge vergelijking van de gevoeligheid van de verschillende statistische grootheden maken de auteurs gebruik van drie begrippen, die als volgt worden omschreven:

1) de gevoeligheidsfactor (R), dit is de verhouding van de procentuele verschuiving van het gemiddelde tot de door de „modificatie gesimuleerde systematische fout.

2) een maat voor de doelmatigheid (efficiëntie), deze wordt verkregen door de standaardafwijking van het gemiddelde (SDM) te delen door de gevoeligheidsfactor (R). <sup>1)</sup>

3) de kleinste systematische fout, die met een waarschijnlijkheid van 0,95 kan worden aangetoond; deze is gelijk aan  $3,6 \times \frac{SDM}{R}$ . <sup>2)</sup>

Uit de resultaten is gebleken, dat berekening van het percentage van de uitkomsten die binnen één van de aangegeven grenzenparen vallen het minst bruikbaar was en dat de diverse gemiddelden onderling weinig verschil in bruikbaarheid vertoonden. Volgens de auteurs is de kleinste systematische fout, die met een waarschijnlijkheid van 0,95 kan worden aangetoond, voor de onderzochte methoden groter dan de overeenkomstige grootte bij het gebruik van een controleserum. Hierbij dient echter te worden opgemerkt, dat zij voor de berekening van de laatstgenoemde grootte gebruik hebben gemaakt van gemiddelden van in de literatuur vermelde standaardafwijkingen voor in één serie verrichte bepalingen en niet van de in hun eigen onderzoek verzamelde uitkomsten van een poolserum.

<sup>1)</sup> Deze formulering is ongelukkig gekozen, aangezien een grootte doelmatiger zal zijn naarmate de waarde van  $SDM/R$  kleiner is.

<sup>2)</sup> De factor 3,6 is opgebouwd uit de termen 2,0 (i.v.m. de afstand van de tolerantiegrenzen tot het gemiddelde) en 1,6 (i.v.m. de geëiste waarschijnlijkheid van 0,95 bij een éézijdige overschrijdingskans.

Dixon en Northam (1970) hebben een onderzoek verricht naar de waarde van de berekening van een daggemiddelde waarbij tevens de invloed van een aantal correcties werd onderzocht. In het onderzoek werd een groep bepalingen betrokken, waarvan dagelijks vrij grote series werden ingezet (gemiddelde aantal 85, spreiding 50 - 140) en die bestond uit natrium, kalium, ureum, totaal eiwit, albumine, calcium, alkalische fosfatase, bilirubine en aspartaat aminotransferase (GOT). Het merendeel van de uitkomsten, die werden verzameld, lag in het betreffende normale gebied. Om de invloed van sterk afwijkende uitkomsten op het daggemiddelde te elimineren werden voor zover nodig afkappingen gehanteerd (zie tabel 2-IV). Om de resultaten van de daggemiddelde methode te kunnen vergelijken met de resultaten van een andere controle methode, werden in elke serie bepalingen controlesera meegenomen.

Uit deelonderzoeken bleek de samenstelling van de groep patiënten, waarvan bloedmonsters werden ingestuurd, van dag tot dag verschillen te vertonen. Zo schommelde het percentage mannen van 35% tot 60% (gem. 46%) en het percentage poliklinische patiënten zelfs van 10% tot 75% (gem. 40%). Tevens werd nagegaan, hoe groot het verschil was tussen gemiddelden, die werden berekend voor een groep klinische en een groep poliklinische patiënten afzonderlijk en voor beide categorieën bij elkaar. Voor alle bepalingen bleken de verschillen significant te zijn. Op grond van deze resultaten werd besloten de daggemiddelden ook te berekenen volgens een procedure waarin een correctie werd opgenomen voor het variabele percentage poliklinische patiënten. Ter vergelijking van de variabiliteit van de aldus berekende met die van de ongecorrigeerde daggemiddelden, werd voor elke bepaling maandelijks het quotient van de standaardafwijking van de reeks ongecorrigeerde en die van de reeks gecorrigeerde gemiddelden berekend. Van deze quotienten werden, voor een periode van 13 maanden, de gemiddelden berekend (tabel 2-V). Hoewel de correctie in het algemeen wel tot een lagere standaardafwijking leidde, werd de verbetering slechts van belang geacht voor albumine, calcium en totaal eiwit. De vergelijking van de variabiliteit van de gecorrigeerde daggemiddelden met die van de uitkomsten van een controleserum werd op dezelfde wijze uitgevoerd (tabel 2-V).

Voor de beoordeling van een methode in het kader van de kwaliteitscontrole is naast de standaardafwijking als maat voor de variabili-

teit, ook de gevoeligheid van belang (zie ook Owen en Campbell, 1968-A). Voor de berekening van de kleinste systematische fout  $m$ , die met een waarschijnlijkheid  $P$  is aan te tonen, gebruiken de auteurs de formule  $m = SD \cdot t/R$  (vergelijk Owen en Campbell, 1968-A). De waarde van  $t$  is afhankelijk van  $P$  en kan worden opgezocht in een 'Students  $t$ ' tabel;  $R$  is de gevoeligheidsfactor. De vergelijking van twee methoden, in dit geval de berekening van het gecorrigeerde daggemiddelde en het meebepalen van een controleserum, kan worden uitgevoerd door middel van de 'relative power' (RP), die berekend kan worden met de formule  $RP = m_2/m_1 = SD_2 R_1 / SD_1 R_2$ . De index 1 slaat hierbij op de gecorrigeerde daggemiddelde methode en de index 2 op de methode met een controleserum. Uit de resultaten (tabel 2-VI) blijkt, volgens de auteurs, dat voor natrium, calcium en albumine de daggemiddelde methode de voorkeur verdient en voor ureum en aspartaat aminotransferase (GOT) de methode met een controleserum. Voor de resterende bepalingen verschillen de twee methoden weinig. Door Reed (1970) werden twee modificaties van de GvN-methode voorgesteld. De eerste modificatie houdt in dat ieder gemiddelde moet worden berekend uit een voor elke bepaling vast te stellen aantal uitkomsten. Hiermee wordt beoogd de vergelijkbaarheid van de gemiddelden te verbeteren. Opgemerkt dient te worden, dat de voorgestelde modificatie er toe zal leiden dat als regel slechts een betrekkelijk klein deel van de in aanmerking komende uitkomsten zal worden gebruikt voor de berekening van een gemiddelde.

De tweede modificatie is gebaseerd op het door de auteur verrichte onderzoek met behulp van mathematische modellen. Een populatie van bepalinguitkomsten werd hierbij opgevat als te zijn samengesteld uit drie, elkaar gedeeltelijk overlappende, deelpopulaties, die zelf een verdeling volgens Gauss bezitten:

- 1) de populatie van abnormaal lage uitkomsten
- 2) de populatie van normale uitkomsten
- 3) de populatie van abnormaal hoge uitkomsten.

De kansdichtheidsfunctie voor de sompopulatie is dan:

$$f_T(x) = \pi_1 f_1(x) + \pi_2 f_2(x) + \pi_3 f_3(x)$$

waarin  $\pi_1$ ,  $\pi_2$ , en  $\pi_3$  de verhouding aangeven van de bijdragen van de deelpopulaties en  $\pi_1 + \pi_2 + \pi_3 = 1$ . Bij de GvN-methode wordt gewerkt met bij de grenzen van het normale gebied afgekapte verdelingen. Als we de afkappgrenzen met  $a$  en  $b$  aanduiden, kan de betreffende kansdicht-

heidsfunctie gekarakteriseerd worden door de elf parameters  $a$ ,  $b$ ,  $\pi_1$ ,  $\mu_1$ ,  $\sigma_1$ ,  $\pi_2$ ,  $\mu_2$ ,  $\sigma_2$ ,  $\pi_3$ ,  $\mu_3$  en  $\sigma_3$ . Van de uitkomsten, die tussen  $a$  en  $b$  vallen, kunnen het gemiddelde ( $\mu$ ) en de standaardafwijking ( $\sigma$ ) worden berekend en deze waarden kunnen worden gebruikt voor de berekening van de waarschijnlijkheid dat een GvN-waarde buiten de grenzen van het tolerantiegebied valt (voor  $\alpha = 0.05$ :  $\mu \pm 1.96\sigma/\sqrt{n}$ ). Deze grenzen kunnen, volgens Hoffmann en Waid (1965), uitgaande van de grenzen van het normale gebied ( $a$  en  $b$ ) worden berekend:

$$\frac{a + b}{2} \pm 1.96 \frac{b - a}{4\sqrt{n}} .$$

Reed toonde aan, dat de overschrijdingskansen van op deze wijze berekende tolerantiegrenzen in zijn model sterk beïnvloed worden door onnauwkeurigheden in de keuze van de grenzen van het normale gebied en beveelt daarom aan op de gebruikelijke manier, door middel van een steekproef uit de betreffende populatie, schattingen van  $\mu$  en  $\sigma$  te berekenen.

Uiteraard zullen deze schattingen slechts goede resultaten kunnen opleveren indien bij de samenstelling van de steekproef aan bepaalde voorwaarden is voldaan (zie hoofdstuk 3).

Begtrup e.a. (1971) hebben de waarde van de GvN-methode onderzocht en vergeleken met het gebruik van een controleserum en met het toepassen van twee verschillende bepalingsmethoden voor het zelfde bestanddeel. Door middel van een computerprogramma, dat door Hoffmann ter beschikking was gesteld, werden, uitgaande van een frekwentieverdeling van een serie uitkomsten, de parameters bepaald van de Gauss-verdeling, die het beste past bij de 'meest voorkomende uitkomsten'. Het gemiddelde wordt aangeduid met  $M_H$  en de standaardafwijking met  $\sigma_H$ ; door Hoffmann wordt het interval  $M_H \pm 2\sigma_H$  het 'normale gebied' genoemd, maar Begtrup e.a. geven de voorkeur aan de term 'gebied van controlewaarden'. Het gemiddelde van de uitkomsten uit een dagserie, die in het gebied van controlewaarden vallen, wordt aangeduid met 'C' (dit is de GvN-waarde van Hoffmann en Waid). De frekwentieverdeling van de waarden van C zal ruwweg overeenkomen met een Gauss-verdeling waarvan het gemiddelde  $M_C$  gelijk is aan  $M_H$  en de standaardafwijking (control standard error):

$$CSE = \sigma_H/\sqrt{n} \tag{I}$$

( $n$  is het aantal bij de berekening van C gebruikte uitkomsten). De grenzen van het tolerantiegebied voor C liggen op een afstand van

2 CSE van  $M_C$  ( $M_H \pm 2\sigma_H/\sqrt{n}$ ). Het optreden van een systematische fout kan gesimuleerd worden door alle uitkomsten van een serie te verhogen met  $\alpha \cdot \sigma_H$ . Dit zal een verschuiving van C tot gevolg hebben, die weergegeven kan worden door  $\gamma(\alpha) \cdot \sigma_H$ . Bij uitbreiding tot meerdere series geldt:

$$M_C^\alpha = M_H + \gamma(\alpha) \cdot \sigma_H \quad \text{II}$$

In het geval dat een populatie van uitkomsten een verdeling volgens Gauss bezit, is het verband tussen  $\alpha$  en  $\gamma(\alpha)$  af te leiden.

De auteurs geven als voorbeeld<sup>1)</sup>:

$\alpha$	0.5	1.0	2.0
$\gamma(\alpha)$	0.47	0.72	1.20

Verder kan voor een bepaalde fout,  $\alpha \cdot \sigma_H$ , worden uitgerekend welk aantal monsters voldoende is om het gemiddelde  $M_C$  buiten de tolerantiegrenzen te laten vallen. Uit I, II en  $M_C = M_H$  volgt:

$$\begin{aligned} \gamma(\alpha) \cdot \sigma_H &= 2\sigma_H/\sqrt{n} \\ n &= (2/\gamma(\alpha))^2 \end{aligned} \quad \text{III}$$

Bij substitutie van de waarde 0.72 voor  $\gamma(\alpha)$  blijkt, volgens de auteurs, dat 8 uitkomsten voldoende zijn om een systematische fout die gelijk is aan  $\sigma_H$  ( $\alpha = \pm 1$ ) aan te tonen. Hierbij dient echter te worden aangetekend, dat dan de kans, dat C buiten de tolerantiegrenzen valt slechts ongeveer 0,5 bedraagt. Deze kans kan berekend worden met de formule:

$$P_{\text{pat}}(n) = \Phi(-2 + \gamma(\pm 1) \cdot \sqrt{n}) + \Phi(-2 - \gamma(\pm 1) \cdot \sqrt{n}) \quad \text{IV}$$

Hierbij is  $\Phi$  de gestandaardiseerde verdelingsfunctie volgens Gauss. Indien de uitkomsten geen verdeling volgens Gauss bezitten, kan met formule IV een redelijke benadering worden verkregen, mits de juiste waarde voor  $\gamma(\pm 1)$  gesubstitueerd kan worden. Voor negen bepalingen zijn  $\gamma(+1)$  en  $\gamma(-1)$  experimenteel bepaald door de verschuiving van het gemiddelde,  $M_C$ , als gevolg van een gesimuleerde fout van  $+1\sigma_H$  en  $-1\sigma_H$  na te gaan (tabel 2-VII).

Hoewel de resultaten meestal vrij duidelijk blijken af te wijken van de voor een Gauss-verdeling te verwachten waarde (in de laatste twee kolommen zou dan overal 72 moeten staan) is over de significantie van de afwijkingen geen uitspraak te doen, omdat de auteurs geen gegevens over de gebruikte aantallen vermelden.

<sup>1)</sup> Op deze problematiek wordt in hoofdstuk 4 nader ingegaan. Hier zij slechts opgemerkt dat de eerste waarde voor  $\gamma(\alpha)$  niet 0.47 maar 0.38 moet zijn.



Gedurende één jaar werd de GvN-methode vergeleken met het toepassen van twee bepalingmethoden voor hetzelfde bestanddeel en met het gebruik van een controleserum. Als voorbeeld van de resultaten van deze vergelijking geven de auteurs voor één bepaling (alkalische fosfatase) een grafiek van een periode van twee maanden met het commentaar, dat de curven van de GvN-methode en die van het controle-serum niet parallel lopen, maar dat dat ook niet te verwachten was. Gezien de hoeveelheid extra werk verbonden aan een dergelijke proef-opzet lijkt de conclusie nogal teleurstellend.

In de discussie vermelden de auteurs de door anderen (Amador, 1968; Kilgariff & Owen, 1968) geconstateerde, sterke variaties van de controlewaarden, die toegeschreven werden aan veranderingen in de samenstelling van het patientenmateriaal. Naar aanleiding hiervan stellen zij:

"Corresponding alterations of  $M_H$  could not be demonstrated in our investigations, in which the monthly calculated averages have only shown insignificant variations".

Het hoeft nauwelijks betoog dat het weinig zin heeft variaties in maandgemiddelden, berekend uit ongeveer duizend uitkomsten, te vergelijken met die in dag- of seriegemiddelden, die berekend zijn uit tien of enkele tientallen uitkomsten. In de verkorte weergave van een eerdere mededeling (Begtrup e.a., 1969) stellen de auteurs trouwens:

"Used as a month-to-month level control 'average of normals' seems to be of greater value, since in that case differences in the composition of the patient material have been partly eliminated".

Voor de vergelijking van de GvN-methode met de controle-serum-methode is een maat voor de gevoeligheid nodig. De auteurs gebruiken hiervoor de kans, dat een controlegrootheid buiten de betreffende tolerantiegrenzen valt bij het optreden (eventueel door simulatie) van een systematische fout van  $1\sigma_H$ . De gevoeligheid van de controle-serum-methode kan, indien als controlegrootheid het gemiddelde van  $n_R$  uitkomsten wordt gebruikt, berekend worden met de formule:

$$P_{\text{ref}}(n_R) = \Phi(-2 + \sqrt{n_R}\sigma_H/\sigma_R) + \Phi(-2 - \sqrt{n_R}\sigma_H/\sigma_R)$$

V

Bij vergelijking van formules IV en V blijkt dat de gevoeligheid van de controle-serum-methode groter is dan die van de GvN-methode indien  $n_R(\sigma_H/\sigma_R)^2 > n(\gamma(\pm 1))^2$ .

Als  $n_R = 2$ , kan het aantal uitkomsten, dat binnen de controlegrenzen moet vallen om de gevoeligheid van de GvN-methode gelijk te maken aan die van de controleserum-methode, berekend worden met:

$$n_{\min} = 2(\sigma_H/\sigma_R)^2/(\gamma(\pm 1))^2.$$

In tabel 2-VIII worden de resultaten weergegeven van deze berekening voor acht bepalingsmethoden.

Voor bijna alle tot nu toe besproken onderzoeken geldt, dat de series bepalingsuitkomsten, waar men onder andere voor het bepalen van tolerantiegebieden van uitgaat, mogelijk beïnvloed zijn door verschuivingen in de nauwkeurigheid (accuracy), veroorzaakt door tijdelijk optreden van (kleine) systematische fouten, terwijl ook de invloed van de fluctuaties niet constant hoeft te zijn.

Dit probleem is onderkend door Lewis & Dixon (1971), die volgens de lijnen van het eerder besproken onderzoek van Dixon & Northam enkele mogelijkheden hebben onderzocht om het onderscheidingsvermogen van de daggemiddelde-methode en dat van de controleserum-methode te verbeteren. Zij gingen hierbij uit van gedurende een jaar verzamelde bepalingsuitkomsten van een controleserum en daggemiddelden. Voor beide series controlegrootheden werden per periode van één maand het gemiddelde, de standaardafwijking en de tolerantiegrenzen (met een overschrijdingskans van 0,01) berekend. Het bleek, dat de standaardafwijkingen en dus ook de afstanden tussen de tolerantiegrenzen vrij grote variaties vertoonden. Deze variaties moeten volgens de auteurs voor een deel aan bovengenoemde verschuivingen worden toegeschreven. Door middel van grafieken en van een computerprogramma werden beide series uitkomsten gecorrigeerd en daarna nogmaals verwerkt. Dit leverde in het algemeen wat kleinere standaardafwijkingen op (zie tabel 2-IX, kolom A1 en A2 en kolom B1 en B2). Voor enkele bepalingen kon nog een verdere verbetering worden verkregen door middel van een procedure, die gericht was op verwijdering van extreme waarden (zie tabel 2-IX, kolom A3 en B3).

Met betrekking tot de controleserum-methode werd voor een periode van twee maanden de standaardafwijking van het gemiddelde van vier controleserum-uitkomsten vergeleken met die van enkelvoudige uitkomsten. Voor het merendeel van de bepalingen bleken de standaardafwijkingen slechts weinig te schelen. Dit wijst er op dat de variaties tussen de series een grotere invloed hebben op de uitkomsten dan de variaties binnen één serie.

Met de door Dixon & Northam gebruikte formule  $m = SD.t/R$  <sup>1)</sup> hebben Lewis & Dixon voor iedere bepaling de waarden van m berekend voor een aantal paren afkapgrenzen. De waarden werden vervolgens in een grafiek uitgezet tegen de gebruikte combinatie van afkapgrenzen. Uit het verloop van deze grafieken is af te lezen voor welk paar afkapgrenzen m het kleinst is.

Tenslotte werd nog het effect onderzocht van een vermindering van het aantal bepalingen, dat per dag wordt verricht, op de standaardafwijking van het daggemiddelde. Hiertoe werden bepalinguitskomsten (gemiddeld 95/dag) een aantal malen verwerkt via een computerprogramma, dat op aselechte wijze een bepaald gedeelte van elke dagserie verwijderde. De standaardafwijkingen van de 'daggemiddelden' van de ingekorte series worden gegeven in tabel 2-X. Hieruit blijkt dat, voor een zelfde bepaling, de standaardafwijking van het daggemiddelde groter is als het aantal verwerkte uitkomsten klein is. Dit heeft tot gevolg, dat de 'relative power' ( $RP = SD_2 \times R/SD_1$  <sup>1)</sup>) een kleinere waarde aanneemt naarmate de serie bepalinguitskomsten korter is.

Door Glick (1972) werden enkele mogelijkheden onderzocht om de uitkomsten van het aangevraagde onderzoek te gebruiken voor het vaststellen van de grenzen van het normale gebied en voor kwaliteitscontrole.

Gedurende één jaar werden van 11 bepalingen alle uitkomsten van mannelijke patienten verzameld. Uit deze verzameling werden op uiteenlopende manieren series steekproeven (van 396, 132 en 66 uitkomsten) samengesteld. Voor elke bepaling werd van 4 steekproeven van 396 uitkomsten een aantal statistische grootheden bepaald, waarna van elke grootheid het gemiddelde en de variatiecoëfficiënt werden berekend. Voor de meeste bepalingen bleek de variatiecoëfficiënt van de mediaan kleiner dan die van het gemiddelde of van de modale waarden. Hoewel dit resultaat in de lijn der verwachting ligt, dient bedacht te worden, dat de schattingen van de variatiecoëfficiënten ieder op slechts 4 waarden zijn gebaseerd. Voor de simulatie van het optreden van een systematische fout werd elk derde element van een steekproef verhoogd met een kwart van het verschil tussen de boven- en ondergrens van het normale gebied. Een dergelijke procedure is als model voor het optreden van een systematische fout naar mijn mening aan bedenkingen onderhevig. Uit de resultaten leidt de auteur

1) Zie het eerder besproken artikel van Dixon & Northam (1970).

af, dat voor de kwaliteitscontrole de mediaan van een serie bepaling-uitkomsten de beste toetsingsgrootte is.

Aangezien de aanpak van Glick nogal sterk verschilt van de eerder besproken onderzoeken, zijn de resultaten niet zonder meer te vergelijken met resultaten, die met één van de besproken methoden zijn verkregen.

Bij het opmaken van een voorlopige balans met betrekking tot het gebruik van uitkomsten van het aangevraagde onderzoek voor kwaliteitscontrole, komen we tot de volgende voor- en nadelen.

Voordelen:

- 1) Geen extra laboratoriumwerk, geen extra kosten en slechts een geringe hoeveelheid eenvoudig rekenwerk; het laatste punt geldt eigenlijk alleen voor korte series.
- 2) Er vindt ook controle plaats op systematische fouten bij de afname en de verwerking van de monsters vóór het serumstadium.
- 3) De methode is ook toepasbaar voor bestanddelen, waarvoor geen goed controleserum beschikbaar is.
- 4) Bewust of onbewust naar een (bekende) waarde van de toetsingsgrootte toewerken is moeilijk.

Nadelen:

- 1) Hoffmann en Waid zijn uitgegaan van de aanvechtbare hypothese, dat voor bijna elke klinisch-chemische bepaling een groot deel van de bepalinguitkomsten een frekwentieverdeling vertoont, die mag worden beschouwd als een verdeling volgens Gauss.
- 2) Vooral voor kleine series zou de methode een geringer onderscheidingsvermogen bezitten dan een methode, die gebruik maakt van een controleserum.
- 3) De toetsingsgrootte wordt behalve door systematische fouten ook beïnvloed door wisselingen in de samenstelling van het patiëntenmateriaal; dit geldt, evenals het vorige punt, vooral voor kleine series.
- 4) De controle betreft slechts een beperkt concentratiegebied.
- 5) Het daggemiddelde kan pas berekend worden als alle uitkomsten bekend zijn en voor grote series is de hoeveelheid rekenwerk niet te verwaarlozen, tenzij de laboratoriumadministratie gecomputeriseerd is.

Een nader onderzoek van enkele aspecten van deze methoden lijkt op grond van het bovenstaande op zijn plaats.

## HOOFDSTUK 3.

### LITERATUURONDERZOEK BETREFFENDE 'NORMALE WAARDEN'.

#### 3.1. INLEIDING.

Bij de interpretatie van de uitkomst van een laboratoriumbepaling in het kader van de medische diagnostiek, wordt deze uitkomst als regel vergeleken met één of meer gegevens, die meestal worden aangeduid met termen als 'normale waarden' of 'grenzen van het normale gebied'. Dergelijke vergelijkingsgegevens worden reeds lange tijd en vrijwel algemeen gebruikt. Toch zijn er, vooral de laatste jaren, nogal wat publicaties gewijd aan problemen verbonden aan het gebruik van normale waarden. Voor een deel hangen de problemen samen met het ontbreken van voldoende standaardisatie van bepalingsmethoden, voor een ander deel met de keuze van statistische methoden, maar ook problemen van meer fundamentele aard spelen een rol (Murphy & Abbey, 1967; Leynse-Ybema & Leynse, 1972; Rümke & Bezemer, 1972; Dybkaer & Gräsbeck, 1973; Siest, 1975; Sunderman, 1975).

De huidige stand van zaken kan, weliswaar gesimplificeerd, weergegeven worden door twee uitspraken tegenover elkaar te stellen. Door Koller (1968) werd de plaats, die het begrip 'normale waarde' in het medische denken inneemt, gekarakteriseerd met de opmerking:

"dass man ihn erfinden müsste, wenn es ihn nicht schon gäbe".

Dybkaer (1972) stelde daarentegen:

"At the same time we should try to do away with some cherished concepts that only serve to obscure basic facts of nature and statistical method. The first concept to discard is 'normal values'".

Bij een poging tot analyse van de problemen blijken deze in drie categorieën te kunnen worden onderscheiden:

1. problemen, die samenhangen met het gebruik van de termen 'normaal' en 'normale waarden';

2. problemen, die de begrippen 'normale waarden', 'normale personen', 'gezondheid' en 'ziekte' betreffen;
3. problemen, die het vaststellen en het gebruik van 'normale waarden' betreffen.

Hoewel het niet de bedoeling is uitvoerig in te gaan op de problemen die tot de eerste twee categorieën behoren, is het wel gewenst enkele opvattingen te bespreken, omdat dit mogelijk verhelderend kan werken bij de behandeling van de problemen in de laatste categorie.

### 3.2. PROBLEMEN, DIE SAMENHANGEN MET DE TERMEN 'NORMAAL' EN 'NORMALE WAARDEN'.

Hiervan dient vermeld te worden, dat door enkele auteurs bezwaar wordt gemaakt tegen het gebruik van de termen 'normaal' en 'normale waarden', omdat daardoor verwarring in de hand wordt gewerkt met de betekenissen, die deze termen bezitten in de statistiek, de metafysica, de ethiek en, niet te vergeten, het normale spraakgebruik. Verschillende alternatieven zijn voorgesteld:

1. 'standaardwaarden' (Albritton, 1952);  
voor laboratoriumuitkomsten evenmin gelukkig gekozen, omdat de term 'standaard' in laboratoria ook al in een andere betekenis wordt gebruikt.
2. 'klinische grenzen' of '-grenswaarden' (Elveback e.a., (1970) deze termen lijken evenzeer tot misverstanden aanleiding te kunnen geven.
3. 'referentie waarden', die nader omschreven dienen te worden (Gräsbeck & Saris, 1969; Gräsbeck, 1972; Dybkaer, 1972; Dybkaer & Gräsbeck, 1973), hierop wordt later nog teruggekomen.

Een poging (Friedman, 1969) om misverstanden te vermijden door een, de bedoelde betekenis aanduidende, toevoeging aan de term 'normaal' verdient vermelding, hoewel er geen voorstel tot algemene invoering aan is verbonden en de gekozen termen (normal-usual, normal-healthy en normal-bell-shaped) daar waarschijnlijk ook niet geschikt voor zijn.

### 3.3. PROBLEMEN, DIE HET BEGRIP 'NORMALE WAARDEN' BETREFFEN.

Door Lindeboom (1957) wordt in een betoog over de begripsvorming in de pathologie opgemerkt, dat de begripsvorming zeker niet de

sterkste zijde van de medische wetenschap is en hij vervolgt:

"Men opereert in haar nog dikwijls met begrippen wier inhoud niet volkomen vaststaat of te voren scherp omschreven is. Daaruit ontstaat natuurlijk niet zelden begripsverwarring en misverstand, dat niet alleen de gedachtenwisseling bemoeilijkt, maar ook het oordeel aan het ziekbed onzuiver kan maken. Want ook de praktische arts, die bij de lijder diens ziekte tracht te herkennen, de diagnose poogt te stellen, gebruikt in de verstandelijke verwerking van de bevindingen van zijn onderzoek, of hij het inziet of niet, voortdurend begrippen; en van de aanwending van bepaalde begrippen hangt zijn therapeutisch handelen af".

Het bovenstaande is zeker ook van toepassing op het begrip 'normale waarden', hoewel dit niet met name door Lindeboom wordt genoemd.

Zo blijken de opvattingen over hetgeen daaronder moet worden verstaan, ondanks het algemeen verbreide gebruik van normale waarden, zo uiteen te lopen, dat een voor ieder aanvaardbare definitie van het begrip 'normale waarden' niet mogelijk is gebleken (Mc Kinlay, 1951; Schneider, 1960; Simonson, 1965; Mainland, 1971).

Door Simonson (1965) wordt gesteld dat we onder 'normaal' niet 'ideaal' moeten verstaan, maar zeker ook niet eenvoudigweg het 'gemiddelde'. Dit laatste zou bijvoorbeeld in het geval van endemische ziekten geen juiste getallen opleveren. Eenzelfde standpunt wordt ook door anderen ingenomen (b.v. Tolls, 1969).

Door Schneider (1960) wordt een omschrijving vermeld van 'normaal' als 'betrekking hebbend op waarden of eigenschappen, die karakteristiek zijn voor gezonde personen'.

Hij heeft echter deze omschrijving zelf reeds ondergraven door op te merken dat de geneeskunde is gebaseerd op de onhoudbare dichotomie van gezondheid en ziekte en dat een algemeen aanvaardbare definitie van gezondheid niet valt te geven. Een omschrijving als 'gezondheid is de afwezigheid van alle ziekten' is in de praktijk niet bruikbaar, omdat criteria voor het vaststellen van de afwezigheid van ziekten ontbreken.

Dat het gevaar in een cirkelredenering te vervallen niet denkbeeldig is, blijkt als we het bovenstaande vergelijken met de beschouwing van Lindeboom (1957) over het begrip 'ziekte':

"Men kan het begrip ziekte ook trachten te benaderen vanuit zijn tegenhanger, dat van gezondheid. Maar ook dit is uiterst moeilijk

te omschrijven. Al spoedig is men geneigd het te vereenzelvigen met dat der normaliteit, met wat normaal is. Maar ook zo verwacht men zich dra in grote moeilijkheden".

Koller (1968) heeft het probleem van de definitie van het begrip 'normaal' tegen de achtergrond van in verschillende takken van wetenschap voorkomende begripscrises geplaatst:

"Diese Begriffe sind gewissermassen von der Mitte eines Variationsbereiches aus entstanden, von den Fällen aus, bei denen kein Zweifel über die Zugehörigkeit zum Typus bestand. Damit kam man Jahrzehnte oder sogar Jahrhunderte aus. Heute bemerkt man mehr und mehr, dass alle diese historischen Begriffe nicht randscharf sind, sondern sich in nicht fassbarer Weise um so mehr auflösen, je weiter man sich vom eigentlichen zentraler Kern des Typus entfernt. Wir wollen heute aber, unserem Streben nach Perfektion entsprechend, gerade möglichst lückenlos und zutreffend zuordnen können und suchen in jedem Falle die richtige Aussage. Das Zuordnungsproblem erfordert aber Abgrenzungen und Randschärfe, die in klassischen Begriffsbildungsprozess, der sich nur den Kern kummerte, nicht enthalten waren. Dies ist der Hintergrund, vor dem sich heute zahllose Begriffskrisen in vielen Wissenschaften abspielen. Auch unser Normproblem gehört, zu einem Teil, dazu".

In recente publicaties (o.a. Elsom, 1971) wordt in samenhang met deze problematiek wel de term 'health-disease continuum' gebruikt.

Het is niet waarschijnlijk, dat een fundamentele benadering van de geschetste problematiek op korte termijn tot een algemeen geldige definitie van 'normale waarden' zou leiden. Een pragmatische aanpak verdient daarom de voorkeur. Men zal er zich dan van bewust moeten blijven dat een dergelijke aanpak beperkingen met zich meebrengt, zodat het bijvoorbeeld geen zin heeft termen als 'de absolute geldigheid van een normaal gebied' (Annino & Williams, 1972) in de discussie hierover te gebruiken.

#### 3.4. PROBLEMEN, DIE HET VASTSTELLEN EN HET GEBRUIK VAN 'NORMALE WAARDEN' BETREFFEN.

Het doel van het vaststellen van 'normale waarden' is als volgt te formuleren:

het mogelijk te maken om, op grond van de ligging van een be-



palingsuitkomst ten opzichte van de betreffende 'normale waarden', steun te verlenen aan het oordeel dat een proefpersoon (patient) in het kader van het verrichte onderzoek al of niet als gezond moet worden beschouwd.

De ideale 'normale waarde' is de uitkomst die de bepaling zou hebben opgeleverd, als de zekerheid had bestaan dat de patient niet ziek was, terwijl het monster op hetzelfde tijdstip en onder dezelfde omstandigheden was afgenomen en verwerkt.

Een dergelijke waarde is niet te verkrijgen. Het komt zelfs slechts zelden voor, dat men de beschikking heeft over één of meer eerder verkregen bepalingen van dezelfde proefpersoon, die als basis voor een betrouwbare schatting van 'zijn normale waarden' kunnen dienen.

Meestal wordt daarom de ideale 'normale waarde' benaderd door na te gaan, in welk gebied de uitkomsten vallen, als 'gezonde' personen worden onderzocht.

Om, afgezien van eventuele fouten die bij het onderzoek gemaakt kunnen worden, altijd juist te kunnen oordelen, zou onder andere voldaan moeten worden aan de volgende voorwaarden:

- 1) De grenzen van het gebied, waarin uitkomsten vallen van bepalingen bij alle 'gezonde' personen, moeten bekend zijn.
- 2) De uitkomsten van bepalingen bij alle 'niet gezonde' personen moeten buiten die grenzen vallen.

Helaas wordt aan deze voorwaarden nooit voldaan en zal een beperkte betrouwbaarheid van dergelijke oordelen geaccepteerd moeten worden. Het is daarom gebruikelijk statistische methoden toe te passen, zodat de beperkte betrouwbaarheid althans in een getal kan worden uitgedrukt. Overigens stelt de toepassing van deze methoden ons zeker niet in staat alle eerder genoemde problemen te omzeilen. Door de normcommissie 70 (Toepassing van statistische methoden) wordt bijvoorbeeld het begrip 'populatie' omschreven (NEN 3117, 2<sup>e</sup> druk 1968) als een:

"Verzameling van operationeel gedefinieerde eenheden ('elementen'), waarop de conclusies van een statistisch onderzoek betrekking hebben".

Onder een operationele definitie wordt verstaan, een voorschrift met behulp waarvan in beginsel kan worden vastgesteld of een object al dan niet een element is van de populatie.

We zullen, met andere woorden, rekening moeten houden met de opmerking van Mc Kinlay (1951), dat onze 'normale waarden' groepswaarden zijn en daarom een zo goed mogelijke beschrijving van de groep, waarvoor ze geldig zijn, vereisen.

Een aantal van de punten, die in een dergelijke beschrijving aan de orde moeten worden gesteld, zijn:

- 1) de eventuele onderverdeling naar geslacht, leeftijd of andere grootheid;
- 2) criteria voor 'gezondheid';
- 3) voorbereiding van de proefpersoon, afname van het materiaal;
- 4) wijze van verwerking.

