



Rijksinstituut voor Volksgezondheid  
en Milieu  
*Ministerie van Volksgezondheid,  
Welzijn en Sport*

**Verkenning van de variatie in ruimte en  
tijd van de incidentie van leukemie bij  
kinderen en jongeren in Nederland  
(1995-2010)**

RIVM rapport 200112002/2014

Y.M. Mulder et al.



Rijksinstituut voor Volksgezondheid  
en Milieu  
*Ministerie van Volksgezondheid,  
Welzijn en Sport*

**Verkenning van de variatie in ruimte en  
tijd van de incidentie van leukemie bij  
kinderen en jongeren in Nederland  
(1995-2010)**

RIVM Rapport 200112002/2014  
Y.M. Mulder et al.

## Colofon

© RIVM 2014

Delen uit deze publicatie mogen worden overgenomen op voorwaarde van bronvermelding: Rijksinstituut voor Volksgezondheid en Milieu (RIVM), de titel van de publicatie en het jaar van uitgave.

Y.M. Mulder  
C.B. Ameling  
O.R.P. Breugelmans  
A. Dusseldorp

Contact:  
Yonne Mulder  
DMG  
yonne.mulder@rivm.nl

Dit onderzoek werd verricht in opdracht van het ministerie van VWS, in het kader van V/200112/14/PC, ondersteuning GGD'en

Dit is een uitgave van:

**Rijksinstituut voor Volksgezondheid  
en Milieu**  
Postbus 1 | 3720 BA Bilthoven  
www.rivm.nl

## Publiekssamenvatting

Het RIVM heeft onderzocht of de mate waarin leukemie bij kinderen en jongeren voorkwam tussen 1995 en 2010 varieerde tussen regio's en jaren. De verschillen blijken klein te zijn.

GGD'en krijgen geregeld meldingen van bezorgde bewoners die vermoeden dat er de laatste jaren in hun buurt opvallend veel kankerpatiënten zijn of die het idee hebben dat een andere ziekte veel voorkomt (clustermelding). Vaak wordt dan een oorzaak in de omgeving verondersteld, zoals industrie, hoogspanningslijnen of bodemverontreiniging. Bij dergelijke clustermeldingen gebruikt de GGD een stappenplan waarbij wordt gekeken of het lokale aantal patiënten afwijkt van het landelijke of regionale gemiddelde. De GGD zoekt uit of blootstelling aan eventuele omgevingsfactoren hier invloed op kan hebben gehad, en besteedt veel aandacht aan de communicatie hierover met de melders.

De resultaten van het RIVM-onderzoek bevestigen dat dit een goede eerste aanpak is bij een clustermelding. Als hiermee niet kan worden uitgesloten dat er sprake is van een cluster, bestaat de mogelijkheid om nader onderzoek uit te voeren. Daarvoor worden dan aanvullende gegevens verzameld en geanalyseerd. In dit RIVM-rapport staan enkele methoden beschreven die daarvoor in aanmerking komen. Er zal naar verwachting niet vaak aanleiding zijn om deze in te zetten, omdat met het stappenplan het merendeel van de meldingen adequaat kan worden afgehandeld.

De in het rapport beschreven methoden zijn geschikt om landelijk in beeld te brengen in welke mate ziekten, waaronder leukemie, voorkomen en of daarin variatie optreedt in ruimte en tijd.

De gegevens en resultaten uit dit onderzoek kunnen op kaarten worden weergegeven, indien daarbij rekening wordt gehouden met privacy regels. De getoonde informatie mag immers niet tot personen herleidbaar zijn. Een dergelijke presentatie van de gegevens kan bijdragen aan de communicatie door de GGD over de mate waarin leukemie en andere vormen van kanker in Nederland variëren.

Kernwoorden: ruimtelijke analyse, leukemie, kankercluster, methoden



## Abstract

The Dutch National Institute for Public Health and the Environment (RIVM) has analyzed the incidence of leukaemia in children and adolescents during the 1995-2010 period, to determine if there were any variations between regions and/or between years. Only minor variations were found.

Municipal health authorities in the Netherlands are frequently contacted by concerned citizens who believe they have observed a suspicious increase in the number of local cancer patients over the past few years, or who believe that certain other diseases occur more frequently in their neighbourhood than elsewhere (so-called 'reported disease clusters'). These reports often assume that environmental factors like heavy industry, high-voltage lines or soil pollution are the cause of higher disease incidence. Municipal health authorities investigate such cluster reports by implementing a phased plan to establish if the number of local patients deviates significantly from expected values based on national or regional data. The health authorities then try to determine if potential environmental exposures might have been a risk factor. A great deal of attention is devoted to careful communication with the citizens who submitted the reports.

The results of the RIVM investigation confirm that this method of dealing with cluster reports is an effective initial response. If the presence of a disease cluster cannot be ruled out, further research may be required to collect and analyze additional data. This RIVM report describes a number of methods that may be used for this purpose. RIVM does not expect that such methods will need to be used frequently, as most cluster reports can be handled adequately using the aforementioned stepwise protocol.

The methods described in the report can be used to analyze the national incidence of diseases as well as the space-time variation in incidence.

Such analyses yield data and results that can be visualized in maps, provided the applicable privacy legislation is adhered to. After all, it must not be possible to use this information to identify individual patients. Municipal health authorities can use such data visualisations during communication with the public on the variation in the incidence of leukaemia and other types of.

Keywords: Spatial Analysis, Leukaemia, Cancer Cluster, Methods



## Inhoudsopgave

### DEEL 1. HOOFDRAPPORT

- 1. Inleidig - 15**
  - 1.1 Aanleiding - 15
  - 1.2 Leukemie bij kinderen - 15
  - 1.3 Doelstelling - 16
  - 1.4 Vraagstelling - 16
  - 1.5 Werkwijze - 16
    - 1.5.1 Begeleiding onderzoek - 17
    - 1.5.2 Rapportage van de bevindingen/leeswijzer - 17
  
- 2. Onderzoekopzet en onderzoeksmethode - 19**
  - 2.1 Dataverzameling - 19
    - 2.1.1 Incidentiegegevens Kankerregistratie IKNL - 19
    - 2.1.2 Bevolkingscijfers op postcodeniveau - 20
  - 2.2 Implicaties van beschikbare data voor keuze onderzoeksgroepen - 21
  - 2.3 Datamanagement - 21
  - 2.4 Statistische analyse - 23
    - 2.4.1 Disease mapping - 23
    - 2.4.2 Scan-analyse - 23
    - 2.4.3 Overeenkomsten en verschillen tussen de methoden - 24
  
- 3. Resultaten analyses - 27**
  - 3.1 Disease mapping - 27
    - 3.1.1 Ruimte- en tijdseffecten - 27
    - 3.1.2 Verkregen SIR-schattingen - 28
  - 3.2 Scananalyse (clusterdetectie) - 28
    - 3.2.1 Ruimtelijke Clustering SatScan - 28
    - 3.2.2 Ruimtelijke clustering FlexScan - 29
    - 3.2.3 Clusteranalyse in ruimte én tijd (SatScan) - 29
    - 3.2.4 Scananalyses samengevat - 29
  - 3.3 Vergelijking disease mapping en scan-analyse - 29
  - 3.4 Vergelijkbaarheid uitkomsten met bestaande clustermeldingen - 29
    - 3.4.1 Gerapporteerd cluster - 29
    - 3.4.2 Clustermeldingen in GGD-registratie - 30
  
- 4. Mogelijkheid tot koppeling met milieufactoren - 31**
  - 4.1 Leukemie in relatie tot omgevingsfactoren - 31
  - 4.2 Nut van aanvullend onderzoek inclusief milieufactoren - 32
    - 4.2.1 Elektromagnetische velden - 32
    - 4.2.2 Bestrijdingsmiddelen - 34
    - 4.2.3 Mogelijke verbeteringen van het clusteronderzoek in de toekomst - 34
  
- 5. Conclusies en discussie - 37**
  - 5.1 Verdeling van de leukemie-incidentie in Nederland in ruimte en tijd - 37
  - 5.2 Bruikbaarheid geavanceerde ruimtelijke analysemethoden bij clustermeldingen - 38
    - 5.2.1 Discussie en conclusies naar aanleiding van data en datamanagement - 38
    - 5.2.2 Bruikbaarheid gevisualiseerde resultaten in de risicocommunicatie - 39
  - 5.3 Bruikbaarheid geavanceerde ruimtelijke methoden voor monitoring - 39
  - 5.4 Koppeling van leukemie bij kinderen met milieufactoren - 40



5.5 Beschouwing - 41

## **DEEL 2. ACHTERGRONDEN**

### **6. Geografische databewerking (achtergrond) - 45**

6.1 Methode databewerking op postcodeniveau - 45

6.1.1 Gebiedsgerelateerde bewerking op PC4-niveau - 45

6.1.2 Gebiedsgerelateerde dataverwerking op PC6-niveau - 49

### **7. Statistische analyse (achtergrond) - 51**

7.1 Disease mapping - 51

7.1.1 Model en modeldiagnostiek - 51

7.1.2 Rekentechnische omstandigheden - 52

7.2 Spatial Scan Statistiek voor clusteranalyse in ruimte en/of tijd - 53

7.2.1 Ruimtelijke analyse met SatScan - 53

### **8. Weergave en visualisatie van de resultaten (achtergrond) - 57**

8.1 Visualisatie disease mapping - 57

8.2 Visualisatie SatScan-analyse - 57

8.3 Visual Inquiry Toolkit - 58

8.4 Visualisatie FleXScan-analyse - 58

8.5 Visualisatie van de data op geaggregeerd niveau - 59

### **9. Literatuur – 61**

Bijlage A Begeleiding en advies - 65

Bijlage B Disease mapping (Bayesiaanse analyse) - 67

Bijlage C Omgevingsgerelateerde risicofactoren kinderleukemie - 69

Bijlage D Afkortingen en toelichting gebruikte termen – 73

## Samenvatting

GGD'en krijgen geregeld vragen van verontruste bewoners over een opvallend groot aantal mensen met kanker in hun omgeving (een veronderstelde kankercluster). Vaak wordt een oorzaak in het milieu vermoed. Bij de behandeling van dergelijke meldingen werkt de GGD volgens een stappenplan, beschreven in de GGD-richtlijn 'Kankerclusters'. Leukemie bij kinderen is een van de kankersoorten waarbij regelmatig clustermeldingen worden gedaan. Jaarlijks wordt bij ongeveer 140 kinderen tussen 0 en 14 jaar in Nederland leukemie geconstateerd. In de onderzoeksperiode (1995-2010) betrof het 2.673 kinderen en jongeren in de leeftijd van 0 tot 19 jaar (verder benoemd als 'kinderen'). Vanwege het kleine aantal ziektegevallen per jaar is een relatie tussen een cluster van zeldzame aandoeningen, zoals leukemie, en milieufactoren lokaal vrijwel nooit te onderzoeken. Daarom is in het verlengde van de richtlijn de vraag gerezen of het RIVM zijn expertise op het gebied van ruimtelijke analyses kan inzetten om inzicht te krijgen in de variatie van leukemie-incidentie op nationale schaal, en om op grond daarvan een uitspraak te doen over de bruikbaarheid van deze technieken en gegevens voor de GGD-praktijk.

### *Doel en vraagstellingen*

Het onderzoek beoogt inzicht te bieden in de vraag of er sprake is van variatie van betekenis in ruimte en tijd van de incidentie van leukemie bij kinderen (0-19 jaar) in Nederland, en op grond daarvan de volgende vragen te beantwoorden:

- Zijn er in Nederland leukemieclusters aan te tonen bij kinderen, in ruimte of tijd?
- Op welke manier zijn de methoden en resultaten van dit onderzoek bruikbaar voor de GGD bij het omgaan met gemelde clusters?
- Zijn de methoden bruikbaar voor landelijke monitoring van leukemie en/of andere aandoeningen?
- Zijn de methoden geschikt om een koppeling te maken tussen het vóórkomen van kankerclusters (en andere aandoeningen) en milieufactoren en is het zinvol om in vervolgonderzoek milieufactoren te betrekken?

### *Data(bewerking)*

Integraal Kankercentrum Nederland (IKNL) heeft voor dit onderzoek gegevens geleverd uit de Nederlandse Kankerregistratie over de incidentie van leukemie in de leeftijdsgroep van 0 tot 19 jaar, in de periode 1989-2010. Over deze periode heeft IKNL ook data verstrekt over hersentumoren en andere vormen van kanker bij deze leeftijdsgroep, bij wijze van referentiegroepen. Deze data zijn op viercijferige postcode (PC4) voor de gehele periode beschikbaar, en op zescijferige postcode (PC6) vanaf 2007 voor het hele land. Landelijke analyses op PC6-niveau zijn daardoor pas volledig landsdekkend mogelijk vanaf 2007. Deze periode is te kort om bij een dergelijke lage incidentie zinvolle analyses uit te voeren. Om over een zo groot mogelijke periode een landsdekkend beeld te krijgen zijn de analyses daarom uitgevoerd op PC4-niveau, vanaf 1995. Vanaf 1995 zijn namelijk ook de bijbehorende bevolkingscijfers (om het aantal verwachte leukemiegevallen in een gebied te kunnen vaststellen) op postcodeniveau beschikbaar. Voorafgaand aan de analyses moesten de data uitgebreid bewerkt worden, omdat de postcodegebieden in de loop van de tijd steeds veranderd zijn. De gebruikte methode is opgenomen als bijlage van dit rapport en is mogelijk bruikbaar in andere projecten waarin wordt gekeken naar veranderingen van aandoeningen in de tijd.

*Ruimtelijke analyses*

Het RIVM heeft de laatste jaren in het SMARHAGT-onderzoeksprogramma (Small Area Health Analysis Geographical Toolkit) expertise ontwikkeld in het toepassen van geavanceerde ruimtelijke analysemethoden bij milieugezondheidsonderzoek. Daarbij lag de nadruk op toepassing van zogenaamde *disease mapping* met behulp van een bayesiaans ruimtelijk regressiemodel. Ook is literatuuronderzoek gedaan naar methoden voor 'clusteranalyse', waaruit '*Scan-Statistics*' (SatScan en FlexScan) als meest relevant naar voren kwamen (scan-analyse). Beide methoden analyseren verschillende aspecten van ruimtelijke incidentiepatronen. Bij *disease mapping* wordt gekeken of het vóórkomen van een ziekte in vooraf gekozen gebieden overeenkomt met het verwachte aantal ziektegevallen. *Scan-analyse* is dynamischer; deze methode inventariseert óf er gebieden zijn te vinden waarin het vóórkomen van de ziekte afwijkt van andere gebieden. Deze gebieden zijn niet van tevoren gedefinieerd. Daardoor kan deze methode clusters van verschillende omvang detecteren (met als kleinste omvang in dit geval het PC4-niveau).

*Geen belangrijke ruimtelijke variatie in leukemie bij kinderen (vraag 1)*

Met *disease mapping* werden geen aanwijzingen gevonden voor belangrijke ruimtelijke variatie in de incidentie van leukemie op PC4-niveau. Het maakte daarbij geen verschil of ook het verloop in de tijd in het model werd opgenomen. Dit betekent niet zonder meer dat er geen variatie aanwezig is. Maar wat er aan variatie is, kan bij het gevonden incidentieniveau en het gekozen analyseniveau (PC4) niet goed worden gedetecteerd.

