

# Een clusterstudie naar perinatale infecties met 4464 Nederlandse MS-patiënten

S. van Buuren, B.M. Zaadstra, J.M. van Noort \*

Verondersteld wordt dat infecties een rol spelen bij het ontstaan van multiplele sclerose (MS). Dit artikel gaat in op de vraag of infecties rondom de zwangerschap, de geboorte of de vroege kindertijd van invloed zijn op de ontwikkeling van MS op latere leeftijd. Gegevens werden verkregen door middel van een enquête onder de leden van de MS Vereniging Nederland. Deze gegevens omvatten onder meer de geboorteplaats en -datum van 4464 MS-patiënten. Met de methode van Jacquez is getoest met er simultane clustering van geboorteplaatsen en -data optreedt, hetgeen een aanwijzing voor een besmettelijke factor rondom de geboorte zou inhouden. De resultaten bevatten geen indicatie dat gelijktijdige clustering rondom geboorteplaats en -datum optreedt. Afgezien van de beperking dat de huidige methodiek niet in staat is een eventuele rol van alledaagse virussen vast te stellen, lijkt het daarom niet zo waarschijnlijk dat een perinatale besmetting een discriminerende bijdrage levert aan het ontstaan van MS.

Trefwoorden: multiplele sclerose (MS), tijd-ruimteclustering, perinatale infecties, k-NN toets

Reeds lang wordt gediscussieerd over de vraag in hoeverre (virus)infecties bijdragen aan het ontstaan van multiplele sclerose (MS), en op welke leeftijden een dergelijke besmetting zou kunnen optreden. Hierbij wordt onder meer gedacht aan de mogelijkheid dat vroege virale infecties een rol spelen in de etiologie van de ziekte, en dat de ziekte verkregen wordt na infectie met een 'langzaam virus' tijdens de kindertijd. Zo zijn er bijvoorbeeld overeenkomsten gevonden tussen het optreden van MS en besmettelijke ziekten als mazelen, polio, waterpokken, bof en kinkhoest.<sup>1,2</sup> Daarnaast spelen alledaagse virussen zoals herpes virus type 1 and 2 en het menselijk herpes type 6 mogelijk een rol.<sup>3-5</sup> MS zou getriggerd kunnen worden door een sterke reactie van het immuunsysteem op een late infectie met een alledaags virus. Zo hebben enkele auteurs gewezen op de periode rondom de puberteit.<sup>6-8</sup> Studies onder migranten laten vrij eenduidig zien dat de kans op het krijgen van MS na de adolescentie niet meer verandert. Dergelijk studies bieden echter weinig aanknopingspunten voor de vraag waar de meest vatbare periode ('window of susceptibility') ligt in de periode tussen conceptie en de vijftiende verjaardag. Het is mogelijk dat een dergelijke vatbare periode zeer kort is, zoals bijvoorbeeld is gevonden voor de relatie tussen mazelen en de ziekte van Crohn, maar een lange termijn is evenzeer mogelijk. Het kennen van een periode waarin men het

meest vatbaar voor MS is, is van belang omdat dit richting zou geven aan het epidemiologisch en etiologisch onderzoek op dit terrein, en bij het ontwikkelen en tijdig implementeren van preventieve maatregelen.

Indien de etiologie van MS inderdaad een besmettelijke component bevat, is het te verwachten dat personen in vatbare leeftijdsgroepen, bijvoorbeeld rondom de geboorte, die zich in de buurt bevinden van reeds geïnfecteerde personen een verhoogd risico op infectie hebben en daarmee op het krijgen van de ziekte. In dat geval zullen de geboorteplaatsen en -data van personen die later MS hebben gekregen dichter bijeen liggen dan op basis van toeval verwacht mag worden. Er bestaan diverse statistische technieken waarmee dit kan worden getoest. De nulhypothese hierbij is dat clustering afwezig is. Geboortedatum en -plaats worden opgevat als indicator van de locatie van elke case gedurende de perinatale periode en de vroege kindertijd.

Eerdere studies op dit gebied hebben weinig aanknopingspunten opgeleverd voor een clustering rondom de geboorte. Veel van deze studies hanteerden echter nogal grove maten voor tijd en ruimte, zoals 'geboren in hetzelfde jaar', of gingen gebukt onder kleine aantallen. In een eerdere studie met andere gegevens vonden we ook geen clustering.<sup>9</sup> De huidige studie is gebaseerd op een aanzienlijk groter aantal patiënten, en kent een completere geografische dekking van de Nederlandse patiëntenpopulatie. Hierdoor is een eventueel aanwezige clustering beter detecteerbaar.

