STICHTING
MATHEMATISCH CENTRUM
2e BOERHAAVESTRAAT 49
AMSTERDAM

STATISTISCHE AFDELING

Rapport S 256

De invloed van een behandeling met sulfapreparaten op het optreden van haemolytische streptococcen in de keel bij kinderen.

door
Hilda A. Kuipers

Augustus 1959

The Mathematical Centre at Amsterdam, founded the 11th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications, and is sponsored by the Netherlands Government through the Netherlands Organization for Pure Research (Z.W.O.) and the Central National Council for Applied Scientific Research in the Netherlands (T.N.O.), by the Municipality of Amsterdam and by several industries.

Inhoud

1.	Inleiding	p. 1
1.1.	Doel van het onderzoek	1
1.2.	Het waarnemingsmateriaal	1
2.	Methode van onderzoek	2
2.1.	Inleiding	2
2.2.	Statistische bewerking van het materiaal	3
3.	Samenvatting der conclusies	4

1. Inleiding.

1.1. Doel van het onderzoek.

Het bij dit onderzoek te behandelen probleem is het volgende:

In de keel voorkomende haemolytische streptococcen kunnen onder bepaalde omstandigheden acuut rheuma veroorzaken. Teneinde de rheuma te bestrijden kan het dus van belang zijn een methode te vinden die het voorkomen van streptococcen in de keel tegengaat. De ons voorgelegde vraag was nu, of het toedienen van sulfapreparaten een dergelijke invloed heeft, dus of bij kinderen die geregeld sulfapreparaten krijgen toegediend, minder streptococcen in de keel worden gevonden dan bij hen, die niet prophylactisch met sulfa worden behandeld.

1.2. Het waarnemingsmateriaal.

Teneinde deze vraag te beantwoorden hadden wij de beschikking over de gegevens betreffende ruim 400 kinderen in de leeftijd van 4 t/m 12 jaar, die geregeld de polikliniek van het Binnengasthuis te Amsterdam bezochten. Hierbij waren zowel jongens als meisjes. Alle kinderen hadden tevoren reeds rheuma gehad. De kinderen waren onder te verdelen in drie groepen:

I een groep die geregeld prophylactisch met sulfa werd behandeld,
II een groep die ongeregeld prophylaxis kreeg en
III een contrôlegroep die geen sulfa kreeg toegediend.

Van elk kind werd bij ieder bezoek aan de polikliniek (hetgeen bij tussenpozen variërend van vijf dagen tot twee maanden plaatsvond) een keeluitstrijkje gemaakt, dat eerst op een tampon 24 uur in een voedingsbodem werd gebracht, en daarna op bloedplasma werd geënt. Het voordeel van deze methode is, dat men een eventuele positieve reactie (dus het voorkomen van streptococcen) beter opmerkt dan wanneer men niet van een voedingsbodem gebruik maakt. Op deze manier waren dus van hetzelfde kind meerdere gegevens beschikbaar, met uitzondering van enkele gevallen waarin slechts één bezoek aan de polikliniek werd gebracht. Alle waarnemingen zijn verricht tussen 1 mei 1955 en 31 december 1956.

2. Methode van onderzoek.

2.1. Inleiding.

Teneinde de invloed van het seizoen zoveel mogelijk te beperken werden de waarnemingen gesplitst in tweemaandelijkse perioden: mei-juni 1955, juli-augustus 1955, ..., november-december 1956. Deze perioden zullen wij in dit rapport verder seizoenen noemen. Dit werd mede gedaan omdat de zomer van 1955 veel slechter was, wat het weer betreft, dan die van 1956, zodat de maanden van deze zomer niet samengenomen konden worden, althans niet direct. Waar nodig werd nog onderscheid gemaakt tussen zomer- en wintermaanden. Daarbij werden als zomermaanden beschouwd mei t/m augustus, als wintermaanden november t/m februari.

De kinderen waren niet onderverdeeld in jongens en meisjes, zodat met het geslacht van de kinderen geen rekening gehouden kon worden en het dus onmogelijk was, na te gaan of jongens anders op het toedienen van sulfa reageren dan meisjes. Onze conclusies gelden dus slechts onder het voorbehoud dat het geslacht geen merkbare invloed heeft, en dat jongens en meisjes aselect over de drie groepen, die een verschillende behandeling ondergingen, zijn verdeeld.