Met de *scan-analyses* werd één statistisch significant cluster aangetroffen bij de tijd-ruimte-analyses, waarbij ook de separate jaren in de analyse zijn opgenomen. Dit betreft een tijdelijke verhoging van een viertal jaren. Bij ruimtelijke analyse over de gehele onderzoeksperiode werden na correctie voor herhaald testen geen statistisch significante clusters gedetecteerd.

*Bruikbaarheid voor de GGD-praktijk (vraag 2)*

De onderzoeksresultaten ondersteunen de werkwijze in de GGD-richtlijn 'Kankerclusters'. Daarin worden lokale cijfers vergeleken met landelijke gemiddelden, wat een goede aanpak is als er een homogene verdeling van de ziekte is. Meestal zal een geavanceerdere en bewerkelijke methode dan ook geen meerwaarde hebben bij een gemeld cluster. In enkele gevallen zou een clusteranalyse of *disease mapping*-methode ingezet kunnen worden om te bezien of het gemelde cluster daadwerkelijk kan worden gedetecteerd en eventueel is af te bakenen in ruimte en tijd.

De gevisualiseerde gegevens uit dit onderzoek zijn bruikbaar bij een lokale clustermelding: de visualisaties maken inzichtelijk hoe de incidentie van leukemie en andere soorten kanker van nature voortdurend varieert in ruimte en tijd. Een instrument en bijbehorend filmpje, gebaseerd op de resultaten uit dit project, wordt in overleg met IKNL beschikbaar gesteld aan de GGD'en.

*Bruikbaarheid voor landelijke monitoring (vraag 3)*

Een mogelijke toepassing is om de methoden in te zetten om (periodiek) in kaart te brengen of er bepaalde vormen van kanker (of andere aandoeningen) clustering vertonen in ruimte en/of tijd. Dit zou gebruikt kunnen worden voor monitoring, in hypothesegenererend onderzoek of ter ondersteuning van beleid.

*Mogelijkheid voor koppeling met milieufactoren (vraag 4)*

In principe is het mogelijk om bij toepassing van ruimtelijke analysemethoden ook potentiële risicofactoren zoals milieufactoren te betrekken. De zin ervan is onder andere afhankelijk van de beschikbare gegevens over blootstelling en het aantal mensen met de betreffende aandoening. In hoofdstuk 4 wordt beargumenteerd dat binnen Nederland vervolgonderzoek naar potentiële risicofactoren voor leukemie bij kinderen op geaggregeerd niveau (zoals PC4) naar verwachting te weinig onderscheidend vermogen zal hebben.



## DEEL 1 HOOFDRAPPORT



# 1 Inleiding

## 1.1 Aanleiding

GGD'en krijgen van tijd tot tijd vragen over het mogelijk verhoogd optreden van leukemie bij kinderen in een bepaalde buurt of wijk (clustermelding) waarbij de melder soms het vermoeden uitspreekt dat dit komt door lokale milieufactoren. In 2012 is de GGD-richtlijn 'Kankerclusters' (Hegger *et al.*, 2012) herzien, waarin de aanpak van clustermeldingen staat beschreven. De werkgroep die de betreffende richtlijn heeft opgesteld, heeft het Centrum Gezondheid en Milieu (cGM) van het RIVM gevraagd om onderzoek te doen naar de variatie in ruimte en tijd in de leukemie-incidentie in Nederland. Het RIVM heeft de afgelopen jaren in het SMARHAGT-onderzoeksprogramma expertise ontwikkeld voor het toepassen van ruimtelijke analysemethoden ten behoeve van milieugezondheidsonderzoek. Verzocht is om ook na te gaan in hoeverre deze onderzoeksmethoden en resultaten bruikbaar zijn bij de ondersteuning van de werkzaamheden van GGD'en bij clustermeldingen.

## 1.2 Leukemie bij kinderen

In Nederland en andere landen in Noordwest-Europa krijgen elk jaar 5 op de 100.000 kinderen leukemie. Het gemiddelde aantal nieuwe gevallen in Nederland bij kinderen van 0 tot 14 jaar lag recent rond de 140 per jaar (Gezondheidsraad, 2012). Uit de literatuur blijkt dat leukemie bij kinderen vaker clustert dan op basis van het toeval te verwachten is (Gezondheidsraad, 2012). De Gezondheidsraad baseert zich daarbij vooral op uitgebreid clusteronderzoek in Engeland door het Committee on Medical Aspects of Radiation in the Environment (COMARE, 2006). In Nederland zijn enkele geverifieerde verhogingen (clusters) geconstateerd van hematopoïëtische maligniteiten (kanker van bloed en beenmerg, waaronder leukemie en lymfomen). Dit had betrekking op de periode 1975-1989 (Mulder *et al.*, 1994) en 1988-2003 (Visser *et al.*, 2005). In het onderzoek van Visser werd in de onderzoekspopulatie 'Groot-Amsterdam' voor de onderzoeksperiode 1988-2003 een statistisch significante verhoging gevonden van hematopoïëtische maligniteiten ten opzichte van landelijke incidentiecijfers (SIR 1,12; 95% CI 1,05-1,19). SIR staat hierbij voor de gestandaardiseerde incidentieratio. Deze verhoging werd in belangrijke mate veroorzaakt door verhoogde incidentie van non-hodgkinlymfomen (NHL) (SIR 1,22; 95% CI 1,12-1,33) en acute lymfatische leukemie (ALL) (SIR 1,34; 95% CI 0,95-1,83). Aangezien toevalsbevindingen inherent zijn aan aandoeningen met een lage incidentie, zoals leukemie bij kinderen, is het van belang om gebruik te maken van methoden die corrigeren voor toevalsbevindingen.

Al lange tijd bestaat de vraag of lokale omgevingsfactoren een rol spelen in het ontstaan van leukemie. Vanwege het kleine aantal ziektegevallen is een dergelijk verband op lokaal niveau vrijwel nooit goed te onderzoeken. Randvoorwaarden voor verder (lokaal) onderzoek naar de invloed van (lokale) omgevingsfactoren zijn: minimaal vijf vergelijkbare ziektegevallen en een circa tien maal verhoogd risico (Neutra, 1990; Rothman, 1990). Geconcludeerd is dan ook dat clusters van zeldzame aandoeningen als leukemie eigenlijk alleen goed (inter)nationaal kunnen worden onderzocht, al dan niet in relatie met mogelijke risicofactoren (Gezondheidsraad, 2001 en 2012; Hegger *et al.*, 2012).



### 1.3 Doelstelling

Het onderzoek beoogt inzicht te bieden in de vraag of er sprake is van variatie van betekenis in ruimte en tijd van de incidentie van leukemie bij kinderen en jongeren (0-19 jaar) in Nederland. Daarbij moet de rapportage bruikbaar zijn voor veldpartijen zoals GGD'en, die met enige regelmaat moeten adviseren over (vermeende) clusters.

#### *Terminologie leeftijdsgroep*

In de rest van dit rapport spreken we van leukemie en kanker bij kinderen, waarmee de leeftijdsgroep van 0 tot 19 jaar wordt bedoeld.

### 1.4 Vraagstelling

De volgende vraagstellingen worden in dit onderzoek beantwoord:

1. Zijn er in Nederland leukemieclusters aan te tonen bij kinderen, in ruimte of tijd? Deze vraag heeft een aantal subvragen:
  - a. Hoe vaak komt de aandoening voor per gebied en is er aantoonbare variatie in ruimte en tijd (is er bijvoorbeeld een gradiënt van oost naar west of een toename in de tijd?)
  - b. Kunnen één of meer ruimtelijke clusters worden gedetecteerd met een verhoogd optreden van leukemie bij kinderen in relatie tot de algemene bevolking of een referentiegroep?
  - c. Kan er aanvullend op het ruimtelijke aspect clustering in tijd worden gedetecteerd?
  - d. Zijn er verschillen in de uitkomsten tussen de gehanteerde analysemethoden?
2. Op welke manier zijn de methoden en resultaten van dit onderzoek bruikbaar voor de GGD bij het omgaan met gemelde clusters?
3. Zijn de methoden bruikbaar voor landelijke monitoring van leukemie en/of andere aandoeningen?
4. Zijn de methoden geschikt om een koppeling te maken tussen het voorkomen van kankerclusters (en andere aandoeningen) en milieufactoren en is het zinvol om in vervolgonderzoek milieufactoren te betrekken?

### 1.5 Werkwijze

De vraag van de GGD is na binnenkomst bij het cGM met de onderzoekers en in overleg met de begeleidingscommissies verder afgebakend (zie paragraaf 1.5.1).

De benodigde data over de incidentie van kanker zijn aangevraagd bij het Integraal Kankercentrum Nederland (IKNL). Het verzoek om op individueel (geanonimiseerd) niveau postcodegegevens te leveren, is medio 2012 besproken in het College van Toezicht van IKNL. In juni 2012 heeft deze de data-aanvraag geaccordeerd, gezien het onderkende belang van het onderzoek. De onderzoekers die met de door het IKNL aangeleverde bestanden werken, hebben een geheimhoudingsverklaring ondertekend. Daarnaast is vastgesteld dat de onderzoeksresultaten niet op tot personen herleidbaar niveau mogen worden gerapporteerd. Dit is geborgd doordat een vertegenwoordiger van IKNL deel heeft uitgemaakt van de begeleidingscommissie van het onderzoek. Bevolkingsgegevens zijn verkregen van het Centraal Bureau voor de Statistiek (CBS). De data hebben een uitgebreide bewerking ondergaan om ze te kunnen koppelen op postcodeniveau. De data zijn vervolgens geanalyseerd met *disease mapping*-methoden om ruimtelijke patronen te bekijken en met Satscan

(*scananalyse*) om clusters te detecteren (zie hoofdstuk 2 voor een beschrijving van deze methoden).

1.5.1 *Begeleiding onderzoek*

Het onderzoek is begeleid door een wetenschappelijke begeleidingscommissie en een externe begeleidingscommissie met vertegenwoordigers van de GGD'en en IKNL. Door deze twee commissies is vanuit verschillende invalshoeken (bijvoorbeeld methodologisch en aansluiting bij de praktijk) gekeken naar het project- en analyseplan, de resultaten en de rapportage. Beide commissies zijn tweemaal bij elkaar gekomen. Daarnaast is in de voorbereiding advies ingewonnen bij een methodologisch adviseur van IKNL. De samenstelling en taken van de begeleidingscommissies is weergegeven in bijlage A.

1.5.2 *Rapportage van de bevindingen/leeswijzer*

Dit rapport beschrijft de methode, resultaten en conclusies van het onderzoek. Het rapport is in twee delen het rapport, hoofdstukken 1 t/m 5), in de achtergronden zijn technische details ten aanzien van databewerking en statistische analyse beschreven (deel 2, hoofdstukken 6 t/m 8). Er is gekeken gesplitst. In het hoofdrapport worden de hoofdzaken weergegeven (deel 1 van in hoeverre onderzoeksresultaten visueel inzichtelijk konden worden gemaakt, in overeenstemming met regelgeving en privacy aspecten en in afstemming met IKNL.



## 2 Onderzoeksopzet en onderzoeksmethode

Dit hoofdstuk beschrijft hoe het onderzoek is opgezet en uitgevoerd en welke gegevens daarvoor zijn gebruikt.

Allereerst wordt in paragraaf 2.1 een toelichting gegeven op de voor het onderzoek beschikbare gegevens en de dataverzameling. Paragraaf 2.2 beschrijft welke onderzoeksgroepen kunnen worden onderscheiden. In paragraaf 2.3 wordt een algemene inleiding gegeven over de statistische analyses, waarna in paragraaf 2.4 en 2.5 beide gehanteerde methoden nader worden toegelicht.

### 2.1 Dataverzameling

Voor de analyses zijn de volgende data gebruikt:

- Incidentiegegevens uit de Nederlandse kankerregistratie (NKR) op postcodeniveau voor leukemie en voor de geselecteerde referentieaandoeningen (zie paragraaf 2.1.1.).
- Bevolkingscijfers (leeftijd, geslacht) op hetzelfde postcodeniveau om rekening te kunnen houden met verwachte aantallen ziektegevallen in een bepaald gebied op basis van de bevolkingsopbouw (zie paragraaf 2.1.2).

#### 2.1.1 *Incidentiegegevens Kankerregistratie IKNL*

Bij IKNL is een aanvraag ingediend voor het verstrekken van de incidentiegegevens van leukemie bij kinderen (0-19 jaar) en twee referentiegroepen (hersentumoren en overige vormen van kanker) uit de NKR. De aanvraag heeft betrekking op de gehele periode waarvoor IKNL data beschikbaar heeft, te weten 1989-2010. De data zijn opgevraagd op het niveau van de viercijferige postcode (PC4), en waar mogelijk volledige postcode (PC6), van het adres van betrokkene op het moment van diagnose.

#### *Referentiegroepen*

Er zijn bij IKNL niet alleen gegevens over leukemie aangevraagd, maar ook over andere vormen van kanker bij kinderen. Dat biedt de mogelijkheid om na te gaan of eventuele clustering van leukemie kenmerkend is voor deze aandoening of eveneens optreedt bij andere diagnosegroepen. Kanker bij kinderen komt weinig voor. In overleg met IKNL is als eerste referentiegroep de groep kinderen met hersentumoren geselecteerd. Dit betreft een groep die net als leukemie een relatief hoge incidentie heeft, zodat deze aandoening als zelfstandige groep opgenomen kan worden. De tweede referentiegroep bestaat uit alle overige vormen van kanker bij kinderen. Die groep is dus behoorlijk heterogeen. Deze groep is geselecteerd omdat in de literatuur soms wordt beschreven dat clustering van kanker bij kinderen optreedt, ook zonder uitsplitsing naar diagnosegroepen (Amin *et al.*, 2010). In groep 3 zitten alle overige vormen van kanker, waaronder ook non-hodgkinlymfomen. Aangezien die mogelijk etiologisch verwant zijn aan groep 1 (leukemie) heeft IKNL ook de morfologiecode meegeleverd. Dat maakt het desgewenst mogelijk om aanvullende analyses uit te voeren. Aangezien de primaire vraagstelling zich richtte op leukemie, is niet op voorhand een combinatiegroep gemaakt van leukemie en non-hodgkinlymfomen. De verdeling van de beschikbare data over de hoofdgroepen staat weergegeven in Tabel 1.

Tabel 1. Onderzoeksgroep IKNL naar hoofdgroepen (en percentage van het totale bestand = 13.260 ziektegevallen over de periode 1989-2010)

Groep	Onderzoeksgroep	Aantal	Percentage
1	Leukemie (ALL, AML en overig)	3.561	27%
2 (ref)	Hersentumoren (Code C71)	1.474	11%
3 (ref)	Overige vormen van kanker	8.225	62%
TOT	Totale onderzoekspopulatie IKNL	13.260	100%

#### *Leeftijdsgroepen bij diagnose*

In overleg met de adviseur van IKNL zijn bij de data-aanvraag vier leeftijdsgroepen onderscheiden: 0-4 jaar, 5-9 jaar, 10-14 jaar en 15-19 jaar.