Dat deze problemen in de praktijk vaak te lichtvaardig terzijde worden geschoven, is te proeven uit de opmerking van Whitehead (1971), dat de manier waarop vele 'normale waarden' zijn vastgesteld en worden gebruikt, de indruk wekt dat een normaal persoon tot het mannelijk geslacht behoort, tussen de twintig en vijfentwintig jaar oud is en medicijnen studeert. Het bezit van grijs haar zou dan op pathologie moeten wijzen, omdat het in de genoemde groep zo weinig voorkomt. Het opstellen van bruikbare criteria voor 'gezondheid' blijft eveneens een probleem, waar men zich vaak onvoldoende in verdiept.

Door Schneider (1960) is een oplossing aangegeven, die in sommige gevallen redelijk lijkt te voldoen (b.v. van Riet & Hoeke, 1968). Binnen het kader van het onderzoek wordt als 'gezond' beschouwd iedereen, die niet in een toestand verkeert waarvan verwacht kan worden dat deze gepaard gaat met pathologische uitkomsten van de betreffende bepaling(en). Dit biedt de mogelijkheid klinische en/of poliklinische patienten op te nemen in de groep 'gezonde personen'. Vrijwel diametraal hiertegenover staat de visie van Jungner & Jungner (1969), die stellen dat zich onder de ogenschijnlijk gezonde personen een aantal bevindt dat, hoewel nog vrij van klinische symptomen, toch feitelijk reeds ziek is en dus eigenlijk niet tot de gezonde groep behoort. De consequentie van deze visie is, dat althans een aantal gebieden van 'normale waarden' zou moeten worden vernauwd. Een pragmatische benadering brengt met zich mee, dat we eerst zullen moeten nagaan of een dergelijke maatregel nuttig zou zijn voor de betreffende personen of eventueel voor de gezondheidszorg in het

algemeen.

Door Cutler e.a. (1969) werden bepalingen<sup>1)</sup>, die werden verkregen in een zogenaamd multi phasic laboratory aan een statistische analyse onderworpen. In het onderzoek werden alle uitkomsten betrokken, die door het laboratorium werden verkregen in de periode van 1 september 1966 tot en met 31 augustus 1967 en die betrekking hadden op monsters afkomstig van blanke personen, met een leeftijd van 20 tot en met 79 jaar (in totaal bijna 35.000 personen). Behalve de acht bepalingen uitkomsten werden van iedere proefpersoon ook een aantal andere gegevens vastgelegd, waaronder de diagnose van de arts die de proefpersoon onderzocht en de uitkomsten beoordeeld had. Als diagnose werd ook opgevat 'No significant abnormality' (NSA-Once). Verder werd door de proefpersoon zelf op een vragenlijst aangegeven of zijn gezondheidstoestand hem activiteitsbeperkingen oplegde of niet. Degenen die deze vraag ontkennend beantwoordden, werden in het kader van het onderzoek als (subjectief) gezond beschouwd (Healthy-Once). Tenslotte werd nagegaan of de proefpersoon eerder het centrum had bezocht en zo dit het geval was, hoe bij die gelegenheid de beoordeling van zijn gezondheid door de arts en door hemzelf was geweest (evt. NSA-Twice en Healthy-Twice).

Op grond van genoemde gezondheidsbeoordelingen werden vier groepen geformeerd, NSA-Once, Healthy-Once, NSA-Twice en Healthy-Twice, die, evenals de totale groep, verder werden onderverdeeld in sub-groepen naar geslacht en leeftijd.

Voor het merendeel van de bepalingen bleken de uitkomsten van de NSA-Twice groep en zijn sub-groepen, iets lagere gemiddelden en standaarddeviaties te vertonen dan de totale groep of de overeenkomstige sub-groepen daarvan. Hoewel de verschillen vaak statistisch significant waren, bleken ze zo klein dat het raadzaam is de 'normale waarden' af te leiden uit de gegevens van de sub-groepen van de totale groep.

In dit verband moet worden vermeld, dat de resultaten van Carlström e.a. (1971; zie ook Nordén, 1972) juist manen tot voorzichtigheid bij de interpretatie van uitkomsten die vallen in het niemandsland tussen het 'normale' en het duidelijk pathologische gebied.

<sup>1)</sup> Van elke proefpersoon werden de volgende bepalingen verricht: glucose (60 min. na oraal toegediende oplossing van 75 g gluc.), cholesterol, totaal eiwit, albumine, urinezuur, calcium, creatinine, SGOT (asp. aminotransferase).

Zij volgden personen die bij een bevolkingsonderzoek een licht verminderde glucose tolerantie vertoonden, gedurende perioden van 3-5 jaren. De proefpersonen (231) waren aselekt in vier groepen verdeeld. Aan de personen uit de eerste groep werd medegedeeld dat zij geen diabetes hadden, aan de anderen dat "they were possibly candidates for manifest diabetes".

Groep I (107 personen) kreeg geen speciale instructies, maar wel werd de orale glucosetolerantietest aan het eind van de onderzoeksperiode herhaald.

Groep II (36 personen) kreeg slechts een dieet advies.

Groep III (47 personen) kreeg een dieet advies en werd behandeld met 1.5 g tolbutamide per dag.

Groep IV (41 personen) kreeg hetzelfde dieet advies en een placebo.

Tegen de verwachting in openbaarden zich gedurende de observatieperiode bij geen van de proefpersonen verschijnselen van diabetes en bleek geen van de glucosetolerantiecurven als diabetisch geklasseerd te moeten worden. De groepsgemiddelden vertoonden slechts geringe verschillen. Hieruit concludeerden Carlström e.a. dat de progressie van de glucose intolerantie, zo daar al sprake van mocht zijn, traag is en dat de observatieperiode verlengd zou moeten worden om na te gaan of de onderzoeksgroep inderdaad een verhoogde kans loopt op klinische diabetes.

Nordén (1972) heeft zelfs gesteld dat het niet verantwoord is op grond van dergelijke bevindingen ongerustheid te zaaien bijvoorbeeld door op een verhoogd risico te wijzen en/of een dieet voor te schrijven. Ook moet gewezen worden op de opvatting van Payne en medewerkers (Payne & Levell, 1968; Cook, Levell & Payne, 1970; Little e.a., 1974), die betoogden, dat voor het vaststellen van 'normale waarden' voor bepalingen, die voor een specifiek diagnostisch doel worden gebruikt, idealiter zou moeten worden uitgegaan van bepalingen uitkomsten van patienten, die niet aan de betreffende ziekte lijden, maar wel verschijnselen vertonen die in dat ziektebeeld zouden kunnen passen. Zeker in dit geval dient van het gebruik van de term 'normale waarden' te worden afgezien. Voorlopig zal realisering van een dergelijke opzet vrijwel onmogelijk zijn, als gevolg van het ontbreken van de benodigde gegevens. Niettemin lijkt de benaderingswijze in principe juist en mogelijk zal deze in de toekomst ook in praktijk kunnen worden gebracht.

In een artikel getiteld "The normal range - A common misuse" werd door Murphy & Abbey (1957) betoogd, dat bij de gebruikelijke vergelijking van een bepaling uitkomst met de grenzen van het normale gebied ten onrechte een aantal aspecten buiten beschouwing wordt gelaten. Naar hun mening dienen, naast de verdeling van uitkomsten bij gezonden, ook de verdeling van uitkomsten bij zieken, de frequentie waarmee de ziekte voorkomt in de groep waartoe de patient behoort en de 'kosten' van een eventuele foutieve beslissing in de beschouwing te worden betrokken. Als aanduiding van de op deze wijze tot stand gekomen grenswaarde worden termen gebruikt (Sunderman, 1975) als kritische waarde (critical value), onderscheidingswaarde (discrimination value) en beslissende waarde (decisive value). Voor het doorvoeren van een dergelijke procedure zijn gegevens nodig, waar nu nog niet of slechts zeer ten dele over beschikt wordt.

Door Gräsbeck en Saris (1969) is, ter vervanging van 'de normale waarden' een systeem voorgesteld, dat bestaat uit nader aan te duiden referentiewaarden. Zowel de selectie van de proefpersonen als de aanduiding moet worden aangepast aan de groep, waarvoor de referentiewaarden bedoeld zijn. Als voorbeelden worden genoemd:

- 1) Healthy reference values (HRV), bepaald door onderzoek van gezonde personen en bedoeld voor gebruik bij periodieke bevolkingsonderzoeken.
- 2) Patient reference values (PRV), bepaald door onderzoek van geselecteerde groepen patienten en bedoeld voor gebruik in ziekenhuizen (van belang i.v.m. de invloed van stress, bedrust e.d.).
- 3) Individual reference values (IRV), bepaald door onderzoek van de betreffende persoon (bijv. bij een aantal periodieke bevolkingsonderzoeken) en bedoeld als benadering van de ideale referentiewaarde.
- 4) Survival reference values (SRV), bepaald door onderzoek van bejaarden, waarbij de verwerking van de gegevens wordt uitgesteld tot een nader te bepalen aantal jaren na het onderzoek en uitkomsten alleen worden gebruikt als deze afkomstig zijn van een proefpersoon, die op dat moment nog in leven is en in de tussenliggende periode geen ernstige ziekte heeft doorge-

maakt.

Dybkaer (1972) definieert referentiewaarden als: "a set of values of a certain type of quantity obtainable from a single individual or a group of individuals corresponding to a stated description". Hieraan knoopt hij de opmerking vast, dat deze beschrijving tot in details vastgelegd moet worden en beschikbaar moet zijn voor anderen, die de referentiewaarden willen gebruiken. Voor iedere grootheid zal een serie referentiegroepen nodig zijn, waarbij rekening gehouden moet worden met leeftijd, geslacht, ras, cyclus, voedingsgewoonten, lichamelijke activiteit, lichaamshouding etc..

Ook door Gräsbeck (1972) wordt gewezen op de invloed, die dergelijke fysiologische en ook andere factoren kunnen hebben op de uitkomst van een bepaling. In het ideale geval zouden de factoren, die de te beoordelen uitkomst hebben beïnvloed, overeen moeten stemmen met die van de referentiewaarden. Dit zou een ver gaande opsplitsing van de referentiegroepen met zich meebrengen. Daarbij komt dan nog dat de auteur voorstelt de eerste twee categorieën (HRV en PRV) uit het voorstel van Gräsbeck en Saris (1969) op een andere wijze in te delen omdat absolute gezondheid niet bestaat. De volgende indeling wordt daarom geprefereerd:

- 1 Individuen
- 2 Populaties
  - 2.1 Niet-gehospitaliseerde populaties
    - 2.1.1 Ongeselecteerde personen
    - 2.1.2 Subjectief gezonde personen
      - 2.1.2.1 Subj. gezonde, niet gescreende personen
      - 2.1.2.2 Subj. gezonde, gescreende personen
        - 2.1.2.2.1 Subj. gezonde, oppervlakkig gescreende personen
        - 2.1.2.2.2 Subj. gezonde, grondig gescreende personen
    - 2.2 Gehospitaliseerde populaties
      - 2.2.1 Subjectief gezonde personen
        - 2.2.1.1 Subj. gezonde, grondig gescreende personen
      - 2.2.2 Zieken (Patienten)
        - 2.2.2.1 Ongeselecteerde patienten
        - 2.2.2.2 Op diagnose geselecteerde patienten
        - 2.2.2.3 Patienten, die (voor één of meer bepalingen) vergelijkbaar zijn met gezonde personen.

Ten aanzien van de voorbereiding van de proefpersoon, de afname van het materiaal en de wijze van verwerken (onder andere de bepalingmethode) kan worden volstaan met de opmerking, dat ook deze procedures zo gedetailleerd mogelijk vastgelegd moeten worden en dat de referentiewaarden strikt genomen alleen gebruikt mogen worden als de uitkomst die beoordeeld moet worden via dezelfde procedures tot stand is gekomen, of als aangetoond is dat afwijking geen invloed op de uitkomst heeft.

Het zal duidelijk zijn, dat een dergelijk systeem slechts ten koste van zeer veel inspanning te realiseren zal zijn, terwijl de vraag of het resultaat al deze moeite waard zal blijken niet *à priori* is te beantwoorden. Het gevaar, dat het grote aantal varianten van de referentiewaarden voor één grootheid de verwarring op dit gebied alleen nog maar zal vergroten, lijkt zeker niet denkbeeldig.

### 3.5. DE KEUZE VAN DE STATISTISCHE METHODE.

Bij de statistische verwerking van verzamelde gegevens tot 'normale waarden' werd in de afgelopen jaren zeer vaak gebruik gemaakt van formules, die gebaseerd zijn op de verdeling volgens Gauss en die, strikt genomen, slechts mogen worden toegepast als de te verwerken gegevens behoren tot een populatie, waarvan de elementen een dergelijke verdeling vertonen. Reeds lang echter was bekend dat niet alle biologische grootheden aan deze voorwaarde voldoen (Rietz, 1927, geciteerd door Elveback e.a., 1970). Bij een onderzoek naar het verdelingstype van de uitkomsten van een aantal klinisch-chemische bepalingen vonden Wootton e.a. (1951) slechts in ongeveer de helft van de gevallen een verdeling volgens Gauss. Desondanks bleef de toepassing van bovengenoemde formules sterk verbreid. Voor Elveback e.a. (1970) was dit aanleiding om, in het kader van een pleidooi voor het gebruik van parameter-vrije methoden, een beroemde uitspraak nogmaals te vermelden:

"Everybody believes in the law of errors, the experimenters because they think it is a mathematical theorem and the mathematicians because they think it is an experimental fact".

Door Reed e.a. (1971) werd een onderzoek verricht naar de invloed van de bij de verwerking gebruikte statistische methode op de daarbij gevonden schattingen van de grenzen van het normale gebied. Bij dit onderzoek werd uitgegaan van een serie van honderd uitkomsten

van haptoglobinebepalingen bij ogenschijnlijk gezonde personen en twee series van honderd door middel van een computerprogramma gegenereerde getallen, waarvan de eerste serie afkomstig was uit een populatie met een verdeling volgens Gauss en de tweede uit een populatie met een log-Gaussiaanse verdeling. Elk van de laatste twee series getallen kon worden beschouwd als een aselechte steekproef uit de betreffende populatie, waarvan de parameters bekend waren. Voor alle drie series werden met verschillende methoden schattingen van de grenzen van het normale gebied (dat wil zeggen het centrale 95%-gebied) berekend. Uit de resultaten, waarvan een deel is samengevat in tabel 3-I, blijkt dat voor gegevens, die uit een verdeling volgens Gauss of uit een log-Gaussiaanse verdeling stammen, het gebruik van een methode die gebaseerd is op een verdelingstype weliswaar tot goede resultaten leidt als dit verdelingstype overeenstemt met dat van de te verwerken gegevens, maar tot vrij grote fouten kan leiden als dit niet het geval is. De parameter-vrije methoden daarentegen leveren in beide gevallen uitkomsten, die redelijk overeenstemmen met de te verwachten waarden. Hieruit valt te concluderen dat voor gegevens, die behoren tot een populatie waarvan het verdelingstype niet bekend is, de parameter-vrije methoden de voorkeur verdienen. Door Rümke en Bezemer (1972) is als globale richtlijn voor de keuze gegeven ( $n$  is het aantal waarnemingen):

$n < 20$  geen normale waarden berekenen

$20 < n < 100$  parameter-vrije methoden toepassen, tenzij men er redelijk zeker van is, dat de populatie waaruit de steekproef stamt bij benadering een Gauss-verdeling bezit

$n > 100$  toetsen op Gauss-verdeling en op grond van de conclusie van de toetsing methode kiezen.

Hoewel ik de bezwaren tegen het ten onrechte toepassen van de op een Gauss-verdeling gebaseerde methoden niet wil bagatelliseren, wil ik er toch op wijzen, dat het toepassen van verdelingsonafhankelijke methoden ons volledig afhankelijk maakt van de mogelijkheid goede criteria voor de selectie van 'normale' personen op te stellen en toe te passen op een voldoende groot aantal personen. Vooral de laatste jaren blijken de hieraan verbonden problemen en onzekerheden voor sommigen zo zwaar te wegen, dat zij de voorkeur geven aan een andere



wijze van selectie. Hiervoor werd de basis gelegd in de reeds vermelde publicatie van Hoffmann (1963). Uitgangspunten zijn de hypothese dat de bepalinguitkomsten van gezonden uit Gauss-verdelingen stammen en het ervaringsfeit dat een groot gedeelte van de uitkomsten van laboratoriumbepalingen in het betreffende 'normale gebied' valt. Het voordeel van deze methoden zou dan ook zijn, dat in principe elk laboratorium voor het vaststellen van zijn 'normale waarden' de eigen bepalinguitkomsten kan gebruiken zonder voorafgaande selectie van proefpersonen op grond van een medische keuring.

Bij de methode van Hoffmann (1963) en bij de hiermee verwante methode van Neumann (1968) wordt de relatieve cumulatieve verdelingscurve uitgezet op z.g. waarschijnlijkheidspapier. Voor een Gauss-verdeling wordt dan een rechte lijn verkregen. Indien men experimenteel verkregen steekproefgegevens, die uit een Gauss-verdeling afkomstig zijn, uitzet op dit papier, zullen de punten niet alle op één rechte liggen. Men kan dan echter op het oog de 'best passende' rechte lijn trekken, waarbij aan de punten, die in de buurt van de 50% liggen een groter gewicht moet worden toegekend. Wanneer we de methode toepassen op in een laboratorium verkregen, ongeselecteerde, bepalinguitkomsten zullen we in het gunstigste geval te maken hebben met een éénzijdig gecontamineerde Gauss-verdeling, waarbij de contaminatie praktisch beperkt is tot de 'staart' van de Gauss-verdeling. Deze contaminatie zal in de figuur tot uiting komen in een afbuiging aan de gecontamineerde kant van de cumulatieve verdelingscurve en in een verandering van de helling van het rechte gedeelte. De werkwijze van Hoffmann komt neer op een extrapolatie van dat rechte gedeelte van de curve. Als grenzen van het normale gebied worden de punten genomen, die overeenkomen met 5% en 95% van de populatie. De keuze van deze punten in plaats van de meer gebruikelijke (2,5% en 97,5%) is bedoeld als correctie voor de verandering van de helling. De aantrekkelijkheid van Hoffmann's methode is vooral gelegen in de simpele uitvoering. Helaas zijn de resultaten, zoals onder andere door Neumann werd aangetoond, niet altijd bevredigend. Neumann's modificaties maken de methode wel meer bewerkelijk, maar volgens hem zijn de resultaten beter. De werkwijze van Neumann kan in de volgende punten worden samengevat (de methode wordt toegepast in hoofdstuk 5):

- 1) De gegevens worden in klassen ingedeeld of gesorteerd naar grootte. Daarna wordt een relatieve cumulatieve frekwentie-

verdeling berekend en uitgezet op waarschijnlijkheidspapier. Het einde van het rechte gedeelte van deze curve wordt gebruikt als afkapgrens.

- 2) Voor de gegevens die onder de afkapgrens vallen wordt een nieuwe cumulatieve frekwentieverdeling berekend en uitgezet.
- 3) Het rechte gedeelte van deze curve wordt geëxtrapoleerd tot de afkapgrens wordt bereikt in punt  $A_n$  (n geeft het aantal doorlopen cycli aan).
- 4) Het bij punt  $A_n$  behorende percentage wordt afgelezen en gebruikt voor de berekening van een gecorrigeerde relatieve cumulatieve frekwentieverdeling, die weer wordt uitgezet. De procedure wordt nu vanaf 3) herhaald tot het verschil in de bij twee achtereenvolgende cycli gevonden percentages voldoende klein geworden is (Neumann suggereert 2% als grens te nemen).

Zowel voor de methode van Hoffmann als voor die van Neumann geldt, dat goede resultaten slechts verwacht kunnen worden als de te verwerken gegevens, eventueel na transformatie, bij voldoende benadering een verdeling volgens Gauss bezitten en de contaminatie uitsluitend éénzijdig is.

Voor de gevallen, waarbij niet aan de laatste voorwaarde is voldaan, kan volgens Gindler (1970) gebruik gemaakt worden van een door Bhattacharya (1967) beschreven methode. Met deze methode kunnen het gemiddelde en de standaardafwijking van een Gauss-verdeling worden gevonden via een grafiek waarin de natuurlijke logaritme van het quotient  $y(x+h)/y(x)$  wordt uitgezet tegen  $x$  (hierin is  $x$  een klassegemiddelde,  $h$  de klassebreedte en  $y$  de frekwentie in de betreffende klasse). Zie verder hoofdstuk 5 voor voorbeelden van toepassing.

Door Harris & DeMets (1972) werd gewezen op de mogelijkheid om voor de verwerking van gegevens, waarvan de verdeling niet als een verdeling volgens Gauss kan worden beschouwd, naast de gebruikelijke ook andere transformaties toe te passen of, zo nodig, een combinatie van transformaties. Van enkele series bepalingen uitkomsten werden gemiddelde, standaardafwijking en de grootheden  $g_1$  en  $g_2$  berekend. De laatste twee als aanduiding van de scheefheid ( $g_1$ ) en kurtose ( $g_2$ ). Voor één van de series bleken de waarden voor  $g_1$  en  $g_2$  zo klein, dat geen transformatie nodig was. Voor de andere series werden naast elkaar de volgende transformaties gebruikt:

$$z_i = \sqrt{x_i}$$

$$z_i = \log x_i$$

$$z_i = \log (x_i + C)$$

Voor elke reeks getransformeerde getallen werden daarna de waarden voor  $g_1$  en  $g_2$  berekend. Voor alle series bepalingsuitkomsten werd door tenminste één van bovenstaande transformaties de waarde van  $g_1$  belangrijk gereduceerd, maar bij drie series bleken de met deze transformaties verkregen waarden voor  $g_2$  nog niet aanvaardbaar. Daarom werd voor deze series nog een transformatie toegepast en wel  $z_i' = \delta \sinh^{-1}(z_i/\lambda)$ . Hierin zijn  $\delta$  en  $\lambda$  constanten, waarvan de waarden verband houden met de waarde van  $\gamma_2(z)$ , de grootheid die de rest-kurtose in de verdeling van  $z_i$  aangeeft. Op deze wijze bleek het mogelijk alle in het onderzoek betrokken series bepalingsuitkomsten om te zetten in reeksen getallen die konden worden beschouwd als afkomstig uit een verdeling volgens Gauss.

Reed en Gaw Tzu Wu (1974) hebben de methode van Harris & DeMets vergeleken met een parameter-vrije methode voor de bepaling van percentielen. Hiervoor werd uitgegaan van door middel van een computer gegenereerde reeksen van honderd getallen, die konden worden beschouwd als aselechte steekproeven uit een aantal populaties, waarvan de, nogal uiteenlopende, verdelingsfuncties bekend waren. Voor elke reeks van honderd getallen werden de onder- en bovengrens van het normale gebied (overeenkomende met 2.5% resp. 97.5%) berekend op twee manieren. Ten eerste met een parameter-vrije methode en ten tweede met de formules voor een Gauss-verdeling, na voorafgaande transformatie door middel van één van de door Harris & DeMets aangegeven mogelijkheden. De selectie van de transformatie gebeurde op basis van één van de volgende criteria:

- 1) Het 'absolute waarde criterium'

Gekozen wordt de transformatie, die leidt tot de kleinste waarde voor  $|2g_1 + g_2|$

- 2) Het Kolmogorov-Smirnov criterium

Gekozen wordt de transformatie, die leidt tot de kleinste waarde voor de toetsingsgrootheid van de Kolmogorov-Smirnov-toets <sup>1)</sup> op normaliteit.

<sup>1)</sup> De toetsingsgrootheid  $D_0$  is de maximale waarde van  $|F_i/n - E_i/n|$ . Hierin is  $F_i/n$  de waargenomen en  $E_i/n$  de verwachte cumulatieve relatieve frekwentie (de Jonge, 1963).

Van enkele populaties werden groepen van 25 reeksen nader bestudeerd. Voor elk van de onder- en bovengrenzen werden berekend of bepaald het groepsgemiddelde, de standaardafwijking, de laagste en de hoogste waarde. Bovendien werden deze resultaten vergeleken met de uit de verdelingsfunctie afgeleide verwachtingswaarden voor de grenzen. Het bleek dat voor steekproeven uit een populatie met een vrijwel symmetrische verdeling de transformatie methode gemiddeld tot betere resultaten leidt dan de parameter-vrije methode. Maar voor steekproeven uit een populatie met een duidelijk asymmetrische verdeling verdient het gebruik van een parameter-vrije methode de voorkeur.

## HOOFDSTUK 4.

### ONDERZOEK NAAR ENKELE EIGENSCHAPPEN VAN DE 'GEMIDDELDE VAN NORMALEN'-METHODE.

#### 4.1. INLEIDING .

Zoals reeds in hoofdstuk 2 werd vermeld, leek het gewenst de 'gemiddelde van normalen'-methode van Hoffmann & Waid (1965) aan een nader onderzoek te onderwerpen. Zo wordt de door Hoffmann & Waid ingevoerde beperking van de te gebruiken bepalingssuitkomsten tot die uitkomsten, die in het 'normale gebied' vallen, door sommige andere auteurs niet toegepast. Om de voor- en nadelen van de keuze van een beperkt gebied beter te kunnen beoordelen, leek het nuttig de invloed van een dergelijke keuze eerst na te gaan aan een eenvoudig model.

Hoffmann & Waid zijn uitgegaan van de arbitraire veronderstelling dat aan de verdeling van bepalingssuitkomsten in het 'normale gebied' een verdeling volgens Gauss ten grondslag ligt. Een dergelijke verdeling wordt, zoals bekend, vaak normale verdeling genoemd. Een normale verdeling wordt gekarakteriseerd door twee parameters, het gemiddelde  $\mu$  en de standaardafwijking  $\sigma$ , en kan daarom worden aangeduid met  $N(\mu, \sigma)$ . In deze aanduiding kunnen de symbolen  $\mu$  en  $\sigma$ , indien deze grootheden bekend zijn, worden vervangen door de betreffende getalwaarden. In principe zijn allerlei berekeningen uit te voeren uitgaande van de verdelings- of van de dichtheidsfunctie van de normale verdeling, als daarin de waarden van  $\mu$  en  $\sigma$  worden gesubstitueerd. Men kan echter ook gebruik maken van de in handboeken (bijv. Beyer, 1968) beschikbare tabellen voor de standaardnormale verdeling  $N(0,1)$ , aangezien elke normale verdeling  $N(\mu, \sigma)$  door standaardisering, dat wil zeggen toepassing van de formule  $\underline{z} = (\underline{x} - \mu)/\sigma$ , is om te zetten in de standaardnormale verdeling. Daarom werd als eerste model een populatie met een gestandaardiseerde normale verdeling gekozen.

## 4.2. MODELPROEVEN.

### 4.2.1. Berekeningen uitgaande van de standaardnormale verdeling.

Voor de standaardnormale verdeling is het gemiddelde van het gedeelte van de populatie, dat tussen de grenzen  $a_1$  en  $a_2$  valt, te berekenen uit:

$$\zeta(a_1 < \underline{z} < a_2) = \frac{\int_{a_1}^{a_2} z \cdot f(z) dz}{\int_{a_1}^{a_2} f(z) dz} = \frac{-f(a_2) + f(a_1)}{F(a_2) - F(a_1)}$$

Hierin is  $F(z)$  de verdelingsfunctie en  $f(z)$  de dichtheidsfunctie van de standaardnormale verdeling.

In tabel 4-I zijn voor enkele combinaties van  $a_1$  en  $a_2$  de waarden van  $\zeta(a_1 < \underline{z} < a_2)$  weergegeven.

In het gekozen model kunnen we het optreden van een systematische fout simuleren door ieder element van de gestandaardiseerde populatie te verhogen met een constante  $\Delta$ . Er wordt dan een populatie verkregen met een verdeling, die we kunnen opvatten als een verschoven standaardnormale verdeling,  $N(\Delta, 1)$ . Ook voor deze populatie is het gemiddelde van het gedeelte, dat tussen de grenzen  $a_1$  en  $a_2$  valt, na standaardiseren van dat gedeelte, te berekenen met behulp van de eerder genoemde tabel. Om de plaats van dit gemiddelde ten opzichte van het oorspronkelijke nulpunt aan te geven, moet daarna de waarde van het gestandaardiseerde gemiddelde worden verhoogd met  $\Delta$ :

$$\xi(a_1 < \underline{x} < a_2) = \frac{-f(a_2 - \Delta) + f(a_1 - \Delta)}{F(a_2 - \Delta) - F(a_1 - \Delta)} + \Delta$$

Als we de, als gevolg van de systematische fout optredende, verschuiving van het gemiddelde van het gedeelte van de populatie, dat tussen de grenzen  $a_1$  en  $a_2$  valt, aanduiden met  $v$  geldt:

$$v = \xi(a_1 < \underline{x} < a_2) - \zeta(a_1 < \underline{z} < a_2)$$

In het kader van de kwaliteitscontrole zal gestreefd worden naar een zo groot mogelijke waarde voor het quotient  $v/\Delta$ , waarbij het in het algemeen gewenst is dat deze waarde niet afhankelijk is van het teken van  $\Delta$ . Uit de bovenstaande formules volgt, dat de grootte  $v$  afhankelijk is van  $a_1$ ,  $a_2$  en  $\Delta$ . Om een indruk te krijgen van de invloed van deze drie grootheden zijn eerst de waarden berekend, die  $v$  aanneemt voor een aantal combinaties van  $a_1$  en  $a_2$  bij  $\Delta=+1$  en bij  $\Delta=-1$ . De resultaten zijn vermeld in tabel 4-II. Het blijkt, dat de absolute waarde van de verschuiving van het gemiddelde en dus ook de waarde van het quotient  $v/\Delta$  groter is naarmate de grenzen ruimer zijn.

Verder blijkt de waarde van het quotient  $v/\Delta$  slechts onafhankelijk te zijn van het teken van  $\Delta$ , indien de grenzen symmetrisch liggen ten opzichte van het midden van de oorspronkelijke populatie. Bij het onderzoek naar het verband tussen  $\Delta$  en  $v$  werden daarom paren symmetrisch gelegen grenzen gebruikt. Voor een reeks waarden van  $\Delta$  werden de daarbij behorende waarden van  $v$  berekend. De resultaten zijn weergegeven in tabel 4-III. Het verband tussen  $v$  en  $\Delta$  blijkt, vooral voor waarden van  $\Delta < 1$ , bij benadering lineair te zijn. Voor hogere waarden van  $\Delta$  neemt  $v$  minder toe dan met een lineair verband overeen zou komen (zie figuur 4-1). Verder laat zich, nog duidelijker dan uit tabel 4-II, aflezen dat de keuze van de grenzen  $a_1 = -3$  en  $a_2 = +3$  voor het gebruikte model een belangrijk gunstiger verhouding tussen  $v$  en  $\Delta$  met zich meebrengt vergeleken met  $a_1 = -2$  en  $a_2 = +2$ . Nog verdere verruiming van het gebied tot  $a_1 = -4$  en  $a_2 = +4$  levert slechts weinig voordeel op.

#### 4.2.2. Steekproeven uit een tweezijdig afgekapte normale verdeling.

Een beperking van de tot nu toe gevolgde werkwijze is dat deze niet de mogelijkheid biedt om het werken met betrekkelijk korte series bepalingen uitkomsten (vaak minder dan honderd) te simuleren. Daarom werd gebruik gemaakt van de mogelijkheid om met de IBM 1800 door middel van het programma 'GAUSS' getallen te verkrijgen, die te beschouwen zijn als willekeurig getrokken getallen uit een normaal verdeelde populatie met een tevoren gekozen gemiddelde en standaardafwijking. Voor deze experimenten werd als populatiegemiddelde 100 gekozen en als standaardafwijking 5. De door het programma geleverde getallen werden geselecteerd op ligging binnen de grenzen 90 en 110. Nadat 1000 getallen de selectie waren gepasseerd, werden 20 series van 10, van 25 en van 50 getallen en 10 series van 100 getallen gevormd. Van elke serie werd het gemiddelde  $M_k$  ( $k=10, 25, 50$  of  $100$ ) berekend en daarna werd per groep van 20 resp. 10 series het gemiddelde en de standaardafwijking van de seriegemiddelden bepaald. Bij het tweede en derde experiment werden de door het programma "GAUSS" geleverde getallen gemodificeerd door er 5 resp. 10 bij op te tellen. Daarna werd de zelfde procedure gevolgd als bij het eerste experiment. De resultaten van deze drie experimenten zijn samengevat in tabel 4-IV. Uit de reeksen B2 en B3 blijkt dat de uitkomsten een goede overeenstemming vertonen met de waarden, die op grond van de bereke-

ningen, uitgaande van de standaardnormale verdeling, te verwachten zijn: voor de  $\bar{M}_k$ -waarden van de B2 en de B3 reeks resp. 0,72 en 1,20 (zie tabel 4-III).

Om ook in dit model de invloed van de grootte van de systematische fout en van de keuze van de afkapgrenzen na te gaan werd voor de steekproefomvang 50 dezelfde procedure, maar uitgebreid met de modificaties +1, +2 en +3 en met andere afkapgrenzen, herhaald. De resultaten van deze experimenten zijn samengevat in tabel 4-V. Ook in dit geval blijkt een goede overeenstemming met de te verwachten uitkomsten: voor de  $\bar{M}$ -waarden van de B5 reeks achtereenvolgens 0,28, 0,72 en 0,94 en voor die van de B6 reeks 0,49, 1,20 en 1,71 (zie tabel 4-III). Het verband tussen de verschuiving van het gemiddelde en de systematische fout, die de verschuiving veroorzaakt, lijkt ook in dit model lineair te zijn voor fouten die niet groter zijn dan de standaardafwijking. De verhouding tussen de verschuiving en de fout wordt ook hier groter naarmate de afkapgrenzen ruimer zijn.

#### 4.2.3. Steekproeven, die zijn samengesteld uit elementen van twee verschillende populaties.

Zoals besproken in hoofdstuk 2 worden de variaties in de waarde van de toetsingsgrootte van de GvN-methode voor een deel toegeschreven aan de van dag tot dag wisselende samenstelling van het patientenmateriaal. Daarom leek het van belang om met het in 4.1.2. gebruikte model ook de invloed te onderzoeken van contaminatie van de steekproeven. Dit kan gebeuren door in die steekproeven een van te voren te bepalen percentage van de oorspronkelijke elementen te vervangen door binnen de afkapgrenzen vallende elementen van een andere (contaminerende) populatie. Bij de bepaling van het contaminatiepercentage zou behalve met de afkapgrenzen ook rekening moeten worden gehouden met een aantal gegevens met betrekking tot de contaminerende populatie. Omdat de betreffende gegevens moeilijk aan de praktijk te ontfemen zijn, werd, enigszins arbitrair, uitgegaan van de volgende veronderstellingen:

1. De contaminerende populatie bezit, bij benadering, een verdeling volgens Gauss, waarvan de parameters bekend zijn.  
In de modelproeven werden als rekenvoorbeelden  $N(130,15)$  en  $N(115,5)$  gebruikt.
2. De verhouding van het totale aantal elementen van de contami-



nerende tot het totale aantal elementen van de oorspronkelijke populatie,  $N(100,5)$ , is bekend.