Verder is ook niet gelet op het sociaal milieu waaruit de kinderen afkomstig zijn. Aangenomen werd echter, dat ook deze factor de uitkomsten niet op systematische wijze ten gunste van een van de onderzochte groepen zal hebben beïnvloed. De conclusies gelden vanzelfsprekend voor de bevolkingsgroep waaruit de kinderen die de polikliniek bezochten afkomstig zijn.

Slechts bij uitzondering werden meerdere kinderen uit één gezin behandeld. Bij het grote aantal patiëntjes, dat aan de proef deelnam moet het echter wel uitgesloten worden geacht dat de resultaten hierdoor merkbaar zullen zijn beïnvloed.

Ook als men echter met deze factor geen rekening zou moeten houden, blijft er nog een moeilijkheid. Deze betreft nl. het aantal waarnemingen dat per kind beschikbaar is. Het is immers niet uitgesloten, dat een kind dat de eerste maanden steeds negatief blijft, op de duur minder frequent de polikliniek bezoekt dan een kind, waarbij af en toe de aanwezigheid van streptococcen is geconstateerd. Het was daarom niet mogelijk om bij de bewerking van het materiaal kinderen die eenzelfde aantal keren geobserveerd waren, onderling vergelijkbaar te achten, en evenmin om meer dan één waarneming per kind in het onderzoek te betrekken, daar dit juist bij de vraag naar de frequentie van negatieve en positieve reacties een schijneffect zou kunnen veroorzaken.

Groep II werd bij de verdere berekeningen buiten beschouwing gelaten, aangezien deze groep te heterogeen geacht moest worden wat het toedienen van sulfa betreft.

2.2. Statistische bewerking van het materiaal.

De kinderen werden zó over de tien tweemaandelijkse perioden verdeeld, dat elk kind slechts in één seizoen werd geplaatst. Hierbij moest er zorg voor worden gedragen dat dit aselect geschiedde, teneinde de onder 2.1. genoemde moeilijkheid te ondervangen. Bij kinderen, waarvan de waarnemingen dus over meer seizoenen waren verdeeld, werd door loting beslist welk van deze waarnemingen in het onderzoek zou worden betrokken. Dit geschiedde uiteraard voor de groepen I en III afzonderlijk. Bleken de aantallen kinderen in de diverse seizoenen te veel te verschillen, dan werd opnieuw een lotingsprocedure toegepast om deze verschillen te corrigeren.

Voor elk van deze tien tweemaandelijkse perioden afzonderlijk werd nu groep I (die geregeld met sulfa werd behandeld) vergeleken met groep III (geen prophylaxis) wat betreft het aantal positieve reacties met behulp van de methode der dubbele dichotomie[1). De toetsing werd steeds eenzijdig uitgevoerd aangezien alleen interesse bestond voor het optreden van meer positieve reacties bij groep III dan bij groep I. Hierbij bleek bij één der tien seizoenen

----

1) Voor een korte uiteenzetting van de gang van zaken bij toetsing van hypothesen, alsmede voor de juiste interpretatie van overschrijdingskansen, zij verwezen naar het aan dit rapport toegevoegde memorandum M 6. De methode der dubbele dichotomie wordt kort beschreven in het eveneens toegevoegde memorandum M 23.

een systematisch verschil op te treden ($k = 0{,}02$), terwijl bij twee andere een kleine aanwijzing in dezelfde richting gevonden werd ($k = 0{,}05$ resp. $k = 0{,}07$). Bij combinatie van de tien seizoensuitkomsten vonden wij echter een overschrijdingskans $k = 0{,}0001$. Hieruit blijkt duidelijk, dat kinderen die geen sulfa kregen toegediend, vaker positief waren dan kinderen die geregeld prophylaetisch met sulfa werden behandeld.

Dezelfde toetsing werd voor zomer- en winterperioden apart nog uitgevoerd, waarbij wij als overschrijdingskansen vonden $k = 0{,}02$ voor de zomer- en $k = 0{,}04$ voor de winterperiode. Hierbij werden de zomermaanden van 1955 en 1956 wèl tezamen genomen.