\* S. van Buuren<sup>1</sup>, B.M. Zaadstra<sup>1</sup>, J.M. van Noort<sup>1</sup>  
1 TNO Preventie en Gezondheid, Leiden

## MATERIAAL EN METHODE

### Gegevens

Gegevens over geboorteplaatsen en -data van MS-patiënten in Nederland zijn verzameld met behulp van een korte schriftelijke enquête onder de leden van de MS Vereniging Nederland (MSVN). De MCSVN heeft in februari 1996 op verzoek van TNO-PG aan 10.504 leden een korte vragenlijst toegestuurd. In totaal hebben 6358 mensen de vragenlijst teruggestuurd, waarvan 93 niet ingevuld (respons van 59,6%). Van de 6265 gaven 4926 (78,6%) aan dat zij MS hadden. *Tabel 1* geeft een samenvatting van enkele achtergrondkenmerken van deze steekproef.

Van 4475 in Nederland geboren MS-patiënten is plaats en datum van geboorte bekend. De naam van de geboorteplaats is gebruikt om aan elke patiënt een x-y coördinaat toe te wijzen. Deze coördinaat wordt gebruikt in de berekening van de afstand tussen de geboorteplaatsen van twee willekeurige MS-patiënten. De koppeling tussen geboorteplaats en x-y coördinaat is als volgt tot stand gebracht. Een bestand met de geografische grenzen van 3941 postcodegebieden in Nederland werd geaggregeerd tot een bestand van 2425 verschillende topografische namen. Elke geboorteplaats in het databestand is vervolgens voorzien van de x-y coördinaat van het geografische centrum van de plaats. Op deze wijze zijn 4464 MS-patiënten aan één van 2425 locaties in Nederland gekoppeld. Bij 11 patiënten kon de opgegeven geboorteplaats niet worden teruggevonden.

### Statistische toets

Voor het vaststellen van tijd-ruimte clustering gebruiken we de recentelijk geïntroduceerde k-nearest neighbour (k-NN) methode van Jacquez.<sup>10</sup> Een belangrijke eigenschap van deze methode is dat het resultaat niet afhankelijk is van de achterliggende populatieverdeling zolang deze verdeling constant is. In vergelijking met de klassieke toetsen van Knox<sup>11</sup> en Mantel<sup>12</sup> hangt de k-NN methode niet af van arbitraire constanten. Uit simulatieonderzoek van Jacquez blijkt dat de statistische power van de k-NN methode superieur is aan die van de twee andere toetsen, maar deze bevinding is nog niet door anderen gestaafd.

In de k-NN methode wordt het aantal paren in de data geteld dat tegelijkertijd buurman is in zowel tijd als ruimte. Eerst wordt per patiënt een set van k (bijvoorbeeld 10) patiënten samengesteld waarvan de geboortedatum zo weinig mogelijk verschilt. Hetzelfde gebeurt voor geboorteplaats. Vervolgens wordt het aantal paren geteld dat in beide sets voorkomt. Dit wordt voor iedere patiënt gedaan. Het totaal aantal paren dat in beide sets wordt aangetroffen noemen we  $J_k$ . Merk op dat per definitie alle paren voor bijvoorbeeld  $k=9$  onderdeel uitmaken van de paren voor  $k=10$ . Er geldt dus  $J_9 \leq J_{10}$ . De grootheid  $\Delta J_k = J_k - J_{k-1}$  geeft het verschil in paren tussen opeenvolgende niveaus van k aan.

De statistische toets bestaat uit het vergelijken van de gevonden  $J_k$  (of  $\Delta J_k$ ) met de referentieverdeling behorend bij nulhypothese van 'geen tijd-ruimteclustering'. Deze

referentieverdeling wordt met een Monte Carlo techniek bepaald door de geboortedata in het databestand te randomiseren en dan opnieuw  $J_k$  te berekenen. Door dit 100 maal te herhalen ontstaat de referentieverdeling voor  $J_k$  waarvoor geldt dat de geboortedata onafhankelijk zijn van de geboorteplaatsen, dat wil zeggen dat er geen clustering optreedt. De referentieverdeling van  $\Delta J_k$  wordt op dezelfde wijze afgeleid. De P-waarde is gelijk aan de propotie waarden uit de referentieverdeling die groter zijn dan  $J_k$ . De benodigde rekentijd van de toets neemt ongeveer kwadratisch toe met het aantal observaties. Het huidige probleem ( $n = 4464$ , 100 simulatieruns) kostte bijna 45 uur CPU-tijd op een Pentium 100 Mhz. Soms komt het voor