Als rekenvoorbeelden werden gekozen 1:2, 1:1 en 2:1.

In figuur 4-2 is de situatie weergegeven voor  $N(130,15)$  als contaminerende populatie en in figuur 4-3 voor  $N(115,5)$ . De berekende contaminatiepercentages worden gegeven in tabel 4-VI. Deze percentages werden, na afronding op een even getal, gebruikt in de volgende modelproeven.

Voor elke combinatie van:

1. één van de contaminerende populaties
2. één van de als voorbeeld gekozen verhoudingen tussen de totale aantallen van de twee populaties
3. één van de paren afkapgrenzen

werd een serie van 20 steekproeven verwerkt. Uitgegaan werd van steekproeven, die bestonden uit 50 binnen de afkapgrenzen vallende elementen uit  $N(100,5)$ . Hierin werd een, met het voor de betreffende combinatie berekende contaminatiepercentage overeenkomend aantal elementen vervangen door binnen de afkapgrenzen vallende elementen uit de contaminerende populatie  $N(130,15)$  of  $N(115,5)$ . Daarna werden enkele statistische grootheden berekend, waaronder de steekproefgemiddelden ( $M_{50}$ ) en, per serie, het gemiddelde ( $\bar{M}_{50}$ ) en de standaardafwijking ( $s_M$ ) van de steekproefgemiddelden. De resultaten zijn weergegeven in tabel 4-VIII. Bij gebruik van de afkapgrenzen 95 en 105 blijkt voor alle combinaties in deze serie de invloed van contaminatie op de steekproefgemiddelden verwaarloosbaar klein te zijn. Bij gebruik van de afkapgrenzen 90 en 110 begint zich een verschil af te tekenen tussen de twee contaminerende populaties. De verschuivingen van  $\bar{M}_{50}$  ten gevolge van contaminatie met elementen uit  $N(115,5)$  zijn duidelijk groter dan de overeenkomstige waarden bij contaminatie met elementen van  $N(130,15)$ . Het verschil is nog duidelijker bij gebruik van de afkapgrenzen 85 en 115.

Ter vergelijking van de bij deze contaminatieproeven optredende verschuivingen met die bij de modificatieproeven (tabel 4-V), zijn de resultaten van beide series samengevat in figuur 4-4. Voor elke combinatie is het gebied van  $\bar{M}_{50} - 2s_M$  tot  $\bar{M}_{50} + 2s_M$  weergegeven (verder aan te duiden als  $M_{50}$ -gebied).

Bij de keuze van afkapgrenzen spelen twee overwegingen een rol. Enerzijds is het gewenst, dat reeds een geringe modificatie van de ele-

menten (overeenkomende met een kleine systematische fout) duidelijk tot uitdrukking komt in de ligging van het steekproefgemiddelde. Dit betekent, dat bij voorkeur de overlapping van het  $M_{50}$ -gebied van de gemodificeerde en dat van de oorspronkelijke populatie bij een zo gering mogelijke modificatie moet zijn opgeheven. Voor de afkapgrenzen 95-105 is dit het geval bij een modificatie van +5, voor 90-110 bij +3 en voor 85-115 bij +2. Het laatste paar afkapgrenzen werkt dus in dit opzicht het gunstigst. Anderzijds dienen de afkapgrenzen ter beperking van de invloed van contaminatie en in dat opzicht werken ze beter naarmate ze minder ver uiteen liggen. Bij beschouwing van figuur 4-4 blijkt, dat de met deze modelproeven verkregen resultaten pleiten voor het gebruik van de afkapgrenzen 90 en 110 ( $\mu \pm 2\sigma$ ). Deze conclusie is, strikt genomen, alleen geldig voor dit modelsysteem en de gekozen getallenvoorbeelden.

#### 4.3. ORIENTERENDE TOEPASSING OP REEKSEN BEPALINGSUITKOMSTEN.

Bij de toepassing van de GvN-methode op reeksen bepalingssuitkomsten, zoals die dagelijks in een klinisch-chemisch laboratorium worden geproduceerd, kunnen zich problemen voordoen die bij de modelproeven geen of een onbelangrijke rol hebben gespeeld. Te denken valt bijvoorbeeld aan de onzekerheid over de mate van afwijking van het verdelingstype van dat van Gauss, zowel wat de oorspronkelijke als wat de contaminerende populatie betreft en aan de daarmee samenhangende onzekerheid omtrent de waarden van de parameters, de optimale afkapgrenzen en de contaminatiegraad.

Hoewel ook deze aspecten mogelijk door middel van een meer gecompliceerd modelsysteem te bestuderen zouden zijn, werd besloten eerst na te gaan of en, zo ja, in welke mate in de praktijk rekening moet worden gehouden met deze problematiek.

Ter orientatie werd daarom voor een aantal bepalingssorten de frequentieverdeling nagegaan van ongeveer duizend grotendeels ongeselecteerde uitkomsten. Hiervoor werd een programma geschreven, dat onder andere de achtereenvolgens ingevoerde bepalingssuitkomsten over dertig klassen verdeelt. Afgezien van de eerste en de laatste klasse, zijn de klassegrenzen zo gekozen dat, voor een bepalingssort de klassebreedte constant is en dat het in de klinische praktijk gehanteerde 'normale gebied' ruimschoots binnen de grenzen valt van het gebied, dat door de resterende 28 klassen wordt bestreken. De

verkregen getallen werden grafisch verwerkt tot gewone en tot relatieve cumulatieve frekwentieverdelingen. Enkele voorbeelden hiervan zijn weergegeven resp. in figuur 4-5 a t/m f en in figuur 4-6 a t/m f. Om alvast een indruk te krijgen van de dagelijkse schommelingen van de toetsingsgrootte werden per bepalingsoort voorlopige waarden gekozen voor de afkapgrenzen en werden de bepalinguitkomsten op datum gegroepeerd ingevoerd. Voor elke bepalingsoort werden van alle uitkomsten die binnen de grenzen vielen het groepsgemiddelde en de standaardafwijking berekend. Bovendien werd voor deze uitkomsten per dag het gemiddelde berekend en de afwijking hiervan ten opzichte van het groepsgemiddelde. Ook hiervan worden enkele voorbeelden weergegeven in figuur 4-7 a t/m f.

De frekwentieverdeling van de kaliumuitkomsten (fig 4-5a) vertoont, afgezien van een enigszins onregelmatig verloop, een redelijk symmetrische verdeling. Uit de cumulatieve verdelingscurve (fig 4-6a) blijkt dat deze uitkomsten bij benadering een verdeling volgens Gauss bezitten.

De frekwentieverdeling van de chlorideuitkomsten (fig 4-5b) is duidelijk asymmetrisch. De schouder in het linkergedeelte van de figuur wordt veroorzaakt door de relatief hoge frekwentie waarmee iets verlaagde chloride-concentraties worden gevonden bij patienten met uiteenlopende aandoeningen. De cumulatieve verdelingscurve (fig 4-6b) vertoont overigens zowel aan de lage als aan de hoge kant van de schaal afwijkingen ten opzichte van een verdeling volgens Gauss.

De frekwentieverdeling van de calciumuitkomsten (fig 4-5c) bezit een vrijwel symmetrische klokvorm. In tegenstelling tot de verwachting die hierdoor zou kunnen worden gewekt, vertoont de cumulatieve verdelingscurve (fig 4-6c) aan beide kanten van de schaal duidelijke afwijkingen ten opzichte van een verdeling volgens Gauss.

De frekwentieverdeling van de ureumuitkomsten (fig 4-5d) vertoont, vooral in het linkergedeelte, een opvallend onregelmatig verloop. Verder blijkt de verdeling duidelijk scheef te zijn. Dit wordt mede veroorzaakt doordat bij patienten vaker een verhoogde ureumconcentratie voorkomt dan een verlaagde.

De frekwentieverdeling van de alkalische fosfataseuitkomsten (fig 4-5e) vertoont enige overeenkomst met die van de ureumuitkomsten. Dit geldt ook voor de cumulatieve verdelingscurven (fig 4-6d en fig 4-6e). De frekwentieverdeling van de cholesteroluitkomsten (fig 4-5f) blijkt

op enkele onregelmatigheden na, een bijna symmetrische figuur op te leveren. Ook de cumulatieve verdelingscurve vertoont slechts geringe afwijkingen ten opzichte van een verdeling volgens Gauss.

De resultaten van de berekening van de daggemiddelden van de binnen de afkapgrenzen vallende uitkomsten zijn weergegeven in figuur 4-7 a t/m f. De berekende afwijkingen van de daggemiddelden ten opzichte van het betreffende groepsgemiddelde werden op twee manieren uitgezet, in de voor de bepaling gebruikelijke eenheid en ook, na vermenigvuldiging met  $\sqrt{n/s}$ , betrokken op de voor de betreffende bepaling en dag berekende standaardfout (standaardafwijking van het gemiddelde). Uit de figuren is af te lezen, dat er, bij met elkaar in overeenstemming gekozen schaalverdelingen, slechts weinig verschil is tussen de twee manieren van uitzetten.

#### 4.4. DISCUSSIE EN CONCLUSIE.

Door middel van de in dit hoofdstuk besproken modelproeven werd getracht inzicht te verkrijgen in enkele eigenschappen van de 'GvN'-methode van Hoffmann en Waid.

Aan de hand van het eerste model, de standaardnormale verdeling, werd de invloed onderzocht van de keuze van de afkapgrenzen op de verandering van het gemiddelde ten gevolge van, door modificatie van alle elementen, gesimuleerde systematische fouten. Hierbij bleek dat naarmate de afkapgrenzen ruimer worden gekozen, het quotient  $v/\Delta$  (hierin is  $v$  de verandering van het gemiddelde en  $\Delta$  de gesimuleerde systematische fout) een grotere waarde aanneemt. Hieruit kan worden geconcludeerd, dat in dit geval het gebruik van afkapgrenzen niet zinvol is.

Om het verwerken van series bepalingsuitkomsten te kunnen nabootsen werd gebruik gemaakt van een tweede model. Dit bestaat uit reeksen getallen, die door middel van een computer kunnen worden gegenereerd en die zijn te beschouwen als aselechte steekproeven uit een populatie, waarvan de verdeling binnen de mogelijkheden van het betreffende programma tevoren kan worden bepaald. In eerste instantie werd hiervoor een tweezijdig afgekapte normale verdeling gekozen. De resultaten van deze experimenten waren in overeenstemming met de uitkomsten van de berekeningen aan de hand van het eerste model. Hieruit mag worden geconcludeerd dat ook voor dit model in deze proefopzet de toepassing van afkapgrenzen niet zinvol is. Deze conclusie geldt

echter niet voor het gebruik van afkapgrenzen bij de GvN-methode, omdat bij de modelexperimenten tot nu toe geen rekening werd gehouden met het voorkomen van 'pathologische' uitkomsten in de reeksen bepalingssuitkomsten. Om ook de invloed van deze contaminatie na te gaan, werden met het tweede model experimenten verricht, waarbij de steekproeven werden samengesteld uit elementen van twee populaties, een hoofdpopulatie en een contaminerende populatie. Als hoofdpopulatie werd gekozen  $N(100,5)$  en als contaminerende populatie één van de populaties  $N(130,15)$  en  $N(115,5)$ . De keuze van de laatste twee berustte op de volgende overwegingen. Indien als bovenste afkapgrens achtereenvolgens 105, 110 en 115 wordt gebruikt, zou een veel groter verschil tussen de contaminerende en de hoofdpopulatie gepaard gaan met een te verwaarlozen contaminatie. Daarentegen zou een veel kleiner verschil leiden tot een zo sterke contaminatie dat de diagnostische interpretatie van de betreffende bepalingssuitkomsten grote moeilijkheden zou opleveren.

De met het tweede model verkregen resultaten zijn samengevat in figuur 4-4. Hieruit blijkt dat het gebruik van afkapgrenzen op een afstand van twee standaardafwijkingen van het gemiddelde van de hoofdpopulatie gunstig zal kunnen werken. Hierdoor wordt de invloed van contaminatie redelijk beperkt, terwijl de gevoeligheid voor modificatie (als simulatie van het optreden van een systematische fout) niet al te sterk wordt aangetast. Strikt genomen geldt deze beschouwing alleen voor het gebruikte modelsysteem en de gekozen getallenvoorbeelden. Maar deze beperking zal, gezien de overwegingen die tot deze keuze hebben geleid en de resultaten van de modelproeven, met enige soepelheid kunnen worden gehanteerd en het leek daarom zeker de moeite waard de GvN-methode in de praktijk op bruikbaarheid te onderzoeken. Bij de oriënterende toepassing op reeksen bepalingssuitkomsten werd geconstateerd, dat de cumulatieve frequentieverdelingen van deze reeksen vaak duidelijk afwijken van een verdeling volgens Gauss. Als voorbeelden van betrekkelijk geringe afwijkingen kunnen de verdelingen van de kalium- en de cholesteroluitkomsten worden genoemd (fig 4-6 a en f). De verdelingen van de uitkomsten van ureum en alkalische fosfatase vertonen een afbuiging aan de hoge kant van de schaal, terwijl die van chloride en calcium aan beide zijden van de schaal afbuigen. Aangezien in elke reeks bepalingssuitkomsten uiteraard ook een onbekend aantal pathologische

uitkomsten was opgenomen, konden de parameters van de verdelingen van de 'normale' uitkomsten niet zonder meer worden berekend. Daarom werd besloten tot onderzoek van enkele, in hoofdstuk 3 genoemde, methoden voor de schatting van parameters.

## HOOFDSTUK 5.

### ONDERZOEK VAN ENKELE METHODEN VOOR DE SCHATTING VAN PARAMETERS.

#### 5.1. INLEIDING.

Zoals in hoofdstuk 3 werd geconstateerd, vormen de 'normale waarden voor klinisch-chemische bepalingssuitkomsten nog steeds een actueel probleem. Ook indien problemen, die samenhangen met de terminologie of met de begripsomschrijving, buiten beschouwing worden gelaten, levert de keuze van de werkwijze voor het vaststellen van de grenzen van het "normale gebied" nog al eens moeilijkheden op.

Door Hoffmann & Waid (1965) is in het kader van de 'gemiddelde van normalen'-methode gewezen op een betrekkelijk eenvoudige, grafische methode, die door Hoffmann (1963) was beschreven en waarbij kan worden uitgegaan van een reeks ongeselecteerde bepalingssuitkomsten (zie p 43). Zowel de aard van het uitgangsmateriaal als de eenvoud van uitvoering maken de methode aantrekkelijk, maar de juistheid van de resultaten is door sommige auteurs (Neumann, 1968; Elveback et al., 1970) in twijfel getrokken. Neumann (1968) heeft een modificatie voorgesteld, die tot betere resultaten zou leiden. Gindler (1970) gebruikte de door Bhattacharya beschreven methode (1967) voor de schatting van de parameters van de 'normale populatie' in een reeks ongeselecteerde bepalingssuitkomsten. Vooral voor tweezijdig gecontamineerde reeksen verdient, volgens Gindler, de laatste methode de voorkeur.

Om een beter inzicht te verkrijgen in de mogelijkheden en beperkingen van de bovengenoemde methoden werden deze eerst toegepast op enkele modellen en daarna op reeksen bepalingssuitkomsten. In een later stadium werd besloten nog een andere methode in het onderzoek te betrekken. Deze laatste methode (Hald, 1952), die kan worden aangeduid als de 'best passende parabool'-methode (bpp-methode), is niet in hoofdstuk 3

genoemd omdat deze, voor zover bekend, nog niet eerder is toegepast op klinisch-chemische bepalingssuitkomsten. In 5.4 zal nader op deze methode worden ingegaan.

## 5.2. MODELPROEVEN.

### 5.2.1. Toepassing op modellen, die gebaseerd zijn op de standaardnormale verdeling.

Voor de eerste groep modelproeven werd gebruik gemaakt van een frequentieverdeling, die als volgt werd verkregen uitgaande van de standaardnormale verdeling. Het gebied van de z-as werd ingedeeld in klassen met een breedte van 0,2. Met behulp van een tabel werd voor iedere klasse de oppervlakte berekend van het gedeelte van de standaardnormale verdeling, dat tussen de klassegrenzen valt. Deze klasseoppervlakten (4 decimalen) werden door vermenigvuldiging met 10.000 omgezet in de overeenkomstige klassefrequenties. De ligging van de klassegrenzen t.o.v. het nulpunt heeft bij deze klassebreedte slechts een geringe invloed op het frekwentiediagram (fig 5-1a t/m c), bij een klassebreedte van 0,4 is deze invloed duidelijk zichtbaar (fig 5-1d t/m f). De cumulatieve frequentieverdeling van dit model geeft, op waarschijnlijkheidspapier uitgezet, een reeks punten, die op een rechte lijn liggen (fig 5-2a). Bij het op deze wijze weergeven van bepalingssuitkomsten wordt vaak gebruik gemaakt van afkapgrenzen om de contaminatie met pathologische uitkomsten te elimineren of althans te beperken. In figuur 5-2b t/m d wordt achtereenvolgens het effect getoond van het gebruik van de afkapgrenzen -3 en +3, -2 en +2 en tenslotte -1 en +1. Het blijkt, dat deze afkapgrenzen, die symmetrisch t.o.v. het midden van de figuur liggen, een afbuiging aan beide uiteinden en een verandering van de helling van het rechte gedeelte tot gevolg hebben. De ligging van het 50%-punt wordt echter niet beïnvloed. Het gebruik van asymmetrisch gelegen afkapgrenzen (-3 en +2, -3 en +1) leidt behalve tot afbuiging en hellingsverandering ook tot een geringe verschuiving van het 50%-punt (fig 5-2e en f).

De door Neumann (1968) voorgestelde modificatie biedt de mogelijkheid iteratief correcties aan te brengen, waardoor de juiste helling en de positie van het oorspronkelijke 50%-punt benaderd kunnen worden. In figuur 5-3 is een gedeelte van figuur 5-2f op een grotere



schaal weergegeven. De extrapolatie van het rechte gedeelte bereikt in  $A_1$  de afkappgrens. Het bij dit punt behorende percentage wordt afgelezen (90%) en gebruikt voor de berekening van de gecorrigeerde frekwentieverdeling. Deze geeft bij uitzetten en extrapolatie van het rechte gedeelte het punt  $A_2$ . Het hierbij behorende percentage (86 %) scheelt meer dan 2% met dat van  $A_1$ , zodat de procedure dient te worden herhaald. Dit levert een relatieve cumulatieve frekwentieverdeling op, waarvan het rechte gedeelte bij extrapolatie het punt  $A_3$  geeft. Het bij  $A_3$  behorende percentage (85%) scheelt minder dan 2% met dat van  $A_2$ , zodat we, volgens Neumann, nu kunnen stoppen. Het 50%-punt van de laatste frekwentieverdeling ligt op  $-0.05$ , wat als een goede benadering van nul kan worden beschouwd. Indien de procedure toch nog een keer wordt herhaald, wordt er een relatieve cumulatieve frekwentieverdeling verkregen, die bij extrapolatie van het rechte gedeelte het punt  $A_4$  (84,2%) oplevert. Het 50%-punt ligt praktisch op nul.

Uit deze resultaten kan worden geconcludeerd, dat de modificatie van Neumann met dit model inderdaad goede resultaten levert.

De methode van Bhattacharya (1967) houdt in, dat voor de opeenvolgende klassen van een frekwentieverdeling de verhouding  $Y_{(x+h)}/Y_{(x)}$  wordt berekend. Hierin is  $y_{(x)}$  gelijk aan de frekwentie in de klasse, waarvan het klassemidden gelijk is aan  $x$  en de klassebreedte aan  $h$ ;  $Y_{(x+h)}$  is dan de frekwentie in de daarop volgende klasse.

Uitgaande van de formule

$$y = \int_{(x-h/2)}^{(x+h/2)} \frac{N}{\sigma_r \sqrt{2\pi}} e^{-(v-\mu_r)^2 / 2\sigma_r^2} dv$$

leidt Bhattacharya af, dat de natuurlijke logaritmen van deze quotienten bij benadering gegeven worden door

$$\ln \frac{Y_{(x+h)}}{Y_{(x)}} = \Delta \ln y = \frac{-h(\sigma_r^2 - h^2/12)(x - \mu_r + h/2)}{\sigma_r^4}$$

Bij uitzetten van  $\Delta \ln y$  tegen  $x$  wordt een rechte verkregen met een negatieve richtingscoëfficiënt. Als  $\lambda$  de waarde van  $x$  is, waarbij  $\Delta \ln y$  gelijk is aan nul, geldt:

$$\lambda - \mu_r + h/2 = 0$$

$$\mu = \lambda + h/2$$

Voor de variantie kan, ook bij benadering, worden afgeleid:

$$\sigma^2 = - \frac{\Delta x}{\Delta(\Delta \ln y)} h - \frac{h^2}{12}$$

In de publicatie van Gindler is het eerste minteken in deze formule (abusievelijk?) weggefallen.

Toepassing van de methode op het eerste model geeft bij een klassebreedte van 0,2 een reeks punten (fig 5-4a), die enige spreiding rond een rechte vertonen. Dit is waarschijnlijk het gevolg van door afronding geïntroduceerde fouten. Bij een klassebreedte van 0,5 is de invloed hiervan geringer als gevolg van de hogere klassefrequenties (fig 5-4b).

Het gebruik van afkapgrenzen zal bij deze methode slechts het aantal punten verminderen, maar de positie van de overblijvende niet beïnvloeden.

Voor de volgende groep modelproeven werden frekwentieverdelingen gebruikt, die werden samengesteld door bij de klassefrequenties van model 1 de binnen dezelfde klassegrenzen vallende klassefrequenties op te tellen van een contaminerende populatie met een normale verdeling. Een soortgelijk model werd reeds eerder toegepast in 4.2.3. (zie ook figuren 4-2 en 4-3). Als voorbeelden van contaminerende populaties werden gekozen  $N(6,3)$  en  $N(3,1)$ . Om een duidelijk effect van de contaminatie te verkrijgen, werd voor de omrekening van oppervlakte naar klassefrequentie voor de contaminerende populaties de faktor 20.000 gebruikt en voor de hoofdpopulatie 10.000.

In figuur 5-5a en b worden de relatieve cumulatieve frekwentieverdelingen gegeven voor dit model met resp.  $N(6,3)$  en  $N(3,1)$  als contaminerende populatie. Voor de toepassing van de correctiemethode van Neumann lezen we uit deze figuren de te gebruiken afkapgrenzen af; deze zijn 1,3 (voor 5-5a) en 0,5 (voor 5-5b). De verdere gang van zaken is weergegeven in de figuren 5-5c en 5-5d. Hierbij valt op dat in 5-5d het punt (0,50) in vier stappen goed benaderd wordt, terwijl in 5-5c als het ware over het doel heen geschoten wordt. Bij de tweede iteratie gaat de rechte dichter langs het punt (0,50) dan bij volgende stappen het geval is. De verklaring hiervoor moet worden gezocht in het verschil in de verdeling van de contaminerende elementen over het beschouwde gebied van -3 tot +3. Terwijl de elementen uit  $N(6,3)$  zich zowel links als rechts van het nulpunt

zullen bevinden, zijn die uit  $N(3,1)$  vrijwel uitsluitend in het gebied rechts van het nulpunt geconcentreerd (verg. fig 4-2 en 4-3). Toepassing van de methode van Bhattacharya op deze frekwentieverdelingen levert eveneens een iets te hoge schatting van  $\mu_1$  (+0.06) voor de frekwentieverdeling die gecontamineerd is met  $N(6,3)$ , terwijl de contaminatie met  $N(3,1)$  de schatting van  $\mu_1$  niet merkbaar beïnvloedt (fig 5-6a resp b).

### 5.2.2. Toepassing op grote steekproeven.

Voor deze groep modelproeven werd weer gebruik gemaakt van de mogelijkheid om door middel van een computer reeksen getallen te genereren, die te beschouwen zijn als a-selecte steekproeven uit een populatie, die een verdeling volgens Gauss bezit en waarvan de parameters vooraf kunnen worden gekozen.

In de programma's, die voor de verdere verwerking van deze getallen werden geschreven, werd de mogelijkheid opgenomen om gecontamineerde steekproeven samen te stellen, door een deel van de elementen van de steekproef afkomstig te laten zijn uit een andere populatie. Verder werden de getallen geselecteerd op ligging binnen de eveneens te voren te kiezen afkapgrenzen, afgerond tot gehele getallen en gesorteerd in klassen.

In de proeven werd gewerkt met een 'hoofdpopulatie'  $N(100,5)$  en eventueel, als bron van contaminatie, één van de populaties  $N(130,15)$  en  $N(115,5)$ . Het contaminatiepercentage werd gekozen in overeenstemming met de waarde, die in 4.2.3. werd berekend voor  $Q = 2$ . Als afkapgrenzen werden gebruikt de combinaties 85.0 en 115.0 en 90.0 en 110.0. Het aantal getallen per experiment bedroeg 1000. Tenslotte werd een frekwentiediagram gemaakt met een klassebreedte  $1 (\approx 0.2 \sigma)$ .

In fig 5-7a t/m f zijn de relatieve cumulatieve frekwentieverdelingen weergegeven van steekproeven uit:

- |                                      |              |
|--------------------------------------|--------------|
| a) 100% $N(100,5)$                   |              |
| b) 76% $N(100,5)$ en 24% $N(130,15)$ | Afkapgrenzen |
| c) 50% $N(100,5)$ en 50% $N(115,5)$  | 85-115       |
| d) 100% $N(100,5)$                   |              |
| e) 84% $N(100,5)$ en 16% $N(130,15)$ | Afkapgrenzen |
| f) 76% $N(100,5)$ en 24% $N(115,5)$  | 90-110       |

De resultaten vertoonden grote overeenkomst met de resultaten die met het tweede model in 5.2.1. werden verkregen. Daarom werd afgezien van het wederom toepassen van de correctiemethode volgens Neumann.

Toepassing van de methode van Bhattacharya op dit model gaf in eerste instantie zeer teleurstellende resultaten. In fig 5-8a en b zijn deze weergegeven voor 1000 getallen uit  $N(100,5)$  met als afkapgrenzen rep. 85-115 en 90-110. De methode blijkt zeer gevoelig te zijn voor kleine onregelmatigheden in de frekwentieverdeling. Van een lineair verband is bij deze niet gecontamineerde steekproeven al niets meer te zien. Naast de te verwachten waarde wordt een aantal schijnbare populatiegemiddelden gevonden, waarbij echter in geen enkel geval meer dan twee opeenvolgende punten op één lijn liggen. Om de invloed van de onregelmatigheden te verminderen, werd de klasbreedte verdubbeld (van  $0,2\sigma$  tot  $0,4\sigma$ ). Hoewel voor beide steekproeven een minder onregelmatig beeld wordt verkregen (fig 5-8c en 8d) en geen extra populatiegemiddelde wordt gevonden, liggen er nog steeds maar twee punten op de rechte, die de twee opeenvolgende ter weerszijden van de X-as liggende punten verbindt. De bij toepassing op een tweetal gecontamineerde steekproeven verkregen resultaten (fig 5-8 e en f) stemmen ruwweg overeen met wat op grond van de voorgaande experimenten kon worden verwacht. Om de resultaten, die met deze methode kunnen worden verkregen, beter te kunnen beoordelen worden enkele op de figuren gebaseerde schattingen van waarden voor  $\mu$  en  $\sigma$  vergeleken met de juiste waarden:

	$\mu$	$\sigma$
Juiste waarde	100.0	5.0
fig 5-8c	99.8	3.7
5-8d	100.0	3.7
5-8e	99.8	5.0
5-8f	100.6	3.8

Hieruit kan worden geconcludeerd, dat deze methode een redelijke schatting van het populatiegemiddelde mogelijk maakt, mits de klassen voldoende breed worden gekozen. De schatting van de standaardafwijking is nog gevoeliger voor de toevallige variaties van de klassefrekwenties dan de schatting van het gemiddelde.

### 5.3. TOEPASSING OP REEKSEN BEPALINGSUITKOMSTEN.

Hiervoor werd uitgegaan van de in 4.3. verkregen frekwentieverdelingen (zie fig 4-5 en -6). Bij toepassing van de methode van Hoffmann op de cumulatieve frekwentieverdeling van de chlorideuitkomsten werden zowel voor de schatting van het gemiddelde als voor die van de standaardafwijking twee waarden verkregen als gevolg van de duidelijke knik in de figuur. Verder verliepen de uitvoering van de methode van Hoffmann en van de door Neumann voorgestelde iteraties geheel volgens de reeds beschreven procedure (zie 5.2.1. en fig 5-3). Daarom wordt hier volstaan met vermelding van de resultaten tesamen met die van de andere methoden (tabel 5-I).

Bij de toepassing van de methode van Bhattacharya werd, zoals min of meer te verwachten was, in elke frekwentieverdeling een aantal schijnbare populatiegemiddelden gevonden. Verdubbeling van de klassebreedte bracht wel enige verbetering (vergelijk fig 5-9 en -10), maar ook hier was er slechts bij uitzondering iets te merken van een lineair verband. Ook de met deze methode verkregen schattingen zijn opgenomen in tabel 5-I. Hierbij dient te worden opgemerkt, dat op grond van de met de modelproeven verkregen resultaten (zie 5.2.2.) aan de schattingen van de standaardafwijkingen slechts weinig waarde moet worden toegekend. Mede in verband hiermee werd besloten nog een methode in het onderzoek te betrekken.

### 5.4. DE 'BEST PASSENDE PARABOOL'-METHODE.

Indien  $n$  elementen uit een normale verdeling, met  $\xi$  en  $\sigma$  als parameters, worden verdeeld over een aantal klassen met klassebreedte  $\Delta t$ , geldt voor de frekwentie  $a_j$  in de klasse  $j$  (met het klassemiddelen  $t_j$ ) volgens Hald (1952) bij benadering:

$$a_j \approx n \int_{t_j - \Delta t/2}^{t_j + \Delta t/2} p\{x\} dx$$

$$a_j \approx n \cdot \Delta t \cdot p\{t_j\}$$

$$a_j \approx \frac{n \cdot \Delta t}{\sigma \cdot \sqrt{2\pi}} e^{-(t_j - \xi)^2 / 2\sigma^2}$$

Het nemen van de logarithme leidt dan tot:

$$\ln a_j \approx \ln\left(\frac{n \cdot \Delta t}{\sigma \cdot \sqrt{2\pi}}\right) - (t_j - \xi)^2 / 2\sigma^2 \quad \text{I}$$

of tot:

$$\log a_j \approx \log\left(\frac{n \cdot \Delta t}{\sigma \cdot \sqrt{2\pi}}\right) - 0.4343 (t_j - \xi)^2 / 2\sigma^2$$

Bij uitzetten van  $a_j$  tegen  $t_j$  op semilogaritmisch papier zal dus de reeks punten gespreid liggen langs een parabool, waarvan de as van symmetrie wordt gegeven door  $t = \xi$ .

Hald geeft aan dat we hiervan gebruik kunnen maken voor een schatting langs grafische weg van  $\xi$  door de 'best passende parabool' te tekenen. Het is op deze wijze zelfs mogelijk de voor een heterogene populatie verkregen figuur op te splitsen in de samenstellende parabolen.

Omdat het tekenen van de op het oog 'best passende parabool' vooral bij een wat grotere spreiding van de punten weinig betrouwbaar leek, meenden wij de voorkeur te moeten geven aan een berekening van de coëfficiënten van de betreffende functie door middel van de methode van de kleinste kwadraten. Aangezien de toepassing van deze methode op kwadratische functies nogal veel rekenwerk met zich meebrengt, werd hiervoor gebruik gemaakt van een computer.

Formule (I) kan ook als volgt worden geschreven:

$$\ln a_j \approx -1/2\sigma^2 \cdot t_j^2 + \xi/\sigma^2 \cdot t_j - \xi^2/2\sigma^2 + \ln\left(\frac{n \cdot \Delta t}{\sigma \cdot \sqrt{2\pi}}\right)$$

Als we over een aantal punten  $(t_j, \ln a_j)$  beschikken, kunnen voor bovenstaande vergelijking met de methode van de kleinste kwadraten de coëfficiënten  $B_1$ ,  $B_2$  en  $B_3$  worden berekend. Hiervoor gelden de volgende relaties:

$$B_1 = -\xi^2/2\sigma^2 + \ln\left(\frac{n \cdot \Delta t}{\sigma \cdot \sqrt{2\pi}}\right)$$

$$B_2 = \xi/\sigma^2$$

$$B_3 = -1/2\sigma^2$$

Achtereenvolgens kunnen dan  $\sigma$ ,  $\xi$  en  $n$  worden berekend:

$$\sigma^2 = -1/2B_3 \quad \rightarrow \quad \sigma = \sqrt{-1/2B_3}$$

$$\xi = B_2 \cdot \sigma^2 = -B_2/2 \cdot B_3$$

$$\ln n = B_1 + \xi^2/2\sigma^2 + \ln \sigma - \ln \frac{\Delta t}{\sqrt{2\pi}}$$

$$\ln n = B_1 - (B_2)^2/4B_3 - \frac{1}{2}\ln(-2.B_3) + \ln\frac{\sqrt{2\pi}}{\Delta t}$$

$$n = e^{\ln n}$$

Bij de toepassing van de bpp-methode op enkele frekwentieverdelingen van het modelsysteem uit 5.2.2. bleek, dat deze methode weliswaar minder gevoelig is voor 'random' variaties van de klassefrequenties dan de methode van Bhattacharya, maar dat de invloed van contaminatie beduidend groter is. Dit geldt zowel voor de schatting van het gemiddelde als voor die van de standaardafwijking. Ter vergelijking zijn de in 5.2.2. met de methode van Bhattacharya verkregen resultaten, tussen haakjes, vermeld achter de resultaten, die met de 'best passende parabool'-methode werden verkregen:

Populatie	schattingen van	
	$\mu$	$\sigma$
100% N(100,5)	99.8	4.8
76% N(100,5) + 24% N(130,15)	101.4 (99.8)	6.0 (5.0)
50% N(100,5) + 50% N(115,5)	102.0 (100.6)	6.2 (3.8)

De resultaten van toepassing op reeksen bepalinguitkomsten zijn opgenomen in tabel 5-I.

#### 5.5. DISCUSSIE EN CONCLUSIE.

Bij vergelijking van de resultaten van de vier toegepaste methoden, die in tabel 5-I zijn weergegeven, blijken deze voor de kalium-, voor de chloride- en voor de calciumuitkomsten een redelijke overeenstemming te vertonen.

Voor de andere bepalingen blijken de met de methode van Hoffmann verkregen schattingen van de betreffende  $\mu$  hoger te zijn dan de waarden, die met de andere methoden zijn verkregen. Dit hangt samen met het asymmetrische karakter van deze frekwentieverdelingen, dat, althans voor een deel, wordt veroorzaakt door éénzijdige contaminatie. Het aantal patiënten met 'hoge' uitkomsten is voor deze bepalingen als regel vele malen groter dan het aantal patiënten met 'lage' uitkomsten.

Wat de met de methode van Bhattacharya verkregen schattingen van de standaardafwijking betreft, is reeds eerder opgemerkt, dat deze sterk worden beïnvloed door al dan niet toevallige onregelmatigheden in de frekwentieverdeling. Dit zou mogelijk op te vangen zijn door

verdere verbreding van de klassen.