#### *Beschikbaarheid IKNL-incidentiegegevens op PC4- en PC6-niveau*

Van alle 13.260 ontvangen records met incidentiegevallen van kanker bij kinderen, waaronder leukemie, blijkt 44% voorzien van een zescijferige postcode (PC6) van het woonadres ten tijde van diagnose. Van de overige records (56%) is een viercijferige postcode (PC4) beschikbaar. Na inventarisatie van de data blijkt de feitelijke beschikbaarheid van gegevens op PC6-niveau te worden bepaald door zowel het incidentiejaar als de IKNL-regio. Integraal Kankercentrum Zuid (IKZ) heeft vanaf het startjaar van de Kankerregistratie (1989) complete data voor PC6. Voor het voormalig Integraal Kankercentrum Oost (IKO), Noord Nederland (IKN) en Limburg (IKL) geldt dit vanaf medio 1998. Data afkomstig van voormalig Integraal Kankercentrum Amsterdam (IKA), West Nederland (IKW), Rotterdam (IKR) en Stedendriehoek Twente (IKST) zijn vanaf 2005 grotendeels compleet voor PC6.

#### *Mogelijkheden incidentiecijfers samengevat*

Voor de gehele onderzoeksperiode zijn gegevens beschikbaar op PC4-niveau. Vanaf 2005 zijn voor grote delen van het land in principe ook landelijke analyses op PC6-niveau mogelijk. Voor 2005 en 2006 moet er daarbij wel rekening mee worden gehouden dat bij respectievelijk 16,2% en 8,8% van de postcodes het alfanumerieke deel ontbrak. Landelijke analyses op PC6-niveau zijn daardoor pas volledig landsdekkend mogelijk vanaf 2007. Dit is een te korte periode om bij een dergelijke lage incidentie zinvolle analyses te doen.

#### 2.1.2

#### *Bevolkingscijfers op postcodeniveau*

Vanaf 1995 (startjaar van de GBA<sup>1</sup>) zijn bij het RIVM voor heel Nederland van het CBS afkomstige bevolkingscijfers op PC4-niveau beschikbaar. Bij het CBS is navraag gedaan of het mogelijk was om ook voor de periode 1989-1994 bevolkingsgegevens op PC4-niveau te verkrijgen. Ondanks een gerichte zoekactie bleek dat niet het geval te zijn; deze zijn alleen te verkrijgen via de individuele gemeenten. Een dergelijke datavergaring viel buiten het bestek van dit onderzoek. Voor de periode vanaf 1995 had het CBS wel bevolkingscijfers op PC6-niveau beschikbaar waar we gebruik van konden maken.

<sup>1</sup> Gemeentelijke Basis Administratie

## 2.2 Implicaties van beschikbare data voor keuze onderzoeksgroepen

De onderzoeksperiode beslaat idealiter de gehele periode van de NKR (zie paragraaf 2.1). Dit betreft de jaren 1989-2010 (het laatste jaar waarvan data beschikbaar waren op het moment van data-aanvraag). Op grond van de beschikbaarheid van de bevolkingscijfers moet bij de analyses echter een onderscheid en keuze worden gemaakt tussen een aantal onderzoeksperiodes:

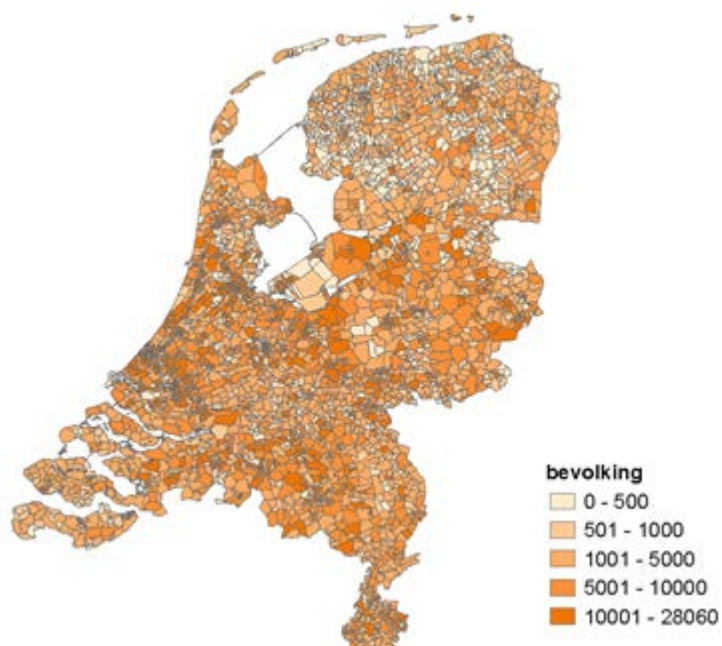
- Voor de periode voor 1989-1994 ontbreken landsdekkende cijfers over de *bevolkingsopbouw* op postcodeniveau.
- Voor de periode 1989-2004 ontbreken landsdekkende *incidentiecijfers* op PC6-niveau.
- Voor de periode 2005-2010 zijn zowel bevolkings- als incidentiecijfers beschikbaar op PC6-niveau, hoewel de laatste pas met ingang van 2007 volledig zijn.

Dit impliceert dat het voor de periode 1989-1994 niet mogelijk is om analyses uit te voeren met de algemene bevolking als vergelijkingsgroep, omdat voor deze periode geen bevolkingsgegevens op PC4-niveau beschikbaar zijn. Voor deze periode (en die daarna) is het in principe alleen mogelijk om de verschillen in incidentie te onderzoeken tussen de primaire onderzoeksgroep (leukemie) en beide IKNL-referentiegroepen. PC6-analyses zijn alleen landsdekkend mogelijk voor de periode 2005-2010. Aangezien ze pas met ingang van 2007 volledig beschikbaar zijn, maakte dat de onderzoeksperiode te klein om binnen dit onderzoek op PC6-niveau te analyseren.

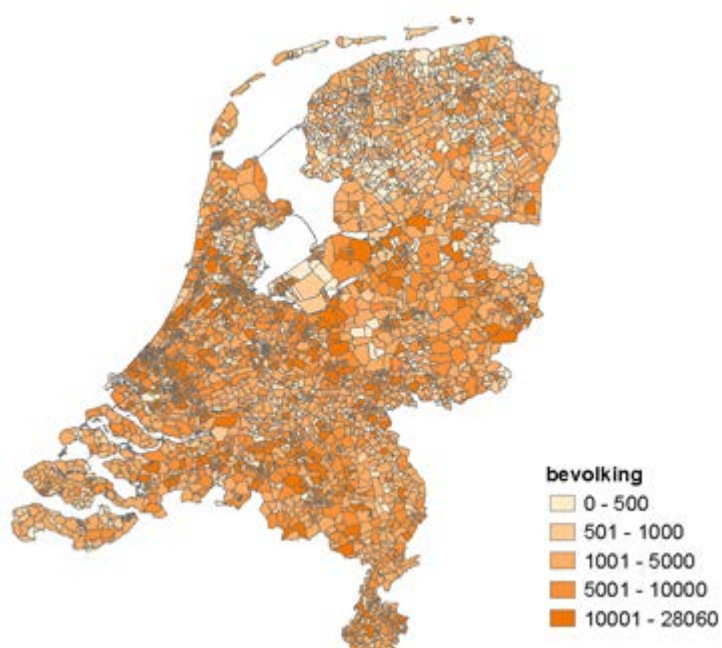
Op basis van de beschikbare data is ervoor gekozen om in dit onderzoek te analyseren op PC4-niveau. Hiermee is analyse mogelijk voor een zo groot mogelijke onderzoeksperiode voor heel Nederland.

## 2.3 Datamanagement

*In dit onderzoek wordt de analyse van de leukemie-incidentie uitgevoerd in ruimte (heel Nederland) en tijd (de periode 1995-2010 of een deel daarvan). Het toewijzen van een leukemiecasse aan een jaar is eenvoudig, maar het toewijzen van een case aan een ruimtelijke eenheid is minder eenvoudig te realiseren. Het ruimtelijke aggregatieniveau dat gekozen wordt, is deels verantwoordelijk voor de gevoeligheid van de statistische analyse en de clustersignalen die kunnen worden opgepikt. Aan de hand van de beschikbare informatie over de leukemiecasses is in eerste instantie gekozen om clustervorming te onderzoeken op PC4-niveau (bijvoorbeeld: 1184). Gemiddeld wonen er in een PC4-gebied ongeveer 4.400 mensen, maar de aantallen lopen enorm uiteen, van enkele honderden tot ruim 25.000 inwoners. De verschillen in aantal inwoners per postcode zijn weergegeven in*



Figuur 1.



Figuur 1. Bevolking per viercijferig postcodegebied in 2010. Bron: CBS

Deze grote spreiding geeft aan dat het verwachte aantal leukemiecases grote verschillen zal vertonen tussen PC4-gebieden. Om de geobserveerde aantallen leukemiecases tussen PC4-gebieden te kunnen vergelijken is het noodzakelijk om informatie over de bevolking naar leeftijd en geslacht aan de PC4-gebieden te koppelen. Deze informatie was voor dit project beschikbaar vanaf 1995.

De postcodekaart van Nederland is niet constant in de tijd: nieuwe postcodegebieden worden gevormd binnen of over de grenzen van oude

gebieden, de grenzen van gebieden worden verlegd om de postbezorging efficiënter te laten verlopen en af en toe worden gebieden opgeheven. Deze 'temporal boundary misalignment' compliceert een adequate analyse van de leukemie-incidentie, doordat zowel de onderzoeksgroep als de referentiegroep in de loop van de tijd aan net andere gebieden moeten worden gekoppeld. Ten behoeve van de analyse is uitgegaan van één basispostcodekaart, de kaart van 2010. Vervolgens zijn de bevolkingsdata, kankerincidentiegegevens en eventuele covariaten toegewezen aan de afzonderlijke postcodegebieden. Dit vereist dat de veranderingen in de postcodegebieden worden vastgelegd en dat de data van de verschillende jaren worden herverdeeld over de postcodegebieden van de basiskaart.

Het proces om de leukemiecases en de bevolking toe te wijzen aan de postcodegebieden wordt uitgebreid beschreven in deel 2 van dit rapport (hoofdstuk 6). Daarbij wordt ook beschreven welke mogelijkheden er zijn om de leukemiecases en de bevolkingsgegevens op andere ruimtelijke aggregatieniveaus te analyseren. Daarbij gaat het onder meer om analyse op PC6-niveau en op 500 bij 500 meter gridniveau.

## 2.4 Statistische analyse

Het RIVM heeft de afgelopen jaren in het SMARHAGT-onderzoeksprogramma expertise ontwikkeld voor het toepassen van geavanceerde ruimtelijke analysemethoden ten behoeve van milieugezondheidsonderzoek. Daarbij is ook de onderzoeksliteratuur op dit vlak gescreend. Daaruit kwamen *disease mapping* (Elliott *et al.*, 2000; Lawson, 2009) en *scan-analyses* zoals de Spatial Scan Statistic (Kulldorff, 1997, 2009) naar voren als geschikte methoden om de spreiding van ziekte in ruimte en tijd te onderzoeken. Deze methoden zijn daarom toegepast bij de beantwoording van de vragen van de GGD'en. Hieronder worden de methoden kort toegelicht; meer uitleg en informatie over de gebruikte modellen is te vinden in deel 2 van dit rapport (hoofdstuk 7).

### 2.4.1 *Disease mapping*

Met behulp van *disease mapping* worden lokale incidenties en de variatie daarvan in ruimte en tijd geschat met behulp van een statistisch model. Zo kan worden bepaald waar incidenties afwijken van verwachte waarden. Daarbij wordt gebruikgemaakt van een vorm van 'smoothing', waarmee wordt gecorrigeerd voor verschillen in variatie die worden veroorzaakt door verschillen in lokale bevolkingsaantallen en bevolkingsopbouw. De incidentiemaat die daarbij wordt gebruikt, is de Standardized Incidence Ratio (SIR).

Er zijn drie verschillende modellen gedraaid:

- basismodel: incidentieschatting per postcodegebied;
- model met toegevoegd ruimtelijk effect;
- model met ruimtelijk effect én tijdseffect.

Met behulp van de diagnostische score 'Deviance Information Criterion' (DIC) wordt bepaald of het model verbetert (de uitkomstmaat wordt beter verklaard) met het toevoegen van bijvoorbeeld het ruimte-effect: de DIC wordt dan lager. Als de DIC niet lager wordt, is er geen ruimtelijk effect aangetoond.

### 2.4.2 *Scan-analyse*

Met behulp van een ruimte- of tijd/ruimte-*scan-analyse* kan worden onderzocht of en waar patronen van clustering in ruimte en tijd optreden ten opzichte van een referentiepopulatie of de verwachte waarden, rekening houdend met de



bevolkingsopbouw. Deze methode kan ook clusters detecteren die groter zijn dan het gekozen analyseniveau (bijvoorbeeld een postcodegebied).

Clusteranalysemethoden zoals scan-analyse blijken een belangrijk hulpmiddel bij de monitoring van het optreden van kanker en andere aandoeningen, om regio's te identificeren met een verhoogd risico en eventueel ten behoeve van het genereren van hypothesen over etiologie (Wheeler, 2007). Belangrijk voordeel van deze methoden is dat toetsing mogelijk is, in tegenstelling tot post-hoctoetsing bij een opgevallen (gemeld) cluster. Dit komt doordat er bij de analyse rekening wordt gehouden met 'preselection bias' en correctie voor meervoudig toetsen (Kulldorff, 1997). Scan-analyses blijken in het algemeen voldoende sensitief te zijn bij zeldzame aandoeningen, zoals leukemie (Amin *et al.*, 2010). Dat is een belangrijke randvoorwaarde voor de toepassing ervan, aangezien clustermeldingen juist bij zeldzame aandoeningen relatief vaak voorkomen. Het toepassingsgebied is echter breder dan zeldzame aandoeningen als leukemie.

Uit de literatuur (Tango, 2010) is gebleken dat het wenselijk kan zijn om de traditionele methode van circulaire scan-'windows' aan te vullen met methoden die flexibeler zijn in de detectie van clusters met een andere vorm (ellips scan of flexibele scan).

Voor dit onderzoek zijn daarom diverse typen *scananalyses* toegepast:

- Ruimtelijke clustering, zowel met cirkelvormige, ellipsvormige als flexibele clusters (FlexScan);
- Ruimte-tijd gecombineerde clustering (SatScan).

De *scan-analyses* leveren als uitkomst een 'Most Likely Cluster' (MLC). Dit is een gebied, bestaande uit één of meer aangesloten gebieden, waarbij de kans het grootst is dat daar clustering optreedt. In aanvulling daarop kunnen ook zogenaamde secundaire of mogelijke clusters worden onderscheiden. Dat zijn andere gebieden dan de MLC waar eveneens een verhoogde kans is op clustering, hoewel ze minder waarschijnlijk zijn dan de MLC. De feitelijke toetsingswaarde bepaalt of de MLC of de secundaire clusters ook statistisch significant afwijken.

#### 2.4.3

##### *Overeenkomsten en verschillen tussen de methoden*

*Disease mapping* en *scan-analyse* zijn twee varianten om de vraagstellingen te kunnen beantwoorden over de verdeling van de incidentie van leukemie in ruimte en tijd. Beide methoden bieden ook de mogelijkheid de eventuele invloed van risicofactoren en andere covariaten (waaronder leeftijd en geslacht) te analyseren. Elke methode heeft daarnaast specifieke accenten en mogelijkheden (zie onder andere Aamodt *et al.*, 2006; Kulldorff, 1997).