**Tabel 1** Achtergrondkenmerken MCSVN patiëntenbestand per februari 1996 ( $N=4926$ )

	Aantal	%
Geslacht		
Mannen	1485	30,2
vrouwen	3428	69,8
Onbekend	13	
Geboortejaar		
Voor 1930	422	8,6
1930-1939	787	16,1
1940-1949	1365	27,9
1950-1959	1462	29,8
1960-1969	766	15,7
Na 1970	88	1,8
Onbekend	36	
Jaar diagnose MS		
Voor 1950	35	0,7
1950-1959	92	1,9
1960-1969	238	4,9
1970-1979	700	14,3
1980-1989	1910	39,1
1990-1996	1910	39,1
Onbekend	41	
Leeftijd bij diagnose MS (jaar)		
Jonger dan 20	119	2,4
20-29	1156	23,8
30-39	1704	35,0
40-49	1300	26,7
Ouder dan 50	588	12,1
Onbekend	59	
Huidige MS vorm		
Aanval-herstel variant	2058	42,8
Geen remissies meer	2034	42,3
Weet niet	717	14,9
Onbekend	117	
Geboorteplaats		
Binnen Nederland	4598	97,1
Buiten Nederland	135	2,9
Onbekend	193	
N	4926	100,0

dat meerdere patiënten in dezelfde plaats geboren zijn, zodat er dan geen unieke eerste (of tweede, derde, ...) buurman is. In deze gevallen is random één van de kandidaten getrokken en is deze in de berekeningen gebruikt.

## RESULTATEN

Tabel 2 bevat de resultaten van de k-NN test, waarbij k van 1 tot 10 oploopt. De rij ' $J_k$ ' bevat het aantal paren in de gegevens dat buurman is in zowel tijd als ruimte. Voor  $k=1$  (dat wil zeggen alleen directe buren tellen mee) blijken er dus twee van dergelijk paren te zijn. De rij ' $E[J_k]$ ' bevat het verwachte aantal onder de nulhypothese. De rij daaronder geeft de P-waarde van het verschil tussen  $J_k$  en  $E[J_k]$ . Zowel in de absolute ( $J_k$ ) als in de differentiële ( $\Delta J_k$ ) test geldt dat het gevonden aantal paren, op één na ( $k=8$ ), niet significant hoger is dan de verwachte waarde onder de nulhypothese. De door Jacquez voorgestelde samenvattende P-waarde is gelijk aan 0,52.

Ter controle hebben we tevens de toetsen van Mantel en Knox uitgevoerd. De Mantel Z-statistic is gelijk aan  $Z=664,0$ , met een verwachte waarde van 663,1 ( $p=0,44$ ). De effecten van de Knox test bleken verwaarloosbaar klein en vielen ruimschoots binnen de grenzen van de steekproefvariatie.

Geen van de uitgevoerde toetsen gaf aanleiding om de nulhypothese te verwerpen dat tijd en plaats van geboorte van MS-patiënten onafhankelijk van elkaar zijn. Wellicht het meest overtuigende bewijs van het tegendeel is dat we na het uitproberen van mogelijke afkappunten zelfs bij deze steekproefgrootte nauwelijks significante resultaten krijgen.

## DISCUSSIE

De gegevens omvatten niet alle MS-patiënten in Nederland. Naar schatting is ongeveer de helft van hen lid van de MSVN en hebben we gegevens van ongeveer één derde van alle patiënten. Wat heeft dat voor gevolgen voor onze conclusies? Een mogelijke selectie bias die een clustering zou kunnen onderdrukken, treedt op wanneer sterk geclusterde MS-patiënten minder vaak MSVN-lid zijn of slechter responderen. Een dergelijk proces lijkt a priori niet erg plausibel. Een ander potentieel verstorend effect treedt op wanneer bijvoorbeeld meer uitval zou plaatsvinden in de lagere SES-groepen, die weer in bepaalde wijken geconcentreerd kunnen zijn. De methode die we gebruikt hebben is voor zo'n concentratie ongevoelig zolang de

sterkte van clustering in geboortedata binnen deze wijken niet afwijkt van die van de algehele populatie. Op voorhand lijkt er geen reden om te twijfelen aan de validiteit van deze laatste voorwaarde. Ook kunnen ontbrekende patiënten de sensitiviteit van de methode aantasten. Stel dat een virus zich verspreidt van case A naar C via B. Wanneer B niet aan het onderzoek deelneemt, wordt de link tussen A en C verbroken. Het gevolg is dat clustering uitvlakt over een grotere tijd- en ruimtespanne en daardoor moeilijker te vinden is.