De met de methode van Hoffmann-Neumann verkregen tamelijk hoge schatting van de standaardafwijking van de chlorideuitkomsten is waarschijnlijk beïnvloed door de vrij sterke contaminatie aan de lage kant van deze frekwentieverdeling.

Tenslotte dient te worden opgemerkt dat de uitkomsten van de 'best passende parabool'-methode nogal sterk af kunnen hangen van de keuze van de punten die voor de berekening worden gebruikt. Om een indruk te verkrijgen van de mate van overeenstemming tussen de empirische en de, op basis van de 'best passende parabool', te berekenen klassefrequenties kunnen beide frequenties worden uitgezet op semilogaritmisch papier.

Als voorbeeld van een redelijke overeenstemming tussen de met behulp van een 'best passende parabool' berekende en de empirische klassefrequenties zijn deze beide voor de calciumuitkomsten weergegeven in figuur 5-11a. In het centrale gebied, van ongeveer 2,15 mmol/l tot ongeveer 2,70 mmol/l, kunnen de verschillen tussen de berekende en de empirische punten als een gevolg van 'random'variatiaties worden beschouwd. Buiten dat gebied kunnen ze worden verklaard door de aanwezigheid van 'pathologische' uitkomsten.

In figuur 5-11b zijn voor de ureumuitkomsten, naast de empirische klassefrequenties, twee berekende parabolen uitgezet. Zoals in de figuur is aangegeven werd voor de berekening van één van de parabolen uitgegaan van de eerste veertien punten van de empirische frekwentieverdeling. Voor de berekening van de andere parabool werd gebruik gemaakt van het derde tot en met het elfde punt van de empirische frekwentieverdeling. Uiteraard zouden er nog vele andere parabolen te berekenen zijn door uit te gaan van andere combinaties van minimaal drie punten. Bij het ontbreken van voldoende informatie omtrent het verdelingstype van 'normale'(niet pathologische) uitkomsten en omtrent de aard en de graad van contaminatie zal de keuze van de punten een arbitrair karakter bezitten. Ook voor de beslissing over verwerping of acceptatie van de berekende parabolen ontbreken theoretisch gefundeerde criteria. We kunnen ons hierbij slechts laten leiden door pragmatische overwegingen, die dan wel beïnvloed zullen zijn door tevoren verworven 'kennis' over de ligging van het 'normale' gebied. In de volgende figuren (5-11c t/m g) worden nog enkele voorbeelden gegeven. In twee gevallen (één berekende parabool



bij de zure fosfatase- en één bij de lactaatdehydrogenaseuitkomsten) is uit het verloop van de berekende parabool ten opzichte van de empirische punten duidelijk af te lezen, dat de keuze van de voor de berekening gebruikte punten niet juist is geweest. De andere voorbeelden illustreren nogmaals het besproken probleem.

Samenvattend kan worden gesteld, dat enerzijds geen van de onderzochte methoden in alle modelproeven goede resultaten opleverde, anderzijds leverde ook geen van de methoden over de hele linie slechte resultaten. Als de parameterschattingen alleen bedoeld zijn voor het vaststellen van afkapgrenzen ten behoeve van de methode van Hoffmann & Waid, lijken alle methoden bruikbaar, met een duidelijk voorbehoud wat de schatting van de standaardafwijking met de methode van Bhattacharya betreft. Indien een grotere nauwkeurigheid van de parameterschattingen gewenst is, lijkt het zinvol tenminste twee methoden toe te passen en bij het optreden van discrepanties tussen de resultaten de oorzaak hiervan op te sporen.

Bij de toepassing op reeksen bepalingsuitkomsten is het voornaamste probleem het ontbreken van goede criteria, waarmee de resultaten zouden kunnen worden getoetst. Een pragmatische benadering lijkt voorlopig aangewezen.

## HOOFDSTUK 6.

### DE 'GEMIDDELDE VAN NORMALEN'-METHODE IN DE PRAKTIJK.

#### 6.1. INLEIDING.

Door middel van modelproeven was enig inzicht verkregen in de eigenschappen van de 'gemiddelde van normalen'- en andere, daarmee verwante methoden. Bij de oriënterende toepassing op reeksen bepalingen-uitkomsten deed zich een praktisch probleem voor, dat bij de modelproeven geen rol had gespeeld. Hoewel de berekening van gemiddelden en, eventueel, standaardafwijkingen eenvoudig van aard is, werd hiervoor gebruik gemaakt van een computer met het oog op het aantal en de lengte van de series bepalingen-uitkomsten. Hieruit vloeide voort dat het uitgangsmateriaal in machinaal leesbare vorm moest worden gebracht. In dit geval betekende het dat de in schriften genoteerde bepalingen-uitkomsten in kaarten moesten worden geponst. Dit bleek zo tijdrovend, dat het de invoering van de methode in de dagelijkse kwaliteitscontrole vrijwel onmogelijk maakte. Ook de voortzetting van het onderzoek over een langere periode werd er door bemoeilijkt. Daarom werd dankbaar gebruik gemaakt van de mogelijkheid om dit gedeelte van het onderzoek te verrichten in samenwerking met het Centraal Klinisch Chemisch Laboratorium van het Academisch Ziekenhuis Dijkzicht, waar de administratieve verwerking van de laboratorium-uitkomsten reeds enige tijd door middel van een computer plaats vond.

Van 20 mei tot 1 december 1974 werden dagelijks voor 16 bepalingen de uitkomsten van dit laboratorium in een ponsband vastgelegd. Omdat de IBM-1800 configuratie van de Centrale computerafdeling van het Academisch Ziekenhuis Utrecht geen ponsbandlezer omvat, werden deze banden verwerkt op de DEC-PDP 15 van de vakgroep Cardiologie van de Rijksuniversiteit Utrecht.

## 6.2. VERWERKING VAN HET MATERIAAL EN BESPREKING VAN DE RESULTATEN.

Begonnen werd met een onderzoek naar de frekwentieverdeling en de parameters. In fig 6-1 worden enkele voorbeelden van frekwentieverdelingen gegeven. Hiervoor zijn dezelfde bepalingssorten gekozen als bij het oriënterende onderzoek. Hierbij moet worden opgemerkt dat niet de absolute, maar de relatieve frekwentie is uitgezet, omdat de absolute totale aantallen nogal sterk verschillen. Verder is van belang, dat in het CKCL-Dijkzicht voor de bepaling van de alkalische fosfatase activiteit een andere bepalingmethode wordt gebruikt dan in ons laboratorium, waardoor de twee reeksen uitkomsten van deze bepaling duidelijk verschillen.

In fig 6-2 zijn de cumulatieve frekwentieverdelingen weergegeven. De schatting van de parameters werd uitgevoerd met de methode van Hoffmann en met de 'best passende parabool'-methode. De resultaten hiervan zijn weergegeven in fig 6-2, fig 6-3 en tabel 6-I. De, met de twee methoden verrichte, schattingen van het gemiddelde voor de ureum- en voor de alkalische fosfataseuitkomsten bleken nogal uiteen te lopen. In verband met het scheve karakter van deze frekwentieverdelingen werd nagegaan of een verbetering kon worden bereikt door de werkwijze volgens Hoffmann te koppelen aan een logaritmische transformatie van de uitkomsten (zie fig 6-4 en tabel 6-I). Hoewel de transformatie in beide gevallen het lineaire gedeelte van de cumulatieve verdelingscurve vergrootte, bleken de nieuwe schattingen van het gemiddelde nauwelijks te verschillen van de waarden die zonder transformatie waren verkregen. In tabel 6-I zijn naast de schatting van  $\mu$  (het 50% punt) in plaats van de standaardafwijking de uit fig 6-4 afgelezen schattingen van het 5%- en 95%punt vermeld. Hierna werden voor elke bepaling afkapgrenzen gekozen en werden de gegevens op bepaling en datum gegroepeerd verwerkt. Voor de alkalische fosfatasebepaling werd de keuze van de afkapgrenzen gebaseerd op de met de 'best passende parabool'-methode gevonden waarden voor het gemiddelde en de standaardafwijking.

Behalve het daggemiddelde D van de uitkomsten, die op de betreffende dag tussen de afkapgrenzen vielen, werd nog een aantal andere gegevens berekend en/of vastgelegd. Hiertoe behoorden de standaardafwijking ( $s_x$ ) en het aantal (N) van de uitkomsten, die voor de berekening van het gemiddelde waren gebruikt. Later werden, voor de uitkomsten

uit de maanden mei (vanaf 20-5) en juni, de gemiddelden van de per dag berekende standaardafwijkingen en ook de gemiddelden en standaardafwijkingen van de aantallen uitkomsten en van de daggemiddelden berekend (resp  $s_x$ ,  $\bar{N}$ ,  $s_N$ , D en  $s_D$ ). Een overzicht van deze (afgeronde) gegevens wordt gegeven in tabel 6-II.

Vervolgens werden op basis van de voor de standaardafwijking ( $s_D$ ) gevonden waarden tolerantiegrenzen voor de grootheid D vastgesteld ( $\bar{D} \pm 2 s_D$ ). Voor alle maanden werd nu nagegaan hoeveel D-waarden onder de lage en hoeveel er boven de hoge tolerantiegrens liggen. In tabel 6-III zijn deze aantallen vermeld in de kolommen L en H. Tevens worden in deze tabel opgegeven de per maand berekende gemiddelden en standaardafwijkingen van de D-waarden.

Wat de kaliumuitkomsten betreft kan worden gesteld, dat deze bepaling, mogelijk met enig voorbehoud t.a.v. de  $\bar{D}$ -waarde voor November, een stabiele indruk maakt.

Uit de resultaten voor de chloridebepaling moet worden geconcludeerd dat het voor de eerste periode gevonden gemiddelde waarschijnlijk iets (ongeveer 0,6 mmol/l) te hoog ligt.

Voor de calciumbepaling geldt, dat de resultaten tot half november een stabiele indruk maken, maar dan een abrupte daling naar een lager niveau aantonen. Hierop zal later nog worden teruggekomen.

De resultaten voor de ureumbepaling geven geen aanleiding tot commentaar.

Uit de resultaten voor de bepaling van de alkalische fosfataseactiviteit kan worden afgeleid, dat het voor de eerste periode gevonden gemiddelde iets lager ligt dan overeen zou stemmen met de andere waarden.

De resultaten voor de cholesterolbepaling doen een ontsporing vermoeden in de maand juli.

In verband met het feit dat eind oktober overgegaan moest worden op een nieuwe batch controleserum werd besloten de uitkomsten uit de maanden oktober en november nader te onderzoeken. Hiertoe werden de D-waarden uit die periode uitgezet samen met de uitkomsten van het controleserum (zie fig 6-5a t/m f). Voor beide reeksen zijn de tolerantiegrenzen aangegeven met stippellijnen.

De uitkomsten van de kaliumbepaling in het controleserum vertonen

van eind oktober tot half november inderdaad een verloop, dat wijst in de richting van tijdelijke moeilijkheden. De D-waarden bleven echter, met één uitzondering begin oktober, binnen de tolerantiegrenzen. Dit geldt niet voor de D-waarden van de chloridebepaling, waarvan er in de periode van eind oktober tot half november een relatief groot aantal onder de lage tolerantiegrens valt. De chloride-uitkomsten van het controleserum zouden, op zich zelf beschouwd, geen reden tot ongerustheid zijn geweest.

De uitkomsten van de calciumbepaling weerspiegelen mogelijk een geringe invloed van de overgang op het nieuwe controleserum, maar tevens een veel duidelijker verschuiving op 14 en 15 november. Deze verschuiving is een gevolg van een wijziging, waartoe werd besloten op grond van, langs andere weg verkregen, informatie over de concentraties van standaard- en controlesera.

De resultaten voor de ureumbepaling geven een kleine verschuiving van de controleserumuitkomsten te zien. De D-waarden blijven de gehele periode binnen de tolerantiegrenzen.

Wat de alkalische fosfatase- en de cholesterolbepaling betreft, is er geen merkbaar effect van de invoering van het nieuwe controleserum.

Tenslotte kan nog worden opgemerkt, dat er slechts bij uitzondering verband lijkt te bestaan tussen de uitkomsten van het controleserum en de D-waarden van dezelfde dag.

### 6.3. SLOTBESCHOUWING.

In hoofdstuk 1 werd als eerste doel gesteld het verkrijgen van inzicht in de achtergronden van methoden voor kwaliteitscontrole, die gebruik maken van uitkomsten van de aangevraagde bepalingen. Wij menen er in geslaagd te zijn, door middel van de in hoofdstuk 4 beschreven modelproeven, inzicht te verwerven in een aantal belangrijke eigenschappen van dit type methoden. Zo kon, bij afwezigheid van contaminatie met pathologische uitkomsten, worden vastgesteld dat het gebruik van afkapgrenzen de gevoeligheid van de toetsingsgrootte voor systematische fouten vermindert. Dit effect is sterker, naarmate het gebied tussen de afkapgrenzen nauwer wordt gekozen. In het bijzonder bleek dit te gelden, indien de afstand van de afkapgrenzen tot het populatiegemiddelde van  $3\sigma$  wordt teruggebracht op  $2\sigma$  (zie fig 4-1). Aangezien in de praktijk slechts bij hoge uitzon-

dering reeksen bepalingssuitkomsten zullen voorkomen waarbij de contaminatie te verwaarlozen is, werd ook het effect van contaminatie op de toetsingsgrootte onderzocht. Uit de resultaten van deze modelproeven, waarbij achtereenvolgens met afkapgrenzen op een afstand van  $1\sigma$ ,  $2\sigma$  en  $3\sigma$  van het populatiegemiddelde werd gewerkt, bleek dat de afstand van  $2\sigma$ , ondanks het bovengenoemde bezwaar, in dit geval toch de voorkeur verdiende (zie fig 4-4). Enig voorbehoud moet hierbij worden gemaakt, omdat bij de modelproeven werd uitgegaan van 'hoofdpopulaties' en 'contaminerende populaties' met een verdeling volgens Gauss, terwijl de empirische frekwentieverdelingen daar nogal sterk van kunnen verschillen. Het zou de moeite waard zijn na te gaan hoe robuust de methode is tegen afwijkingen van het verdelingstype ten opzichte van een verdeling volgens Gauss. Van de mogelijkheid om een empirische verdeling door middel van één of meer transformaties om te zetten in een verdeling volgens Gauss (Harris & DeMets, 1972) werd in het kader van dit onderzoek afgezien. Dit gebeurde op grond van de overweging dat, als gevolg van de heterogeniteit van het materiaal en het ontbreken van een theoretische fundering voor de keuze van de transformatie(s), wederom een element van willekeur zou worden geïntroduceerd bij de keuze van het gebied dat op deze wijze wordt genormaliseerd. Een zelfde redenering geldt uiteraard ook voor het in hoofdstuk 5 besproken onderzoek van de methoden voor de schatting van parameters. De met deze methoden verkregen resultaten lijken daarom slechts beperkt bruikbaar, bijvoorbeeld voor de vaststelling van de afkapgrenzen. Extra voorbehoud moet worden gemaakt ten aanzien van de methode van Bhattacharya, die erg gevoelig bleek te zijn voor onregelmatigheden in de frekwentieverdeling.

Bij de toepassing van de GvN-methode op reeksen bepalingssuitkomsten moeten tolerantiegrenzen voor de toetsingsgrootte worden vastgesteld. In dit onderzoek gebeurde dat op basis van de van 20 mei tot eind juni verkregen waarden voor de toetsingsgrootte. Deze procedure heeft het nadeel, dat de ligging van de tolerantiegrenzen kan zijn beïnvloed door eventuele ontsporingen in de betreffende periode. Dit nadeel is echter inherent aan de GvN-methode, die uitsluitend gericht is op het aantonen van veranderingen in de grootte van een systematische fout; over de grootte zelf kan met de GvN-methode geen informatie worden verkregen. Overigens zal ook het optreden van inci-

dentele fouten, of het toenemen van de invloed van fluctuaties de toetsingsgrootte niet in belangrijke mate beïnvloeden. Deze beperkingen houden in dat de GvN-methode zeker niet geschikt is om als enige methode in het kader van de kwaliteitscontrole te worden toegepast, maar indien met het bovenstaande rekening wordt gehouden, kan de methode toch als welkome aanvulling van de mogelijkheden op het gebied van de kwaliteitscontrole worden beschouwd.

## SAMENVATTING.

In hoofdstuk 1 wordt in het kort toegelicht dat het in dit proefschrift beschreven onderzoek kan worden gezien als een schakel in het streven naar vergroting van de bruikbaarheid van klinisch-chemische uitkomsten door toepassing van elektronische informatieverwerking. Als eerste doelstelling is gekozen het verkrijgen van inzicht in de achtergronden van methoden van kwaliteitscontrole, die gebruik maken van uitkomsten van aangevraagde bepalingen. In hoofdstuk 2 wordt de literatuur betreffende dit onderwerp besproken, waarbij vooral aandacht is geschonken aan de door Hoffmann & Waaid (1965) geïntroduceerde 'gemiddelde van normalen'-methode.

Daarop aansluitend wordt in hoofdstuk 3 de literatuur besproken met betrekking tot een aantal facetten van de problematiek rond 'normale waarden'. Hierbij wordt kort ingegaan op de problemen die samenhangen met de termen en/of met de begrippen 'normaal' en 'normale waarden'. Daarna worden problemen die samenhangen met het vaststellen en met het gebruik van 'normale waarden' besproken. In het bijzonder worden enkele methoden (Hoffmann, 1963; Neumann, 1968 en Bhattacharya, 1967) behandeld, die het mogelijk maken voor de schatting van parameters en/of het vaststellen van grenzen van het 'normale gebied' uit te gaan van reeksen bepalingensuitkomsten zonder voorafgaande selectie van proefpersonen.

Voor het in hoofdstuk 4 beschreven onderzoek naar de eigenschappen van de 'gemiddelde van normalen'-methode werd gebruik gemaakt van verschillende modellen.

Als eenvoudigste model werd de standaardnormale verdeling gekozen. Een belangrijk argument hiervoor was, dat voor deze verdeling uitgebreide tabellen beschikbaar zijn. Een nadeel van dit model is, dat



het verwerken van series, bestaande uit enige tientallen bepalingsuitkomsten, niet eenvoudig kan worden nagebootst. Om dit bezwaar te ondervangen werd gebruik gemaakt van een tweede model. Dit bestond uit reeksen, door middel van een computer gegenereerde, getallen, die zijn te beschouwen als aselechte steekproeven uit een populatie met een tweezijdig afgekapte verdeling volgens Gauss, waarvan de parameters tevoren konden worden gekozen.

Het optreden van een systematische fout kan worden gesimuleerd door modificatie van alle elementen van één of meer steekproeven. Omdat series klinisch-chemische bepalingsuitkomsten als regel ook pathologische uitkomsten bevatten, werd ook een model gebruikt waarbij in een steekproef naast de elementen afkomstig uit een hoofdpopulatie tevens elementen werden opgenomen die afkomstig waren uit een 'contaminerende' populatie. Uit de resultaten van de modelproeven kan worden geconcludeerd, dat het gebruik van afkapgrenzen, die op een afstand van ongeveer tweemaal de standaardafwijking ter weerszijden van het gemiddelde van de hoofdpopulatie liggen, gunstig zal werken (zie figuur 4-4). Ter orientatie werd de 'gemiddelde van normalen'-methode toegepast op reeksen van ongeveer duizend uitkomsten. Voorbeelden hiervan worden besproken.

In hoofdstuk 5 wordt het onderzoek beschreven naar de eigenschappen van methoden voor de schatting van parameters. Ook hier werd begonnen met toepassing van de methoden op modellen. In eerste instantie werden de drie eerder genoemde (hoofdstuk 3) methoden in het onderzoek betrokken. In een later stadium werd ook de 'best passende parabool'-methode onderzocht. Uit de resultaten van de modelproeven kan worden afgeleid, dat geen van de onderzochte methoden in alle gevallen goede uitkomsten levert, terwijl niet a priori vast te stellen is welke methode in een bepaald geval het beste resultaat zal leveren. Hieruit vloeit voort, dat we bij toepassing van deze methoden op reeksen bepalingsuitkomsten eigenlijk behoefte hebben aan objectieve criteria voor de toetsing van het resultaat. Het ontbreken daarvan blijft een principieel nadeel van dit type methoden.

Tenslotte werd, zoals beschreven in hoofdstuk 6, de 'gemiddelde van normalen'-methode toegepast op de gedurende een langere periode per dag geproduceerde bepalingsuitkomsten van het CKCL van het Academisch Ziekenhuis Dijkzigt. Voor de schatting van parameters voor dit materiaal werd gebruik gemaakt van de methode van Hoffmann en van de

'best passende parabool'-methode. Voorbeelden van de bij deze toepassing verkregen resultaten worden besproken.

Als slotconclusie wordt gesteld, dat de 'gemiddelde van normalen'-methode als welkome aanvulling van de mogelijkheden op het gebied van de kwaliteitscontrole kan worden beschouwd, mits rekening wordt gehouden met een aantal beperkingen.

## SUMMARY.

The investigations described in this thesis can be looked upon as part of our efforts to improve the usefulness of the test results rendered by the clinical chemistry laboratory through the application of electronic data processing. The initial objective of this study, which was started during the implementation period of our data-acquisition and laboratory administration system, was to clarify the the basis of those quality control methods which make use of results of requested determinations.

Chapter two reviews the literature on this subject. Special attention is paid to the 'Average of Normals'-method introduced by Hoffmann & Waid (1965). In connection with the discussion of this method, chapter three reviews some of the literature on problems regarding 'normal values'. Problems related to terminology and to such concepts as 'normal' and 'normal values' are briefly discussed. Problems connected with the establishment and the use of 'normal values' are discussed more extensively. Among these are some methods that enable us to base estimates of the parameters of the 'normal' population and/or the 'normal range' on series of test results as produced by carrying out the determinations requested by the clinicians (Hoffmann, 1963; Neumann, 1968; Bhattacharya, 1967).

The properties of the 'average of normals'- and related methods were studied (chapter four) by means of several models, the simplest being the standard normal distribution. The latter model was included mainly on account of the availability of extensive tables for this distribution. A disadvantage of this model is however, that it does not allow simulation of the daily processing of series of test results. In order to cope with this problem a second model was used, consisting of series of computer-generated numbers. These series can be considered

as random samples from a truncated Gaussian distribution. The parameters of the Gaussian distribution and the truncation limits can be chosen beforehand.

The occurrence of a systematic error can be simulated by modification of every element of one or more samples. Series of clinical laboratory results mostly contain varying amounts of pathological values. To study the effect of such contamination on the 'average of normals' a third model was used. This model consisted of a main population and a contaminating population. The samples were formed by taking elements from both populations that fell between the truncation limits. Experiments were performed in which the truncation limits were set on either side of the mean of the main population at a distance of successively one, two and three times the standard deviation of that population.

From the results of these model studies it can be concluded that the application of truncation limits lying at a distance of about twice the standard deviation from the mean of the main population will diminish the influence of contamination, whilst the sensitivity towards a systematic error will not be unduly impaired.

Terminating chapter four is a report on preliminary employment of the 'average of normals'-method to some series of about one thousand laboratory results.

Chapter five describes the study, by means of models, of methods for the estimation of population parameters and, subsequently, of their application to series of laboratory results.

At first the study was confined to the three methods already mentioned (chapter three). Later on a fourth method was included, which was a computer-adapted modification of the 'best fitting parabola'-method described by Hald (1952). From the results of the model experiments we must conclude that none of the methods studied will invariably lead to correct values for the parameters of the main population. It is suggested that at least two methods should be used when estimating parameters of series of laboratory results. In case of discrepancies objective criteria would be needed to select the better estimate. The unavailability of such criteria remains a serious disadvantage of this type of methods.

Chapter six reports on the application of the 'average of normals'-method to most of the daily output of the clinical chemistry

laboratory of the University Hospital Dijkzigt - Erasmus University Rotterdam. For the estimation of population parameters from this data, Hoffmann's method and the 'best fitting parabola'-method were used. Examples of the results are discussed. The final conclusion is that the 'average of normals'-method can be considered to be a valuable addition to the range of quality control methods provided certain restrictions are kept in mind.

L I T E R A T U R L I J S T

ALBRITTON EC

STANDARD VALUES IN BLOOD.  
(1952) SAUNDERS PHILADELPHIA

AMADOR E

QUALITY CONTROL BY THE REFERENCE SAMPLE METHOD.  
(1968) AMER J CLIN PATH 50 P 360

AMADOR E, HSI BP & MASSOD MF

AN EVALUATION OF THE 'AVERAGE OF NORMALS' AND RELATED METHODS OF  
QUALITY CONTROL.  
(1968) AMER J CLIN PATH 50 P 369

ANNINO JS & WILLIAMS LA

ON NORMAL VALUES.  
(1972) CLIN CHEM 18 P 488

BARNARD AJ, JOY EF & MICHELOTTI FW

CHARACTERIZATION OF HIGH-PURITY CHEMICALS FOR USE AS CLINICAL  
LABORATORY STANDARDS.  
(1971) CLIN CHEM 17 P 841

BEGTRUP H, LEROY S & WALLOE-HANSEN P

A COMPARISON OF SOME METHODS FOR QUALITY CONTROL.  
(1969) SCAND J CLIN LAB INVEST SUPPL 110 P 102

BEGTRUP H, LEROY S, THYREGOD P & WALLOE-HANSEN P

'AVERAGE OF NORMALS' USED AS CONTROL OF ACCURACY AND A COMPARISON  
WITH OTHER CONTROLS.  
(1971) SCAND J CLIN LAB INVEST 27 P 247

BEYER WH - EDITOR

HANDBOOK OF TABLES FOR PROBABILITY AND STATISTICS.  
(1968) THE CHEM RUBBER CO CLEVELAND OHIO 2ND EDITION

BHATTACHARYA CG

A SIMPLE METHOD OF RESOLUTION OF A DISTRIBUTION INTO GAUSSIAN  
COMPONENTS.  
(1967) BIOMETRICS 23 P 115

BUTTNER H

STATISTISCHE QUALITATSKONTOLLE IN DER KLINISCHEN CHEMIE.  
(1967) Z KLIN CHEM KLIN BIOCHEM 5 P 41

BUTTNER H

INT SYMPOSIUM ON STATISTICAL QUALITY CONTROL IN THE ANALYTICAL  
LABORATORY.  
(1968) Z ANAL CHEMIE 243 P 751

- BUTTNER H  
QUALITATSKONTROLLE IM AUTOMATISIERTEN KLINISCH-CHEMISCHEN  
LABORATORIUM.  
(1969) Z KLIN CHEM KLIN BIOCHEM 7 P 89
- BUTTNER H, HANSERT E & STAMM D  
QUALITATSKONTROLLE. IN - METHODEN DER ENZYMATISCHEN ANALYSE -  
BERGMEYER HV  
(1970) 2E DRUK P 331
- BUTTNER H & STAMM D  
QUALITATSKONTROLLE KLINISCH-CHEMISCHER UNTERSUCHUNGEN.  
(1966) Z KLIN CHEM 4 P 303
- CAMPBELL DG & OWEN JA  
CLINICAL LABORATORY ERROR IN PERSPECTIVE.  
(1967) CLIN BIOCHEM 1 P 3
- CARAWAY WT  
ACCURACY IN CLINICAL CHEMISTRY.  
(1971) CLIN CHEM 17 P 63
- CARLSTROM S, LUNDQUIST A, LUNDQUIST I, NORDEN A, SCHERSTEN B &  
WOLLMARK G  
STUDIES IN SUBJECTS WITH POSITIVE POST PRANDIAL CLINISTIX TEST.  
III SPECIAL STUDIES AND FOLLOW-UP OF CASES WITH BORDERLINE  
GLUCOSE TOLERANCE.  
(1971) ACTA MEDICA SCAND 189 P 415
- COOK MG, LEVELL MJ & PAYNE RB  
A METHOD FOR DERIVING NORMAL RANGES FROM LABORATORY SPECIMENS  
APPLIED TO URIC ACID IN MALES.  
(1970) J CLIN PATH 23 P 778
- COPELAND BE  
A QUALITY CONTROL SYSTEM FOR CLINICAL CHEMISTRY AND ITS  
APPLICATIONS.  
(1968) Z ANAL CHEM 243 P 754
- COPELAND BE, BLAKE WJ, MUELLING RJ & SKENDZEL LP  
A REPORT OF THE STANDARDS COMMITTEE OF THE COLLEGE OF AMERICAN  
PATHOLOGISTS, NATIONAL COMPREHENSIVE LABORATORY SURVEY 1965,  
CHEMISTRY SECTION.  
(1967) AMER J CLIN PATH 48 P 104
- CUTLER JL, COLLEN ME, SIEGELAUB AB & FELDMAN R  
NORMAL VAUES FOR MULTIPHASIC SCREENING BLOOD CHEMISTRY TESTS.  
(1969) ADVAN AUTOM ANALYSIS - TECHNICON INT CONGR VOL III P 67
- DIXON K & NORTHAM BE  
QUALITY CONTROL USING THE DAILY MEAN.  
(1970) CLIN CHIM ACTA 30 P 453
- DYBKAER R  
CONCEPTS AND NOMENCLATURE IN THEORY OF REFERENCE VALUES.  
(1972) SCAND J CLIN LAB INVEST 29 SUPPL 126 P 19.1

- DYBKAER R & GRASBECK R  
EDITORIAL - THEORY OF REFERENCE VALUES.  
(1973) SCAND J CLIN INVEST 32 P 1
- ELSOM KO  
ELEMENTS OF THE MEDICAL PROCESS. THEIR PLACE IN MEDICAL CARE  
PLANNING.  
(1971) J AMER MED ASS 217 P 1226
- ELVEBACK LR, GUILLER CL & KEATING FR  
HEALTH, NORMALITY AND THE GHOST OF GAUSS.  
(1970) J AMER MED ASS 211 P 69
- FRIEDMAN GD  
AN EPIDEMOLOGIST'S ASSESSMENT OF THE REPORTED LABORATORY FINDINGS.  
(1969) ADVAN AUTOM ANALYSIS - TECHNICON INT CONGR VOL III P 111
- GINDLER EM  
CALCULATION OF NORMAL RANGES BY METHODS USED FOR RESOLUTION OF  
OVERLAPPING GAUSSIAN DISTRIBUTIONS.  
(1970) CLIN CHEM 16 P 124
- GLICK D  
STATISTICS OF PATIENT TEST VALUES. APPLICATION TO INDIRECT NORMAL  
RANGE AND TO QUALITY CONTROL.  
(1972) CLIN CHEM 18 P 1504
- GRANNIS GF, EDISON JA & GRUMER HD  
LABORATORY MISTAKES.  
(1971) CLIN CHEM 17 P 646
- GRANNIS GF, GRUMER HD, LOTT JA, EDISON & MCCABE WC  
PROFICIENCY EVALUATION OF CLINICAL CHEMICAL LABORATORIES.  
(1972) CLIN CHEM 18 P 222
- GRASBECK R  
TYPES OF REFERENCE GROUPS.  
(1972) SCAND J CLIN LAB INVEST 29 SUPPL 126 P 19.2
- GRASBECK R & SARIS NE  
ESTABLISHMENT AND USE OF NORMAL VALUES.  
(1969) SCAND J CLIN LAB INVEST SUPPL 110 P62
- HALD A  
STATISTICAL THEORY WITH ENGINEERING APPLICATIONS.  
(1952) JOHN WILEY & SONS INC NEW YORK P 156 & 157
- HARRIS EK & DEMETS DL  
ESTIMATION OF NORMAL RANGES AND CUMULATIVE PROPORTIONS BY TRANS-  
FORMING OBSERVED DISTRIBUTIONS TO GAUSSIAN FORM.  
(1972) CLIN CHEM 18 P 605
- HOFFMANN RG  
STATISTICS IN THE PRACTICE OF MEDICINE.  
(1963) J AMER MED ASS 185 P 864



- HOFFMANN RG  
ESTABLISHING NORMAL RANGES.  
(1971) CLIN CHEM 17 P 456
- HOFFMANN RG & WAID ME  
THE NUMBER PLUS METHOD OF QUALITY CONTROL OF LABORATORY ACCURACY.  
(1963) AMER J CLIN PATH 40 P 263
- HOFFMANN RG & WAID ME  
THE 'AVERAGE OF NORMALS' METHOD OF QUALITY CONTROL.  
(1965) AMER J CLIN PATH 43 P 134
- HOWARTH AT, NOBLE RL, PAYNE RB, STEEL AE & TEASDALE PR  
THE LEEDS REGIONAL QUALITY CONTROL SCHEME FOR CLINICAL BIOCHEMISTRY.  
A PROGRESS REPORT.  
(1973) J CLIN PATH 26 P 875
- JONGE H DE  
INLEIDING TOT DE MEDISCHE STATISTIEK.  
(1963) WOLTERS-NOORDHOFF NV GRONINGEN 2E DRUK P 275-279
- JUNGNER G & JUNGNER J  
INTERPRETATION OF DATA OBTAINED IN LABORATORY SCREENING PROGRAM.  
IN - MULTIPLE LABORATORY SCREENING - BENSON ES & STRANDFORD PE  
(1969) ACADEMIC PRESS NEW YORK - LONDON P 257
- KILGARIFF M & OWEN JA  
AN ASSESSMENT OF THE 'AVERAGE OF NORMALS' QUALITY CONTROL METHOD.  
(1968) CLIN CHIM ACTA 19 P 175
- KOLLER S  
DIE PROBLEMATIK DER NORMALWERTE IN DER MEDIZIN. IN - AUTOMATI-  
SIERUNG DES KLINISCHEN LABORATORIUMS - GRIESSER G & WAGNER G  
(1968) FK SCHATTAUER VERLAG P 141
- LEIJNSE-YBEMA HJ & LEIJNSE B  
DE 'NORMALE WAARDEN' IN DE KLINISCHE CHEMIE.  
(1972) PHARM WEEKBL 107 P 337
- LEVEY S & JENNINGS ER  
THE USE OF CONTROL CHARTS IN THE CLINICAL LABORATORY.  
(1950) AMER J CLIN PATH 20 P 1059
- LEWIS PW & DIXON K  
ACTION LIMITS FOR INTERNAL QUALITY CONTROL.  
(1971) CLIN CHIM ACTA 35 P 21
- LINDEBOOM GA  
BEGRIPSVORMING IN DE PATHOLOGIE. IN - DE VORMING VAN WETENSCHAP-  
PELIJKE BEGRIPPEN - DONNER AM EN ANDEREN  
(1957) JH KOK KAMPEN P 20
- LITTLE AJ, WILLIAMS RB, PARKER SD & PAYNE RB  
THE DERIVATION OF BIOCHEMICAL NORMAL RANGES FROM A HOSPITAL  
OUTPATIENT POPULATION.  
(1974) CLIN CHIM ACTA 57 P 91

MAINLAND D

NORMAL VALUES IN MEDICINE.

(1969) ANN NY ACAD SCI 161 ART 2 P 527

MAINLAND D

REMARKS ON CLINICAL 'NORMS'

(1971) CLIN CHEM 17 P267

MCKINLAY PL

THE MEASUREMENT OF NORMALITY.

(1951) BRIT MED BULL 7 P 275

MCSWINEY RR & WOODROW DA

QUALITY CONTROL OF NON-CHEMICAL ERRORS.