3. Samenvatting der conclusies.

Uit dit onderzoek is gebleken dat door het prophylaetisch toedienen van sulfa de kans op het optreden van haemolitische streptococcen in de keel kleiner wordt. Dit effect treedt zowel op in de zomer als in de winter.

Hierbij moet men dan nog in aanmerking nemen dat het grootste deel van de beschikbare waarnemingen niet in het onderzoek kon worden betrokken (zie par. 2.1.); wellicht kan bij een meer uitgebreid waarnemingsmateriaal dit effect nog duidelijker worden aangetoond.

Ten aanzien van deze conclusies moeten wij in verschillende opzichten een voorbehoud maken:

1. Aangezien het geslacht van de onderzochte kinderen niet was vermeld, kan niet worden vastgesteld of dit van enige invloed is.
2. Bij het onderzoek waren alleen kinderen van 4 t/m 12 jaar betrokken; de conclusies kunnen dus niet zonder meer tot andere leeftijdsgroepen worden uitgebreid.
3. De conclusies gelden strikt genomen alleen voor bevolkingsgroepen met dezelfde samenstelling als die waaruit de onderzochte kinderen afkomstig zijn. Over deze samenstelling zijn geen nadere gegevens voorhanden.

4. In dit onderzoek zijn alleen kinderen betrokken die reeds rheuma gehad hebben. Het is dus niet zeker dat de conclusies ook voor andere kinderen geldigheid bezitten.
5. Ondersteld is, dat de verdeling van de kinderen over de groepen I, II en III aselect is geweest ten aanzien van de andere, hier niet onderzochte, mogelijk van invloed zijnde factoren, zoals de leeftijd en in verband daarmee het al of niet schoolgaan.

MATHEMATISCH CENTRUM
Amsterdam
Statistische Afdeling
S47(M6)

Algemene gang van zaken bij het toetsen van een hypothese.[1]

De toetsing van een hypothese H_0 berust steeds op een aantal waarnemingen $x_1, x_2, \ldots, x_n$ van één of meer stochastische grootheden[2], of op enige groepen van waarnemingen (bv. twee steekproeven).

Bij een toets behoort een toetsingsgrootheid $\underline{u}$ (soms meer dan één), die een functie is van bovengenoemde stochastische grootheden en die, voor de waargenomen waarden $x_1, x_2, \ldots, x_n$ een waarde aanneemt, die berekend kan worden (bv.: het gemiddelde der waarnemingen, of de spreiding, of het verschil van de gemiddelden van **2 groepen waarnemingen.**

De toetsingsgrootheid wordt steeds zo gekozen, dat men, op grond van de onderstelling, dat H_0 juist is, de waarschijnlijkheidsverdeling van deze grootheid kan berekenen.

Vervolgens kiest men een verzameling Z van mogelijke uitkomsten van $\underline{u}$, en wel op zodanige wijze, dat de kans, dat $\underline{u}$ een in Z gelegen waarde aanneemt, onder de hypothese H_0, gelijk is aan een gegeven getal α, zodat Z dus van α afhankelijk is.[3] Z heet de kritieke zône van de toets, α de onbetrouwbaarheidsdrempel (Engels: level of significance). Voor α neemt men veelal de waarde 0,05 of 0,01.

Men verwerpt nu H_0 op grond van de waarnemingen $x_1, x_2, \ldots, x_n$, indien de bij deze waarnemingen behorende

---

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

2) Een stochastische grootheid is een grootheid, die een waarschijnlijkheidsverdeling bezit, of, anders gezegd, een grootheid, die voor de elementen van een collectie (universum, populatie) gedefinieerd is en daarop allerlei waarden aanneemt. Stochastische grootheden worden aangegeven door onderstreepte letters.

3) Soms kan men slechts bereiken, dat deze kans $< \alpha$ is.

waarde van *u* in Z ligt. Dit wordt vaak uitgedrukt door te zeggen, dat het resultaat van het experiment "significant" is. De waarde van α moet dan echter worden vermeld. De kans, dat dit zal gebeuren, is, indien H_0 juist is, gelijk aan α. Derhalve is α de *kans op ten onrechte verwerping van de juiste hypothese*, ook de kans op een fout van de eerste soort genoemd. Indien men deze methode toepast, met $\alpha = 0{,}05$ resp. 0,01, zal men in gemiddeld ongeveer één op 20 resp. op 100 van de gevallen, waarin de hypothese die men toetst juist is, deze toch verwerpen.