*Disease mapping* schat het incidentieniveau per afzonderlijk – vooraf gekozen - gebied (postcodegebied, gemeente, provincie). Uit het kaartbeeld van de incidentieschattingen blijkt dan of er clustering van verhoogde niveaus over meerdere gebieden plaatsvindt. Met *disease mapping* kan ook worden bekeken in hoeverre er sprake is van ruimtelijke en ruimte-tijd-variatie in de incidentie.

*Scan-analyse* is zowel geschikt voor monitoring van de variatie in ruimte en tijd bij bepaalde aandoeningen als om te onderzoeken of gemelde en vermeende clusters daadwerkelijk verhoogd voorkomen. Scan-analyse is daarbij flexibel bij het selecteren van gebieden of perioden waartussen mogelijk variatie in het optreden van een aandoening bestaat; deze methode inventariseert óf er aparte of gegroepeerde gebieden zijn die wat incidentieniveau betreft significant afwijken van de andere gebieden, nadat is gecorrigeerd voor meervoudig

toetsen. *Scan-analyse* kan detecteren of er clusters zijn in *niet* vooraf gedefinieerde gebieden binnen een heel land of regio (proactief). Met *scan-analyse* kan ook gekeken worden naar de significantie van een clustering in een specifiek gebied na een clustermelding (reactief). De omvang van de meest waarschijnlijke clusters wordt bepaald door eventuele patronen in de dataset zelf en kan beduidend groter zijn dan het oorspronkelijke analyseniveau van de data (in dit onderzoek het postcodeniveau). *Scan-analyse* biedt dus meer mogelijkheden om de meest waarschijnlijke en de significante clustergebieden in ruimte en tijd af te bakenen.



## 3 Resultaten analyses

### 3.1 Disease mapping

In deze paragraaf staan de resultaten beschreven van de analyses met behulp van het bayesiaanse conditioneel autoregressiemodel (CAR) van Besag (Botella-Rocamora *et al.*, 2013) op PC4-niveau. Voor deze analyse zijn drie modellen gedraaid (zie paragraaf 2.4) voor vier uitkomstgroepen en zes leeftijdscategorieën. Elk model is dus 24 keer uitgevoerd.

#### 3.1.1 Ruimte- en tijdseffecten

Voor 19 van de 24 ruimtelijke modellen blijkt dat de score van de DIC niet beduidend (minstens 5 punten) lager wordt ten opzichte van het basismodel (zie Tabel 2). Dit betekent dat toevoegen van een ruimtelijk effect bij de meeste modellen geen beter model oplevert, met andere woorden: er is meestal geen belangrijk ruimtelijk effect aanwezig in de gemodelleerde respons. Voor de toevoeging van een tijdseffect geldt min of meer hetzelfde. Dit levert bij een van de 24 modellen een relevante (minstens 5 punten) verlaging van de DIC-score op. Voor leukemie kan worden gesteld dat er geen ruimtelijk effect wordt gedetecteerd. De relevante veranderingen van de DIC bevinden zich in groepen met de hoogste incidentie (alle respectievelijk overige kankersoorten) en niet in die van alleen leukemie.

Tabel 2. Overzicht van de DIC-waarden bij de modellen (*cursief geeft verbetering ten opzichte van eenvoudiger model aan - minstens 5 punten verlaging*)

Diagnosegroep	Leeftijd	Model1: basis	Model 2: ruimtelijk	Model 3: ruimte-tijd
Alle kankersoorten	0-19 jaar	27122,98	27113,05	27118,27
	0 jaar	4899,12	490704	4907,23
	1-4 jaar	12279,38	12280,58	12285,29
	5-9 jaar	9578,37	9575,89	9581,74
	10-14 jaar	9967,75	9961,78	9967,74
	15-19 jaar	14644,9	14644,21	14648,72
G1: Leukemie	0-19 jaar	12957,2	12958,8	12964,25
	0 jaar	1388,44	1388,15	1394,07
	1-4 jaar	6979,47	6980,15	6986,39
	5-9 jaar	4780,43	4780,89	4786,81
	10-14 jaar	3746,9	3747,43	3753,43
	15-19 jaar	3240,25	3241,21	3247,18
G2: Hersentumoren	0-19 jaar	6758,4	6759,92	6765,92
	0 jaar	660,47	660,06	665,94
	1-4 jaar	2686,1	2685,77	2691,79
	5-9 jaar	2667,73	2668,79	2674,83
	10-14 jaar	2080,41	2080,45	2086,41
	15-19 jaar	1699,43	1699,21	1705,17
G3: Overig	0-19 jaar	21392,67	21386,74	21391,94
	0 jaar	3715,79	3722,83	3716,36
	1-4 jaar	7115,93	7119,38	7126,8
	5-9 jaar	5474,89	5463,27	5469,08
	10-14 jaar	7267,6	7262,33	7268,29
	15-19 jaar	13041,62	13040,69	13044,98

### 3.1.2 *Verkregen SIR-schattingen*

Gezien de gehanteerde betrouwbaarheidsgrens van 95% zou alleen al door toeval bij ongeveer tweehonderd gebieden per model een statistisch significante afwijking van de SIR worden verwacht (24 modellen). Met het ruimtelijke model werd in totaal slechts 38 maal een significante verhoging van de SIR in een gebied aangetroffen, uit meer dan 96.000 SIR-schattingen (ruim 4000 gebieden in 24 modellen). Er werden geen statistisch significante verlagingen aangetroffen. Dit duidt erop dat de incidentie te laag is om in dit model een ruimtelijk patroon te kunnen aantonen. Bij deze lage incidenties is het ook te verwachten dat in een Poisson-model geen verlagingen worden gevonden.

Met het ruimte-tijdmodel werd in totaal slechts 42 maal een significante verhoging aangetroffen, uit vier keer meer SIR-schattingen dan voor het ruimtelijk model (384.000). Er werden geen significante verlagingen aangetroffen. Gezien de gehanteerde betrouwbaarheidsgrens van 95% zou alleen al door toeval bij ongeveer achthonderd gebieden per model een statistisch significante afwijking van de SIR worden verwacht (24 modellen). Dit alles maakt dat het niet zinvol werd geacht te proberen eventuele trends in de data aan te tonen.

Uit de *disease mapping*-analyses blijkt daarom, over heel Nederland bezien, dat de variatie tussen PC4-gebieden verwaarloosbaar is. Het aantal gebiedjes waar lokaal significante verhogingen werden gevonden lag ver onder het aantal dat op basis van de gekozen significantiegrens door toeval mag worden verwacht.

#### *Opmerking*

Te recent voor opname in dit rapport is bij het RIVM de rekencapaciteit beschikbaar gekomen om ook de afzonderlijke jaren als tijdseffect in Bayesiaanse modellen op te kunnen nemen (we konden alleen voor perioden van vier jaar rekenen). Een van de modellen is nog wel als proef op die manier gedraaid, dit leverde geen nieuwe inzichten. Er is geen reden om aan te nemen dat het bij de gebleken lage incidenties voor deze data zin heeft ook de resterende modellen zo uit te voeren. Bij een volgende toepassing is het wel gewenst om per jaar te analyseren, om een zo hoog mogelijke power te verkrijgen.

## 3.2 **Scananalyse (clusterdetectie)**

### 3.2.1 *Ruimtelijke Clustering SatScan*

Gezien de lage incidentie in de primaire onderzoeksgroep (leukemie bij kinderen) zijn de analyses in eerste instantie uitgevoerd voor de gehele periode 1995-2010. Het gaat daarbij om 2670 personen binnen 4042 postcodes. Vier postcodes die geen aangrenzend postcodegebied hebben, zijn buiten de analyses gelaten.

#### *Cirkelvormige clusters*

Het meest waarschijnlijke cluster (MLC) werd gevonden in het midden van het land en bestond uit zeven aansluitende postcodegebieden. Het relatieve risico (RR) bedroeg 2,85 met daarbij een 'Log Likelihood Ratio' van 7,14. Deze bleek echter niet statistisch significant ( $p=0,77$ ).

In aanvulling op het primaire cluster werd een aantal andere zogenaamde secundaire ('mogelijke') clusters gevonden met relatieve risico's tussen 1,4 en 8,2. Na toetsing vallen deze mogelijke clusters echter alle binnen het normale betrouwbaarheidsinterval en zijn daarom niet significant (kunnen op toeval gebaseerd zijn). Dat gold eveneens voor een leukemiecluster waarvan melding

is gedaan bij een GGD en die door de GGD is onderzocht. Dit cluster werd wel geïdentificeerd als potentieel secundair cluster (evenals de melding beperkt tot één postcode), maar ook deze bevinding bleek na correctie voor meervoudige toetsing niet statistisch significant verhoogd te zijn.

#### *Ellipsvormige clusters*

Ook uit de analyse met de ellipsvormige scan bleek dat geen van de gevonden regio's met potentiële ruimtelijke clustering statistisch significant afweek van de natuurlijke variatie die kan optreden bij zeldzame aandoeningen.

#### 3.2.2 *Ruimtelijke clustering FleXScan*

De FleXScan-analyses leidden op hoofdlijnen tot dezelfde uitkomsten als de SatScan-analyse. De SatScan-analyse maakt ook onderdeel uit van het pakket FleXScan. Het toepassen van de aangepaste 'Likelihood Ratio'-test heeft niet geleid tot substantiële verschillen met de uitkomsten uit de SatScan-Analyse.

#### 3.2.3 *Clusteranalyse in ruimte én tijd (SatScan)*

Uit de ruimte-tijdanalyse met perioden van vier jaar in het model kwamen vijf secundaire of mogelijke clusters naar voren, waarvan drie in de laatste onderzoeksperiode (2007-2010). Het gebied waar de meeste clustering optrad (het gebied met het 'Most Likely Cluster') betrof een groep postcodes binnen een straal van twintig kilometer. In de onderzoeksperiode werden hier 46 cases gevonden. Het relatief risico bedroeg 2,4, hetgeen bij een 'Log Likelihood' van 12,95 statistisch significant was ( $p=0,03$ ), ook na correctie voor meervoudig toetsen. De overige gevonden mogelijke clusters bleken niet statistisch significant.

Uit de ruimte-tijdanalyse met alle jaren apart in het model werd een 'Most Likely Cluster' gevonden in een periode van twee jaar (2005-2006) met 29 cases ( $RR=3,2$ ). Deze bevinding was niet significant ( $p=0,17$ ).

#### 3.2.4 *Scananalyses samengevat*

In de ruimte-tijdanalyse werd één statistisch significant leukemiecluster gevonden in een gebied met een straal van twintig kilometer. De periode waarin de verhoogde incidentie van leukemie optrad, was beperkt tot een viertal jaren. Doordat het een betrekkelijk klein cluster ( $n= 46$ ) gedurende een relatief korte periode is, werd bij analyse over de gehele onderzoeksperiode geen significant resultaat gevonden bij de ruimtelijke analyses.

### 3.3 **Vergelijking disease mapping en scan-analyse**

De resultaten van de ruimtelijke analyses met *disease mapping* en *scan-analyse* bleken op hoofdlijnen vergelijkbaar, ondanks de verschillen in de methodiek.

### 3.4 **Vergelijkbaarheid uitkomsten met bestaande clustermeldingen**

#### 3.4.1 *Gerapporteerd cluster*

De uitkomsten uit de *scananalyses* zijn vergeleken met de rapportage over een lokaal clusteronderzoek waarbij de methodiek is gevolgd van de richtlijn 'Kankerclusters' voor GGD'en. Uit dit lokale clusteronderzoek concludeert de GGD dat in een bepaalde woonwijk in regio Rotterdam meer kinderen leukemie hebben gekregen dan was te verwachten (Hegger *et al.*, 2013). Gezien de zeer kleine onderzoekspopulatie is geconstateerd dat statistische toetsing niet verantwoord is en dat de bevinding ook op toeval kan berusten. Ook in de *scananalyse* over de gehele beschikbare onderzoeksperiode blijkt het betreffende postcodegebied voor te komen in de gedetecteerde mogelijke clusters (MLC). Dit vormt een bevestiging van het signalerend vermogen van

bevolking en de GGD op lokaal niveau. Ook bij de tijd-ruimteanalyses komen de bevindingen van de GGD terug in de landelijke analyse over dezelfde onderzoeksperiode. Voordeel van de uitgevoerde landelijke (en dus niet signaalgestuurde) analyse is dat de bevindingen beter kunnen worden getoetst. De lokale bevinding bleek niet statistisch significant verhoogd te zijn en kan daardoor ook niet als geverifieerd cluster worden aangemerkt.

#### 3.4.2 *Clustermeldingen in GGD-registratie*

De GGD'en registreren alle meldingen die zij over milieu en gezondheid krijgen in een landelijk systeem. Hierin zijn ook clustermeldingen opgenomen. Deze gegevens worden tweejaarlijks gerapporteerd door het RIVM (onder andere Dusseldorp *et al.*, 2013). In de landelijke database staat de datum van melding, de GGD waar het cluster is gemeld, het onderscheid naar kanker- en ziektecluster, en of de melder een bepaalde oorzaak vermoedt. Het type kanker is niet opgenomen in de database. Vaak weten de melders dat ook niet precies. Om de onderzoeksresultaten te vergelijken met deze meldingen is nadere informatie van de GGD'en nodig die een clustermelding hebben ingevoerd. Dat viel buiten de reikwijdte van dit project.

## 4 Mogelijkheid tot koppeling met milieufactoren

In een advies van de Gezondheidsraad wordt het toepassen van geavanceerde methoden voor clusteronderzoek, al dan niet geïnitieerd op grond van een bevestigd cluster, nuttig bevonden voor landelijk onderzoek naar (oorzaken voor) clustering van betrekkelijk zeldzame aandoeningen, zoals leukemie bij kinderen. Dergelijk onderzoek heeft volgens de Gezondheidsraad meer zin dan het ter plaatse instellen van een epidemiologisch onderzoek naar de invloed van lokale milieufactoren die ook elders voorkomen (Gezondheidsraad, 2001; Hegger *et al.*, 2012).

Het huidige onderzoek richtte zich op de vraag of er sprake is van clustering in tijd en/of ruimte van leukemie in Nederland, zonder daarbij rekening te houden met potentiële risicofactoren. De GGD'en zijn geïnteresseerd in een antwoord op de vraag of er in een vervolgonderzoek met de gebruikte analysemethoden al dan niet een relatie gelegd kan worden met milieufactoren, in het bijzonder hoogspanningslijnen en nieuwbouwwijken. Ten behoeve daarvan wordt in dit hoofdstuk ingegaan op de volgende drie vragen:

- Welke (milieugerelateerde) risicofactoren worden in de literatuur beschreven voor leukemie bij kinderen?
- Is het, mede op geleide van de uitkomsten van het huidige en vergelijkbaar onderzoek, zinvol om in Nederland aanvullend onderzoek uit te voeren waarbij ook milieufactoren (potentiële risicofactoren) in beschouwing worden genomen?
- Zo ja, welke risicofactoren komen daarvoor in aanmerking en welke randvoorwaarden (zoals het schaalniveau) zijn hierbij van belang?