(1969) 7TH INT CONGR CLIN CHEM P 1.9.05

MURPHY EA & ABBEY H

THE NORMAL RANGE - A COMMON MISUSE.

(1967) J CHRONIC DISEASE 20 P 79

NEUMANN GJ

THE DETERMINATION OF NORMAL RANGES FROM ROUTINE LABORATORY DATA.

(1968) CLIN CHEM 14 P 979

NORDEN A

DISCRIMINATION INTO HEALTHY AND SICK IN HEALTH SCREENING.

(1972) SCAND J CLIN LAB INVEST 29 SUPPL 126 P 21.3

NORMCOMMISSIE 70

STATISTISCHE TERMEN.

(1968) NEDERLANDS NORMALISATIE INSTITUUT. NEN 3117 - 2E DRUK

O'HALLORAN MW, STUDLEY-RUXTON J & WELLBY ML

A COMPARISON OF CONVENTIONALLY DERIVED NORMAL RANGES WITH THOSE OBTAINED FROM PATIENTS' RESULTS.

(1970) CLIN CHIM ACTA 27 P 35

OWEN JA & CAMPBELL DG

LABORATORY QUALITY CONTROL USING PATIENTS' RESULTS.

(1968A) CLIN CHIM ACTA 20 P 327

OWEN JA & CAMPBELL DG

A COMPARISON OF PLASMA ELECTROLYTE AND UREA VALUES IN HEALTHY PERSONS AND IN HOSPITAL PATIENTS.

(1968B) CLIN CHIM ACTA 22 P 611

PAYNE RB & LEVELL MJ

REDEFINITION OF THE NORMAL RANGE FOR SERUM SODIUM.

(1968) CLIN CHEM 14 P 172

RADIN N

WHAT IS A STANDARD.

(1967) CLIN CHEM 13 P 55

- REED AH  
USE OF PATIENT DATA FOR QUALITY CONTROL OF CLINICAL LABORATORY TESTS.  
(1970) CLIN CHEM 16 P 129
- REED AH, HENRY RJ & MASON WB  
INFLUENCE OF STATISTICAL METHOD USED ON THE RESULTING ESTIMATE  
OF NORMAL RANGE.  
(1971) CLIN CHEM 17 P 275
- REED AH & WU GT  
EVALUATION OF A TRANSFORMATION METHOD FOR ESTIMATION OF NORMAL RANGE.  
(1974) CLIN CHEM 20 P 576
- RICHTERICH R & COLOMBO JP  
ULTRA MIKROMETHODEN IM KLINISCHEN LABORATORIUM. II DIE BESTIMMUNG  
DER ZUVERLASSIGKEIT VON LABORATORIUMSMETHODEN.  
(1962) KLIN WSCHR 40 P 529
- RIET HG VAN & HOEKE JOO  
AMYLASE AND LYPASE VALUES IN NORMAL SUBJECTS.  
(1968) CLIN CHIM ACTA 19 P 459
- RUMKE CL & BEZEMER PD  
METHODEN VOOR DE BEPALING VAN NORMALE WAARDEN (1 EN 2).  
(1972) NED T GENEESK 116 P 1224 & 1559
- SCHNEIDER AJ  
SOME THOUGHTS ON NORMAL, OR STANDARD, VALUES IN CLINICAL MEDICINE.  
(1960) PEDIATRICS 26 P 973
- SIEST G, HOUOT O & HENNY J  
MATERIALS AND METHODS FOR HEALTH TESTING.  
(1975) 6TH INT SYMP QUALITY CONTROL, GENEVA P 143
- SIMONSON E  
THE CONCEPT AND DEFINITION OF NORMALITY.  
(1965) ANN NY ACAD SCI 134 P 541
- STAMM D  
RINGVERSUCHE IN DER KLINISCHEN CHEMIE.  
(1971) SCHWEIZ MED WSCHR 101 P 429
- STAMM D  
CALIBRATION AND QUALITY CONTROL MATERIALS.  
(1974) Z KLIN CHEM KLIN BIOCHEM 12 P 137
- STAMM D & BUTTNER H  
RINGVERSUCH ZUR QUALITATSKONTROLLE.  
(1969) Z KLIN CHEM KLIN BIOCHEM 7 P 393
- SUNDERMAN FW  
CURRENT CONCEPTS OF 'NORMAL VALUES', 'REFERENCE VALUES' AND  
'DISCRIMINATION VALUES' IN CLINICAL CHEMISTRY.  
(1975) CLIN CHEM 21 P 1873

- TOLLS RE, WERNER M, HULTIN JV & MELLECKER J  
SEX AND AGE DEPENDENCE OF SEVEN SERUM CONSTITUENTS IN A LARGE  
AMBULATORY POPULATION.  
(1969) ADVAN AUTOM ANALYSIS - TECHNICON INT CONGR VOL III P 9
- TONKS DB  
A STUDY OF THE ACCURACY AND PRECISION OF CLINICAL CHEMISTRY  
DETERMINATIONS IN 170 CANADIAN LABORATORIES.  
(1963) CLIN CHEM 9 P 217
- WAID ME & HOFFMANN RG  
THE QUALITY CONTROL OF LABORATORY PRECISION.  
(1955) AMER J CLIN PATH 25 P 585
- WHITBY LG, MITCHELL FL & MOSS DW  
QUALITY CONTROL IN ROUTINE CLINICAL CHEMISTRY.  
(1967) ADVAN CLIN CHEM 10 P 65
- WHITEHEAD TP  
MULTIPLE ANALYSES AND THEIR USE IN THE INVESTIGATION OF PATIENTS.  
(1971) ADVAN CLIN CHEM 14 P 389
- WHITEHEAD TP  
QUALITY CONTROL TECHNIQUES IN LABORATORY SERVICES.  
(1974) BRIT MED BULL 30 P 237
- WHITEHEAD TP & MORRIS LO  
METHODS OF QUALITY CONTROL.  
(1969) ANN CLIN BIOCHEM 6 P 94
- WHITEHEAD TP, BROWNING DM & GREGORY A  
A COMPARATIVE SURVEY OF THE RESULTS OF ANALYSES OF BLOOD SERUM IN  
CLINICAL CHEMISTRY LABORATORIES IN THE UNITED KINGDOM.  
(1973) J CLIN PATH 26 P 435
- WOOTTON IDP, KING EJ & MACLEAN SMITH J  
THE QUANTITATIVE APPROACH TO HOSPITAL BIOCHEMISTRY.  
(1951) BRIT MED BULL 7 P 307
- WOOTTON IDP & KING EJ  
NORMAL VALUES FOR BLOOD CONSTITUENTS.  
(1953) LANCET I P 470
- YOUNG DS & MEARS TW  
MEASUREMENT AND STANDARD REFERENCE MATERIALS IN CLINICAL CHEMISTRY.  
(1968) CLIN CHEM 14 P 929
- YOUNG DS, HARRIS EK & COTLOVE E  
BIOLOGICAL AND ANALYTICAL COMPONENTS OF VARIATION IN LONG-TERM  
STUDIES OF SERUM CONSTITUENTS IN NORMAL SUBJECTS.  
(1971) CLIN CHEM 17 P 403
- ZIEF M & MICHELOTTI FW  
CLINICAL CHEMISTRY - A CHALLENGE FOR HIGH-PURITY STANDARDS AND  
REAGENTS.  
(1971) CLIN CHEM 17 P 833

## CURRICULUM VITAE.

In 1948 eindexamen HBS-B aan het Lorentzlyceum te Haarlem, daarna scheikunde studie aan de Universiteit van Amsterdam.

In maart 1952 candidaatsexamen (g) en in maart 1956 doctoraalexamen met physiologische chemie als hoofdrichting en toxicologie als bijvak.

Van april 1956 tot half december 1957 in militaire dienst, waarvan het laatste jaar gedetacheerd op het laboratorium van het Centraal Militair Hospitaal Utrecht.

Van januari 1958 tot eind 1960 werkzaam bij Poviet Producten N.V. als research chemicus.

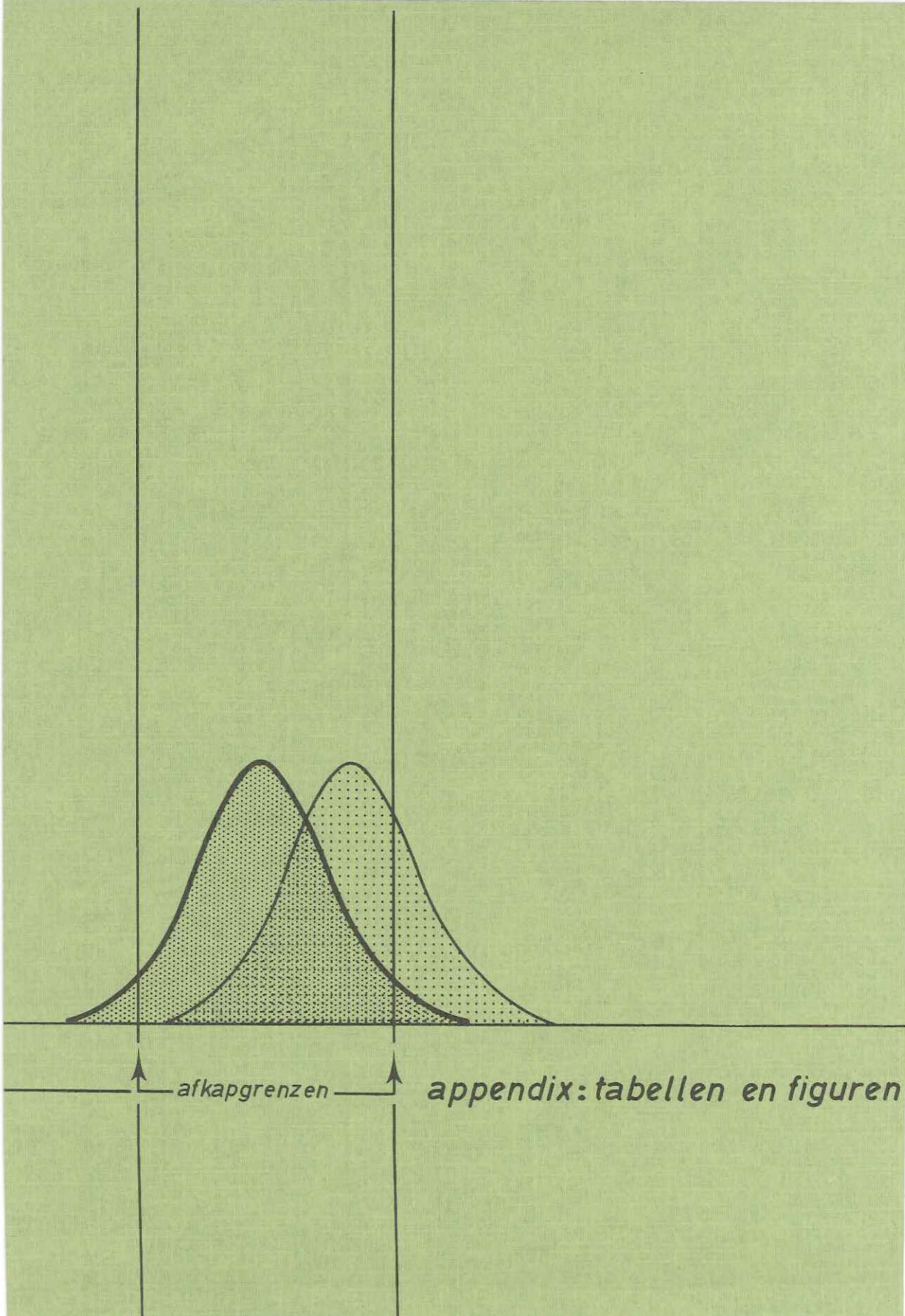
Vanaf januari 1961 verbonden aan de Rijksuniversiteit Utrecht.

Opgeleid tot klinisch chemicus en ook daarna werkzaam op het Laboratorium voor Klinische Biochemie (Hoofd: dr. J. de Wael) van de Kliniek voor Inwendige Geneeskunde van het Academisch Ziekenhuis Utrecht. De laatste jaren vooral betrokken bij de computerisering van het laboratorium.

Sinds october 1975 tevens lid van het bestuur van bovengenoemde kliniek.

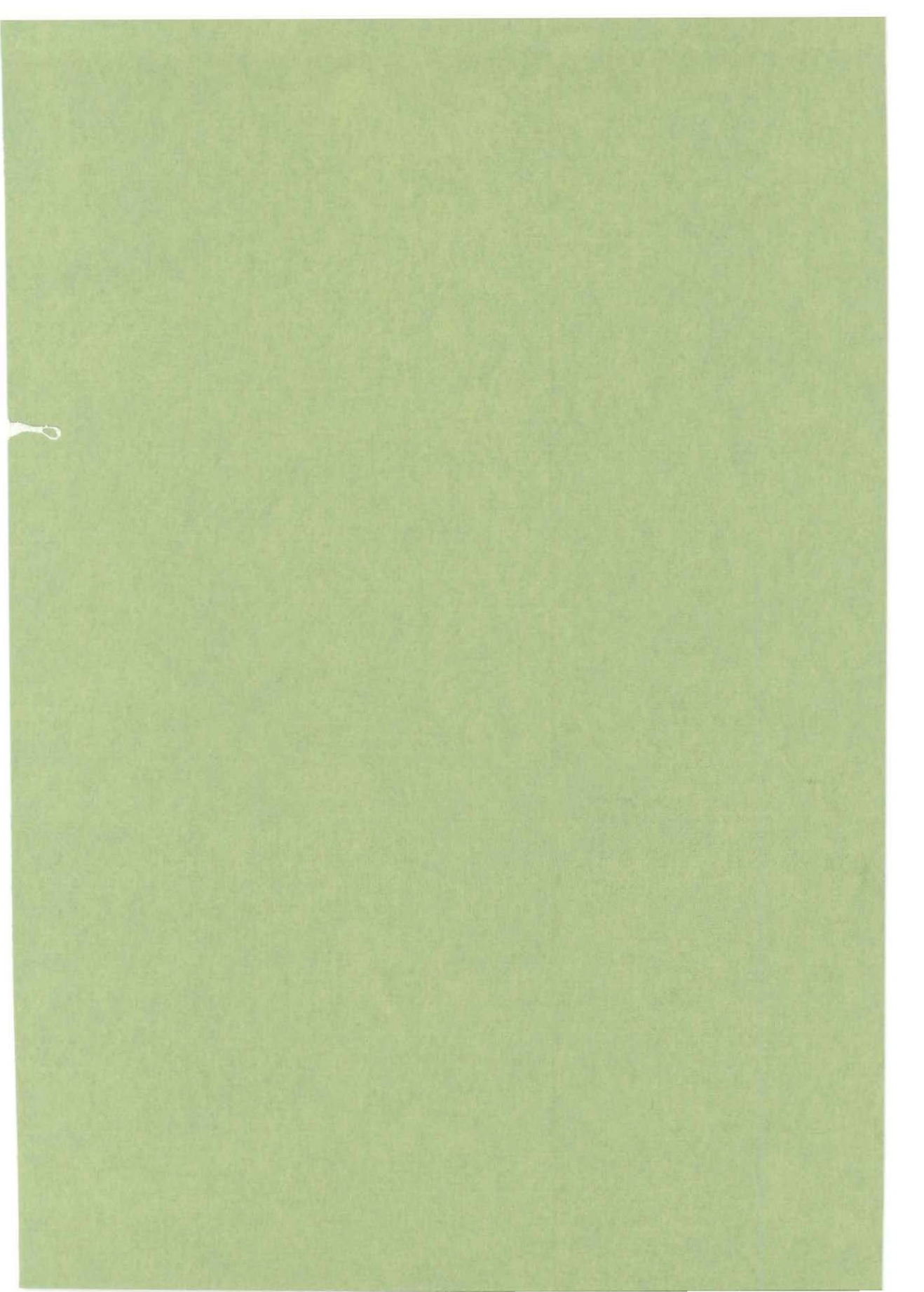
Bij het in dit proefschrift beschreven onderzoek werden, naast bepalingsuitkomsten van het Laboratorium voor Klinische Biochemie, ook uitkomsten verwerkt van de afdeling Klinische Chemie van het Academisch Ziekenhuis Rotterdam - Dijkzigt, terwijl in de loop van het onderzoek vaak gastvrijheid werd verleend door de Centrale Computerafdeling van het A.Z.U. en door de Computergroep van de vakgroep Cardiologie van de Rijksuniversiteit Utrecht.





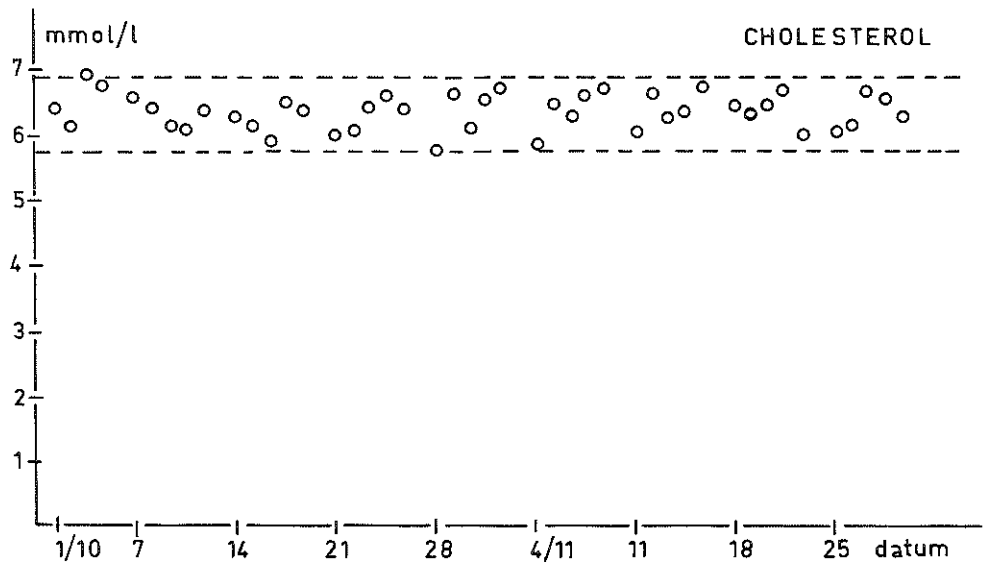
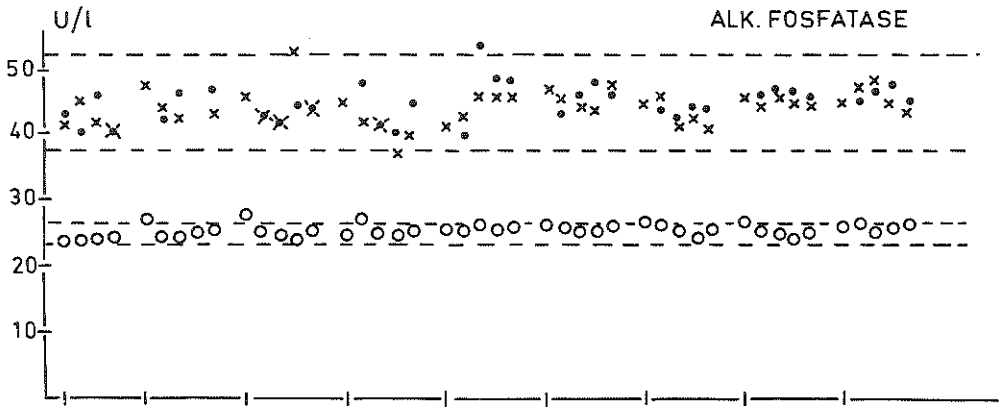
afkapgrenzen

appendix: tabellen en figuren



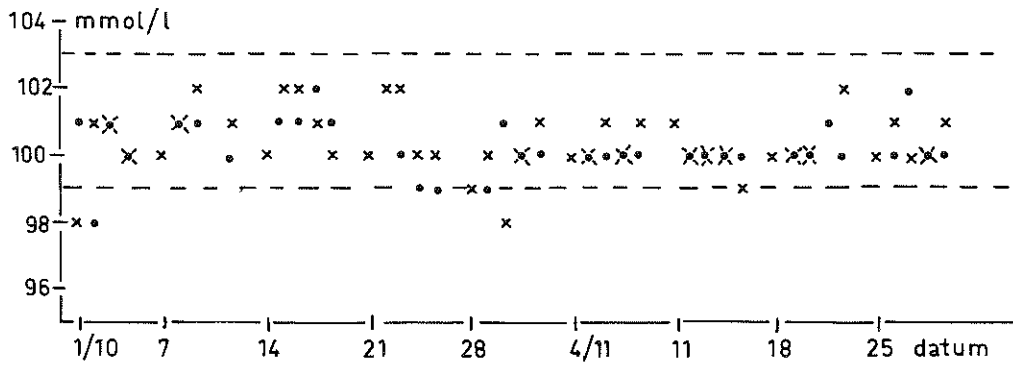
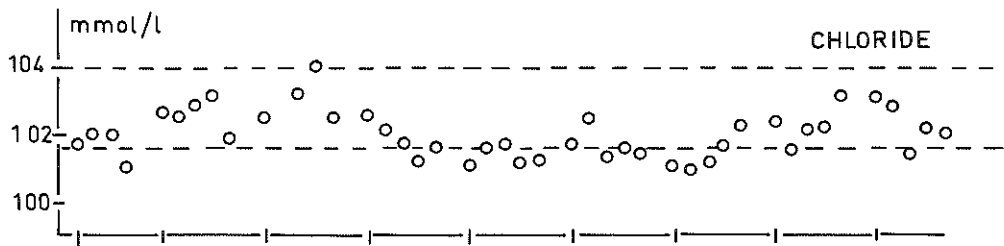
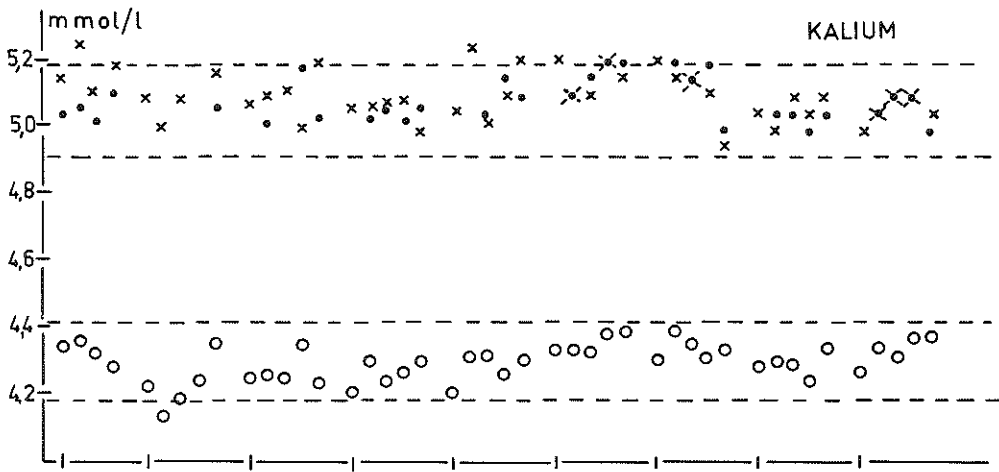


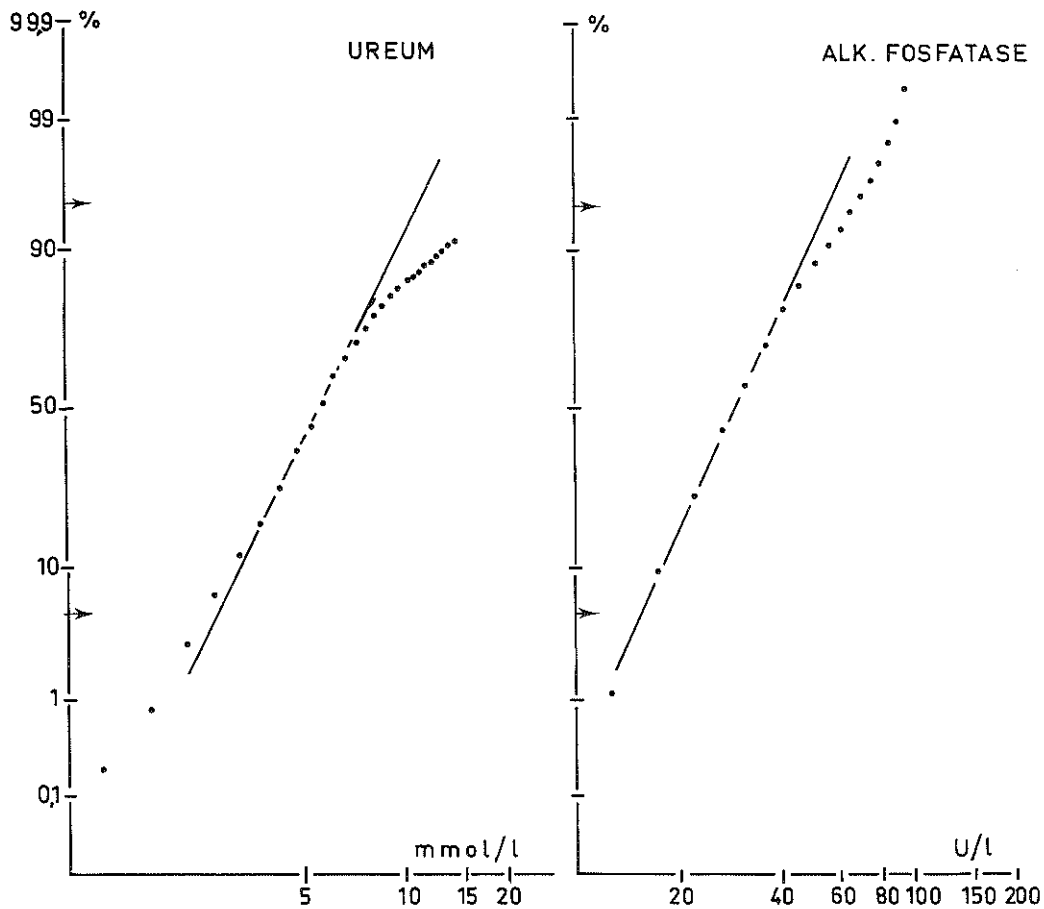




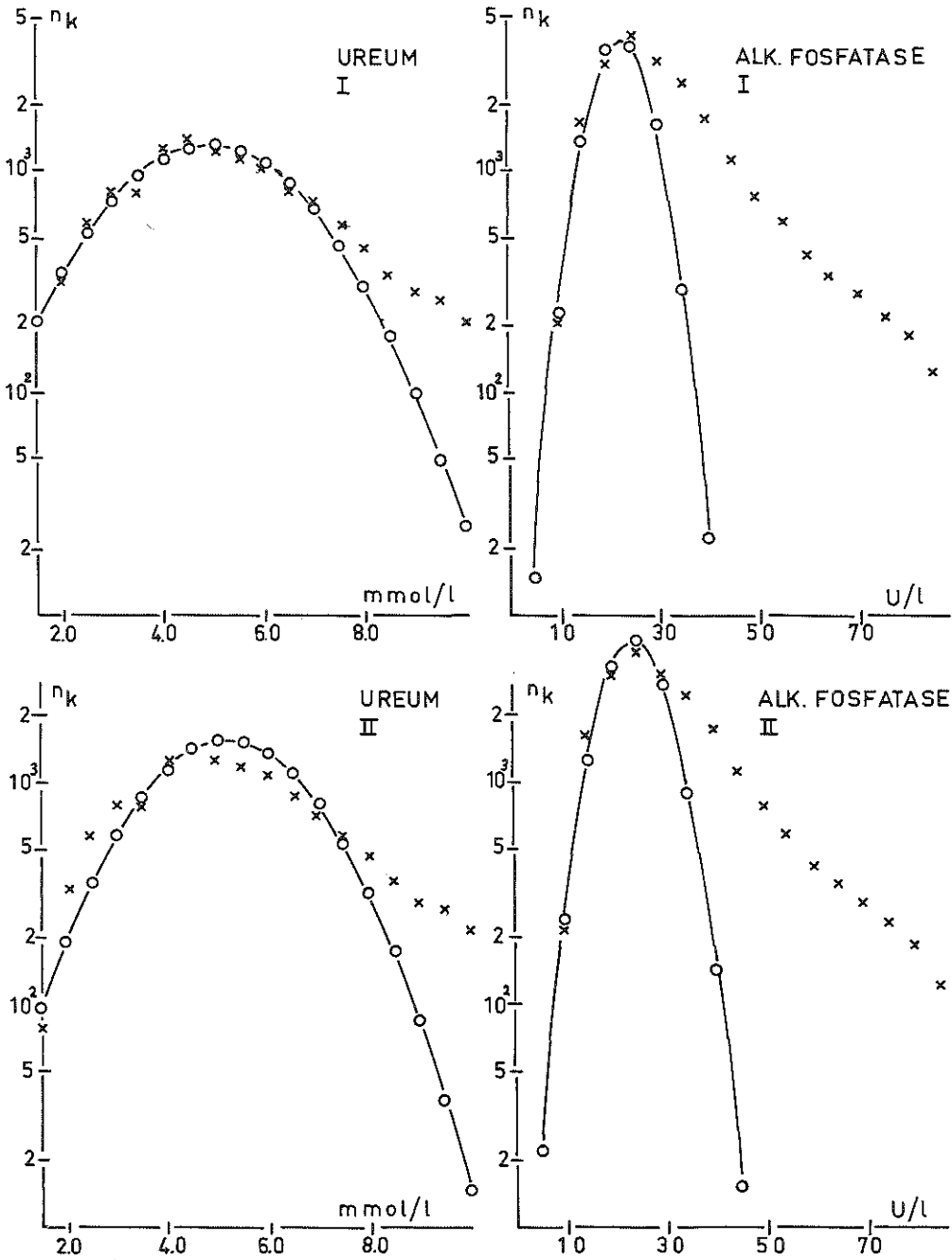
Figuur 6-5.  
 Verloop van de toetsingsgrootte van de 'GvN'-methode (o) en van de uitkomsten van het controleserum (• en \*) in de periode van 1-10 tot 1-12-1974. De tolerantiegrenzen worden door gebroken lijnen aangegeven. Zie p 70.







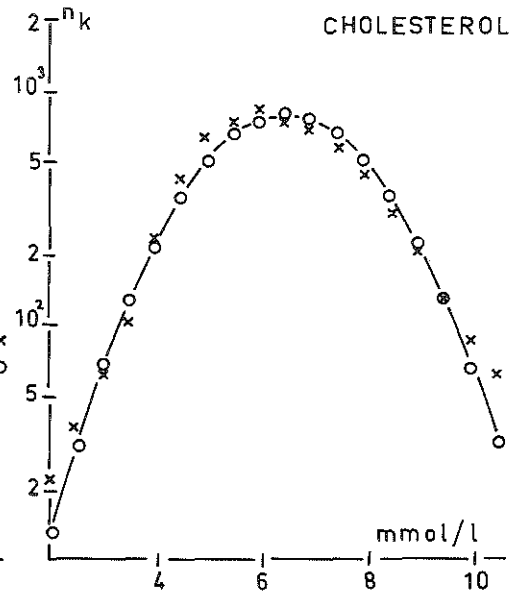
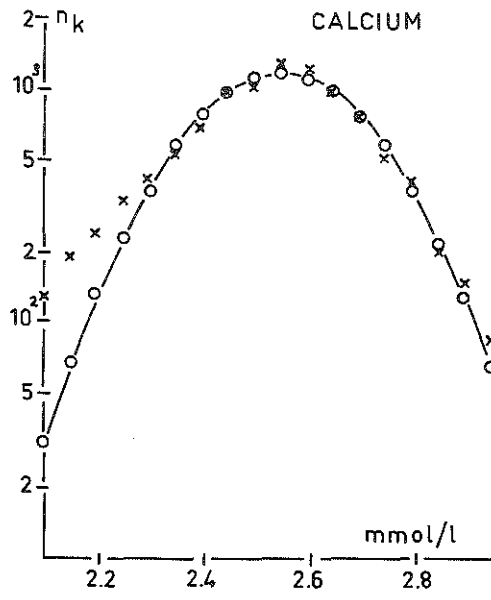
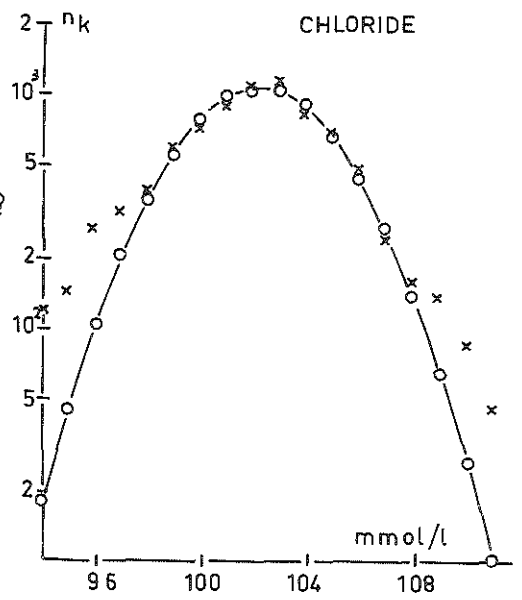
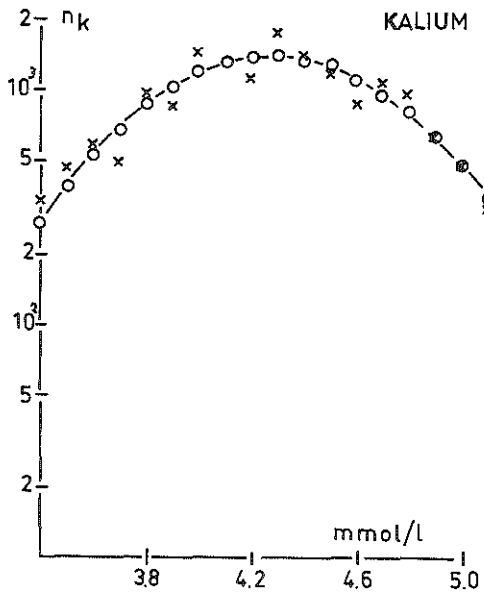
Figuur 6-4.  
 Relatieve cumulatieve frekwentieverdeling, weergegeven met gebruik-  
 making van een logarithmische schaal.  
 Zie p 69.