De toetsingstheorie biedt in het algemeen *geen* mogelijkheid om tot *aanvaarding* van een hypothese te komen. Indien een bepaalde hypothese H_0 niet verworpen kan worden, is dit gewoonlijk met een hele verzameling van hypothesen tegelijk het geval. Niet-verwerpen staat dus niet gelijk met aanvaarden.

Wel zal men vaak in de loop van een statistische analyse bepaalde onderstellingen, die plausibel schijnen en voor de verdere analyse van nut zijn, toetsen, alvorens ze bij de verdere bewerking van het materiaal te gebruiken. Worden zij dan op grond van de toets niet verworpen, dan houdt dit in zo verre een rechtvaardiging van die onderstellingen in, dat een grote afwijking door de toets veelal wel zou zijn ~~ge~~ontdekt. Indien men dan verder de onderstellingen gebruikt, verwaarloost men eventueel aanwezige afwijkingen van onbekende grootte, die echter niet zo groot zijn, dat zij door de toets zijn ontdekt.

Vele toetsen gelden zelf alleen onder bepaalde onderstellingen omtrent de waarschijnlijkheidsverdelingen der stochastische grootheden, waarvan waarnemingen zijn verricht. Deze *nevenvoorwaarden* dienen steeds uitdrukkelijk te worden vermeld en, zo mogelijk, zelf te worden getoetst.

In plaats van de onbetrouwbaarheidsdrempel α wordt vaak bij de uitslag van een toetsing de *overschrijdingskans* k opgegeven; dit is de kleinste waarde van α, waarbij in het betrokken geval, nog tot verwerping van H_0 zou zijn overgegaan; anders gezegd: de kleinste α, waarvoor de gevonden waarde der toetsingsgrootheid nog juist in de (bij α behorende) kritieke zône Z ligt. Wordt dus de waarde k opgegeven en werkt men met onbetrouwbaarheidsdrempel α , dan wordt H_o verworpen, indien $k \leqq \alpha$ is.

Voor het onderscheid tussen één- en tweezijdige toetsing en de keuze tussen deze twee mogelijkheden vergelijke men bv. de tweede hieronder gegeven literatuurplaats. Wij moeten hier volstaan met de opmerking, dat éénzijdige toetsing veelal eerder tot verwerping van H_o leidt, maar dat deze slechts onder bijzondere omstandigheden kan worden toegepast.

Literatuur

J. Neyman, First course in probability and statistics, New York, 1950, Chapter 5.

J. Hemelrijk en H.R. van der Vaart, Het gebruik van één en tweezijdige overschrijdingskansen voor het toetsen van hypothesen, Statistica 4 (1950) p.54-66.

MATHEMATISCH CENTRUM
2e Boerhaavestraat 49
A m s t e r d a m - O

Statistische Afdeling
S 53 (M 23)
(Herzien: aug. 1959)

Toetsing van de hypothese $p_1 = p_2$ met behulp van een 2 x 2-tabel [1])

Wij beschouwen twee reeksen van onafhankelijke experimenten, waarbij ieder experiment van de ene reeks één van de twee resultaten A of $\bar{A}$ (non-A) heeft en ieder experiment van de tweede reeks één van de beide resultaten B of $\bar{B}$ (hierbij kan A=B zijn). Daarbij wordt ondersteld, dat bij ieder der experimenten van de ene reeks de kans op A gelijk aan p_1 (en dus de kans óp $\bar{A}$ gelijk aan $1-p_1$) is en bij ieder der experimenten van de tweede reeks de kans op B gelijk aan p_2 (en dus de kans op $\bar{B}$ gelijk aan $1-p_2$). De te toetsen hypothese luidt nu:

$$H_o : p_1 = p_2.$$

Indien de eerste reeks uit n en de tweede reeks uit m waarnemingen bestaat, waaronder a (resp. b) maal A (resp. B) voorkomt, kunnen deze gegevens in de volgende 2 x 2-tabel worden samengevat:

