

EXPERIMENTELE PROEFOPZET EN STATISTISCHE ANALYSE VAN VERDUNNINGS-
REEKSEN¹

door

LEO W.G. STRIJBOSCH², RONALD J.M.M. DOES³ en WIM A. BUURMAN⁴SAMENVATTING

Verdunningsreeksen (limiting dilution assays) worden gebruikt om de frequentie van een bepaalde cel in een populatie van cellen te bepalen. Behalve de louter technische problemen die zich bij deze methode voordoen in het laboratorium zijn experimentele proefopzet en statistische analyse van belang. In dit artikel wordt op beide aspecten ingegaan. Er wordt een nieuwe methode voorgesteld om het aantal verschillende verdunningen in het experiment en het aantal cellen per verdunning te kiezen, afhankelijk van een door de onderzoeker op te geven onder- en bovengrens voor de te schatten parameter en van de hoeveelheid "informatie" die de resultaten moeten bevatten. Drie statistische analyse methoden, te weten de minimum chikwadraat methode, de maximum likelihood methode en de jackknife versie van de laatste methode worden vergeleken m.b.v. simulatie resultaten. De conclusie is dat de jackknife methode de voorkeur verdient.

- 1) Lezing gehouden op de Statistische Dag d.d. 3 april 1986, KH Tilburg.
- 2) Vakgroep Econometrie, KU Brabant, Postbus 90153, 5000 LE Tilburg, tel.: 013-662478
- 3) Vakgroep Medische Informatica en Statistiek, RU Limburg, Postbus 616, 6200 MD Maastricht, tel.: 043-434410 tst. 43
- 4) Vakgroep Algemene Heelkunde, RU Limburg, Postbus 616, 6200 MD Maastricht, tel.: 043-888590

INLEIDING

Een verdunningsreeks of in het engels limiting dilution assay (LDA) is een laboratoriumtechniek voor de bepaling van de relatieve frequentie van een bepaald type cel in een populatie van cellen. Als voorbeeld kan daarbij worden gedacht aan de proportie witte bloedcellen die in staat is ("immunocompetent") om bepaalde vreemde indringers (bijv. bacteriën) te vernietigen. Deze proportie is afhankelijk van de tijd: de immunocompetente cellen zullen zich vermeerderen en de indringers vernietigen totdat de strijd gestreden is. De snelheid waarmee dit proces zich voltrekt is interessant en kan in principe worden bepaald door een reeks LDA's. Deze experimenten zijn echter duur, kosten veel tijd en zijn vaak niet herhaalbaar. Naast de kwaliteit van de statistische analyse is derhalve ook de proefopzet van groot belang: hoe moet het laboratorium onderzoek worden opgezet zodanig dat de kans op een "mislukte" proef zo klein mogelijk is ?.

Dit artikel bestaat uit twee delen: in deel I wordt een voorstel gedaan voor een nieuwe methode om LDA's op te zetten en in deel II zullen resultaten worden gepresenteerd van Monte Carlo experimenten en theoretische overwegingen die tot doel hebben om drie statistische analyse-methoden te vergelijken.

I. DE EXPERIMENTELE PROEFOPZET

Laboratoriumopstelling en uitgangspunten.

Het eigenlijke laboratoriumwerk is zeer ingewikkeld en ten dele niet interessant voor het onderwerp van dit artikel. Daarom zal slechts schematisch worden aangegeven wat er gebeurt en welke veronderstellingen worden gemaakt.