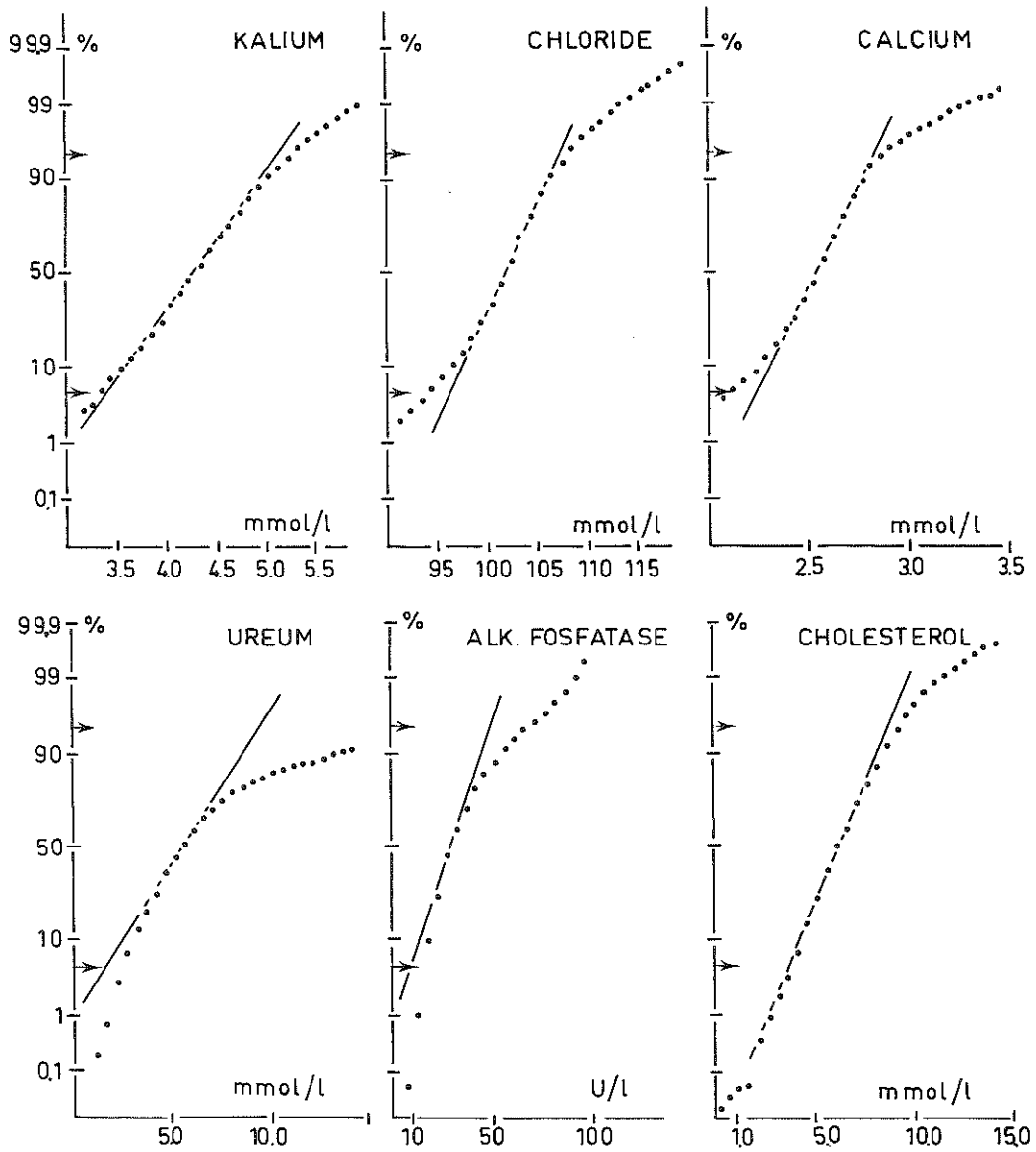


Figuur 6-3.

Resultaten van de 'best passende parabool'-methode. Zie p 69.

x gevonden klassefrequenties en o berekende klassefrequenties



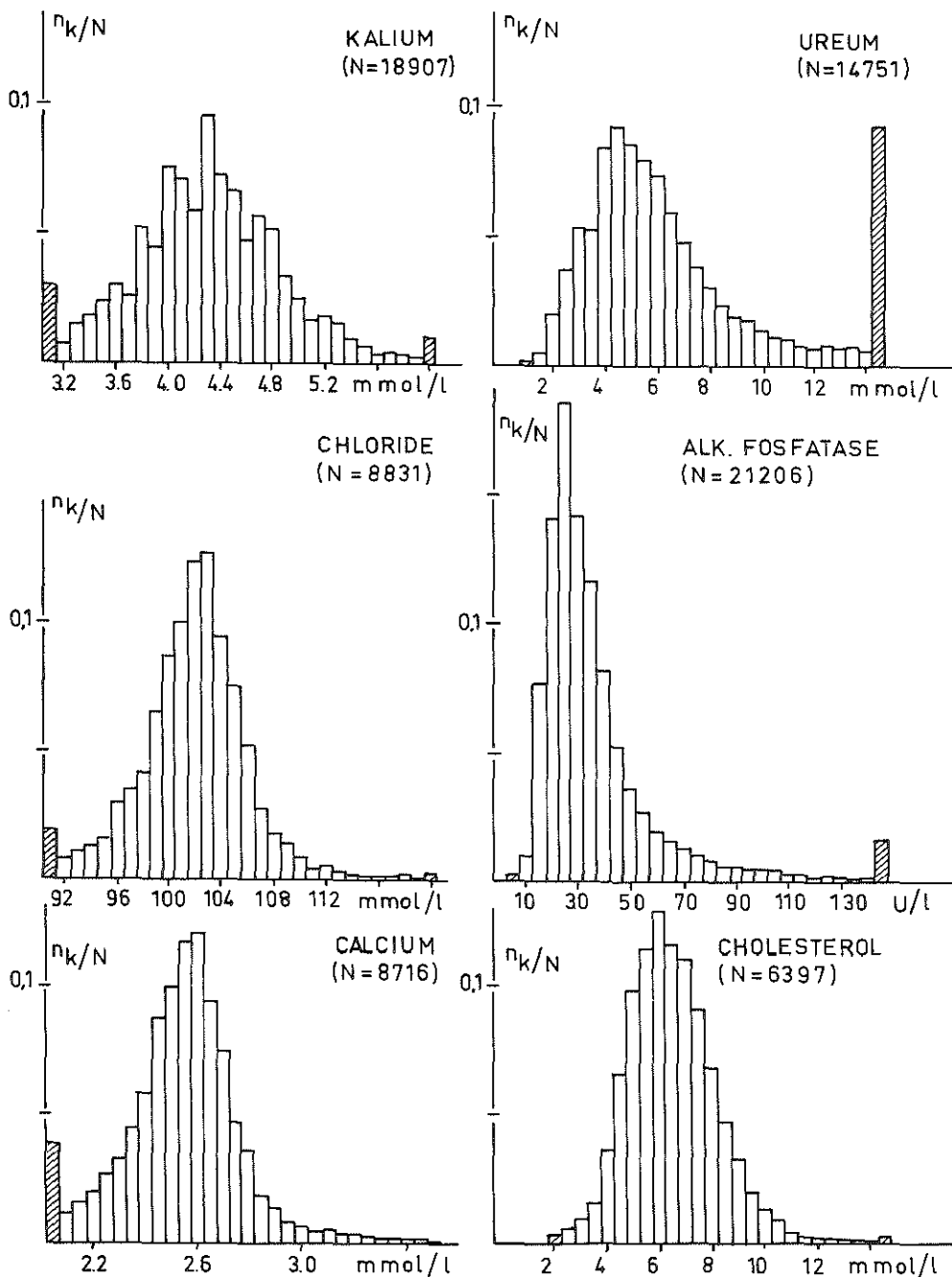


Figuur 6-2.

Relatieve cumulatieve frekwentieverdelingen van bepalingssuitkomsten.

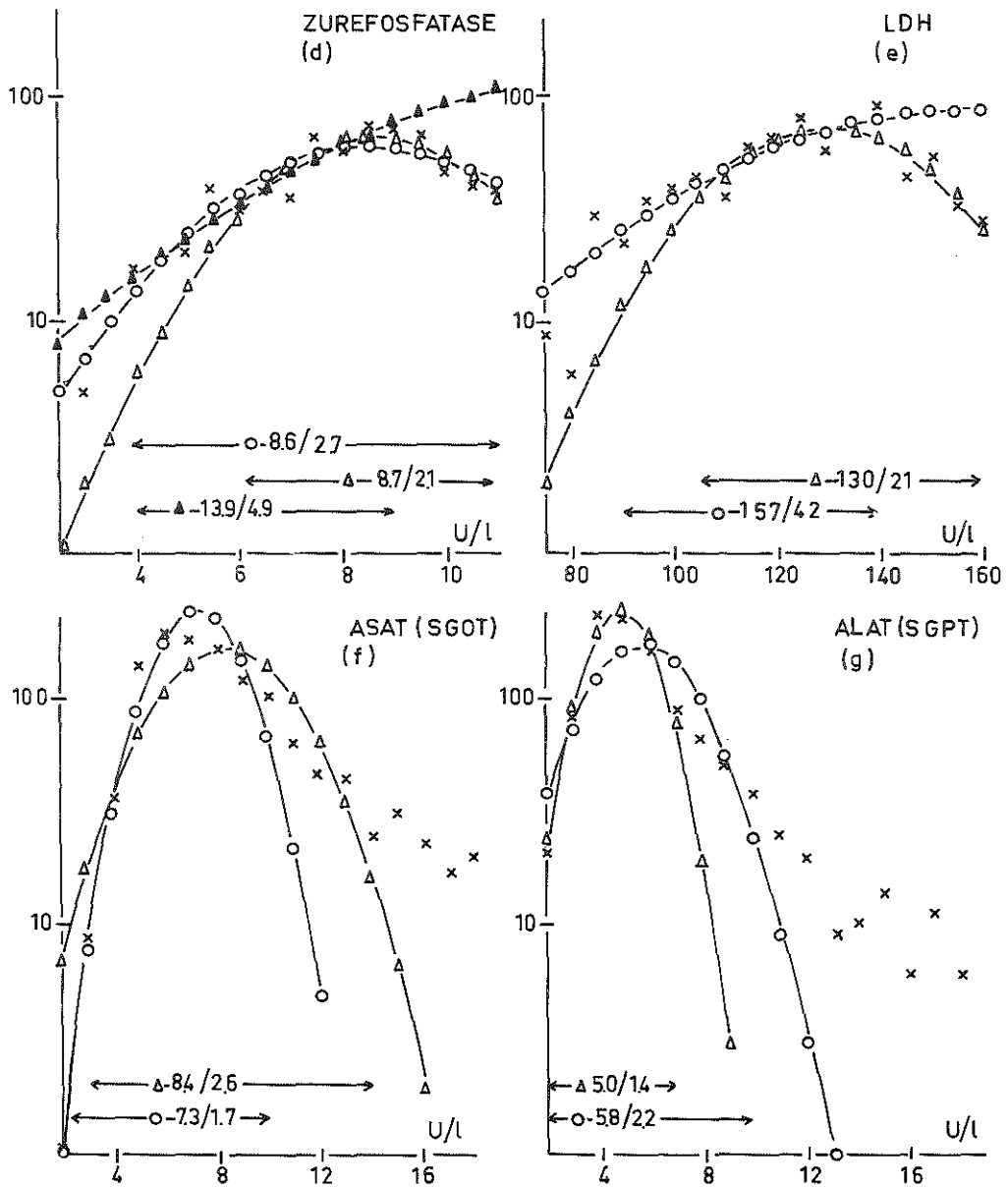
Zie p 69.





Figuur 6-1.

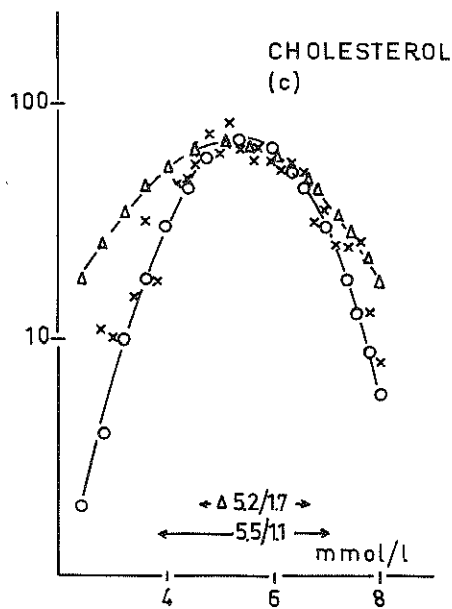
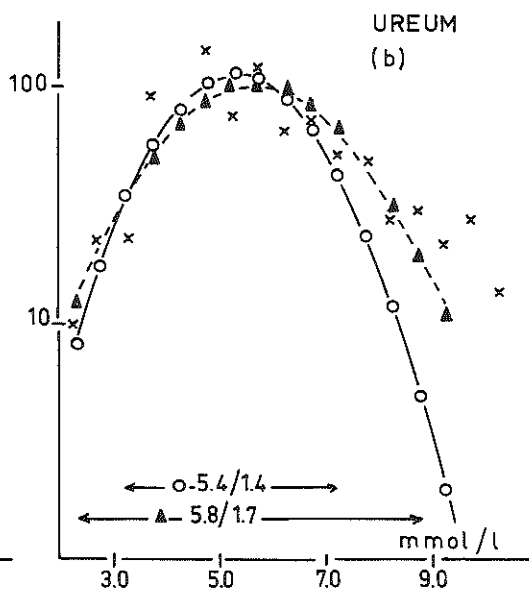
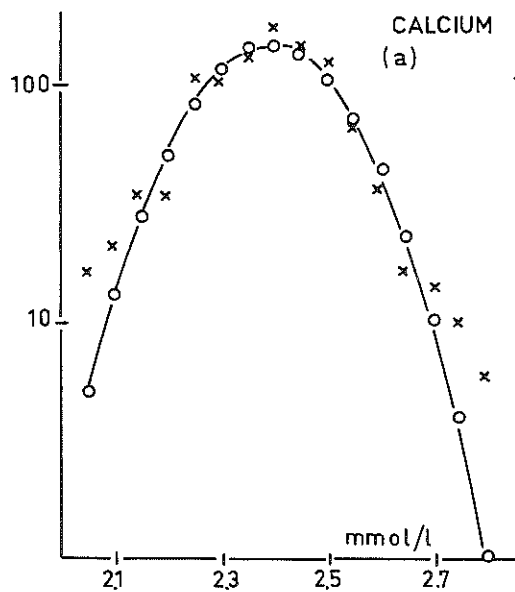
Relatieve frekwentieverdelingen van bepalinguitkomsten. Het verwerkte aantal (N) is per bepaling aangegeven. Zie p 69.

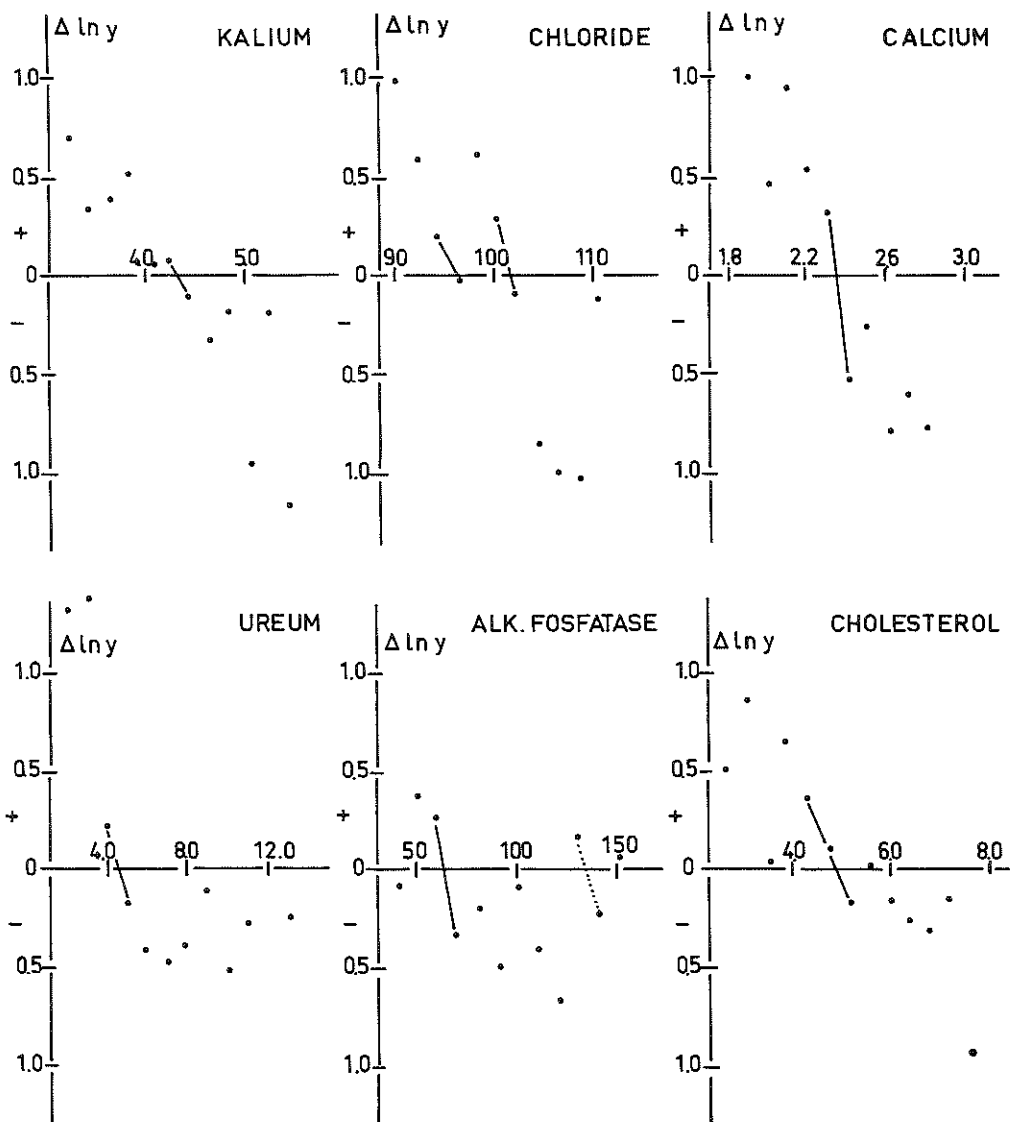


Figuur 5-11.

Resultaten van de 'best passende parabool'-methode. Zie p 66 en 67.

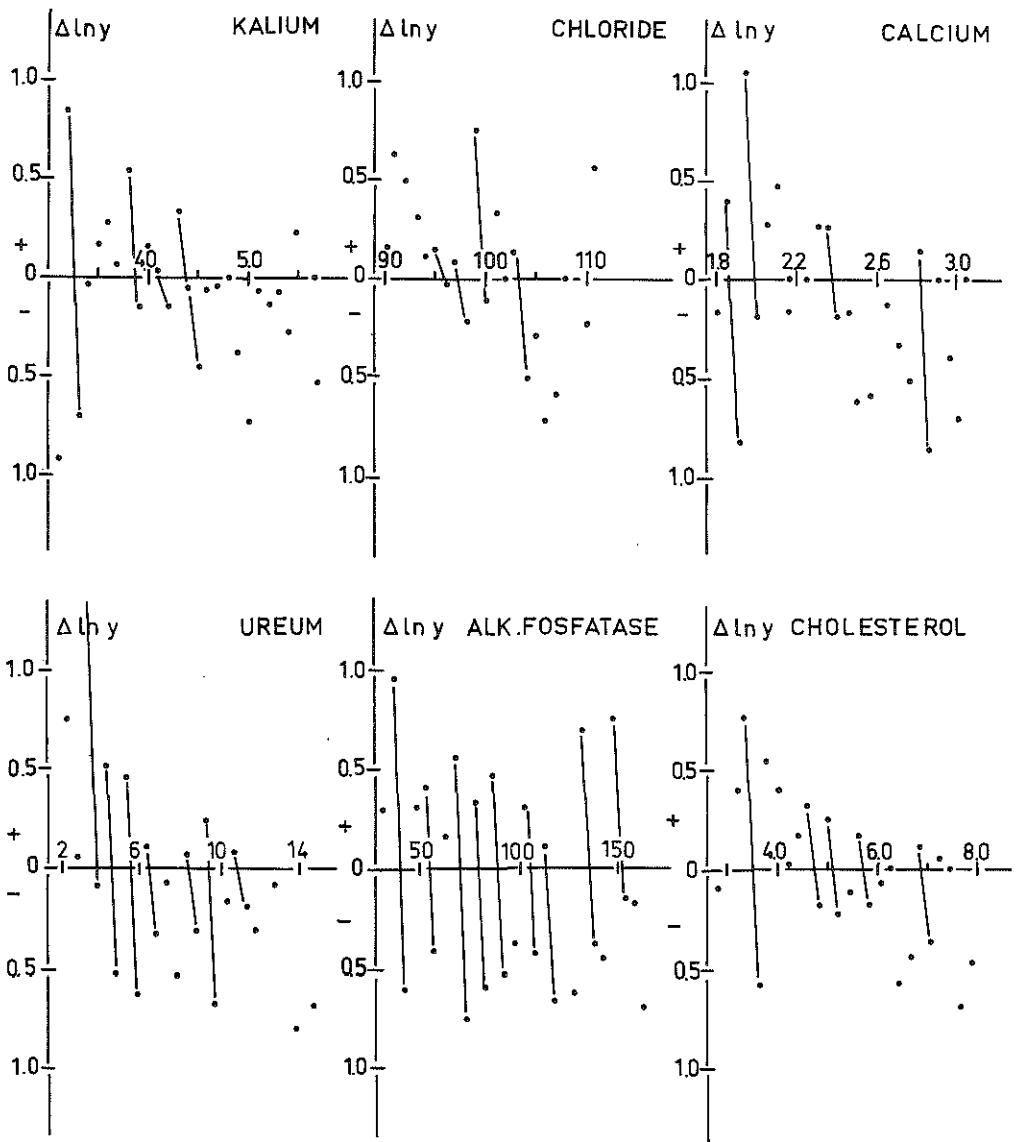
x gevonden klassefrequenties en o,  $\Delta$  en  $\Delta$  berekende klassefrequenties





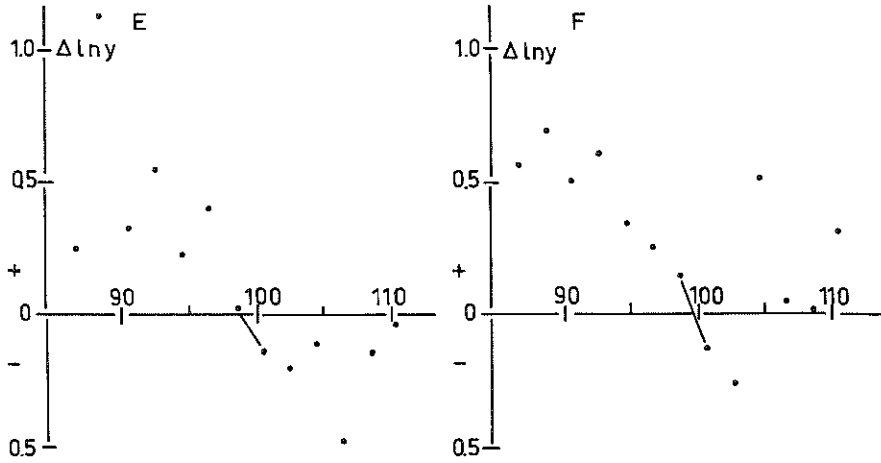
Figuur 5-10.

Toepassing van de methode van Bhattacharya op reeksen bepaling-  
uitkomsten, na verdubbeling van de klassebreedte. Zie p 63.



Figuur 5-9.

Toepassing van de methode van Bhattacharya op reeksen bepalingen-  
uitkomsten. Zie p 63.

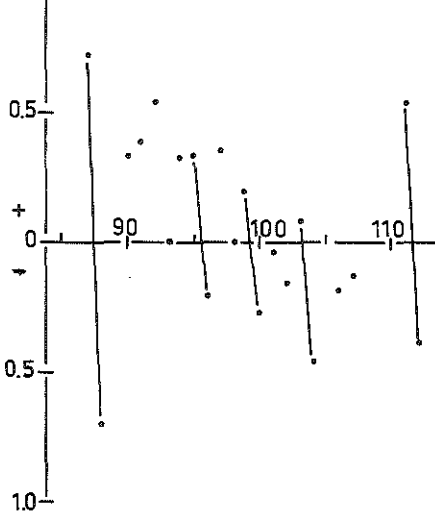


Figuur 5-8.

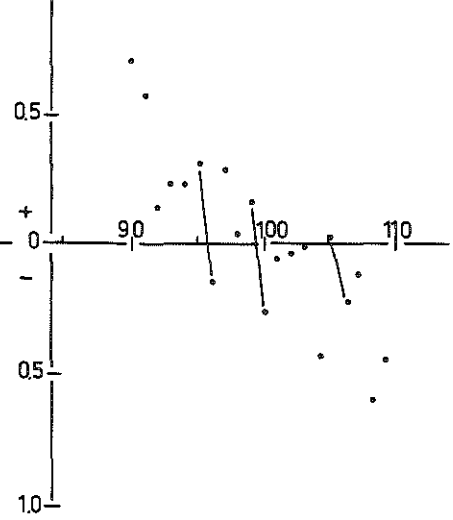
Toepassing van de methode van Bhattacharya op steekproeven.

Zie p 62.

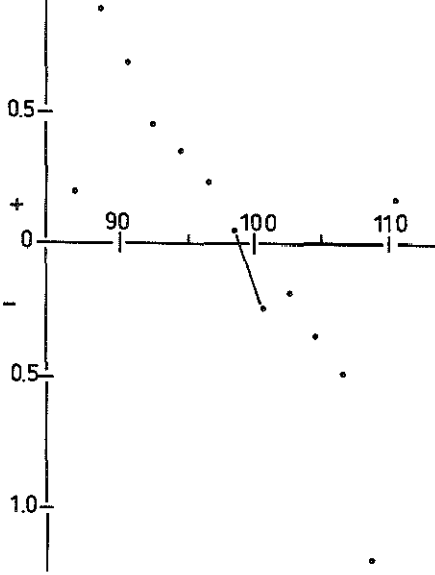
1.0- $\Delta \ln y$  A



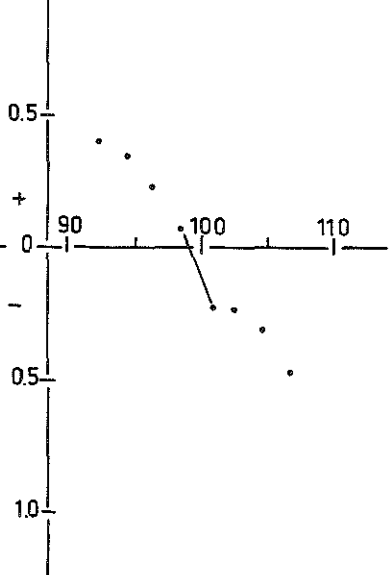
1.0- $\Delta \ln y$  B

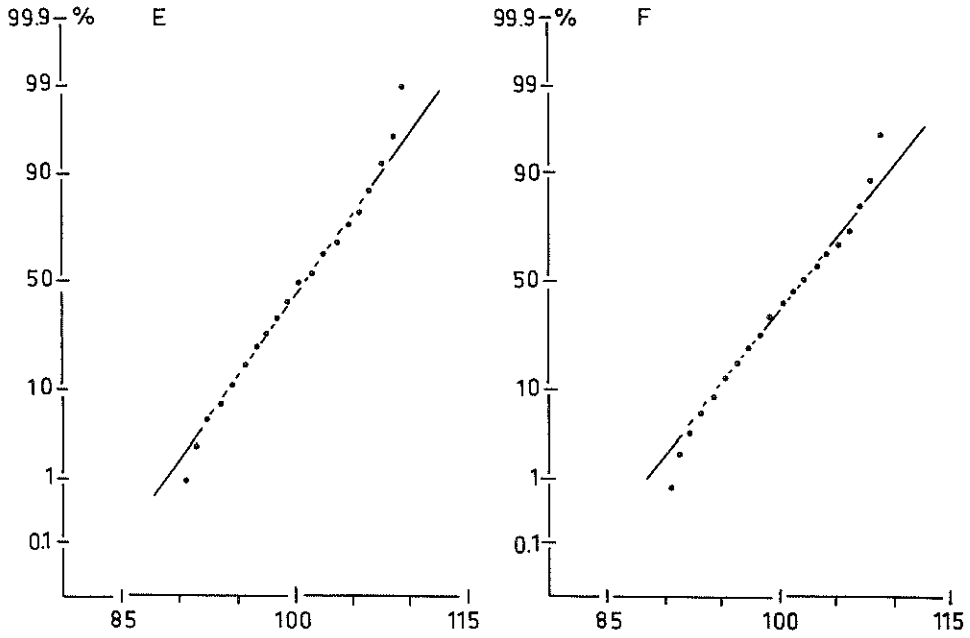


1.0- $\Delta \ln y$  C



1.0- $\Delta \ln y$  D

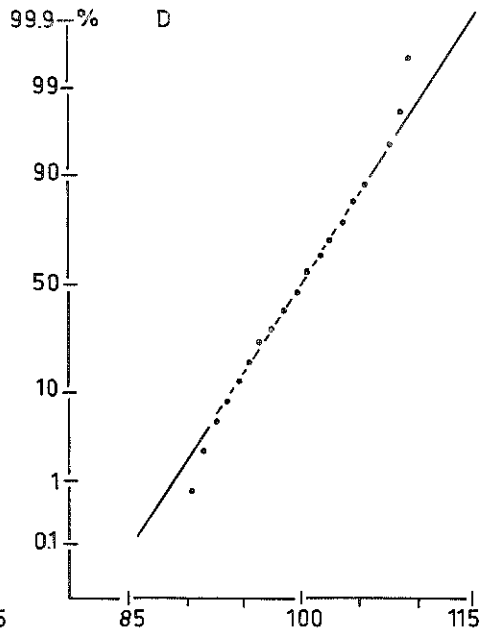
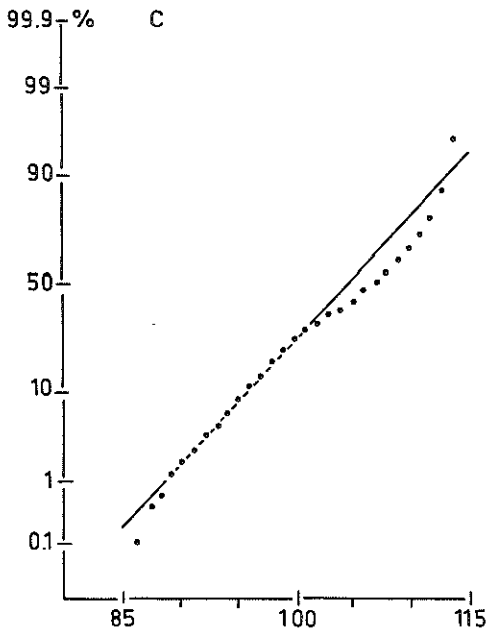
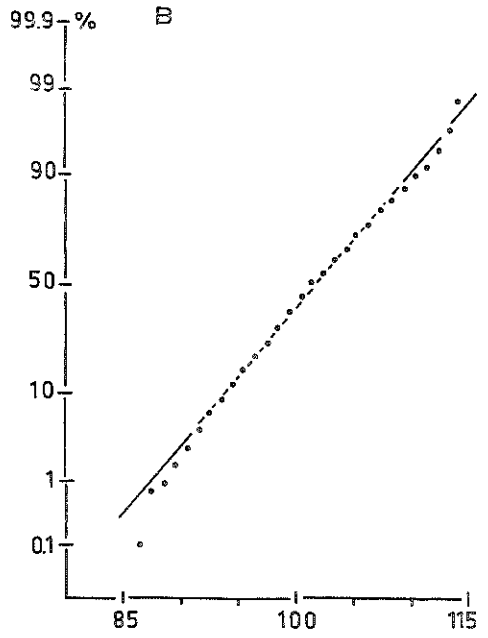
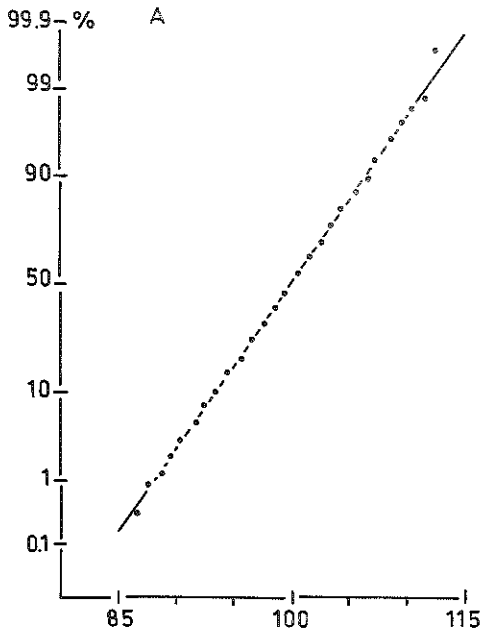


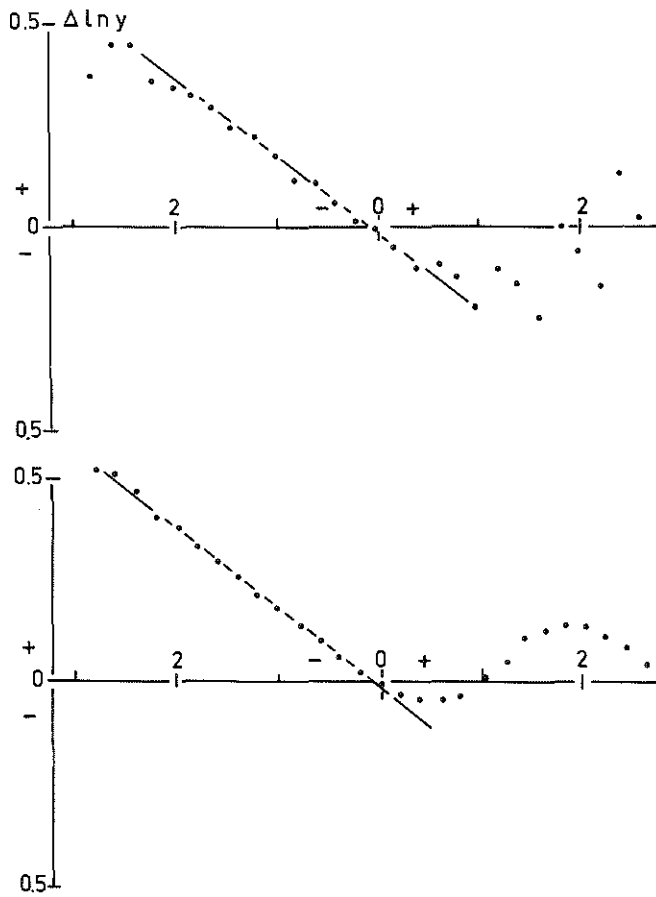


Figuur 5-7.

Relatieve cumulatieve frekwentieverdelingen van steekproeven bestaande uit 1000 elementen, afkomstig uit al of niet gecontamineerde populaties. Zie p 61.