	A resp. B	$\bar{A}$ resp. $\bar{B}$	totaal
eerste reeks	a	c	n
tweede reeks	b	d	m
totaal	r	s	N

Als toetsingsgrootheid wordt a, het aantal malen A in de eerste reeks waarnemingen, gebruikt. Indien H_o, juist is bezit deze grootheid onder de voorwaarde, dat r de bij het experiment gevonden waarde aanneemt, de volgende waarschijnlijkheidsverdeling: de kans, dat een bepaalde waarde a aangenomen wordt, is gelijk aan:

----

1) Dit memorandum is slechts bedoeld ter oriëntatie en streeft niet naar volledigheid of volledige exactheid.

$$\frac{\binom{n}{a}\binom{m}{b}^{1)}}{\binom{N}{r}} .$$

Als kritieke zone worden de waarden van a met de kleinste waarschijnlijkheden bijeengezocht, tot de gekozen onbetrouwbaarheidsdrempel het toevoegen van een nieuwe waarde verhindert (bij éénzijdige toetsing bestaat de kritieke zone uitsluitend uit grote of uitsluitend uit kleine waarden van a). De overschrijdingskans, behorende bij de gevonden waarde van a, is gedefinieerd als de som van alle waarschijnlijkheden van bovenstaande verdeling, die hoogstens gelijk aan de waarschijnlijkheid van de gevonden waarde zijn (bij éénzijdige toetsing echter gelijk aan de som van de waarschijnlijkheden van alle waarden die groter of gelijk aan de gevondene, of van alle waarden, die kleiner of gelijk aan de gevondene zijn). Deze exacte toetsingsmethode voor H_o is afkomstig van R.A. FISHER.

Indien n en m zo groot zijn, dat deze exacte berekening te omslachtig wordt, maakt men gebruik van de volgende benadering:

Gemiddelde en spreiding van de grootheid a zijn (indien H_o juist is):

$$\frac{nr}{N} \text{ resp. } \sqrt{\frac{nmrs}{N^2(N-1)}} .$$

Men gebruikt dan in plaats van de exacte waarschijnlijkheidsverdeling van a de normale verdeling met hetzelfde gemiddelde en dezelfde spreiding.

Bij linkseenzijdige toetsing berekent men dan

$$u_1 = \frac{a-\frac{nr}{N}+\frac{1}{2}}{\sqrt{\frac{nmrs}{N^2(N-1)}}} = \frac{aN-nr+\frac{1}{2}n}{\sqrt{\frac{nmrs}{N-1}}} = \frac{ad-bc+\frac{1}{2}N}{\sqrt{\frac{nmrs}{N-1}}} ,$$

bij rechtseenzijdige toetsing

---

1) $\binom{n}{a} = \frac{n!}{a!(n-a)!} = \frac{(n-a+1)(n-a+2)\ldots n}{1.2 \ldots a}$

$$u_2 = \frac{ad-bc-\frac{1}{2}N}{\sqrt{\frac{nmrs}{N-1}}}$$

en bij tweezijdige toetsing

$$u_3 = \frac{ad-bc \ -\frac{1}{2}N}{\sqrt{\frac{nmrs}{N-1}}} \quad .$$

De term $\pm\frac{1}{2}$ in de ~~noemer~~ teller is de z.g. continuïteitscorrectie, die bij toenemende n en m weldra verwaarloosd kan worden.

De overschrijdingskans wordt opgezocht in een tabel van de normale verdeling met gemiddelde 0 en spreiding 1; bij links (resp. rechts) eenzijdige toetsing is deze overschrijdingskans het oppervlak links (resp. rechts) van u_1 (resp. u_2). Bij tweezijdige toetsing is de overschrijdingskans tweemaal het oppervlak rechts van u_3.

Literatuur.

R.A. Fisher, Statistical Methods for Research Workers, London 1948, p. 96. Opmerking: Fisher gebruikt hier de éénzijdige overschrijdingskans.

J. Hemelrijk, Waarschijnlijkheidsrekening en Statistiek, Vacantiecursus Mathematisch Centrum, Amsterdam 1950, § 4.