Een startoplossing bevat zeer veel gelijksoortige cellen. Een onbekende fractie ϕ (de te schatten parameter) hiervan is immunocompetent. Via een verdunningsproces worden J verschillende verdunningen geproduceerd en wordt de j -de verdunning ($j=1, \dots, J$)

verdeeld over n_j kleine bakjes, zodanig dat mag worden aangenomen, dat het totaal aantal immunocompetente cellen in een bakje van de j -de verdunning een Poisson variabele is met gemiddelde ϕx_j , en dat de aantallen cellen in de n_j bakjes onafhankelijk zijn. Vervolgens worden alle bakjes besmet en wordt na verloop van tijd per bakje de reactie geïnventariseerd, d.w.z. hetzij negatief hetzij positief. Hierbij wordt aangenomen dat de aanwezigheid van tenminste één immunocompetente cel zeker leidt tot een positieve reactie, terwijl het ontbreken van immunocompetente cellen zeker leidt tot een negatieve reactie. Het hier beschreven model staat bekend als het "single-hit Poisson model".

Voordat het laboratoriumonderzoek kan worden uitgevoerd moet een beslissing genomen worden over de design-parameters J (het aantal verschillende verdunningen), x_j (het gemiddeld aantal cellen per bakje van de j -de verdunning), en n_j (het aantal bakjes van de j -de verdunning, $j=1,2,\dots,J$). In feite is dit een complex probleem waarvoor verschillende benaderingswijzen denkbaar zijn.

Finney (1978) geeft m.b.t. serial dilution assays algemene principes om tot een goede proefopzet te komen. Deze adviezen zijn mutatis mutandis ook toepasbaar op LDA's. Hij vermeldt onder meer dat bij $J=1$ (er is slechts één verdunning), x_1 het beste zodanig gekozen kan worden dat de kans op een negatieve respons ($=\exp(-\phi x_1)$) gelijk is aan .203 ofwel $\phi x_1 = 1.59$, hetgeen betekent dat per bakje gemiddeld 1.59 immunocompetente cellen aanwezig zijn. Immers de Fisher informatie $I(\phi) = n x_1^2 / (\exp(\phi x_1) - 1)$ wordt voor deze waarde $\phi x_1 = 1.59$ gemaximaliseerd (zie ook Fazekas de St Groth (1982)).

Als er al een tamelijk betrouwbare puntschatting van ϕ bestaat is deze proefopzet optimaal, anders adviseert Finney uit te gaan van een intervallschatting $[\phi_1, \phi_2]$ voor ϕ waarbij wordt aangenomen dat $\phi_1 \leq \phi \leq \phi_2$, en voor de x_j 's een meetkundige rij te kiezen zodanig dat er gemiddeld maximaal 2 en minimaal .5 immunocompetente cellen per bakje aanwezig zijn, d.w.z. $.5 \leq \phi x_j \leq 2$ (of $.14 \leq \exp(-\phi x_j) \leq .61$), $j=1,2,\dots,J$. Als dat lukt is de hoeveelheid informatie op grond van $I(\phi)$ ten minste 60% van het

maximaal haalbare (zie ook Fazekas de St. Groth (1982)). Voor de reden van meetkundige rij wordt aanbevolen een zo klein mogelijke waarde te nemen, waarbij wordt gesuggereerd deze tussen 2 en 10 te nemen.

Een onderzoeker die met deze richtlijnen een proefopzet wil bepalen weet eigenlijk nog niet precies wat hij moet doen. Belangrijke vragen als "hoe groot moet J zijn" en - als $x_j = x_1 c^{j-1}$ - "welke waarden moeten c en x_1 hebben" worden niet opgelost. De aanbevelingen van Finney (1978) helpen hierbij niet.

Een nieuw idee voor de proefopzet.

In dit artikel geven wij voor J, c en x_1 formules, die een onderzoeker kan gebruiken bij het bepalen van zijn proefopzet nadat hij een beslissing genomen heeft over de waarden van de volgende parameters:

1. P_1 en P_2 (in de aanbevelingen van Finney (1978) resp. .14 en .61) : een onder- en een bovengrens voor de kans op een negatieve reactie ($P = \exp(-\phi x)$, $0 < P_1 < P_2 < 1$). De onderzoeker geeft hiermee aan dat wanneer van n_j bakjes het aantal negatieve reacties ligt tussen $P_1 n_j$ en $P_2 n_j$ hij de proef m.b.t. de j -de verdunning voldoende "informatief" vindt, $j=1, 2, \dots, J$.
2. d : het aantal verdunningen waarvoor hij wil dat de proportie negatieve reacties ligt tussen P_1 en P_2 , ($d \geq 1$).
3. ϕ_1 en ϕ_2 : een onder- en een bovengrens voor de onbekende ($0 < \phi_1 \leq \phi_2 < 1$). De onderzoeker gaat ervan uit dat de onbekende ϕ minimaal ϕ_1 is en maximaal ϕ_2 .