### 4.1 Leukemie in relatie tot omgevingsfactoren

Binnen het samenwerkingsverband van het European Science Advisory Network for Health (EuSANH) hebben de Nederlandse Gezondheidsraad en de Belgische Hoge Gezondheidsraad eind 2012 de wetenschappelijke kennis in kaart gebracht over een breed scala van omgevingsfactoren die mogelijk gerelateerd zijn aan leukemie bij kinderen. Jaarlijks wordt bij circa 140 kinderen in Nederland en tachtig kinderen in België leukemie geconstateerd. In het laatste decennium van de twintigste eeuw liet het aantal gevallen van kinderleukemie een stijging zien, wat aanleiding was voor de vraag of blootstelling aan bepaalde omgevingsfactoren daarin een rol zou kunnen spelen. Hoewel de trend inmiddels afgevlakt lijkt te zijn, en misschien zelfs gekeerd, was deze vraag aanleiding voor de risicoanalyse en aanbevelingen van de Gezondheidsraad van beide landen gezamenlijk. Daarbij hebben zij zich laten leiden door het voorzorgprincipe. Dit betekent dat zij suggesties hebben gedaan om, waar dat kan, de blootstelling aan mogelijk schadelijke omgevingsfactoren te beperken.

Over het verband tussen omgevingsinvloeden en kinderleukemie is vooralsnog veel onduidelijk of betrekkelijk onzeker. Alleen in het geval van ioniserende straling is een verband met kinderleukemie beoordeeld als 'aangetoond'. Een verband tussen kinderleukemie en blootstelling aan benzeen is beoordeeld als 'waarschijnlijk'. Een verband met roken door de ouders en met blootstelling aan bestrijdingsmiddelen en bepaalde andere chemische stoffen, zoals PCB's, is beoordeeld als 'mogelijk tot waarschijnlijk'. Voor alle andere bestudeerde omgevingsfactoren is het bestaan van een verband beoordeeld als 'mogelijk' (bijvoorbeeld voor elektromagnetische velden (EMV) in de buurt van



hoogspanningslijnen, en nieuwbouwwijken), 'onzeker' (bijvoorbeeld voor infecties) of 'onbekend' (voor onder meer veel chemische stoffen).

Tabel 3 geeft een selectie uit de tabel van de Gezondheidsraad van de omgevingsfactoren die een mogelijke rol spelen bij het ontstaan van leukemie bij kinderen en die in aanmerking zouden kunnen komen voor verder onderzoek met de analysemethoden die in dit rapport beschreven worden. Het gaat hierbij om omgevingsfactoren die door de Gezondheidsraad als 'mogelijke' of 'waarschijnlijke' risicofactoren voor leukemie bij kinderen worden aangemerkt, niet samenhangen met persoonlijke kenmerken (zoals roken) en waarvoor een milieublootstelling zou kunnen worden gedefinieerd. De gehele tabel met risicofactoren zoals opgenomen in de rapportage van de Gezondheidsraad staat opgenomen in Bijlage C.

*Tabel 3. Selectie van omgevingsfactoren als mogelijke risicofactor voor het ontstaan van leukemie bij kinderen. Bron: Gezondheidsraad, 2012.*

Omgevings-factor	Waarschijnlijkheid van causaliteit	Mogelijke blootstelling	Mogelijke indicator blootstelling
Elektro-magnetische velden	Mogelijk	Extreem laagfrequente elektromagnetische velden	Afstand/veldsterkte hoogspanningslijnen
Chemicaliën	Mogelijk tot waarschijnlijk	Bestrijdingsmiddelen	Afstand tot (type) landbouwgrond
Nieuwbouw-wijken	Mogelijk	Vluchtige Organische Stoffen; Population mixing	Locatie nieuwbouwwijken

## 4.2 Nut van aanvullend onderzoek inclusief milieufactoren

Voor elektromagnetische velden en bestrijdingsmiddelen is oriënterend bekeken wat de mogelijkheden zijn om nader onderzoek te verrichten met de in dit rapport beschreven statistische analysemethoden<sup>2</sup>. De volgende vragen zijn daarbij in ogenschouw genomen:

- Is het aantal kinderen dat wordt blootgesteld aan de genoemde omgevingsfactor voldoende groot om het verband te onderzoeken?
- Is het mogelijk om een blootstellingsindicator te formuleren die de blootstelling op het gekozen aggregatieniveau voor de beschouwde tijdsperiode kan specificeren?
- Zijn er andere factoren die van invloed zijn op de beoordeling van de zinvolheid van aanvullend onderzoek?

### 4.2.1 Elektromagnetische velden

Extreem laagfrequente elektromagnetische velden (ELF-EMV) zijn de meest bekende veronderstelde risicofactor voor leukemie bij kinderen. Het mogelijke verband tussen leukemie bij kinderen en blootstelling aan ELF-EMV is voornamelijk onderzocht rond bovengrondse hoogspanningslijnen. Meta-analyses van de beschikbare studies geven aan dat kinderen die langdurig worden blootgesteld aan een jaargemiddeld magneetveld hoger dan 0,4 microtesla een verhoogd risico lopen op het ontstaan van leukemie. Een

<sup>2</sup> In tabel 3 staan ook nieuwbouwwijken genoemd als omgevingsfactor. Ook hiervoor zou goed moeten worden gekeken naar de (on)mogelijkheden en zinvolheid.

causaal verband is echter nog niet aangetoond (Gezondheidsraad, 2012). Op basis van de beschikbare wetenschappelijke kennis hanteert de Nederlandse overheid het voorzorgsprincipe en heeft besloten om rond hoogspanningslijnen een magneetveldzone te laten berekenen waarbinnen het jaargemiddelde van 0,4 microtesla wordt overschreden. Gemeenten hebben een aantal beperkingen binnen deze zones om te voorkomen dat kinderen langdurig worden blootgesteld (Kelfkens *et al.*, 2003). Onzekerheid over het causale verband wordt daarbij zo goed mogelijk ondervangen met maatregelen om blootstelling te minimaliseren.

Op basis van informatie over de mogelijke blootstelling in Nederland (Van de Plas *et al.*, 2001, Tabel 4) heeft het RIVM indertijd een oriënterende poweranalyse uitgevoerd (persoonlijke mededeling Houthuijs, 2014). Destijds is aangenomen dat kinderen woonachtig bij magnetische velden hoger dan 0,2 microtesla een incidentieratio voor leukemie van 1,65 hebben ten opzichte van kinderen die zijn blootgesteld aan velden onder de 0,2 microtesla. Uitgaande van een onderzoeksperiode van 25 jaar en heel Nederland als onderzoeksgebied zou met een (retrospectief) cohortonderzoek als onderzoeksopzet een relatief risico van circa 1,5 tot 2,2 kunnen worden gedetecteerd. Voor het detecteren van een risico van 1,5 zouden circa 34.000 en voor een risico van 2,2 circa 14.000 kinderen aan het verhoogde risico moeten zijn blootgesteld. Voor een case-controlonderzoek als onderzoeksopzet zijn eveneens oriënterende berekeningen uitgevoerd. Een case-controlonderzoek heeft minder onderscheidend vermogen om een 'odds ratio' van 1,5 vast te stellen dan een cohortonderzoek. Een cohortonderzoek met geheel Nederland als onderzoeksgroep lijkt incidentieratio's te kunnen detecteren die in dezelfde orde van grootte liggen als in de literatuur worden gerapporteerd, maar de uitvoering ervan is sterk afhankelijk van de beschikbaarheid van goede informatie over de blootstelling op het woonadres, mogelijke verstorende variabelen en het verhuisgedrag. Ook is dergelijk onderzoek vanuit privacy-oogpunt complex. Er werd daarom destijds geconcludeerd dat de uitkomsten van de oriënterende powerberekeningen dergelijk onderzoek binnen de Nederlandse context niet ondersteunen.

De analysemethodieken die in dit rapport zijn beschreven worden in ecologisch onderzoek toegepast. In dit type onderzoek wordt het aantal leukemiecases op geaggregeerd niveau geanalyseerd. De mogelijke aggregatieniveaus zijn het PC4-niveau, PC6-niveau en het CBS-grid van 500 x 500 meter. Bij onderzoek op PC4-niveau krijgen alle personen binnen het PC4-gebied dezelfde potentiële blootstelling aan het elektromagnetisch veld toegewezen. Dit heeft een grote mogelijke misclassificatie van de blootstelling in de omgeving van de hoogspanningslijnen tot gevolg, zowel op het PC4 als op het 500 x 500 m-gridniveau. De blootstellingsmisclassificatie kan voor een deel ondervangen worden door statistische analyse op PC6-niveau (betreft 10-15 adressen) uit te voeren, maar de verwachting is dat, vooral in landelijk gebied, er nog steeds een grote misclassificatie van de veldsterkte kan bestaan.

De consequentie van deze misclassificatie is dat het vermogen van studies om een bepaald relatief risico te detecteren terugloopt. Het is daarom aannemelijk dat een ecologisch onderzoek minder 'power' zal hebben dan een cohortonderzoek. De relatieve risico's die kunnen worden gedetecteerd zullen hoger liggen dan de hierboven beschreven range van 1,5 tot 2,2. De uitvoering van een onderzoek met de in dit rapport gepresenteerde methodieken ligt daarmee niet voor de hand.

Wanneer de uitvoering van een onderzoek naar de samenhang tussen de veldsterkte van hoogspanningslijnen en de incidentie van leukemie onder kinderen wenselijk zou zijn, ligt het in de rede om allereerst de veldsterkte van hoogspanningslijnen in geheel Nederland te modelleren. Vervolgens kan de blootstellingspopulatie worden vastgesteld en kan een 'update' van de powerberekening worden uitgevoerd, waarna beoordeeld kan worden of een (retrospectief) cohortonderzoek of case-controlonderzoek mogelijk is. Daarnaast zou verkend kunnen worden of het mogelijk is uit verschillende registratiesystemen een cohort voor geheel Nederland samen te stellen met leukemiecasses uit de Kankerregistratie en waarin het verhuisgedrag van personen op adresniveau kan worden gevolgd. Dit laatste is in principe vanaf 1994 met CBS-gegevens mogelijk.

De conclusie is daarom vooralsnog dat het, op basis van power en het belang van statistische analyse op fijnschalig aggregatieniveau (bij voorkeur adresniveau), nu niet zinvol is om op dit vlak onderzoek in Nederland te starten. Eventuele activiteiten zouden zich eerst moeten richten op het vergaren van informatie om te kunnen beoordelen of een dergelijk onderzoek mogelijk zou zijn (modellering van veldsterktes, verkenning van koppeling en nieuwe powerberekeningen). De GR-commissie heeft daarnaast een pleidooi gehouden om nader etiologisch onderzoek bij voorkeur in internationaal verband te laten plaatsvinden.

#### 4.2.2 *Bestrijdingsmiddelen*

Op basis van literatuuronderzoek van de Gezondheidsraad (2012) in de rapportage over leukemie bij kinderen en omgevingsfactoren wordt een relatie tussen leukemie en (vooral beroepsmatige) blootstelling aan bestrijdingsmiddelen beoordeeld als 'mogelijk tot waarschijnlijk'.

In het advies van de Gezondheidsraad over gewasbeschermingsmiddelen en omwonenden wordt eveneens geconcludeerd dat buitenlands onderzoek naar mogelijke gezondheidseffecten bij omwonenden enige aanwijzingen oplevert dat bepaalde aandoeningen, waaronder leukemie bij kinderen, wellicht vaker voorkomen bij omwonenden van landbouwpercelen die met gewasbeschermingsmiddelen zijn behandeld (Gezondheidsraad, 2014). Bij beroepsmatige blootstelling is deze associatie consistent.

In het kader van het hieraan gerelateerde onderzoek naar blootstelling van omwonenden aan gewasbeschermingsmiddelen is geconcludeerd dat in heel Nederland ongeveer 30.000 mensen binnen vijftig meter van een perceel met bollenteelt en ongeveer 58.000 mensen binnen vijftig meter van een perceel met fruitteelt wonen (Bogers *et al.*, 2014). Op basis van de beschikbare gegevens is het in principe mogelijk om per postcodegebied een proxy voor potentiële blootstelling van omwonenden aan bestrijdingsmiddelen vast te stellen.

Gezien het relatief beperkte aantal omwonenden, en de genoemde resultaten van de powerberekeningen voor kinderen, is het gewenst een eventuele onderzoeksopzet voor gewasbeschermingsmiddelen zoveel mogelijk te optimaliseren om de power te verhogen. Ook voor deze milieufactor geldt dat het wenselijk is om eerst te onderzoeken of de blootstelling op het woonadres kan worden vastgesteld en of de incidentie op individueel niveau in de tijd met behulp van registratiesystemen via een cohortonderzoek kan worden gevolgd.

#### 4.2.3 *Mogelijke verbeteringen van het clusteronderzoek in de toekomst*

Bij het toewijzen van de bevolking en de kankercases aan de PC4-gebieden is een zekere mate van misclassificatie niet uit te sluiten. Het 500 x 500 meter-

grid met bevolkingsgegevens dat sinds 2000 door het CBS beschikbaar wordt gesteld biedt mogelijkheden om kankerclusters op een lager aggregatieniveau en daarmee nauwkeuriger te analyseren. Voorwaarde voor analyse op gridniveau is wel dat door IKNL vanuit de NKR-gegevens op PC6-niveau ter beschikking worden gesteld, aangezien deze goed kunnen worden toegewezen aan de grids. Voordeel is eveneens dat vergelijking over de tijd een stuk eenvoudiger wordt, omdat er op deze manier geen verschuivingen plaatsvinden op postcodeniveau (CBS, 2012). Aangezien de door IKNL beschikbaar gestelde dataset vanaf 2007 landsdekkend is op PC6-niveau, kan de toepassing van grids pas met ingang van 2007 betrouwbaar plaatsvinden voor het gehele land. Voor eerdere jaren zijn de kankercases op PC4-niveau niet eenvoudig toe te wijzen aan de gridcellen, met de mogelijkheid van misclassificatie tot gevolg. Inzicht in de omvang van deze misclassificatie is van belang om vast te stellen met welke gevoeligheid een clustersignaal kan worden gedetecteerd.



## 5 Conclusies en discussie

In dit hoofdstuk worden de conclusies uit het inventariserend onderzoek bediscussieerd en in context geplaatst. De vier paragrafen sluiten aan bij de vraagstellingen uit de inleiding en beginnen met de belangrijkste conclusies in enkele bullets. Voor de overzichtelijkheid nemen we de vragen hier nogmaals op:

- Zijn er in Nederland leukemieclusters aan te tonen bij kinderen, in ruimte of tijd? (zie paragraaf 5.1)
- Op welke manier zijn de onderzoeksresultaten bruikbaar voor de GGD bij het omgaan met gemelde clusters? (zie paragraaf 5.2)
- Zijn de methoden bruikbaar voor landelijke monitoring van leukemie en/of andere aandoeningen? (zie paragraaf 5.3.)
- Zijn de methoden geschikt om een eventuele relatie te onderzoeken tussen het vóórkomen van kankerclusters en milieufactoren? (zie paragraaf 5.4)

### 5.1 Verdeling van de leukemie-incidentie in Nederland in ruimte en tijd

- Er kon met disease mapping en scananalyses voor de gehele periode 1995-2010 geen significante clustering van leukemie worden aangetoond bij ruimtelijke analyses over 4 x 4 jaar in Nederland.
- Bij de scananalyse waarbij alle onderzoeksjaren apart in het model (zestien jaar) werden opgenomen werd één statistisch significante bevinding gedaan over een periode van een viertal jaren.