In deze studie is geen direct lichamelijk onderzoek verricht. Men kan zich daarom afvragen of de diagnose MS juist is. Aangezien MS een ernstige ziekte is, zal de diagnose MS niet licht worden gesteld. Tegenwoordig zet de neuroloog de MRI-scan routinematig in om MS eenduidig te diagnosticeren. Het is daarom aannemelijk dat degenen die zeggen MS te hebben, ook werkelijk de ziekte hebben. Het omgekeerde geldt ook. MS heeft dermate ernstige gevolgen op het dagelijks functioneren dat er weinig of geen mensen zijn die wel MS hebben, maar waarbij niet de diagnose is gesteld.

De gehanteerde methodiek is niet in staat alle vormen van besmetting rondom de geboorte op te sporen. Overdracht tussen zuigelingen zou kunnen plaatsvinden via een derde, bijvoorbeeld een volwassene die alleen tijdelijk drager is en zelf geen MS ontwikkelt. Als dit dragerschap kortdurend is, verwacht men, net als bij een directe overdracht tussen zuigelingen onderling, een clustering in geboortedata en -plaatsen. Indien de drager echter lang besmettelijk blijft, kunnen ook op andere tijden en plaatsen zuigelingen worden besmet, eventueel zelfs via infecties van andere dragers die op hun beurt weer zuigelingen besmetten. In die omstandigheden speelt besmetting wel degelijk een rol binnen de etiologie van MS, maar kan als gevolg van verdunning weinig of geen clustering in de gegevens aanwezig zijn. De methodiek levert slechts informatie over die processen of infectievormen die tot een detecteerbare clustering in de geboorteplaatsen en -data leiden. De aanwezigheid van clustering is een sterk argument voor de infectiehypothese, terwijl de afwezigheid van clustering maar in beperkte mate iets zegt over de afwezigheid van een infectieuze component.

Toetsen voor tijd-ruimte-interactie worden regelmatig bekritiseerd vanwege hun lage statistische power. Uit simulatiestudies blijkt dat de zwakste toets, de Knox test, een bevredigende power ( $>0,80$ ) heeft voor

**Tabel 2 Resultaten van de k-NN test op tijd-ruimteclustering voor  $k=1,\dots,10$  ( $n=4464$ )**

	k									
	1	2	3	4	5	6	7	8	9	10
$J_k$	2	5	10	22	27	33	39	62	81	100
$E[J_k]$	0,95	3,98	9,45	16,6	25,4	36,7	49,6	64,9	82,1	101
p	0,25	0,41	0,46	0,13	0,46	0,66	0,90	0,67	0,59	0,56
$\Delta J_k$	2	3	5	12	5	6	6	23	19	19
$E[\Delta J_k]$	0,95	3,03	5,47	7,11	8,87	11,3	12,9	15,3	17,2	19
p	0,25	0,59	0,56	0,08	0,91	0,97	0,99	0,03	0,41	0,54

steekproefgroottes groter dan 125.<sup>10</sup> De huidige studie kent een steekproefgrootte van 4464 personen. Bij dergelijke aantallen treedt eerder het omgekeerde probleem op ('alles is significant') en is het vrij uitzonderlijk wanneer géén effect wordt gevonden.

Een aanname van de gebruikte methode is dat de geografische verdeling van het aantal geboorten in de achtergrondpopulatie constant is in de tijd. Strikt genomen gaat deze assumptie niet op. Groeikernen zoals Zoetermeer of Hoofddorp leveren na verloop van tijd een steeds groter aandeel aan het totale aantal geboorten. Om de aanname te controleren hebben we analyses ook uitgevoerd per decennium van geboorten. Ook op deze wijze hebben we niet het bestaan van clustering kunnen vaststellen, zodat de aanname van tijdinvariantie in onze hoofdanalyse niet in ernstige mate lijkt te worden geschonden.

Het ruimtelijk oplossend vermogen van de analyse hangt samen met de omvang van de geboorteplaats. Mensen die zijn geboren in gehuchten en dorpen, kunnen tot binnen 200 à 300 meter van de werkelijke geboortelocatie op de analysekaart geplaatst worden. Degenen die in grotere plaatsen zijn geboren, kunnen echter niet van elkaar worden onderscheiden en worden beschouwd alsof ze geboren zijn in het centrum van de gemeente of stad. In grote steden kan de discrepantie tussen de werkelijke en toegezwezen geboorteplaats oplopen tot maximaal zo'n vijf kilometer. Omdat deze onnauwkeurigheid een eventueel aanwezige clustering zou kunnen maskeren hebben we de analyse tevens uitgevoerd op een substeekproef ( $n=3144$ ) waaruit alle patiënten zijn verwijderd met geboorteplaatsen in de grote steden. De conclusie uit deze analyse was dezelfde als die voor de volledige steekproef. Het bestaan van onvoldoende resolutie in de grote steden biedt daarom geen verklaring voor de afwezigheid van clusters.