Als functie van deze 5 parameters, die samen eigenlijk de kwaliteit van het design bepalen geven de volgende formules waarden voor J, c en x_1 , die nodig zijn om het design vast te stellen.

4.
$$J = d + \text{ENTIER}(^c \log(\phi_2 / \phi_1)),$$
5.
$$x_1 = \{(\ln P_1)(\ln P_2) c^{1-J} / (\phi_1 \phi_2)\}^{1/2},$$

$$6. \quad x_j = x_1 c^{j-1}, \quad j = 2, \dots, J$$

en

$$7. \quad c = \{(\ln P_1)/(\ln P_2)\}^{1/d}.$$

De eigenschappen van het aldus gevonden design kunnen als volgt geformuleerd worden:

Voor elke mogelijke waarde van ϕ in het interval $[\phi_1, \phi_2]$ zijn er precies d waarden van de meetkundige rij $\{x_j\}_{j=1, \dots, J}$, die voldoen aan de eigenschap dat

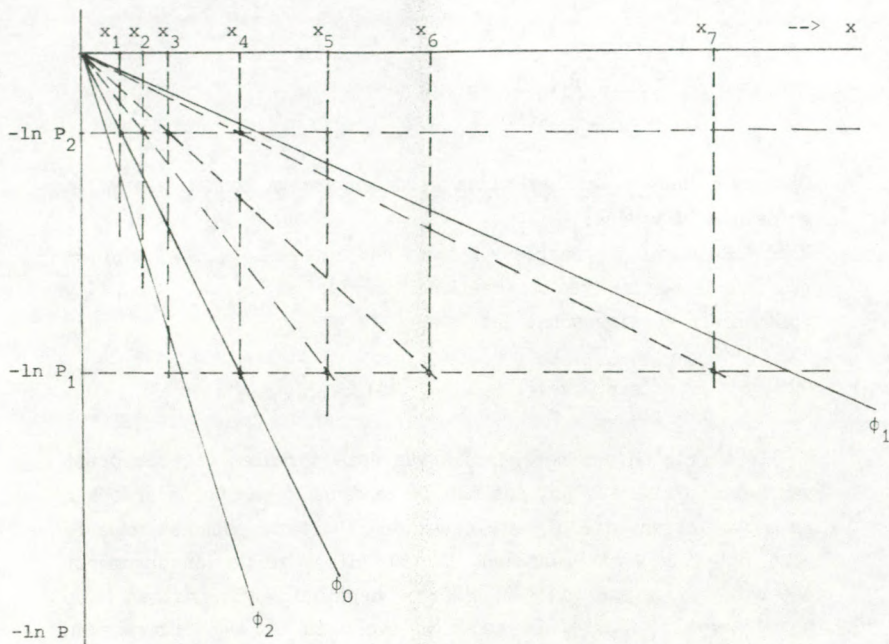
$$8. \quad P_1 \leq \exp(-\phi x_j) < P_2, \quad j=1, \dots, J.$$

De voordelen van het gebruik van deze formules voor de opzet van het experiment zijn, dat met de voorkennis van de onderzoeker en met de eisen, die hij stelt aan de resultaten van het experiment rekening wordt gehouden. Hierbij dient te worden opgemerkt dat wordt uitgegaan van een uniforme a priori verdeling van ϕ op het interval $[\phi_1, \phi_2]$. De reden hiervoor is dat een onderzoeker veelal een globaal idee heeft tussen welke waarden ϕ moet liggen. Aangezien bij dit soort experimenten, waarbij veelal proefdieren worden gebruikt, herhalingen niet vaak mogelijk zijn, beschikt de onderzoeker niet over gedetailleerde a priori informatie. Een ander voordeel is, dat door toepassing van de formules de experimenten meer zelfstandig door het laboratoriumpersoneel kunnen worden uitgevoerd.