Met *disease mapping* kon geen ruimtelijke variatie worden aangetoond in het vóórkomen van leukemie. Bij de groepen met de hoogste incidentie (alle respectievelijk overige kankersoorten) werden wel wat ruimtelijke effecten gezien. Gezien de diversiteit van de aandoeningen kunnen hier in deze rapportage geen conclusies uit getrokken worden.

De *scananalyses* laten zien dat er in de periode 1995-2010 geen statistisch significante ruimtelijke clusterpatronen worden gevonden als de tijdscomponent buiten beschouwing wordt gelaten (ruimtelijke analyse). Door toevoeging van de factor tijd op jaarniveau in de analyses bleek in een periode van enkele jaren gecombineerde tijd-ruimteclustering op te treden in een specifiek gebied, ook na correctie voor meervoudig toetsen.

Het onderzoek leverde geen bevestiging op van eerder onderzoek door Visser *et al.* (2005) over verhoging van hematopoïëtische maligniteiten, waaronder leukemie, in de regio Schiphol. De verklaring daarvoor kan liggen in de toepassing van een bredere diagnostische afbakening (Visser *et al.* onderzochten niet alleen leukemie maar ook lymfomen) en een eerdere onderzoeksperiode. Mogelijk zou ruimtelijke clustering met de gebruikte methoden wel kunnen worden aangetoond bij analyse op een lager aggregatieniveau dan PC4. Vanwege de beperkte beschikbaarheid van PC6-gegevens (pas sinds 2007 compleet voor heel Nederland) was het in dit onderzoek niet zinvol om analyses op PC6-niveau uit te voeren. In de toekomst zou dit wel gedaan kunnen worden, met de PC6-gegevens of eventueel gegevens op gridniveau (500 bij 500 meter), die sinds kort beschikbaar zijn gekomen (zie verder paragraaf 6.1.2).

## 5.2 Bruikbaarheid geavanceerde ruimtelijke analysemethoden bij clustermeldingen

- De databewerking (het koppelen van incidenties aan de veranderende PC4-gebieden) voorafgaand aan de analyses is tijdrovend, waardoor de methoden minder geschikt zijn om bij een lokale clustermelding in te zetten.
- Bij een clustermelding waarin veel vragen blijven bestaan, kan overwogen worden met deze geavanceerde methoden te onderzoeken of het gemelde cluster significant is en hoe het is afgebakend in ruimte en tijd.
- De resultaten laten zien dat de aanpak uit de GGD-richtlijn 'Kankerclusters' goed bruikbaar is bij clustermeldingen.
- De gevisualiseerde resultaten van dit onderzoek kunnen bij een clustermelding door de GGD worden gebruikt bij de informatievoorziening over (de variatie in) het vóórkomen van kanker bij kinderen in Nederland.

Het uitgevoerde onderzoek biedt een epidemiologische aanvulling op de recent uitgebrachte GGD-richtlijn medische milieukunde over 'Kankerclusters' (Hegger *et al.*, 2012). Deze richtlijn beschrijft drie stappen voor het omgaan met een clustermelding: oriëntatiefase, inventarisatie en kwantitatieve analyse. De melding kan na elke stap eventueel worden afgesloten, als de melder voldoende informatie heeft en/of bijvoorbeeld blijkt dat het aantal kankergevallen in de verwachte range ligt (Hegger *et al.*, 2012).

De voor dit rapport gebruikte technieken komen in beeld als na de kwantitatieve fase aanleiding is voor nader onderzoek. Het rapport biedt inzicht in het al dan niet optreden van statistisch significante clustering van leukemie in ruimte en tijd waarbij optimaal gebruik is gemaakt van alle beschikbare Nederlandse registratiegegevens vanuit IKNL en CBS. De resultaten uit het onderzoek onderschrijven dat de richtlijn 'Kankerclusters' voor GGD'en goed bruikbaar is voor de eerste afhandeling van (leukemie)clusters. Daarin worden, zoals hierboven beschreven, lokale cijfers vergeleken met landelijk gemiddelden, wat een goede aanpak is als er een homogene verdeling van de ziekte is. Dat laatste blijkt uit onze analyses.

Met behulp van de *scananalyses* bleek het mogelijk om een eerder gemeld en door de GGD onderzocht lokaal cluster te detecteren. Het kankerclustersignaal dat voor lokale ongerustheid heeft gezorgd kon door de aanvullende statistische analyses over een groot gebied beter worden afgebakend in ruimte en tijd, en bovendien statistisch getoetst. Het betreffende clustersignaal bleek na correctie voor meervoudig toetsen niet statistisch significant te zijn. Ook in de toekomst kan scananalyse eventueel worden ingezet voor verificatie en nadere afbakening in ruimte en tijd van een clustersignaal. De resultaten kunnen tevens een bijdrage leveren aan communicatie op lokaal niveau over de betekenis van het gemelde cluster (zie paragraaf 5.2.1 en 5.3).

### 5.2.1 *Discussie en conclusies naar aanleiding van data en datamanagement*

Ten behoeve van het onderzoek kon gebruik worden gemaakt van de NKR. Dit is een zeer complete registratie; naar schatting betreft het aandeel ontbrekende gegevens, bij vergelijking met de SKION-registratie, niet meer dan 1 procent. Dit wordt vooral veroorzaakt door kleine verschillen in diagnostiek op het raakvlak van goedaardige en kwaadaardige aandoeningen (Visser, persoonlijke mededeling 2013).

De keuze voor het ruimtelijke aggregatieniveau van de analyse wordt veelal gedreven door de beschikbare informatie. In dit onderzoek is geanalyseerd op

PC4-niveau, om analyse over een relatief lange tijdsperiode mogelijk te maken (1995-2010). PC6-gegevens zijn pas vanaf 2007 landelijk compleet. Gezien de korte periode waarover deze gegevens landsdekkend beschikbaar zijn en de lage incidentie was een analyse op PC6-niveau niet zinvol.

Postcodegebieden zijn ingedeeld om efficiënte postbezorging mogelijk te maken en zijn niet optimaal voor epidemiologisch onderzoek. Echter, vaak zijn gegevens op postcodeniveau de enige toegestane manier om gegevens aan elkaar te koppelen (ziektegevallen, bevolkingsopbouw). Het belang van het juist toewijzen van de bevolking en de kankercases aan de PC4-gebieden is voor dergelijke analyses groot. De veranderingen in de postcodekaart van Nederland gedurende de onderzoeksperiode 1995-2010 vormden daardoor een drempel voor het snel uitvoeren van een clusteronderzoek op PC4-niveau.

#### *Mogelijke verbeteringen in de toekomst*

Bij het toewijzen van de bevolking en de kankercases aan de PC4-gebieden is een zekere mate van misclassificatie niet uit te sluiten. Het 500 x 500 meter-grid met bevolkingsgegevens dat sinds 2000 door het CBS beschikbaar wordt gesteld biedt mogelijkheden om kankerclusters op een lager aggregatieniveau en daarmee nauwkeuriger te analyseren. Voorwaarde voor analyse op gridniveau is wel dat door IKNL vanuit de NKR gegevens op PC6-niveau ter beschikking worden gesteld, aangezien deze goed kunnen worden toegewezen aan de grids. Voordeel is dat vergelijking over de tijd een stuk eenvoudiger wordt, omdat er op deze manier geen verschuivingen plaatsvinden op postcodeniveau (CBS, 2012). Aangezien de door IKNL beschikbaar gestelde dataset vanaf 2007 landsdekkend is op PC6-niveau, kan de toepassing van grids pas met ingang van 2007 betrouwbaar plaatsvinden voor het gehele land. Voor eerdere jaren zijn de kankercases op PC4-niveau niet eenvoudig toe te wijzen aan de gridcellen, met de mogelijkheid van misclassificatie tot gevolg. De aanbeveling is daarom om de ruimtelijke analyse nogmaals uit te voeren als over een voldoende groot aantal jaren (indicatief tien jaar) informatie op PC6-niveau beschikbaar is.

#### 5.2.2 *Bruikbaarheid gevisualiseerde resultaten in de risicocommunicatie*

Bij de visualisatie van de resultaten moet goed rekening worden gehouden met privacyaspecten. Uitgangspunt is dat geen tot op persoon herleidbare informatie op postcodeniveau wordt gepresenteerd. De ruimtelijke patronen kunnen het best in beeld worden gebracht op grond van data in een niet van tevoren afgebakend gebied (dus niet naar aanleiding van een melding), zeker bij aandoeningen met een lage incidentie. De mogelijke weergave in kaartvorm is een van de mogelijkheden tot visualisatie. In overleg met IKNL wordt erop ingezet om voor GGD'en een instructiefilmpje beschikbaar te stellen waarin de ruimtelijke en geografische variatie in de incidentie van kinderkanker in beeld wordt gebracht, met in acht name van privacyaspecten. Dit kan desgewenst door de GGD worden gebruikt als achtergrondmateriaal bij de risicocommunicatie over een clustermelding.

### 5.3 **Bruikbaarheid geavanceerde ruimtelijke methoden voor monitoring**

- Ruimtelijke analyses zoals toegepast in dit onderzoek zijn bruikbaar om proactief periodiek in kaart te brengen of bepaalde vormen van kanker of andere aandoeningen clusteren in ruimte en/of tijd.
- Bij zeldzame aandoeningen zoals kanker bij kinderen worden wel extra eisen gesteld aan de data (schaalniveau en hoeveelheid) om bij de analyses voldoende onderscheidend vermogen te hebben.



- Het CBS biedt in samenwerking met andere aanbieders sinds enkele jaren mogelijkheden om populatie en blootstellingsgegevens te verkrijgen op 500 bij 500-gridniveau. Dat biedt voor de toekomst mogelijkheden om eenvoudiger en betrouwbaarder analyses van ruimtelijke patronen mogelijk te maken.

Landelijk onderzoek in lijn met het uitgevoerde onderzoek is bruikbaar om te inventariseren of bevindingen uit de internationale literatuur ook geldig zijn in Nederland. Uit de literatuur blijkt dat leukemie bij kinderen vaker clustert dan op basis van het toeval verwacht zou worden (COMARE, 2011; Gezondheidsraad, 2012). In verschillende landen zijn geverifieerde leukemieclusters beschreven. Vooral in Engeland is veel onderzoek gedaan op dit vlak, ook in relatie met kernenergiecentrales. In een recent onderzoek in Duitsland kon over de periode 1987-2010 echter geen clustering van leukemie worden aangetoond (Schmiedel *et al.*, 2010).

Landelijk onderzoek op basis van monitoringgegevens, zoals de in dit onderzoek uitgevoerde analyses, kan achtergrondinformatie opleveren die voor de GGD bruikbaar kan zijn bij de communicatie en afhandeling van het lokaal gemelde clustersignaal. Deze informatie (of en, zo ja, hoe vaak statistisch significante clustering van aandoeningen in Nederland voorkomt) kan daarmee dienen als communicatiemateriaal, zowel tekstueel als visueel (spreiding van het optreden van aandoeningen). De opgedane ervaringen in dit project, onder andere met databewerking, de statistische technieken en de methoden voor visualisatie, kunnen hierbij worden benut. Deze methodiek is eveneens toepasbaar voor ruimtelijke analyse van andere aandoeningen en gezondheidsproblemen.

#### 5.4 Koppeling van leukemie bij kinderen met milieufactoren

- Aanvullend onderzoek met de huidige methoden naar het verband tussen leukemie en milieufactoren op postcodeniveau (PC4) in de Nederlandse context is naar verwachting weinig zinvol.
- Een gebrek aan onderscheidend vermogen door het relatief kleine aantal leukemiepatiënten en aantal 'hoog' blootgestelde kinderen in Nederland en de kans op misclassificatie van de blootstelling op postcodeniveau liggen hieraan ten grondslag.

Van de milieufactoren die als risico worden genoemd voor leukemie zijn de extreem laagfrequente elektromagnetische velden (ELF-EMV) het meest onderzocht, voornamelijk in de buurt van hoogspanningslijnen. Uit eerdere powerberekeningen is gebleken dat de blootstellingskarakterisering van de veldsterkte van hoogspanningslijnen in een ecologische studie (zoals de door ons gebruikte methoden) op PC4-niveau of 500 x 500 meter-grid tot misclassificatie zal leiden. Daardoor neemt het vermogen van de studie af om de risico's aan te tonen. Analyses op PC6-niveau zullen deze misclassificatie slechts ten dele op kunnen lossen. Als onderzoeksopzet zou een (retrospectief) cohortonderzoek met heel Nederland als onderzoeksgebied eerder in aanmerking komen als geschikte onderzoeksmethode. Ook dan moeten activiteiten zich echter eerst richten op het vergaren van informatie om te kunnen beoordelen of een dergelijk onderzoek mogelijk zou zijn (modellering van veldsterktes, verkenning van mogelijkheden tot koppeling en nieuwe powerberekeningen). De GR-commissie heeft een pleidooi gehouden om nader etiologisch onderzoek bij voorkeur in internationaal verband te laten plaatsvinden.

In aanvulling op de uit de literatuur bekende relatie met EMV zijn in de rapportage van de Gezondheidsraad in 2012 nog enkele andere potentiële risicofactoren geïdentificeerd, zoals bestrijdingsmiddelen en mogelijke blootstelling aan emissies uit nieuwe materialen. Dit zijn factoren die in relatie tot leukemie veel minder goed zijn onderzocht. Er is niet zonder meer te zeggen of de hier gebruikte methoden ingezet kunnen worden om daar zinvol onderzoek naar te doen. Daarvoor is eerst een specifieke poweranalyse nodig, waarmee wordt ingeschat of met beschikbare data, het aggregatieniveau en gegeven het aantal blootgestelde personen, een eventuele relatie kan worden aangetoond. Op basis van 'expert judgement' (interne begeleidingscommissie RIVM) wordt voorsnog ingeschat dat de groep kinderen die hoog kan zijn blootgesteld aan bestrijdingsmiddelen te klein is om een eventuele significante associatie aan te kunnen tonen bij onderzoek met gegevens over ziekte en blootstelling op geaggregeerd niveau.

## 5.5 **Beschouwing**

Een onderzoek op deze schaal naar de clustering van leukemie bij kinderen met gebruik van deze methodiek is voor zover ons bekend niet eerder in Nederland uitgevoerd. Eerder internationaal onderzoek naar clusters, bijvoorbeeld Euroclus, maakte wel gebruik van geografische gegevens en gebruikte een methode die toetste of er clustering in algemene zin (inhomogeniteit) optrad en de mogelijke relatie met omgevingsfactoren (Alexander, 1999). De resultaten van de Euroclus-studie leverden echter geen inzichten op die direct relevant zijn voor praktijkvragen, zoals de beoordeling van de betekenis van nieuwe clusters op specifieke locaties en het bieden van referentiegegevens over de situatie in Nederland. Dit onderzoek biedt daar voor de periode sinds 1995 wel inzicht in. De Gezondheidsraad heeft in een eerder advies ter overweging gegeven expertise op dit vlak te bundelen, zodat veldpartijen daarvan gebruik kunnen maken (Gezondheidsraad, 2003). Afgaand op de behoorlijk grote inspanning die nodig is gebleken voor de voorbereidende activiteiten (dataverzameling, datamanagement) kan dit GR-advies op basis van de ervaringen in het huidige onderzoek worden onderschreven.