Onze resultaten leveren geen ondersteuning aan het idee dat een infectieuze gebeurtenis rondom de geboorte een factor van belang is voor de ontwikkeling van MS. Hierbij tekenen we aan dat in de praktijk overdrachtskan- sen niet altijd evenredig zijn met de inverse van de geografische afstand. Verder sluit, zoals gezegd, het ontbreken van clustering niet uit dat perinatale infecties toch een rol spelen. Een lage gevoeligheid voor MS bij geïnfecteerden kan clustering maskeren. Het infectieuze agens kan zeer algemeen zijn en het werkelijk beloop kan mede afhangen van de aanwezigheid van andere factoren. Afgezien van deze beperkingen zijn onze bevindingen in overeenstemming met studies waarin geopperd wordt dat eerder de puberteit dan de perinatale periode het tijdsinterval bevat waarin een mogelijke infectie de ontvankelijkheid voor MS kan beïnvloeden.

## ONDERSTEUNING

Deze studie is mogelijk gemaakt met financiële steun van het Praeventiefonds.

De auteurs zijn zeer erkentelijk voor de medewerking van de MS Vereniging Nederland en van mevr. dr. A.M. Kremer voor de bewerking van de gegevens.

## ABSTRACT

*A cluster study for perinatal infections in 4464 Dutch MS patients*

Infections may play a role in the development of multiple sclerosis (MS). The paper studies the question whether perinatal infections are likely to be of relevance to MS pathogenesis. Birthplace and date of birth of 4464 Dutch MS patients was obtained by means of a questionnaire sent to the members of the MS Vereniging Nederland (Dutch Association of MS patients). The k-NN method of Jacquez was used to test whether simultaneous clustering in date and place of birth occurs, which would indicate a perinatal contagious factor. The results contain no evidence for the existence of such a clustering. Apart from the fact that the method cannot determine a possible role for common and widespread viruses, this suggests that perinatal infectious events have no distinctive contribution to MS susceptibility.

*Key words: multiple sclerosis (MS), perinatal infections, k-NN method, space-time clustering*

## LITERATUUR

- 1 Sadovnick AD, Ebers CE. Epidemiology of multiple sclerosis: a critical overview. *Can J Neurol Sci* 1993;20:17-29.
- 2 Martyn CN. Infection in childhood and neurological diseases in adult life. *Brit Med Bull* 1997;53:24-39.
- 3 Herndon RM. Herpes viruses in multiple sclerosis. *Arch Neurol* 1996;53:123-4.
- 4 Sanders VJ, Waddell AE, Felisan SL, Li X, Conrad AJ, Tourtellotte WW. Herpes simplex virus in postmortem multiple sclerosis brain tissue. *Arch Neurol* 1996;53:124-33.
- 5 Challoner PB, Smith KT, Parker JD, et al. Plaque-associated expression of human herpesvirus 6 in multiple sclerosis. *Proc Natl Acad Sci USA* 1995;92:7440-4.
- 6 Wolfson C, Wolfson DB. The latent period of multiple sclerosis: a critical review. *Epidemiology* 1993;4:464-70.
- 7 Poser CM. The epidemiology of multiple sclerosis: a general overview. *Ann Neurol* 1994;36:S180-93.
- 8 Riise T, Grønning M, Klauber MR, Barrett-Connor E, Nyland H, Albrektsen G. Clustering of residence of multiple sclerosis patients at age 13 to 20 years in Hordaland, Norway. *Am J Epidemiol* 1991;133:932-9.
- 9 Van Buuren S, Zaadstra BM, Zwanikken CP, Buljevac D, Van Noort JM. Space-time clustering of multiple sclerosis cases around birth. *Acta Neurol Scand* 1998;97:351-8.
- 10 Jacquez GM. A k nearest neighbour test for space-time interaction. *Stat Med* 1996;15:1935-49.
- 11 Knox EG. Detection of space-time interactions. *Appl Stat* 1964;13:25-9.
- 12 Mantel N. Detection of disease clustering and a generalized regression approach. *Cancer Res* 1967;27:209-20.

## CORRESPONDENTIEADRES

Dr. S. van Buuren, Vakgroep Statistiek, TNO Preventie en Gezondheid, Postbus 2215, 2301 CE Leiden, tel. 071-5181802, fax 071-5181920, e-mail: s.vanbuuren@pg.tno.nl

Voor publicatie aanvaard op 11 september 1998