Afleiding van de formules.

De correctheid van de formules van de meetkundige rij m.b.t. de bovenvermelde eigenschap (8) is wiskundig te bewijzen (zie

Strijbosch e.a. (1986)). We geven hier niet dat wiskundige bewijs, maar een intuïtieve afleiding van de formules op basis van een grafische voorstelling (zie figuur 1).



Figuur 1

De relatie $P = \exp(-\phi x)$ kan worden voorgesteld als een rechte lijn in een assenstelsel met $-\ln P$ en x als coördinaat-assen. De parameter ϕ is dan de richtingscoëfficiënt van de lijn $-\ln P = \phi x$. Stel $d=3$, en x_1, x_2, x_3 zijn gegeven waarden die voldoen aan de eigenschap $-\ln P_2 < \phi_2 x_1 < \phi_2 x_2 < \phi_2 x_3 \leq -\ln P_1$. We proberen nu vanuit x_1, x_2, x_3 de rest van de meetkundige rij te vinden door ϕ te laten afnemen van ϕ_2 tot ϕ_0 , bepaald door het punt $(x_1, -\ln P_2)$. Voor waarden van ϕ in $(\phi_0, \phi_2]$ geldt dat $-\ln P_2 < \phi x_1 < \phi x_2 < \phi x_3 < -\ln P_1$, terwijl voor $\phi \leq \phi_0$, geldt dat $\phi_0 x_1 \leq -\ln P_2$.

Hieruit volgt dat x_4 wordt gegeven door de waarde $(-\ln P_1)/c_0$. Omdat $c_0 = (-\ln P_2)/x_1$ blijkt dat $x_4 = x_1 (\ln P_1)/(\ln P_2)$. Algemeen kan nu worden gesteld dat $x_{j+d} = x_j (\ln P_1)/(\ln P_2)$, zodat de keuze voor een meetkundige rij $x_j = x_1 c^{j-1}$ en $c = ((\ln P_1)/(\ln P_2))^{1/d}$ voor de hand ligt. De enige nog onbekende parameters J en x_1 kunnen vervolgens worden afgeleid uit de volgende eisen:

$$9. \quad x_d / ((-\ln P_1)/\phi_2) = ((-\ln P_2)/\phi_1) / x_{J-d+1},$$

$$10. \quad x_{J-d+1} > (-\ln P_2)/\phi_1.$$

De eerste eis (9) zorgt ervoor dat x_d en x_{J-d+1} op logaritmisch equidistante afstanden aan weerszijden liggen van het interval $[(-\ln P_1)/\phi_2, (-\ln P_2)/\phi_1]$ (symmetrie-eis); de eerste en tweede eis (9 en 10) samen garanderen dat x_{J-d+1} groot genoeg en x_d klein genoeg is. Hieruit kan worden afgeleid dat

$$11. \quad x_1 = \{(\ln P_1)(\ln P_2)c^{1-J}/(\phi_1\phi_2)\}^{1/2}.$$

De kleinste waarde van J die voldoet aan alle voorwaarden wordt gevonden als

$$12. \quad J = d + \text{ENTIER} (c \log(\phi_2/\phi_1)).$$

Voor het aantal herhalingen n_j van de j -de verdunning (meestal wordt alle n_j 's gelijk genomen) geeft de procedure geen waarde. In de praktijk worden hiervoor een beperkt aantal waarden gebruikt: meestal een veelvoud van 12, t.g.v. het beschikbare laboratoriummateriaal. Veel gebruikte waarden zijn 24 en 48 (zie ook Fazekas de St Groth (1982)); hogere waarden resulteren uiteraard in kleinere intervallschattingen voor ϕ , maar zijn vanwege de hoge laboratoriumkosten niet gangbaar.