Daarnaast kan proactieve monitoring van incidentiepatronen behulpzaam zijn bij de zorgvuldige afhandeling van en communicatie over lokaal gemelde kankerclusters, op voorwaarde dat de data en databewerking die hiervoor nodig is van goede kwaliteit en fjnschalig beschikbaar is of kan komen. Dit inventariserend onderzoek heeft de methodiek voor het hiervoor benodigde datamanagement in kaart gebracht, beschreven en toegepast.



## DEEL 2 ACHTERGRONDEN



## 6 Geografische databewerking (achtergrond)

Voor dit onderzoek is uitgebreid gekeken naar mogelijkheden om data in te delen op grond van de – in de tijd veranderende- postcodes. Omdat dit bruikbare informatie is voor projecten waar ook een dergelijke exercitie nodig is, wordt deze methode hier apart gedocumenteerd.

Het toewijzen van bevolking en kankercases is in dit onderzoek op een deterministische manier gerealiseerd. Dit betekent dat op basis van de beschikbare informatie de waargenomen en verwachte aantallen toegewezen zijn aan gebieden. Het is ook mogelijk om te kiezen voor een aanpak op basis van kansberekening, waarbij een onderliggend incidentiepatroon als een soort deken over Nederland wordt gemodelleerd en naar afwijkingen in dit patroon wordt gezocht. Deze aanpak is in dit rapport niet nader uitgewerkt.

### 6.1 Methode databewerking op postcodeniveau

#### 6.1.1 *Gebiedsgerelateerde bewerking op PC4-niveau*

In dit rapport is gezocht naar kankerclusters die zich voordoen op PC4-niveau gedurende de periode 1995-2010. Om de ruimtelijke analyses mogelijk te maken is het noodzakelijk om de kankercases te koppelen aan het PC4-gebied waarin ze optreden. Daarnaast is informatie over de leeftijds- en geslachtsopbouw van elk PC4-gebied nodig om het aantal verwachte kankercases per PC4-gebied te kunnen vaststellen.

#### *Geografische indeling*

De postcodekaart van Nederland is niet constant in de tijd: nieuwe postcodegebieden worden gevormd binnen of over de grenzen van oude gebieden, de grenzen van gebieden worden verlegd om de postbezorging efficiënter te laten verlopen en af en toe worden gebieden opgeheven. Deze 'temporal boundary misalignment' compliceert een adequate analyse van de leukemie-incidentie, doordat zowel delen van de onderzoeksgroep als de referentiegroep in de loop van de tijd aan net andere gebieden moeten worden gekoppeld. Dit vereist dat de veranderingen in de postcodegebieden worden vastgelegd en dat de data van de verschillende jaren worden herverdeeld over de postcodegebieden van de basiskaart. Er bestaat nog geen consensus over de methode die gehanteerd moet worden om het probleem van verschuivende postcodegrenzen te ondervangen. De volgende drie opties zijn overwogen:

1. *Uitgaan van de postcodekaart bij de start van de onderzoeksperiode.*  
Nieuwe postcodegebieden worden voornamelijk gevormd binnen de grenzen van bestaande postcodegebieden. Bevolkings- en ziektegegevens van de nieuwe PC4-gebieden van latere jaren kunnen dan worden toegewezen aan het oorspronkelijke PC4-gebied. Nadeel is het gebruik van een verouderde postcodekaart (uit 1995!) die geen rekening houdt met verschuivingen in de ruimtelijke spreiding van de bevolking (denk aan Vindex-locaties als IJburg en Leidsche Rijn). Deze methode is daarmee ook minder geschikt indien het van belang wordt geacht bij de analyse aandacht te besteden aan eventuele clustering binnen nieuwe woonlocaties. De methode is relatief eenvoudig, maar arbeidsintensief.
2. *Uitgaan van een postcodekaart aan het einde van de onderzoeksperiode.*  
Het grote voordeel van deze methode is dat de resultaten beschikbaar

komen voor de meest recente situatie (in dit geval 2010) en dat daarmee ook nieuwbouwlocaties op een adequate wijze in beeld worden gebracht. Nadeel is het toewijzen van de bevolking aan PC4-gebieden die in latere jaren aan verandering onderhevig zijn (een groter postcodegebied gaat bijvoorbeeld in een later jaar op in een of meer kleinere gebieden).

### 3. Gebruik van een geostatistisch model.

Er zijn verschillende pogingen ondernomen om geostatistische modellen te ontwikkelen die rekening trachten te houden met verschuivingen in de onderliggende (postcode)kaart (Kelsall en Wakefield, 2002; Hund *et al.*, 2012). Het voordeel van deze modellen is dat het probleem modelmatig op een uniforme wijze wordt opgelost. Het nadeel is de complexiteit van de methode en de grote rekencapaciteit die nodig is.

Op basis van de beschikbare data is gekozen voor optie 2. Als referentiejaar voor de geografische indeling is het jaar 2010 gekozen, het laatste jaar van de onderzoeksperiode. Daarbij zijn zowel de bevolkingsgegevens als de incidentiegegevens toebedeeld aan de geografische indeling van de postcodes in 2010. Van deze kaart is een burenmatrix gemaakt die de ruimtelijke relatie tussen de postcodegebieden beschrijft. De burenmatrix is een simpele voorstelling van de relatie tussen postcodegebieden: PC4-gebieden met een gemeenschappelijke grens krijgen de waarde 1, de overige PC4-gebieden de waarde 0. In de statistische analyse wordt er rekening mee gehouden dat de kankerincidenties van postcodegebieden die dicht bij elkaar liggen meer met elkaar samenhangen dan van ver uit elkaar gelegen postcodegebieden. Wanneer de beschikbare informatie in een postcodegebied klein is, wordt informatie 'geleend' uit naburige postcodegebieden om de analyse te verbeteren.

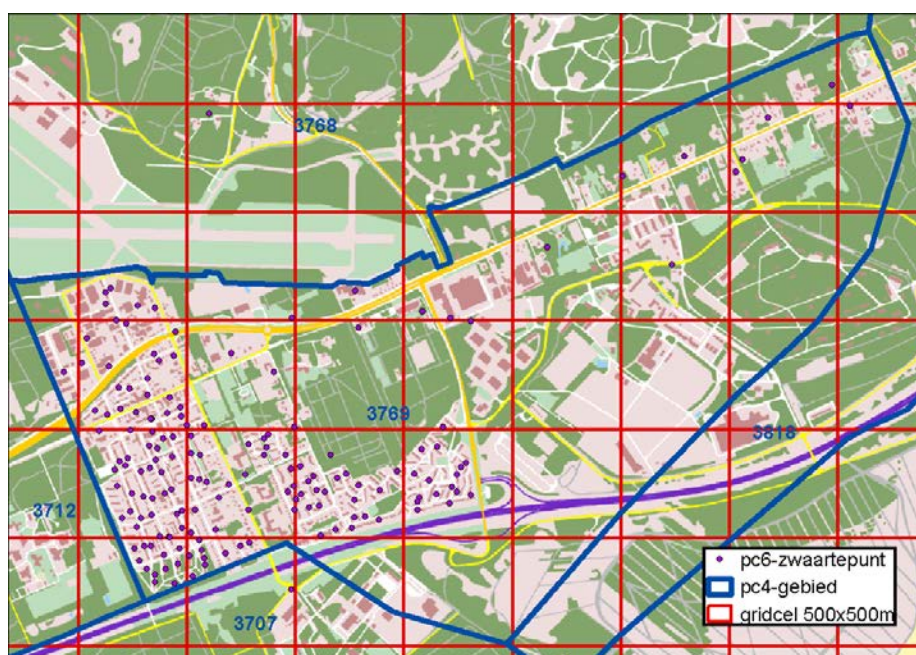
In Figuur 2 is het probleem van verschuiving en nieuwvorming van postcodegebieden inzichtelijk gemaakt.



Figuur 2. Verschuiving en nieuwvorming van postcodegebieden tijdens de onderzoeksperiode in de Leidsche Rijn, Vinx-locatie aan de rand van Utrecht.

### Bevolkingsgegevens per PC4-gebied

Het verwachte aantal kankercases per PC4-gebied en jaar wordt vastgesteld op basis van de leeftijds- en geslachtsverdeling van de bevolking in de afzonderlijke PC4-gebieden. Deze gegevens zijn in afgeronde vorm beschikbaar in de online Statline-database van het CBS (<http://statline.cbs.nl/statweb/>). De gegevens worden afgerond om herleidbaarheid naar personen te vermijden. Het RIVM heeft deze gegevens in niet afgeronde vorm beschikbaar voor de onderzoeksperiode. Voor 2010 levert de koppeling van de CBS-gegevens aan de PC4-gebieden geen problemen op omdat deze kaart als referentie is gekozen. Voor de overige jaren (1995-2009) moeten de gegevens worden gekoppeld aan de PC4-gebieden in 2010, hiervoor is een herberekening noodzakelijk. Dit is uitgevoerd door gebruik te maken van additionele informatie die voor het onderzoek beschikbaar was over de PC6-gebieden in Nederland. Er zijn in Nederland ongeveer 445.000 PC6-gebieden. Een PC6-gebied bevat gemiddeld tussen de vijftien en twintig adressen. Van de jaren 1995-2010 is het aantal mensen per PC6-zwaartepunt beschikbaar. Een zwaartepunt is het geografische middelpunt (x-coördinaat) van een PC6-gebied. De locatie van de PC6-zwaartepunten geeft inzicht in de verdeling van de bevolking over een PC4-gebied, het ontstaan van nieuwe bevolkingskernen in de tijd en de impact van veranderende postcodegrenzen. Figuur 3 geeft de verschillende manieren van aggregeren weer.



*Figuur 3. Onderscheid tussen drie manieren van aggregeren van incidentie en populatiegegevens: PC4-gebieden, PC6 zwaartepunten en gridcellen (500 bij 500 meter) voor de situatie in 2010.*

De volgende stappen zijn uitgevoerd om de bevolking naar leeftijd en geslacht per jaar en PC4-gebied vast te stellen:

1. De referentiekaart van PC4-gebieden in 2010 is vastgesteld in ArcGIS. De gebieden zijn geëxporteerd naar het statistische pakket SAS.
2. De locatie van de PC6-zwaartepunten van de periode 1995-2010 is in ArcGIS gekoppeld aan de PC4-gebieden van de 2010-referentiekaart. Dit resulteert in zestien databestanden met de bevolkingsaantallen per PC6-gebied in jaar x en het nummer van het PC4-gebied uit 2010 waarin het PC6-gebied valt.



3. De bevolkingstotalen per PC4-gebied bleken enigszins af te wijken van de som van de bevolking van de PC6-gebieden. Dit is opgelost door de bevolking per PC6-gebied te kalibreren met de PC4-randtotalen van het CBS. Hierbij is aangenomen dat de bevolkingsaantallen van het CBS per PC4-gebied het meest nauwkeurig zijn.
4. De leeftijds- en geslachtsverdeling van de bevolking per PC4-gebied in alle onderzoeksjaren is ingedeeld analoog aan de indeling die het IKNL gemaakt heeft van de kankerregistratie. Het gaat daarbij om achttien groepen en de totale bevolking:
  - totale bevolking;
  - 0-19 jaar (totaal/man/vrouw);
  - 0 jaar (totaal/man/vrouw);
  - 1-4 jaar (totaal/man/vrouw);
  - 5-9 jaar (totaal/man/vrouw);
  - 10-14 jaar (totaal/man/vrouw);
  - 15-19 jaar (totaal/man/vrouw).
5. De negentien leeftijds- en geslachtsgroepen zijn uitgedrukt als een fractie van de totale bevolking in een PC4-gebied en jaar.
6. Deze fracties zijn voor elk onderzoeksjaar gekoppeld aan de PC6-zwaartepunten van dat jaar. Vervolgens zijn de fracties vermenigvuldigd met de gekalibreerde bevolkingsaantallen per PC6-gebied.
7. Elk PC6-zwaartepunt van een onderzoeksjaar is gekoppeld aan het bijbehorende nummer van het PC4-gebied waar het in 2010 in zal vallen. Hiermee wordt de bevolking per PC4-gebied uitgerekend.

#### *Kankerincidentie per PC4-gebied*

De kankercases zijn voor de periode 1995-2010 door IKNL geleverd met een bijbehorend PC4-nummer. Deze nummers zijn rechtstreeks gekoppeld aan de PC4-gebieden in het referentiejaar 2010. Een aantal cases kon niet worden toegewezen aan een PC4-gebied, omdat de opgegeven adressen een postbus betreffen. Deze cases zijn uitgezocht en uit de dataset verwijderd. Daarnaast bevat de IKNL-database een klein aantal PC4-nummers die niet te lokaliseren zijn. Deze zijn ook verwijderd. Het betrof hierbij in totaal negen records, waarvan vijf uit groep 1 (leukemie), één uit groep 2 (hersentumoren) en drie uit groep 3 (overige vormen van kanker). Dat betekent dat van de 9.921 cases gedurende de periode 1995-2010 er 9.912 beschikbaar waren ten behoeve van de analyse.

De uitgevoerde stappen zijn:

1. Voor de periode 1995-2010 zijn de kankercases toegewezen aan een PC4-gebied van de postcodekaart in 2010.
2. De cases zijn per PC4-gebied ingedeeld naar geslacht en leeftijdscategorie (zie vorige paragraaf).
3. Het indelen van de cases naar leeftijd en geslacht gebeurt vier keer. Eerst voor alle cases in de IKNL-dataset. Vervolgens voor de cases van de primaire onderzoeksgroep en de twee referentiegroepen apart.
4. De cases naar leeftijd, geslacht, onderzoeksgroep en totaal zijn gesommeerd per PC4-gebied.
5. De teller (het aantal kankercases per PC4-gebied) en de noemer (bevolking per PC4-gebied) zijn gekoppeld.

De dataset is nu gereed voor gebruik ten behoeve van *scan-analyses*. Voor de ruimtelijke *disease mapping* is nog een directe standaardisatie noodzakelijk.

Door te standaardiseren wordt inzicht verkregen in de vraag hoeveel kankercases mogen worden verwacht in een PC4-gebied, indien de leeftijds- en geslachtsverdeling in het PC4-gebied gelijk zou zijn aan een standaardpopulatie. Als standaardpopulatie wordt in dit geval de Nederlandse bevolking van 0 tot 19 jaar in het gekozen onderzoeksjaar genomen. De directe standaardisatie maakt een vergelijking van de kankerincidentie tussen PC4-gebieden mogelijk. Hiertoe wordt de SIR (Standardized Incidence Ratio) uitgerekend, dit is het aantal geobserveerde cases per PC4-gebied (observed) gedeeld door het aantal verwachte cases per PC4-gebied (expected). Het aantal verwachte cases wordt berekend uit de leeftijds- en geslachtsspecifieke cijfers van de gekozen standaardpopulatie.

Standaardisatie vindt plaats voor de vier onderzoeksgroepen (totaal, primair, referentie 1 en referentie 2) per leeftijdsgroep en voor alle leeftijdsgroepen samen (0 tot 19 jaar).

#### 6.1.2 *Gebiedsgerelateerde dataverwerking op PC6-niveau*

In deze paragraaf staat op hoofdlijnen beschreven welke stappen moeten worden doorlopen om analyses op 500 x 500 meter-gridniveau mogelijk te maken. Daarvoor zouden vanaf 2007-2010 PC6 gegevens kunnen worden gebruikt. Voor de periode daarvoor (2000-2006) zijn landsdekkend alleen PC4-gegevens van IKNL beschikbaar. Om een gridanalyse mogelijk te maken moeten deze toegewezen worden aan gridcellen.