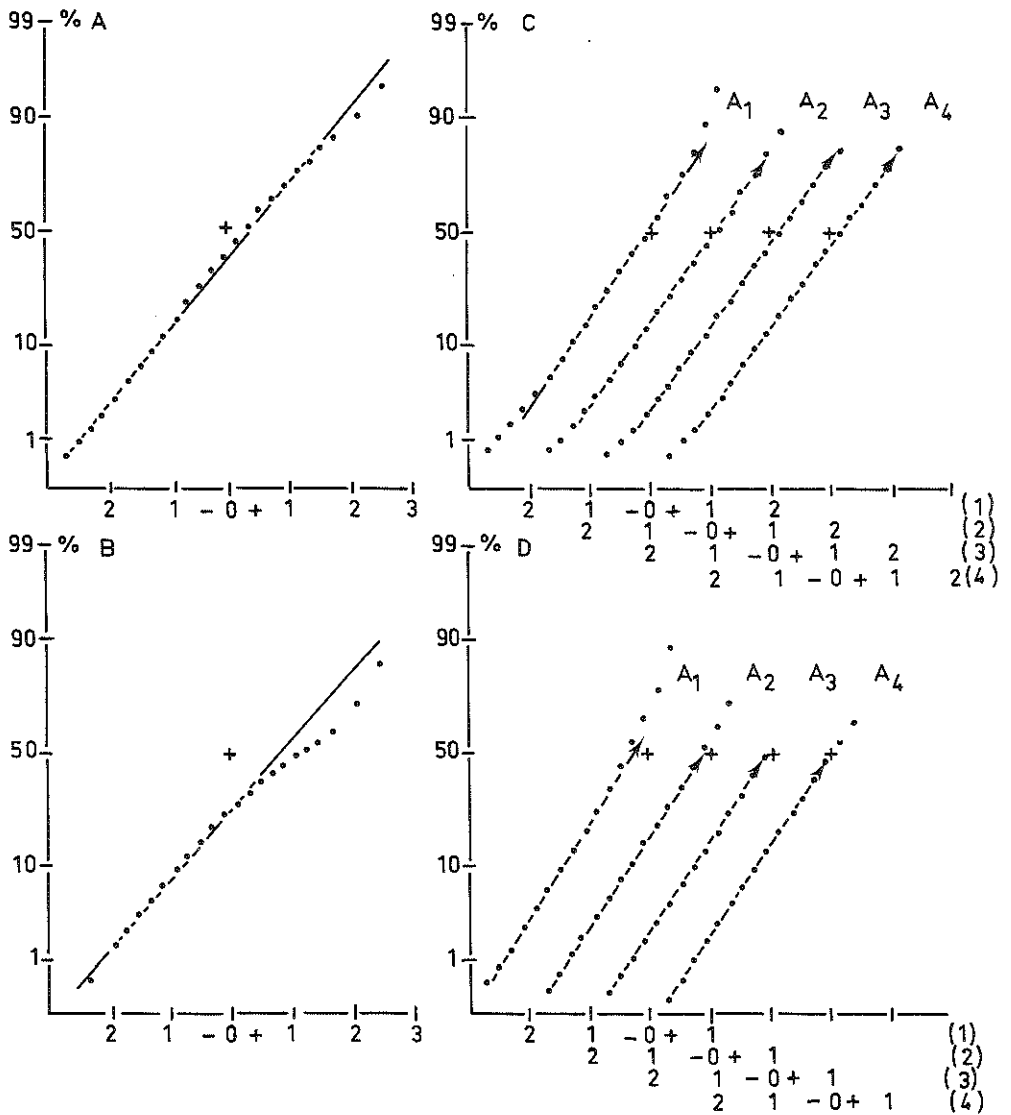






Figuur 5-6.

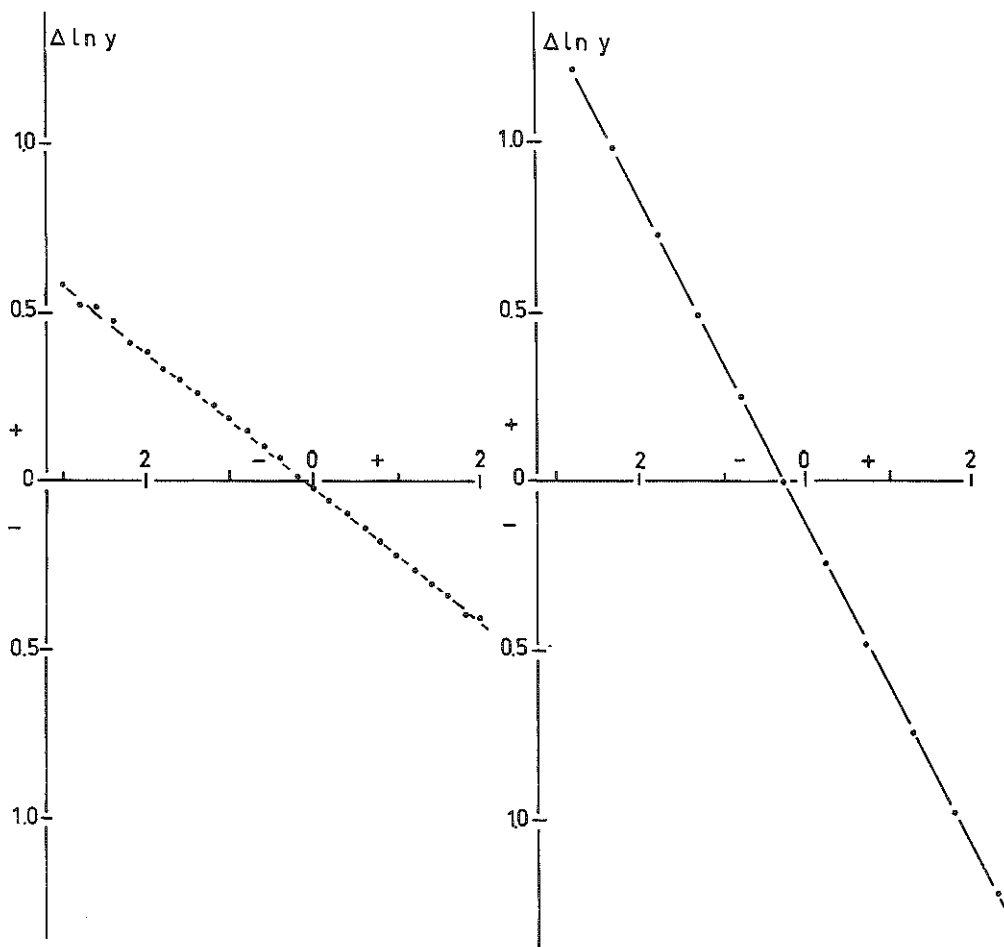
Toepassing van de methode van Bhattacharya op gecontamineerde populaties. Hoofdpopulatie in beide gevallen  $N(0,1)$ ; contaminerende populatie in de bovenste figuur  $N(6,3)$  en in de onderste  $N(3,1)$ . Zie p 61.



Figuur 5-5.

Toepassing van de methode van Neumann op gecontamineerde populaties. In C) en D) is bij elke volgende iteratie de horizontale schaal over een afstand van één eenheid naar rechts verschoven.

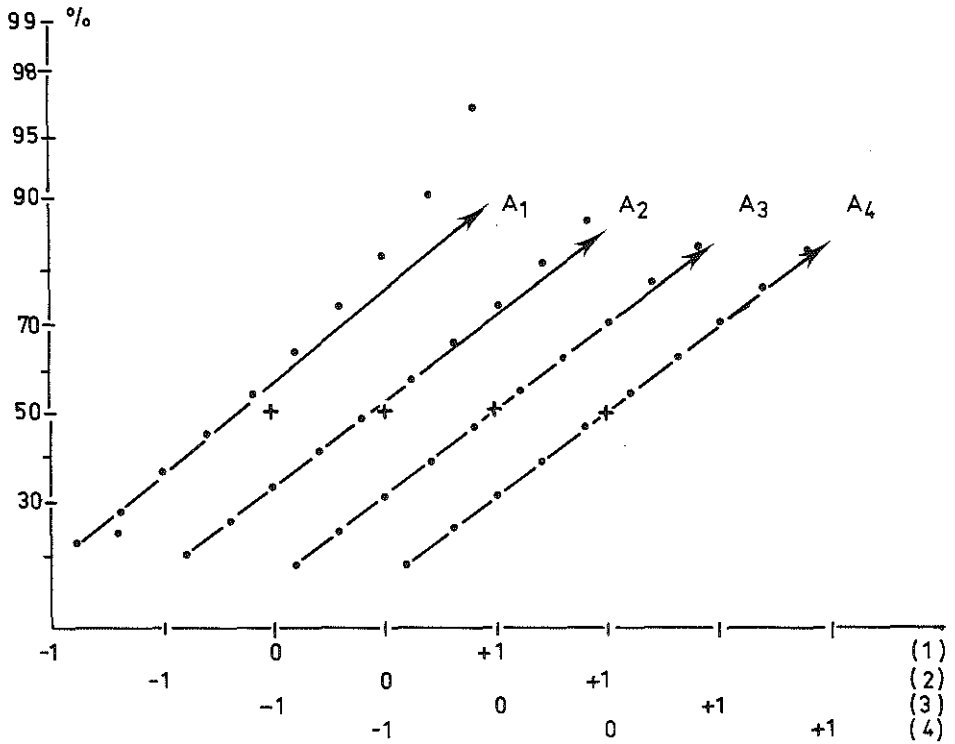
Zie p 60.



Figuur 5-4.

Toepassing van de methode van Bhattacharya op een populatie met standaardnormale verdeling, ingedeeld in klassen met een klassebreedte van 0.2 (links) en 0.5 (rechts).

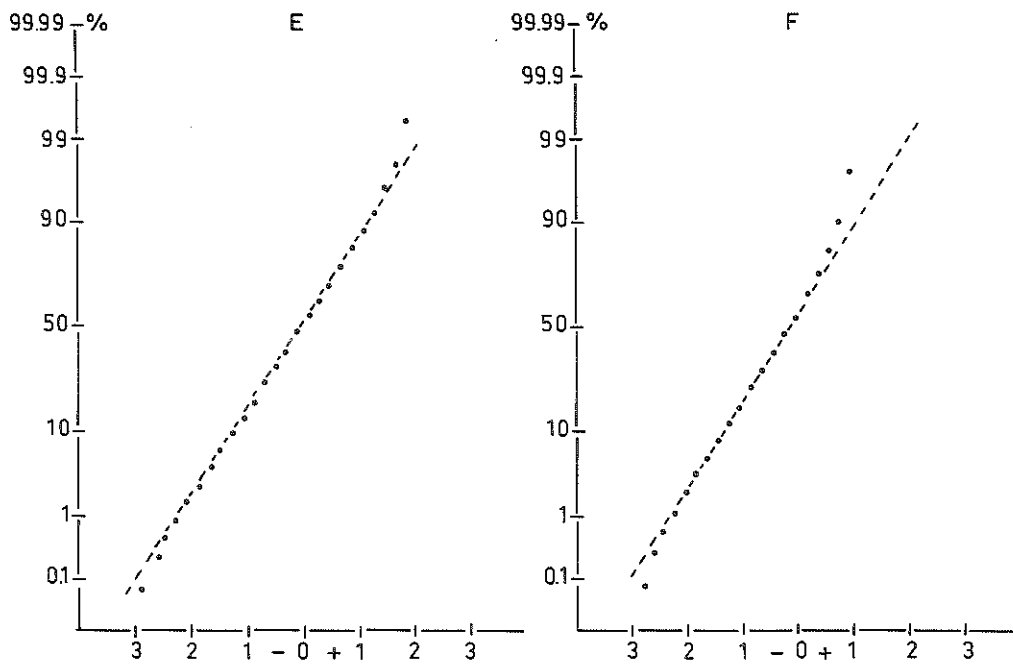
Zie p 60.



Figuur 5-3.

Toepassing van de methode van Neumann op een bij +1 afgekapte standaardnormale verdeling. Bij elke volgende iteratie is de horizontale schaal over een afstand van 1 eenheid naar rechts verschoven.

Zie p 58 en 59.



Figuur 5-2.

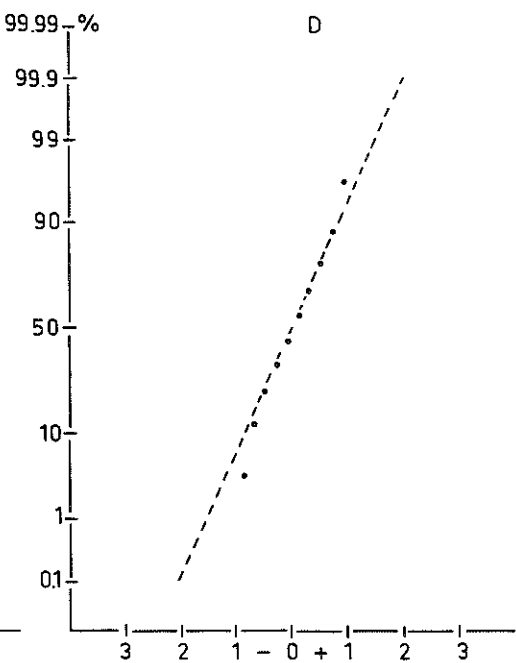
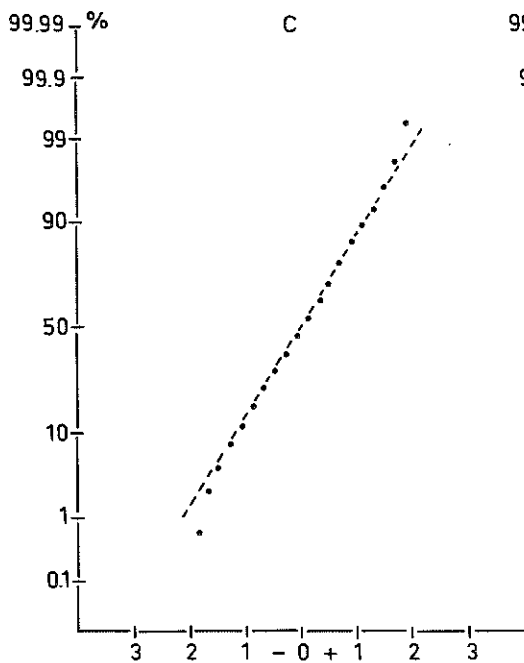
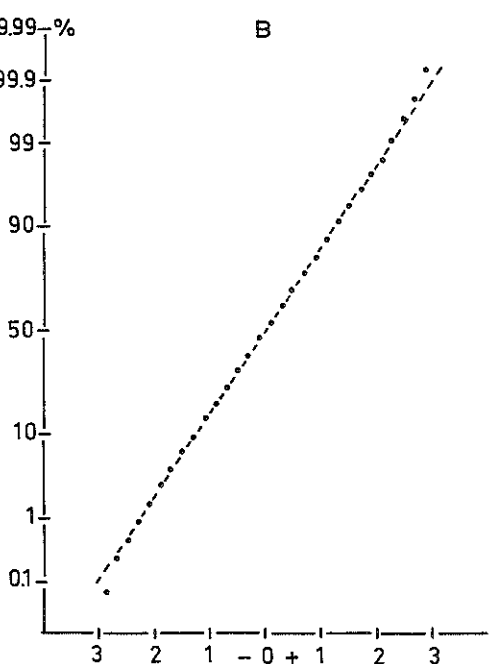
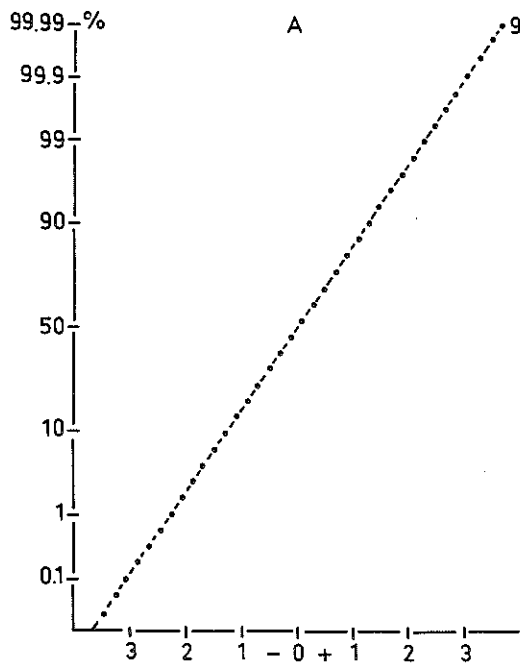
Relatieve cumulatieve frekwentieverdelingen voor:

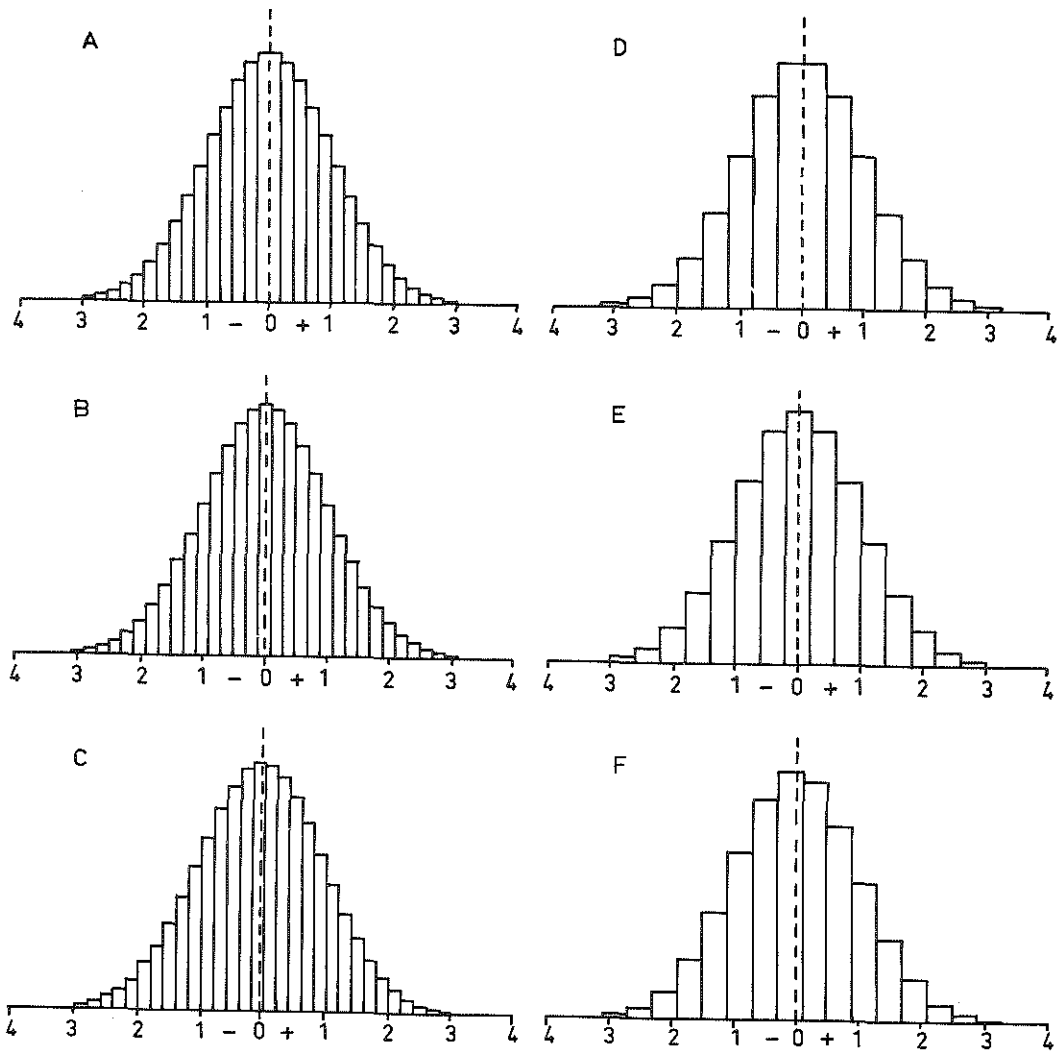
A) de standaardnormale verdeling

B, C en D) symmetrisch afgekapte standaardnormale verdelingen

E en F) asymmetrisch afgekapte standaardnormale verdelingen.

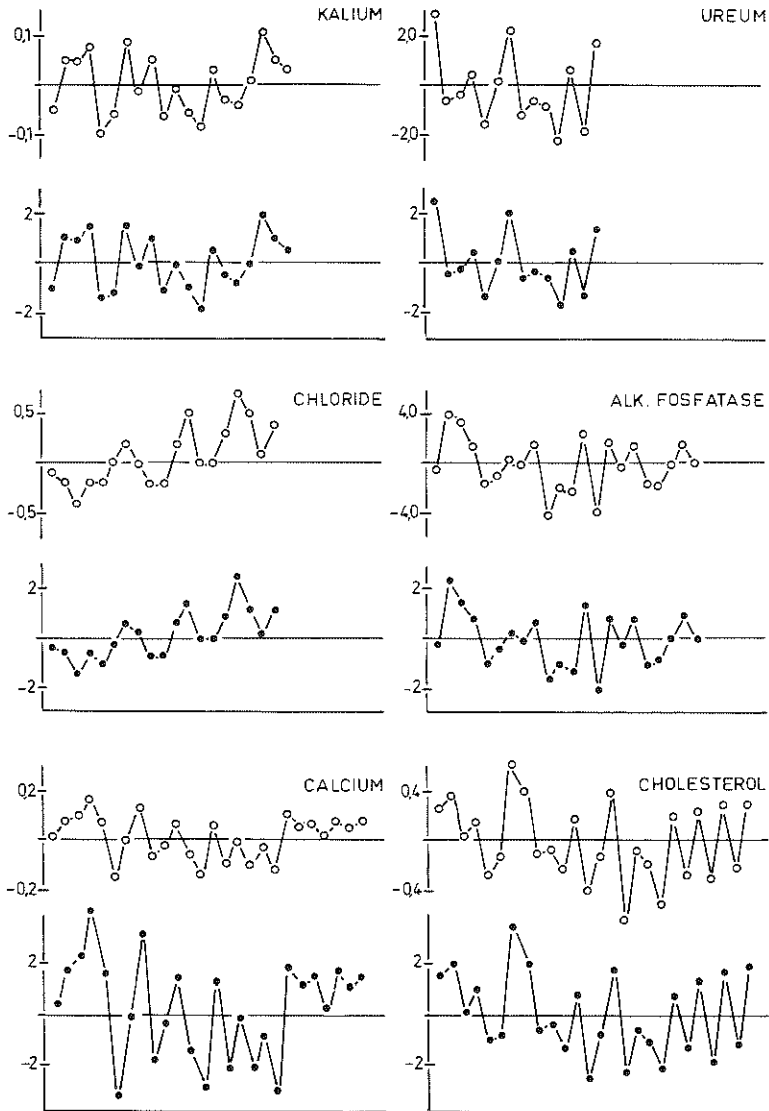
Zie p 58.





Figuur 5-1.  
 Frekwentiediagrammen, berekend voor de standaardnormale verdeling,  
 bij gebruik van verschillende klassegrenzen. Voor A, B en C is de  
 klassebreedte 0.2 en voor D, E en F 0.4. Zie p 58.





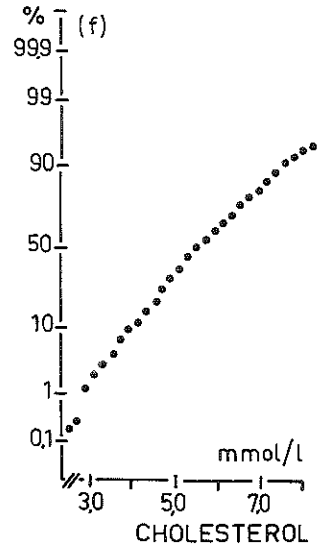
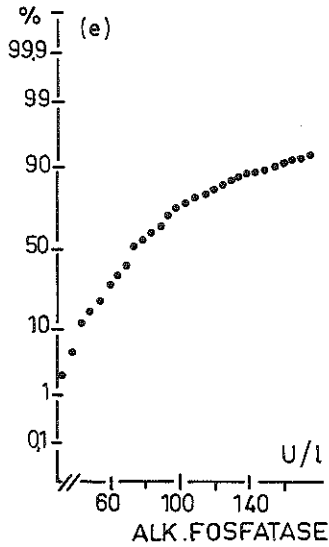
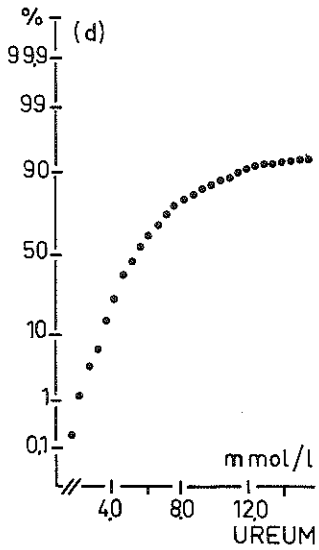
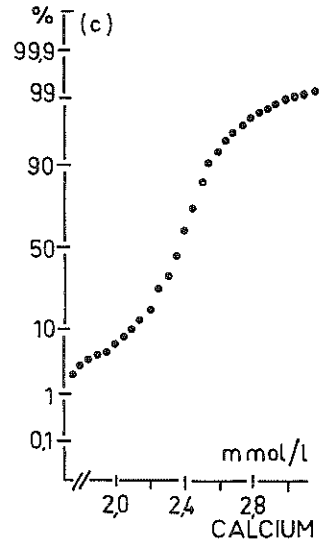
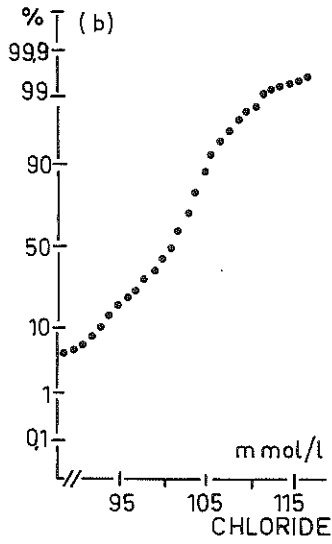
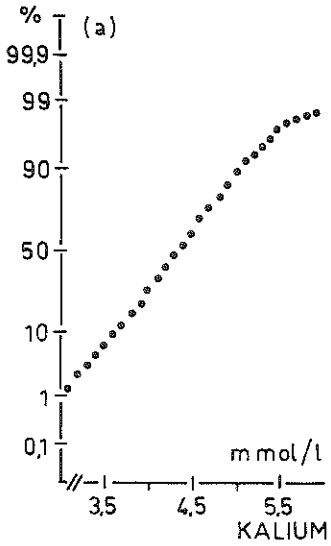
Figuur 4-7.

Het verloop van:

o) het verschil van de per dag berekende GvN-waarde met het groepsgemiddelde, uitgedrukt in de voor de betreffende bepaling gebruikelijke eenheid.

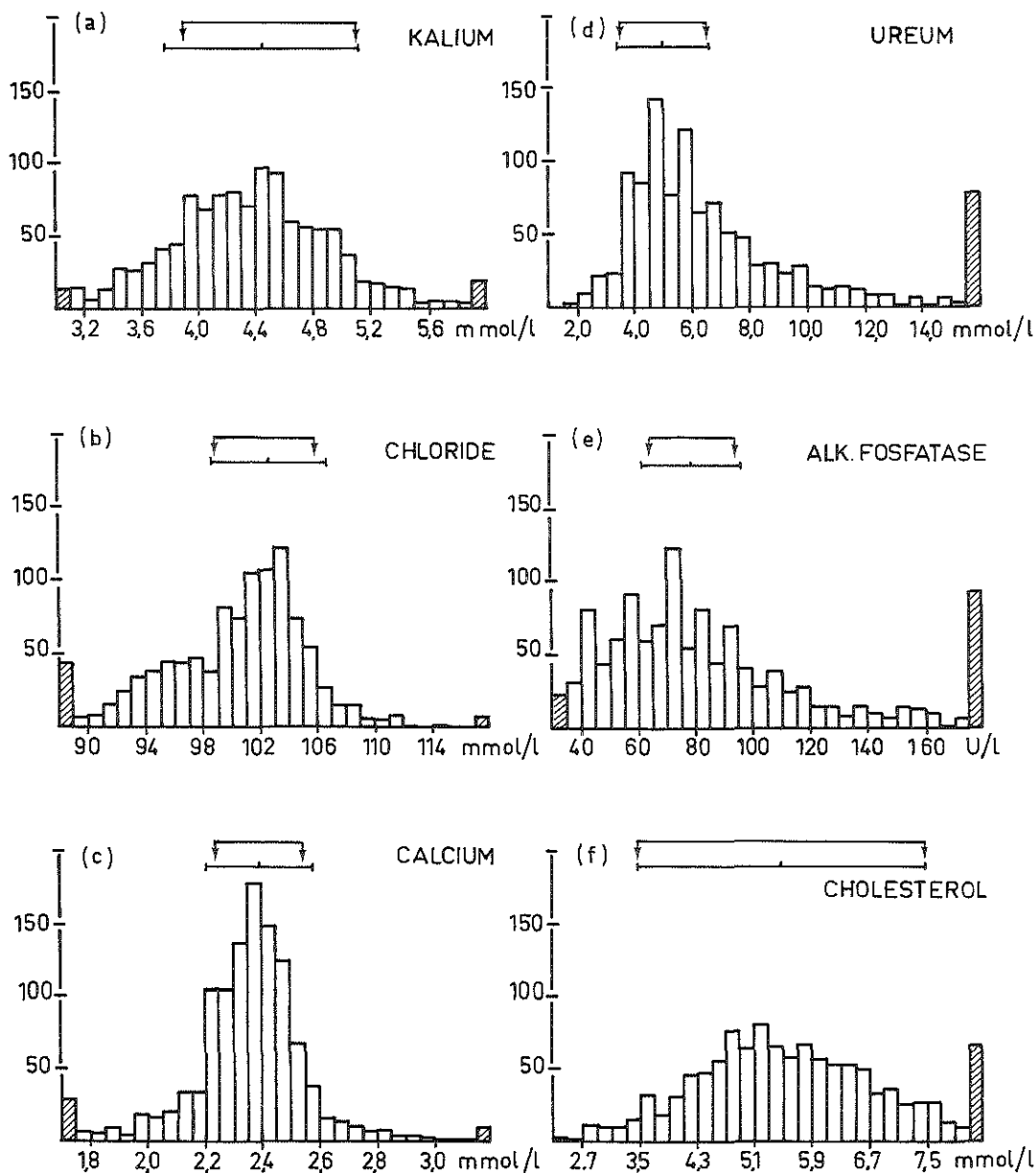
●) hetzelfde verschil, maar nu uitgedrukt in de standaardfout.

Zie p 54.



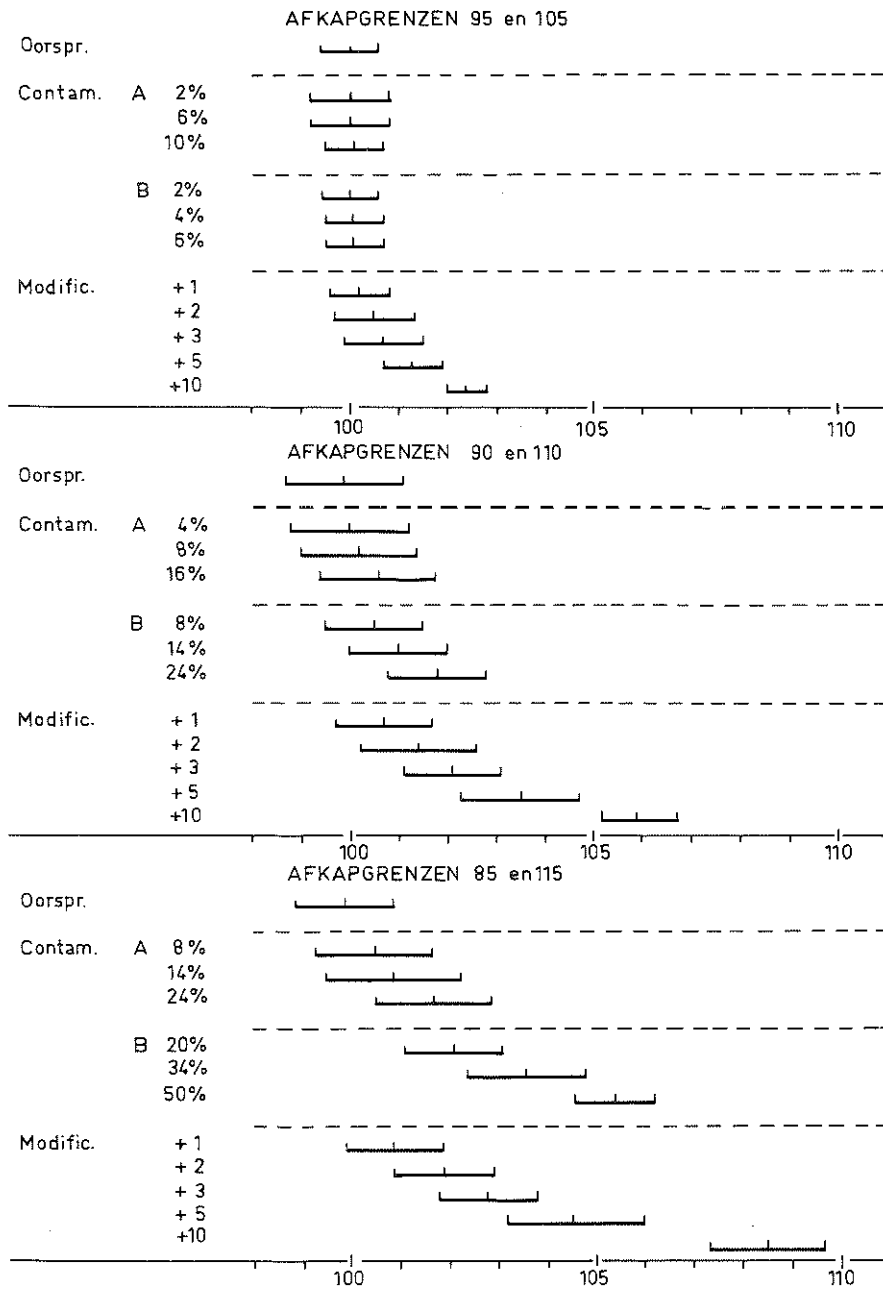
Figuur 4-6.

Voorbeelden van relatieve cumulatieve frekwentieverdelingen.  
Zie p 53 en 54.



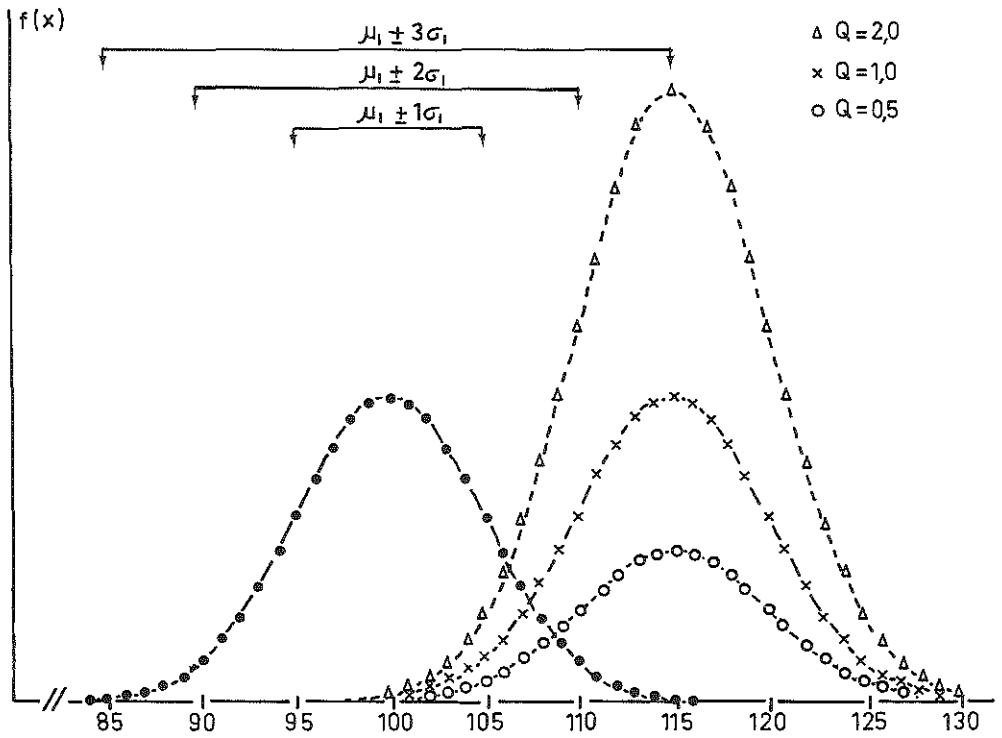
Figuur 4-5.