Ten aanzien van de keuzes voor P_1 , P_2 en d kan het volgende nog worden opgemerkt. Op grond van een groot aantal Monte Carlo experimenten blijkt dat $P_1 = .15$ en $P_2 = .70$ voldoen. Verder wordt door praktische ervaring de keuze $d=3$ geadviseerd.

II. DE KEUZE VAN DE STATISTISCHE METHODE

Voor het uitvoeren van LDA zijn, zoals we reeds eerder hebben vermeld, de volgende design parameters noodzakelijk: het aantal verdunningen J ; het gemiddeld aantal cellen x_j per verdunning en het aantal herhalingen n_j per verdunning j , $j=1,2,\dots,J$. Bij de uitvoering wordt dan het aantal negatieve reacties, notatie r_j , geregistreerd, $j=1,2,\dots,J$. De kans op een negatieve reactie voor verdunning j is gelijk aan $\exp(-\phi x_j)$. Deze kans kan voor iedere verdunning j geschat worden door het quotiënt $p_j = r_j/n_j$. In het artikel van Taswell (1981) worden een viertal schattingsmethoden voor ϕ met elkaar vergeleken.

A. De kleinste kwadraten (KK) methode

Een formule voor de schatter van ϕ d.m.v. de KK methode wordt verkregen door toepassing van de standaardformule voor ongewogen KK van de regressie lijn $\ln p = -\phi x$. D.w.z. de KK schatting $\hat{\phi}_{KK}$ wordt bepaald door de waarde van ϕ die $\sum (\ln p_j + \phi x_j)^2$ minimaliseert.

B. De gewogen gemiddelde (GG) methode

Voor iedere verdunning j kan een schatting van ϕ verkregen worden d.m.v. $f_j = (-\ln p_j)/x_j$. Door elke schatting te wegen met de reciproke van zijn variantie en te delen door de som van de gewichten verkrijgt men de GG schatting.

C. De maximum likelihood (ML) methode

De ML schatting voor ϕ wordt bepaald door het maximaliseren van de log likelihood vergelijking

$$\ln L(\phi) = \sum_{j=1}^J \left[\ln \binom{n_j}{r_j} - r_j \phi x_j + (n_j - r_j) \ln(1 - \exp(-\phi x_j)) \right].$$

D. De minimum chikwadraat (MC) methode

De MC schatting wordt verkregen door het minimaliseren van de chikwadraat grootheid:

$$\chi^2_{J-1}(\phi) = \sum_{j=1}^J (r_j - n_j \exp(-\phi x_j))^2 / (n_j \exp(-\phi x_j) (1 - \exp(-\phi x_j))).$$

Tot 1981 was de KK methode de meest gebruikte schattingsmethode. Zowel Taswell (1981) als Fazekas de St Groth (1982) geven een aantal belangrijke bezwaren die pleiten tegen het gebruik van deze methode. Ook het aanpassen van de KK methode door over te gaan naar de gewogen KK methode wordt door Fazekas de St Groth (1982) verworpen. De conclusie van Taswell (1981) luidt dat de MC methode op grond van de kleinste onzuiverheid en variantie de beste methode is om ϕ te schatten. Fazekas de St Groth (1982) daarentegen beweert dat de ML methode te preferen is boven de MC methode. De keuze tussen ML en MC is binnen de mathematische statistiek meerdere malen aanleiding tot discussie, zie ter illustratie het artikel van Berkson (1980) en in het bijzonder de bijbehorende discussie. Op grond van de tegenstrijdigheden hebben wij beide methoden opnieuw met elkaar vergeleken.

Naar aanleiding van een artikel van Reeds (1978) is bovendien de jackknife versie van de ML methode bij het onderzoek betrokken. Het effect van jackknife procedures is om de onzuiverheid te reduceren en om robuuste interval schattingen te verkrijgen (zie Miller (1974) voor een algemene inleiding). De jackknife versie (JML) van de ML methode wordt als volgt bepaald:

Zij gegeven de ML schatter $\hat{\phi}_{ML}$, dan wordt de JML schatter $\hat{\phi}_{JML}$ verkregen door

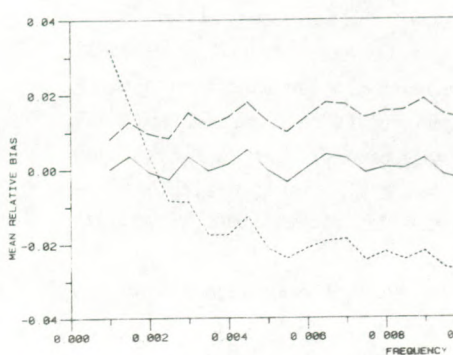
$$\hat{\phi}_{JML} = \hat{\phi}_{ML} - \frac{n-1}{n} \sum_{k=1}^n D_k,$$

waarbij $D_k = \hat{\phi}_{ML}^{(k)} - \hat{\phi}_{ML}$; $\hat{\phi}_{ML}^{(k)}$ is de ML schatter van ϕ gebaseerd op

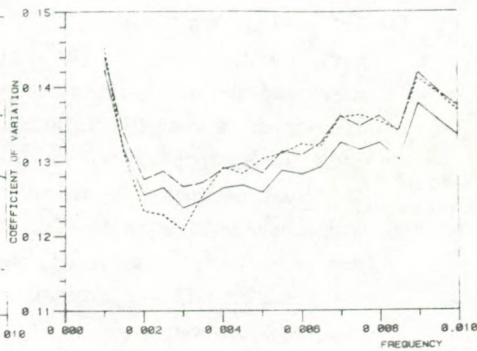
de data waarbij de k -de herhaling van elke verdunning is wegge-
laten. Hierbij merken we op dat de JML methode alleen gebruikt
kan worden als $n_1 = \dots = n_j = n$, hetgeen veelal in de praktijk het
geval is.

De drie methoden (MC, ML en JML) werden met elkaar vergeleken
d.m.v. Monte Carlo experimenten. De simulaties werden alle uitge-
voerd op een VAX 11/780 computer met Fortran programma's, die ge-
bruik maakten van IMSL subroutines GGPOS (voor het trekken van
aselecte steekproeven uit een Poisson verdeling) en GGBIR (voor
het trekken van aselecte steekproeven uit een binomiale verde-
ling). De empirische studie bestond uit het simuleren van LDA's
in 19 verschillende situaties gespecificeerd door waarden van ϕ_1 ,
 ϕ_2 , P_1 , P_2 , d en ϕ ($\phi_1 \leq \phi \leq \phi_2$) waarbij ϕ de werkelijke waarde
van de frequentie is. Ter bepaling van J en $x_1, \dots, x_j$ werd de
procedure 4 t/m 7 uit deel I gebruikt. In het bijzonder werd
gekozen: $\phi_1 = .001$, $\phi_2 = .01$, $P_1 = .15$, $P_2 = .70$, $d = 3$ en ϕ
lopend van .001 tot .01 in stapjes van .0005. Hierbij merken wij
op dat de uitslag van het Monte Carlo onderzoek geldig is voor
iedere combinatie ϕ_1 en ϕ_2 met $\phi_2/\phi_1 = 10$. De omvang van elk
experiment was ($n_j = n =$) 24. Voor elk van de 19 experimentele
situaties werd 2000 maal een LDA gesimuleerd.