Voorbeelden van bij het oriënterende onderzoek verkregen frekwentie-diagrammen. Boven de diagrammen zijn de (a priori gekozen) afkapgrenzen aangegeven en het gemiddelde, met naar links en rechts tweemaal de standaardafwijking van de binnen de afkapgrenzen vallende uitkomsten. Zie p 53 en 54.



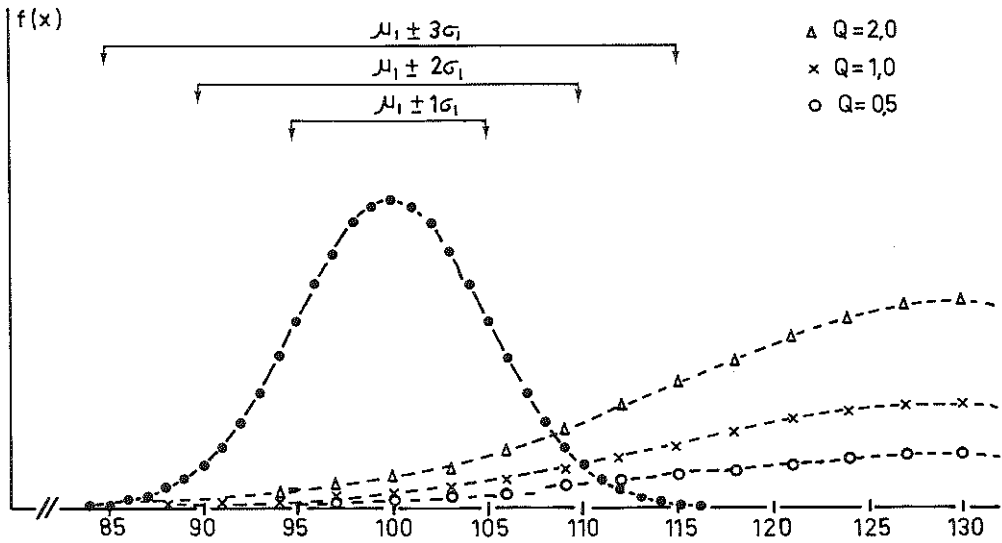
Figuur 4-4.

Overzicht van de met modelproeven verkregen resultaten. Zie p 51.



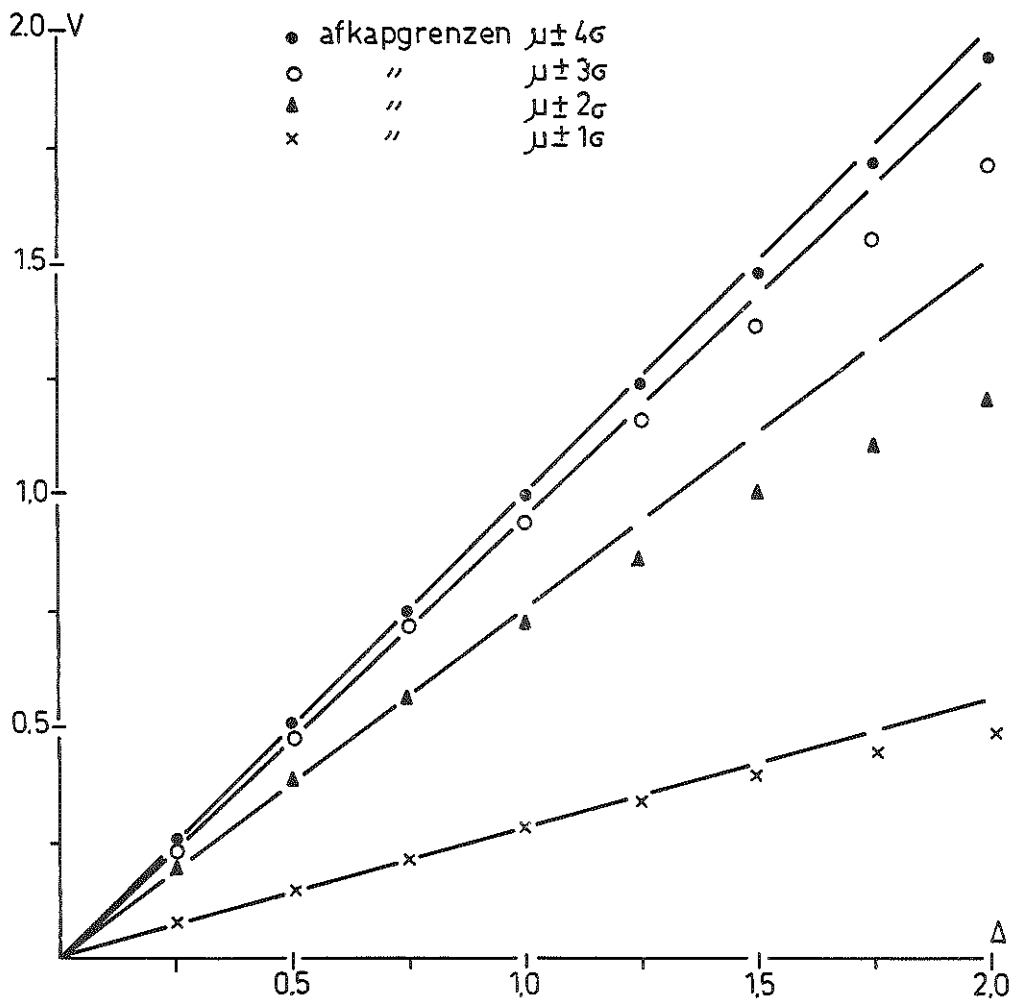
Figuur 4-3.

Dichtheidsfuncties van de hoofdpopulatie  $N(100,5)$  en van de contaminerende populatie  $N(115,5)$ . De afkapgrenzen zijn boven de hoofdpopulatie aangegeven. Zie p 50 en 51.



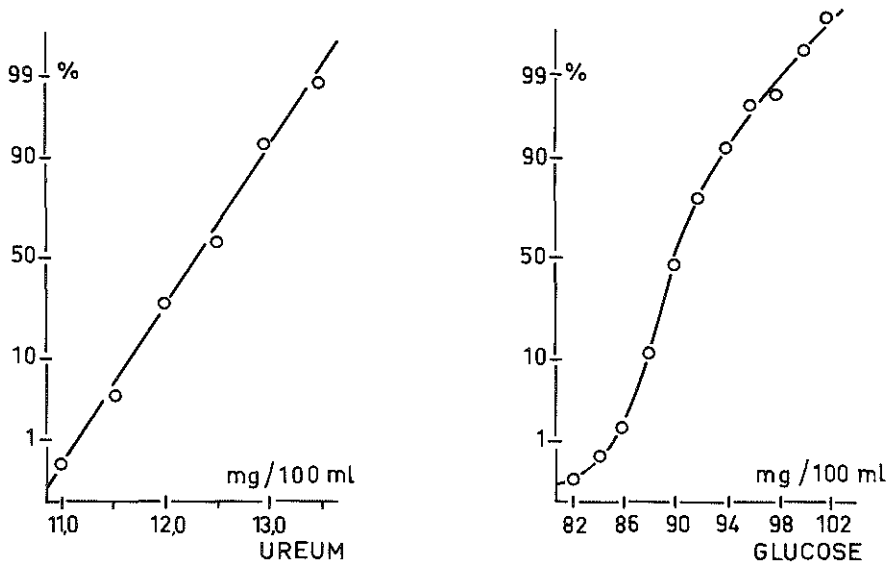
Figuur 4-2.

Dichtheidsfuncties van de hoofdpopulatie  $N(100,5)$  en van de contaminerende populatie  $N(130,15)$ . De afkapgrenzen zijn boven de hoofdpopulatie aangegeven. Zie p 50 en 51.



Figuur 4-1.

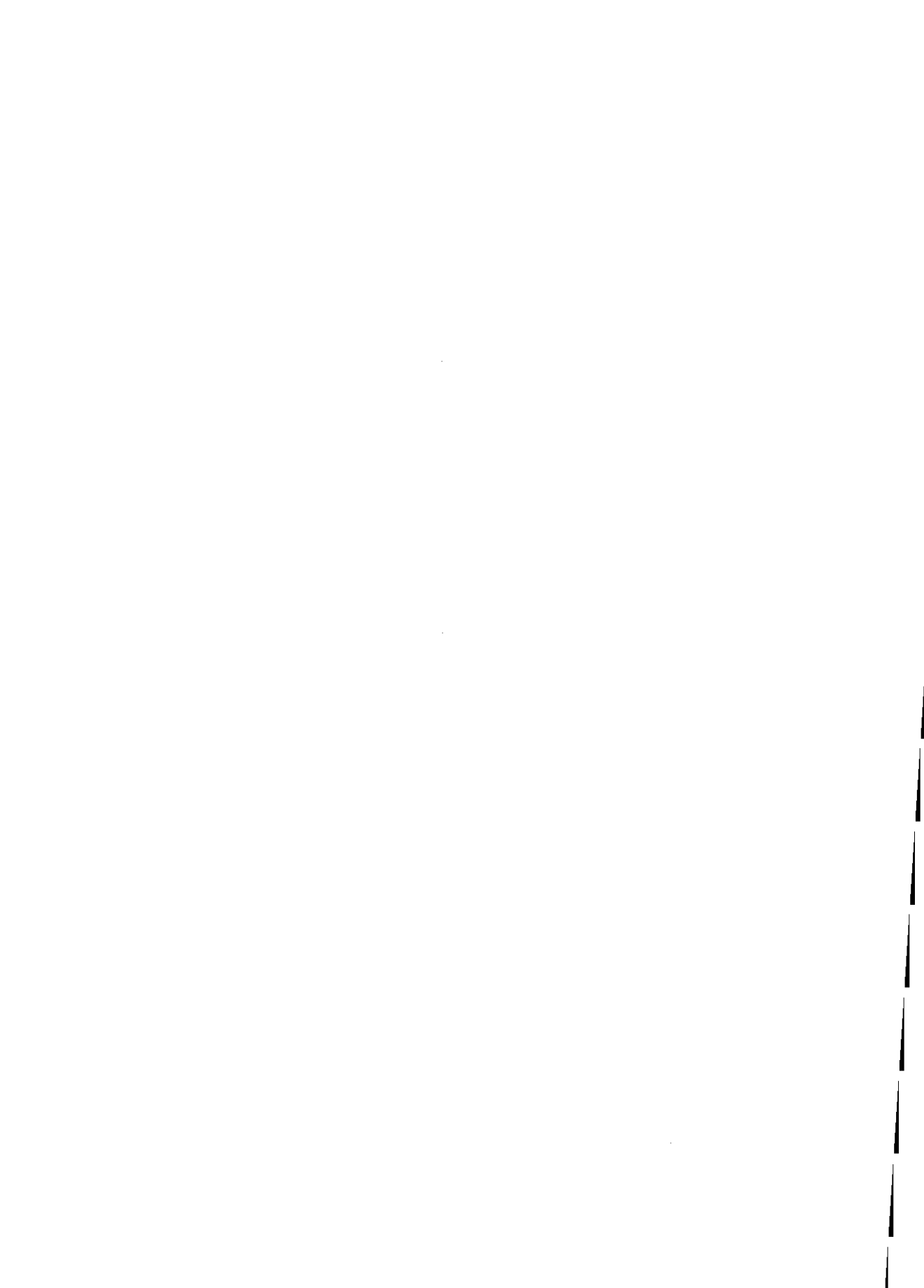
Het verloop van de verschuiving (V) van het gemiddelde van het tussen de afkapgrenzen liggende gedeelte van de populatie, ten gevolge van een modificatie ( $\Delta$ ) van alle elementen van een populatie met standaardnormale verdeling. Zie p 49.



Figuur 2-1.

Cumulative frekwentieverdelingen van bepalingsuitkomsten; gegevens ontleend aan Büttner (1969), zie p 14.





TABEL 6-III

Tolerantie- grenzen	Kalium				Chloride				Calcium			
	4.18-4.42				101.6-104.0				2.45-2.61			
	L	$\bar{D}$	$s_D$	H	L	$\bar{D}$	$s_D$	H	L	$\bar{D}$	$s_D$	H
Mei/Jun	-	4.30	0.06	-	-	102.8	0.6	-	-	2.53	0.04	1
Jul	-	4.31	0.06	-	1	102.3	0.6	-	-	2.53	0.04	-
Aug	-	4.29	0.05	-	3	102.1	0.5	-	-	2.52	0.03	-
Sep	-	4.27	0.04	-	-	102.6	0.8	-	-	2.55	0.03	-
Oct	1	4.27	0.06	-	4	102.2	0.8	-	-	2.55	0.03	1
Nov	-	4.34	0.04	-	7	101.9	0.6	-	I -	2.56	0.03	-
									II 7	2.44		

Tolerantie- grenzen	Ureum				Alk.Fosf.				Cholesterol			
	4.73-5.69				23.3-26.5				5.75-6.87			
	L	$\bar{D}$	$s_D$	H	L	$\bar{D}$	$s_D$	H	L	$\bar{D}$	$s_D$	H
Mei/Jun	-	5.2	0.2	-	1	24.9	0.8	-	2	6.3	0.3	-
Jul	-	5.3	0.2	-	verwerking mislukt				6	6.0	0.3	-
Aug	-	5.2	0.2	-	-	25.7	1.0	6	1	6.2	0.3	1
Sep	-	5.1	0.2	-	-	25.4	1.0	2	-	6.5	0.2	1
Oct	-	5.2	0.2	-	-	25.4	1.0	3	-	6.4	0.3	1
Nov	-	5.4	0.2	-	-	25.6	0.8	2	-	6.4	0.3	-

Overzicht voor de periode van 20-5 tot 1-12-'74.

L aantal van de D-waarden, die onder de lage tolerantiegrens vielen.

$\bar{D}, s_D$  gemiddelde en standaardafwijking van de binnen de tolerantiegrenzen vallende D-waarden.

H aantal van de boven de hoge tolerantiegrens vallende D-waarden.  
Zie p 70 en 71.

TABEL 6-II

Bepaling	Kalium	Chloride	Calcium
Afkapgrenzen	3.40-5.20	97-109	2.20-2.80
$\frac{s_x}{\bar{N}}$	0.45	3.0	0.15
$\frac{s_N}{\bar{D}}$	114	55	65
$s_D$	19	13	15
	4.30	102.8	2.53
	0.06	0.6	0.04
Bepaling	Ureum	Alk.fosf.	Cholesterol
Afkapgrenzen	2.5-8.5	15-35	3.3-9.3
$\frac{s_x}{\bar{N}}$	1.5	6	1.5
$\frac{s_N}{\bar{D}}$	78	93	46
$s_D$	12	20	8
	5.21	24.9	6.31
	0.24	0.8	0.28

Resultaten voor de periode van 20-5 tot 1-7-'74.

$s_x$  afgeronde schatting van de standaardafwijking van de binnen de afkapgrenzen vallende uitkomsten.

$\bar{N}, s_N$  gemiddelde en standaardafwijking van de aantallen uitkomsten, die per dag binnen de afkapgrenzen vielen.

$\bar{D}, s_D$  gemiddelde en standaardafwijking van de per dag berekende GvN-waarde.

Zie p 70.

TABEL 6-I

Methode	<u>Kalium</u>		<u>Chloride</u>		<u>Calcium</u>	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Hoffmann	4.28	0.42	101.9	2.8	2.55	0.15
Best passende parabool	4.28	0.49	102.2	2.9	2.55	0.16

Methode	<u>Ureum</u>		<u>Alk.fosf.</u>		<u>Cholesterol</u>	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Hoffmann	5.8	1.9	30	10	6.4	1.3
Best passende parabool I	4.9	1.8	23	5	6.5	1.6
II	5.2	1.6	24	6		
Log transformatie- Hoffmann	5.6	(2.8-11)	30	(16-58)		

Schattingen van  $\mu$  en  $\sigma$ , zie p 69.

TABEL 5-I

Methode	Kalium		Chloride		Calcium	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Hoffmann	4.43	0.46	101.9(102.4)	2.8(5.6)	2.38	0.13
Hoffmann Neumann	4.40	0.53	101.7	3.5	2.38	0.16
Bhattacharya	4.42	0.48	103.0(97.3)	3.2	2.41	0.11
Best passende parabool	4.40	0.51	102.7	2.6	2.39	0.13

Methode	Ureum		Alk.fosfatase		Cholesterol	
	$\mu$	$\sigma$	$\mu$	$\sigma$	$\mu$	$\sigma$
Hoffmann	5.7	1.18	80	25.0	5.65	1.05
Hoffmann Neumann	5.1	1.2	70	22.5	5.25	1.08
Bhattacharya	5.2	1.6	70	13.0	5.10	0.8
Best passende parabool	5.4	1.4	68	30	5.53	1.13

Schattingen van  $\mu$  en  $\sigma$ , zie p 63 t/m 65.

TABEL 4-VII

Contaminerende populatie	Q	Afkapgrenzen								
		95-105			90-110			85-115		
		C%	$\bar{M}_{50}$	$s_M$	C%	$\bar{M}_{50}$	$s_M$	C%	$\bar{M}_{50}$	$s_M$
N(130,15)	0.5	2	100.0	0.4	4	100.0	0.6	8	100.5	0.6
	1.0	6	100.0	0.4	8	100.2	0.6	14	100.9	0.7
	2.0	10	100.1	0.3	16	100.6	0.6	24	101.7	0.6
N(115,5)	0.5	2	100.0	0.3	8	100.5	0.5	20	102.1	0.5
	1.0	4	100.1	0.3	14	101.0	0.5	34	103.6	0.6
	2.0	6	100.1	0.3	24	101.8	0.5	50	105.4	0.4

Gemiddelde ( $\bar{M}_{50}$ ) en standaardafwijking ( $s_M$ ) van 20 steekproefgemiddelden (50 elementen per steekproef) verkregen voor combinaties van 3 paren afkapgrenzen met contaminatiepercentages (C%), die waren gebaseerd op de in tabel 4- VI vermelde resultaten.

Zie p 51.

TABEL 4-VI

Contaminerende populatie	Q	Afkapgrenzen		
		95-105	90-110	85-115
M(130,15)	0.5	2.7	4.4	7.3
	1.0	5.2	8.4	13.6
	2.0	9.9	15.6	24.0
M(115,5)	0.5	1.5	7.7	20.0
	1.0	3.2	14.3	33.4
	2.0	6.3	25.0	50.1

$$Q = \frac{\text{Totale aantal elementen van de contaminerende populatie}}{\text{Totale aantal elementen van de oorspronkelijke populatie}}$$

Contaminatiepercentages voor combinaties van Q en paren afkapgrenzen.

Zie p 51.

TABEL 4-V

Nr	Populatie	$a_1=95, a_2=105$		$a_1=90, a_2=110$		$a_1=85, a_2=115$	
		$\bar{M}_{50}$	$s_M$	$\bar{M}_{50}$	$s_M$	$\bar{M}_{50}$	$s_M$
<u>A</u>							
1	N(100,5)	100.0	0.3	99.9	0.6	99.9	0.5
2	N(101,5)	100.2	0.3	100.7	0.5	100.9	0.5
3	N(102,5)	100.5	0.4	101.4	0.6	101.9	0.5
4	N(103,5)	100.7	0.4	102.1	0.5	102.8	0.5
5	N(105,5)	101.3	0.3	103.5	0.6	104.6	0.7
6	N(110,5)	102.4	0.2	105.9	0.4	108.5	0.6
<u>B</u>							
1	N(0,1)	0.00	0.06	-0.02	0.12	-0.02	0.10
2	N(0.2,1)	0.04	0.06	0.14	0.10	0.18	0.10
3	N(0.4,1)	0.10	0.08	0.28	0.12	0.38	0.10
4	N(0.6,1)	0.14	0.08	0.42	0.10	0.56	0.10
5	N(1,1)	0.26	0.06	0.70	0.12	0.92	0.14
6	N(2,1)	0.48	0.04	1.18	0.08	1.70	0.12

$a_1$  en  $a_2$  = gebruikte afkapgrenzen

$\bar{M}_{50}$  = gemiddelde van de steekproefgemiddelden van de uit 50 elementen bestaande steekproeven.

Onder A worden de gegevens vermeld zoals die door het programma geleverd zijn en onder B na standaardisering van de eerste populatie. Zie p 50.



TABEL 4-III

a <sub>1</sub>	a <sub>2</sub>	Δ							
		0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00
-4	+4	0.25	0.50	0.75	1.00	1.24	1.48	1.72	1.94
-3	+3	0.24	0.48	0.72	0.94	1.16	1.36	1.55	1.71
-2	+2	0.19	0.38	0.56	0.72	0.86	1.00	1.10	1.20
-1	+1	0.07	0.14	0.21	0.28	0.34	0.39	0.44	0.49

Waarden van  $v$  voor vier grenzenparen  $a_1$  en  $a_2$  in combinatie met 8 waarden van  $\Delta$ . Zie p 49.

TABEL 4-IV

Nr	Populatie	$\bar{M}_{10}$		$\bar{M}_{25}$		$\bar{M}_{50}$		$\bar{M}_{100}$		
		$s_M$	$s_M$	$s_M$	$s_M$	$s_M$	$s_M$			
<u>A</u>										
1	N(100,5)	99.6	1.4	100.0	1.1	99.9	0.6	99.9	0.3	
2	N(105,5)	103.5	1.3	103.5	0.7	103.5	0.6	103.4	0.4	
3	N(110,5)	105.7	1.2	105.9	0.5	105.9	0.4	105.9	0.3	
<u>B</u>										
1	N(0,1)	-0.08	0.28	0.00	0.22	-0.02	0.12	-0.02	0.06	
2	N(1,2)	+0.70	0.26	0.70	0.14	0.70	0.12	0.68	0.08	
3	N(2,1)	+1.14	0.24	1.18	0.10	1.18	0.08	1.18	0.06	

$\bar{M}_k$  = gemiddelde van de steekproefgemiddelden van de uit  $k$  elementen bestaande steekproeven.

Gebruikte afkapgrenzen :  $a_1 = 90$  en  $a_2 = 110$ .

Onder A worden de gegevens vermeld zoals die door het programma geleverd zijn en onder B na standaardisering van de eerste populatie. Zie p 49.

TABEL 4-II

$\Delta$	$a_1 \backslash a_2$	-3	-2	-1	0	+1	+2	+3	+4
+1	-4	0.00	0.01	0.15	0.27	0.50	0.77	0.95	0.99 <sup>6</sup>
-1	-4	-0.01	-0.14	-0.26	-0.48	-0.76	-0.94 <sup>5</sup>	-0.99 <sup>2</sup>	-0.99 <sup>6</sup>
+1	-3		0.01	0.14	0.26	0.48	0.76	0.94 <sup>5</sup>	0.99 <sup>2</sup>
-1	-3		-0.06	-0.21	-0.44	-0.72	-0.90	-0.94 <sup>5</sup>	-0.95
+1	-2			0.06	0.21	0.44	0.72	0.90	0.94 <sup>5</sup>
-1	-2			-0.08	-0.28	-0.54	-0.72	-0.76	-0.77
+1	-1				0.08	0.28	0.54	0.72	0.76
-1	-1				-0.08	-0.28	-0.44	-0.48	-0.50
+1	0					0.08	0.28	0.44	0.48
-1	0					-0.08	-0.21	-0.26	-0.27
+1	+1						0.08	0.21	0.26
-1	+1						-0.06	-0.14	-0.15
+1	+2							0.06	0.14
-1	+2							-0.01	-0.01
+1	+3								0.01
-1	+3								-0.00

Waarden van  $v$  voor  $\Delta = +1$  en  $\Delta = -1$  en combinaties van  $a_1$  en  $a_2$  .

Zie p 48.

TABEL 4-I

$a_1 \backslash a_2$	-3	-2	-1	0	+1	+2	+3	+4
-4	-3.31	-2.37	-1.53	-0.80	-2.29	-0.05 <sup>5</sup>	-0.00 <sup>4</sup>	0.00
-3	-	-2.32	-1.51	-0.79	-0.28	-0.05	0.00	+0.00 <sup>4</sup>
-2	-	-	-1.38	-0.72	-0.23	0.00	+0.05	+0.05 <sup>5</sup>
-1	-	-	-	-0.46	0.00	+0.23	+0.28	+0.29
0	-	-	-	-	+0.46	+0.72	+0.79	+0.80
+1	-	-	-	-	-	+1.38	+1.51	+1.53
+2	-	-	-	-	-	-	+2.32	+2.37
+3	-	-	-	-	-	-	-	+3.31

Waarden van  $\zeta(a_1 < \underline{z} > a_2)$  voor combinaties van  $a_1$  en  $a_2$ .

Zie p 48.

TABEL 3-I

<u>Veronderstelde verdelingstype</u>	<u>Gebruikte methode</u>	<u>Haptoglobineuitk.</u>		<u>Gauss</u>		<u>Log-Gauss</u>	
		<u>?</u>	<u>?</u>	<u>12</u>	<u>179</u>	<u>31</u>	<u>231</u>
Gauss	95-50 tolerantie- interval	11	180	13	180	-6	202
Log-Gauss	95-50 tolerantie- interval	31	233	16	306	32	233
Geen	95-50 tolerantie- interval r=2,s=3	21	191	-5	183	26	245
Idem	95-50 tolerantie- interval r=3,s=2	21	199	15	184	32	245
Idem	95-50 tolerantie- interval gemidd.	21	195	5	184	29	245
Idem	Percentielschat- ting	21	195	5	184	29	245

Gegevens ontleend aan Reed e.a. (1971), zie p 41 en 42.

Met '95-50 tolerantieinterval' wordt bedoeld de schatting, met een betrouwbaarheid van 50%, van de grenzen, die 95% van de populatie omvatten.

TABEL 2-X

Gem. aantal p.dag	Standaardafwijking van het daggemiddelde								
	Albu- mine	Alk. fosfat.	Asp.am. transf.	Bili- rubine	Cal- cium	Ka- lium	Na- trium	Tot. eiw.	Ureum
90	0.10	0.55	1.48	0.05	0.12	0.11	1.03	0.11	2.00
75	0.11	0.53	1.49	0.06	0.12	0.11	1.11	0.12	2.08
65	0.11	0.57	1.53	0.07	0.11	0.11	1.09	0.11	2.09
60	0.12	0.57	1.68	0.07	0.11	0.11	1.10	0.11	2.16
50	0.12	0.59	1.73	0.07	0.11	0.11	1.12	0.12	2.41
40	0.13	0.64	1.78	0.07	0.12	0.12	1.15	0.13	2.66
35	0.14	0.67	1.96	0.07	0.13	0.13	1.18	0.14	2.89
25	0.16	0.75	2.11	0.07	0.14	0.14	1.27	0.16	3.40
17	0.21	0.91	2.64	0.08	0.19	0.15	1.33	0.19	3.86
10	0.27	1.32	3.15	0.09	0.26	0.21	1.58	0.24	5.34

Gegevens ontleend aan Lewis & Dixon (1971), zie p 29.

TABEL 2-VIII

Bepaling	Eenh.	$M_H$	$\sigma_R$	$\sigma_H$	$\sigma_H/\sigma_R$	$+\sigma_H^{n_{\min}} - \sigma_H$	
Creatinine	$\mu\text{mol/l}$	87.5	8.84	14.06	1.59	7.0	12.3
Kalium	$\text{mmol/l}$	3.8	0.11	0.477	4.34	72.7	129.2
Natrium	$\text{mmol/l}$	136	2.8	3.60	1.29	10.2	10.2
Chloride	$\text{mmol/l}$	104	2.5	5.14	2.06	15.5	18.4
Tot.eiwit	$\text{g/l}$	73	3.8	5.87	1.55	8.5	10.1
St.bicarb.	$\text{mmol/l}$	22	1.8	1.85	1.03	5.7	6.5
Alk.fosfatase	$\text{U/l}$	36.6	3.84	10.31	2.68	35.1	70.9
Asp.am.trf. (GOT)	$\text{U/l}$	16.6	1.83	2.839	1.55	6.7	12.1

Gegevens ontleend aan Begtrup e.a. (1971), zie p 28.

TABEL 2-IX

Bepaling	A) <u>Stand.afw. daggem.</u>			B) <u>Stand.afw. controleserum</u>		
	1	2	3	1	2	3
Albumine	0.13	0.10	0.10	0.13	0.10	0.10
Alk.fosfatase	0.79	0.76	0.76	1.03	0.74	0.70
Asp.am.trf. (GOT)	2.61	2.28	1.92	1.60	1.44	1.44
Bilirubine	0.07	0.07	0.07	-	-	-
Calcium	0.13	0.12	0.12	0.21	0.17	0.16
Kalium	0.12	0.10	0.10	0.13	0.11	0.11
Natrium	1.58	1.11	1.11	2.12	1.96	1.83
Tot.eiwit	0.18	0.11	0.11	0.18	0.16	0.15
Ureum	2.17	2.00	2.00	1.44	1.29	1.19

Gegevens ontleend aan Lewis & Dixon (1971), zie p 28.

TABEL 2-VI

Bepaling	Relative power
Albumine	1.37
Alkalische fosfatase	0.90
Asp.aminotransferase (GOT)	0.83
Bilirubine	1.14
Calcium	1.45
Kalium	1.03
Natrium	1.71
Totaal eiwit	1.02
Ureum	0.71

Gegevens ontleend aan Dixon & Northam (1970), zie p 24.

TABEL 2-VII

Bepaling	Eenh.	$M_H$	$\sigma_H$	$\gamma(1)\sigma_H$	$\gamma(-1)\sigma_H$	$\gamma(1)$ x100	$\gamma(-1)$ x100
Hemoglobine	mmol/l	8.303	0.938	+0.559	-0.565	+59	-60
Creatinine	$\mu$ mol/l	87.52	14.06	+12.02	-9.11	+85	-64
Kalium	mmol/l	3.8	0.477	+0.346	-0.259	+72	-54
Natrium	mmol/l	136	3.60	+2.05	-2.05	+57	-57
Chloride	mmol/l	104	5.14	+3.83	-3.53	+74	-68
Tot.eiwit	g/l	73.0	5.87	+4.42	-4.04	+75	-69
St.bicarb.	mmol/l	22	1.85	+1.12	-1.07	+61	-57
Alk.fosfatase	U/l	36.6	10.31	+6.60	-4.75	+64	-45
Asp.am.trf.	U/l	16.6	2.839	+2.44	-1.81	+85	-63

Gegevens ontleend aan Begtrup e.a. (1971), zie p 26.

TABEL 2-IV

Bepaling	Ondergrens	Bovengrens
Albumine	-	-
Alkalische fosfatase (KAeeh/100ml)	-	30
Asp.aminotransferase (GOT,RFeeh/ml)	-	100
Bilirubine (mg/100ml)	-	2,9
Calcium (mg/100ml)	-	-
Globuline	-	-
Kalium (meq/l)	-	6,5
Natrium (meq/l)	-	-
Totaal eiwit (g/100ml)	5,0	9,0
Ureum (mg/100ml)	-	79

Door Dixon & Northam (1970) toegepaste afkapgrenzen, zie p 23.

TABEL 2-V

Bepaling	SD ongecorr.gem.	SD contr. serum
	SD gecorr. gem.	SD gecorr. gem.
Albumine	1.43	1.37
Alk.fosfatase	1.05	1.12
Asp.aminotransferase (GOT)	1.15	0.92
Bilirubine	1.12	1.10
Calcium	1.23	1.45
Kalium	1.02	1.14
Natrium	1.08	1.71
Totaal eiwit	1.31	1.28
Ureum	1.01	0.79

Gegevens ontleend aan Dixon & Northam (1970), zie p 23.



TABEL 2-III

Bloedsuikeruitkomsten												
Methode	L & J			W & H			H & W(Np)			H & W(GvN)		
	0%	+5%	-5%	0%	+5%	-5%	0%	+5%	-5%	0%	+5%	-5%
Aantal boven												
bovengrens	0	18	0	2	4	1	1	8	0	4	5	1
Aantal boven												
gemiddelde	8	20	2	11	15	8	12	20	4	25	33	14
Aantal onder												
gemiddelde	12	0	18	9	5	12	8	0	16	15	7	26
Aantal onder												
ondergrens	1	0	12	0	0	2	0	0	3	0	0	2

SGOTuitkomsten												
Methode	L & J			W & H			H & W(Np)			H & W(GvN)		
	0%	+10%	-10%	0%	+10%	-10%	0%	+10%	-10%	0%	+10%	-10%
Aantal boven												
bovengrens	0	14	0	0	0	0	3	7	1	0	1	0
Aantal boven												
gemiddelde	10	19	0	7	12	6	14	18	7	15	21	5
Aantal onder												
gemiddelde	10	1	20	13	8	14	5	1	12	25	19	35
Aantal onder												
ondergrens	1	0	10	0	0	0	1	0	1	4	0	4

Gegevens ontleend aan Amador e.a. (1968), zie p 20 en 21.

Betekenis van de afkortingen:

L & J	Levey & Jennings (1950)	}	H & W(Np)	Hoffmann & Waid (1963)
W & H	Waid & Hoffmann (1955)		H & W(GvN)	Hoffmann & Waid (1965)

TABEL 2-I

Bepaling	Uitkomsten				Uitkomsten			
	zonder transformatie				na log. transformatie			
	gem.	st.afw.	$g_1$	$g_2$	gem.	st.afw.	$g_1$	$g_2$
Chloride	100.1	1.93	0.197	0.480	4.6058	0.196	0.118	0.293
CO <sub>2</sub>	14.8	1.31	0.400	0.787	2.6878	0.0904	0.038	0.411
Kalium	4.01	0.10	0.188	0.528	1.3893	0.0248	0.079	0.302
Natrium	138.0	2.64	0.137	0.384	4.9270	0.0192	0.069	0.106
Ureum	62.0	2.18	0.360	0.380	4.1272	0.3530	0.232	0.332

Gegevens ontleend aan Owen & Campbell (1968-B). Resultaten van analyses van 420 monsters van een poolserum, verricht in een periode van 6 maanden. Zie p 14.

TABEL 2-II

Bepaling	Aangebrachte modificatie
Calcium	+0,5 mg/100 ml
Chloride	+5 meq/l
CO <sub>2</sub>	+2 meq/l
Magnesium	-0,2 meq/l
Kalium	-0,4 meq/l
Natrium	+5 meq/l
Ureum	-5 mg/100 ml
Urinezuur	-1 mg/100 ml

Gegevens ontleend aan Kilgariff & Owen (1968). Zie p 20.