Het verschil tussen de werkelijke waarde van ϕ en de geschat-
te waarde van ϕ in elke simulatie werd gebruikt als maat voor de
fout van de desbetreffende schattingsprocedure. Door alle 2000
fouten op te tellen en te delen door 2000 werd de gemiddelde
onzuiverheid voor een specifieke experimentele situatie verkre-
gen. De drie schattingsprocedures werden vergeleken met de gemid-
delde relatieve onzuiverheid, d.w.z. de gemiddelde onzuiverheid
gedeeld door ϕ . De waarden van de gemiddelde relatieve onzuiver-
heid worden voor alle drie methoden gepresenteerd in figuur 2. De
niet onderbroken lijn (—) is gebruikt voor de JML methode, de
gestippelde lijn (...) voor de MC methode en de streepjes lijn
(---) voor de ML methode. Bestudering van figuur 2 laat zien dat
de ML en MC methode min of meer vergelijkbaar zijn op een groot
interval van ϕ , alhoewel tegengesteld van teken; op een klein



Figuur 2



Figuur 3

interval, gedraagt de MC methode zich beter dan de ML methode, terwijl in de buurt van de ondergrens ϕ_1 , de gemiddelde relatieve onzuiverheid van de MC methode sterk toeneemt. De JML methode is over het hele interval van ϕ veruit superieur t.o.v. zowel de MC als de ML methode, vanwege een gemiddelde relatieve onzuiverheid van bijna gelijk aan 0. Zoals reeds eerder is opgemerkt is het nut van een jackknife procedure o.a. dat de onzuiverheid (sterk) gereduceerd wordt. T.o.v. de ML methode blijkt dit dus te kloppen.

Een ander veel gebruikt criterium om schatters met elkaar te vergelijken is de variantie. In figuur 3 wordt de geschatte standaardafwijking gedeeld door ϕ uitgezet tegen ϕ . De drie methoden blijken alle drie min of meer vergelijkbare varianties te hebben waarbij de laagste waarde in het algemeen bij de JML methode wordt gevonden.

De conclusie van Taswell (1981) dat de MC methode te prefereren is boven de ML methode op grond van een kleine onzuiverheid en variantie wordt niet ondersteund door onze Monte Carlo experimenten. Indien men op grond van figuur 2 een keuze moet maken tussen de MC en ML methode dan kiezen wij in overeenstemming met Fazekas de St Groth (1982) voor de ML methode. Naar onze mening is het verschil te verklaren uit het feit dat Taswell (1981) een

te beperkte Monte Carlo studie heeft uitgevoerd: alleen LDA's werden gebruikt waarbij het aantal cellen per verdunning verdeeld zijn rond een dosis die correspondeert met een kans op een negatieve reactie van 50%. Inderdaad volgt uit de figuren 2 en 3 dat de MC methode op een klein interval beter is dan de ML methode. Beschouwt men het hele interval dan krijgt men een lichte voorkeur voor de ML methode. Indien men de drie methoden vergelijkt dan is de beste keuze de JML methode.

Tot slot zij nog opgemerkt dat een volledig verslag van dit onderzoek gegeven wordt in Strijbosch e.a. (1986).

REFERENTIES

1. Berkson, J. (1980), Minimum chi-square, not maximum likelihood! (with discussion), *Ann. Statist.* 8, 457-487.
2. Fazekas de St Groth, S. (1982), The evaluation of limiting dilution assays, *J. Immun. Methods* 49, R11 - R23.
3. Finney, D.J. (1978), *Statistical Method in Biological Assay*, Third Edition, Griffin, London.
4. Miller, R.G. (1974), The jackknife - a review, *Biometrika* 61, 1-15.
5. Reeds, J.A. (1978), Jackknifing maximum likelihood estimates, *Ann. Statist.* 6, 727 - 739.
6. Strijbosch, L.W.G., W.A. Buurman, R.J.M.M. Does, P.H. Zinken en G. Groenewegen (1986), Limiting dilution assays: experimental design and statistical analysis, *Medical Informatics and Statistics Report* 6, University of Limburg, Maastricht. Ter publicatie aangeboden.
7. Taswell, C. (1981), Limiting dilution assays for the determination of immunocompetent cell frequencies I: data analysis, *J. Immun.* 126, 1614 - 1619.