*Beschikbaarheid PC6-data*Ten behoeve van het onderzoek heeft IKNL geaccordeerd dat ook beschikbare PC6-gegevens gebruikt kunnen worden om de analyse nauwkeuriger uit te kunnen voeren. Deze PC6-gegevens waren echter niet voor de gehele onderzoeksperiode beschikbaar.

#### *Toewijzen PC6-zwaartepunten aan het CBS-grid*

De geografische beschrijving van analyses op PC6-niveau bestaat uit de zwaartepunten van de PC6-gebieden. Voor de analyse van PC6-gegevens kunnen de attribuutdata van de PC6-gebieden worden toegewezen aan het regelmatige 500 x 500 meter-grid dat recent in gebruik is genomen door het Centraal Bureau voor de Statistiek (CBS). Overstappen op het 500 x 500 meter-grid is een grote stap voorwaarts, omdat het probleem van 'temporal boundary misalignment' bij deze werkwijze niet optreedt. Voor elk onderzoeksjaar worden de attribuutdata van de PC6-zwaartepunten toegekend aan de gridcel waarbinnen het zwaartepunt zich bevindt. Het maakt dan niet meer uit of PC6-gebieden verschijnen of verdwijnen in een onderzoeksjaar, zolang ze maar toegewezen kunnen worden aan een gridcel. Gezien de arbeidsintensiteit van de gegevensbewerking op PC-niveau en de uitval van data als gevolg van niet-integrale beschikbaarheid van PC6-gegevens voor de gehele onderzoekspopulatie en -periode, zijn de analyses op PC6-niveau nog niet uitgevoerd.



## 7 Statistische analyses (achtergrond)

### 7.1 Disease mapping

Voor de uitgevoerde *disease mapping* is gebruikgemaakt van het bayesiaans conditioneel autoregressiemodel (CAR) van Besag. Voor deze analyse is gebruikgemaakt van het statistische softwarepakket R en het R-INLA-pakket (Rue *et al.*, 2009). Met behulp van het model wordt een schatting gemaakt van de ziekte-incidentie in ruimte en tijd. De incidentiegegevens waarop de analyse is uitgevoerd betreffen tellingen van incidentie leukemiegevallen per postcodegebied. Gezien de lage incidentie van leukemie is ervan uitgegaan dat het incidentiepatroon gekenmerkt wordt door een Poisson-verdeling. Zoals in hoofdstuk 3 wordt beschreven, zijn allereerst de postcodegebieden geprojecteerd naar het referentiejaar (2010). Voor de analyse is voorafgaand een directe standaardisatie van de verwachte incidenties per postcodegebied noodzakelijk vanwege de verschillen in populatieomvang en -samenstelling tussen de postcodegebieden. Dit gestandaardiseerde aantal verwachte cases wordt berekend aan de hand van de leeftijds- en geslachtsspecifieke incidentiecijfers van de totale Nederlandse bevolking (bron: IKNL). Vervolgens kan de SIR (Standardized Incidence Ratio) per PC4-gebied worden uitgerekend, oftewel het aantal ziektegevallen per PC4-gebied (observed) gedeeld door het aantal verwachte cases in dit gebied (expected). Het gebruikte model baseert zich op deze SIR.

#### 7.1.1 Model en modeldiagnostiek

Voor modellering van een ruimtelijk effect is het 'proper Besag'-model uit het R-INLA pakket gebruikt. Het bij disease mapping meer gangbare model van Besag-York-Mollie is niet geschikt voor de grote gebiedsomvang van dit onderzoek, omdat daarin als gevolg van de vaste ruimtelijke correlatiefunctie de correlatie lokaal toeneemt als functie van het aantal gebieden in de analyse. In het 'proper Besag'-model is de correlatiefunctie aan te passen aan de gebiedsgrootte.

Het aantal geobserveerde kankergevallen  $y_i$  binnen postcodegebied  $i$  wordt verondersteld Poisson verdeeld te zijn:

$$y_i \sim \text{Poisson}(E_i \theta_i)$$

Waarbij  $E_i$  het verwachte aantal kankergevallen is en  $\theta_i$  de ruimtelijke gesmooth SIR van postcodegebied  $i$  weergeeft. Het gevonden verschil tussen het aantal geobserveerde en het aantal verwachte kankergevallen is vervolgens gemodelleerd met een log-lineair gemengd model. Om daarbij te kunnen bepalen of er een ruimtelijk en/of een tijdseffect optrad, moest dat model driemaal gedraaid worden:

1.  $\log \theta_i = \beta$  (basismodel)
2.  $\log \theta_i = \beta + u_i + v_i$  (ruimtelijk model)
3.  $\log \theta_i = X_i^T \beta + u_i + v_i$  (ruimte-tijdmodel)

waarbij de  $X_i$  een matrix is met daarin een constante (intercept) en de tijdsvariabele,  $\beta$  de vector met regressiecoëfficiënten en  $u_i$  en  $v_i$  de random effecten voor de ruimtelijk gestructureerde variatie en de overige ongestructureerde variatie voorstellen.

Het basismodel zonder ruimtelijk effect levert per uitkomst en leeftijdsgroep een schatting van de gemiddelde gestandaardiseerde incidentieratio (SIR) in Nederland over de hele periode van zestien jaar. De SIR-schattingen uit het basismodel zonder ruimtelijk effect zijn niet van belang, het basismodel is alleen nodig om een uitgangswaarde voor de DIC ('Deviance Information Criterion') vast te stellen waaraan de relevantie van de uitbreiding tot ruimtelijke en ruimte-tijdmodellen kan worden afgemeten.

Het ruimtelijk model berekent een gecorrigeerde gemiddelde SIR per PC4-gebied. Informatie uit dit soort kleine gebieden verschilt onderling namelijk in statistische betrouwbaarheid als gevolg van verschillen in inwoneraantallen. Incidenties worden in het ruimtelijk model gewogen door gebruik te maken van de informatie uit naburige gebieden. Versturende toevalsvariatie als gevolg van lage inwoneraantallen in dunbevolkte PC4-gebieden wordt uitgefilterd. Dit moet eerst gebeuren voordat kan worden vastgesteld of er systematische verschillen zijn tussen PC4-gebieden.

Het ruimte-tijdmodel levert aanvullend per PC4-gebied voor elke periode van vier jaar een schatting van de ruimtelijk gecorrigeerde gemiddelde SIR over die periode. Er kan mee worden getoetst of er systematische verschillen zijn tussen de perioden.

De drie genoemde modellen zijn gedraaid voor alle aanwezige combinaties van uitkomsten en leeftijdsgroepen. Er zijn de vier uitkomstgroepen uit Tabel 3. Ook zijn er zes leeftijdscategorieën: 0-jarigen apart, de groepen 1-4, 5-9, 10-14, 15-19 jaar en de hele groep samen (0-19 jaar). Dit betekent dat elk van de drie modellen 24 maal is gedraaid.

Er zijn verschillende criteria aan de hand waarvan een model beoordeeld kan worden. De DIC-score is daar één van. Dit is een grootte waarmee, analoog aan de AIC voor lineaire modellen, voor een en dezelfde uitkomst kan worden bepaald of uitbreiding van een model met extra verklarende variabelen een verbetering van het model oplevert. Hierbij geldt in principe: hoe lager, hoe beter. Indien bij het toevoegen van een effect aan een model de DIC-score niet wordt verlaagd, is de toevoeging niet zinvol.

### 7.1.2 *Rekentechnische omstandigheden*

Bij het draaien van een uitgebreider ruimte-tijdmodel met alle jaren apart trad een probleem op met de rekencapaciteit. In verband met de vertrouwelijkheid van de gegevens was het niet zonder meer mogelijk om gebruik te maken van externe rekencapaciteit.

Tijdens de eerste runs trad een probleem op met de berekening van de DIC. Dit bleek het gevolg te zijn van de aanwezigheid van 'eilanden' in het onderzoeksgebied in combinatie met lage incidentiecijfers. Daarom zijn alle eilanden uit het onderzoeksbestand verwijderd door een verbinding te definiëren met relevante buurgebieden. Voor de Waddeneilanden is gekozen voor een koppeling tussen de PC4-gebieden via de veerboothavens. Zeeuws-Vlaanderen is aan Walcheren gekoppeld via de Westerscheldetunnel enzovoort. Hiermee wordt het rekenprobleem opgelost dat optreedt wanneer er helemaal geen kankergevallen op een 'eiland' hebben plaatsgevonden. Ook komt zo informatie

uit de dichtstbijzijnde gebieden beschikbaar om op dezelfde manier als voor alle andere gebieden in Nederland de meest betrouwbare schatter voor de incidentie te berekenen.

## 7.2 Spatial Scan Statistiek voor clusteranalyse in ruimte en/of tijd

Met behulp van de zogenaamde scan-statistiek kan worden onderzocht of, en zo ja wanneer en waar, er clustering van leukemie optreedt. Daarbij kan zowel ruimtelijke clustering worden onderzocht als gecombineerde clustering naar ruimte én tijd. Er zijn twee soorten scan-statistiek toegepast: SatScan en FleXScan.

### 7.2.1 Ruimtelijke analyse met SatScan

De Spatial Scan Statistiek (Kulldorff, 2009) is een veelgebruikte en effectief gebleken test voor het onderzoeken van clustering in ruimte en/of tijd van ziekten of overige gebeurtenissen. Daarbij wordt de onderzoeksregio (heel Nederland) verdeeld in een groot aantal gebieden ('windows') van diverse omvang, waarbij op elk moment in plaats en/of tijd wordt berekend hoe de verhouding ligt tussen de incidentie van de aandoening binnen en buiten de onderzoeks-'windows'. Door het systematische 'scannen' van regio's en of perioden van verschillende grootte is onderzocht of er regio's en/of perioden getraceerd kunnen worden met een statistisch verhoogde (of desgewenst verlaagde) clustering van leukemie in vergelijking met de beschikbare referentiegroepen. De primaire referentiegroep betreft de algemene bevolking van dezelfde leeftijd (Poisson-verdeling). Daarnaast kan ook onderzocht worden of er clustering optreedt ten opzichte van beide referentiegroepen binnen de NKR. Daarbij kan onderzocht worden of er verschillen zitten in mate en regio van clustering tussen de diagnosegroepen. Bij deze clusteranalyse kunnen leeftijd, geslacht en desgewenst andere variabelen in de analyse worden opgenomen als covariaten. Bij de inputspecificaties kan worden aangegeven hoe groot de maximaal te traceren cluster grootte kan zijn. Als defaultwaarde voor de analyses is daarbij gekozen voor vijftig procent van de totale onderzoekspopulatie, aangezien dit voor deze SatScan-test de meest gangbare waarde is (Han, 2011). Dat impliceert dat de minimale grootte van een eventueel gevonden cluster gelijkstaat aan de kleinste gebiedsindeling (in geval van dit onderzoek dus PC4-niveau) en het maximale cluster overeenkomt met de helft van de totale onderzoekspopulatie. Bij de beoordeling van mogelijke clusters en de significantie daarvan wordt gecorrigeerd voor meervoudig toetsen.

*Uitkomstmaten* Als uitkomst van de scanstatistiek wordt het zogenaamde 'Most Likely Cluster' (MLC) gedefinieerd. Dat is een gebied, bestaande uit een of meer aangesloten gebieden, waarbij de kans het grootst is dat daar substantiële clustering optreedt na correctie voor verstoringe factoren. In aanvulling daarop kunnen ook zogenaamde secundaire of mogelijke clusters worden onderscheiden. Dat zijn gebieden waar eveneens een verhoogde kans is op clustering, hoewel ze minder waarschijnlijk zijn dan het Most Likely Cluster. Identificatie als MLC of secundair cluster wil overigens niet per definitie zeggen dat de geïdentificeerde clustering statistisch significant is.

Als uitkomstmaat wordt per geïdentificeerde mogelijke cluster het Relatief Risico (RR) berekend. Daarbij wordt ook de 'Log Likelihood Ratio' (LLR) vastgesteld met de daarbij behorende significantie (p-waarde). Hierbij wordt door toepassing van Monte Carlo-simulatie gecorrigeerd voor herhaald testen. Op basis van de uitgevoerde toetsen kan worden geconcludeerd of er clustering in ruimte en/of tijd optreedt en, zo ja, of deze clustering significant is.

De uitkomstmaten van de *scan-analyses* hebben betrekking op de geografische afbakening van een of meer aaneengesloten regio's en niet op individuele postcodes (tenzij binnen een enkele postcode een zodanig substantiële verhoging van het risico bestaat dat een enkele postcode voldoende is om een effect te identificeren). Met de scananalyse is het ook mogelijk om gebieden te detecteren met een verlaagde incidentie. Indien de resultaten het toelaten om ruimtelijke patronen in kaart te brengen is het daarom meestal mogelijk om dat te doen zonder weergave van individuele postcodegebieden. Als er sprake is van clustering, wordt tevens een outputfile opgeleverd, die het mogelijk maakt om dit cluster via een GIS-pakket in kaart te brengen.

#### *Ellipsvormige Clusters*

SatScan biedt sinds versie 6 ook de mogelijkheid om de *scan-analyse* uit te voeren met een ellipsvormig cluster-'window'. Deze optie komt tegemoet aan de behoefte om te kunnen onderzoeken of de clustering meer lijnvormig dan cirkelvormig is. De traditionele analyse met een cirkelvormig 'window' is soms niet sensitief genoeg. In vergelijking met de reguliere cirkelvormige scan-analyse wordt bij de uitkomsten ook de hoek en vorm van de potentiële cluster aangegeven.

#### *Flexibele Scan Statistiek (FlexScan)*

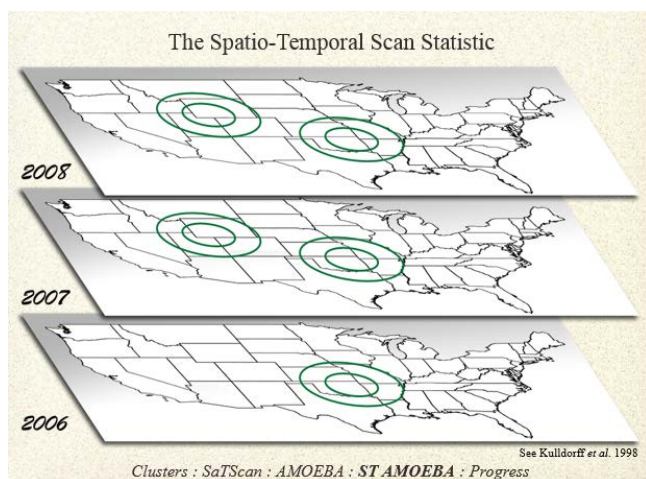
In eerste instantie zijn cirkelvormige scan-'windows' als uitgangspunt gehanteerd voor de scananalyses. In aanvulling daarop is ook getest met een tweede pakket: FlexScan (Takahashi *et al.*, 2010). FlexScan onderscheidt zich van Statscan doordat het ook clusters kan detecteren met een flexibele vorm (niet alleen cirkelvormig zoals *oorspronkelijk* het geval was bij SatScan). Daarnaast worden bij FlexScan de inputspecificaties zodanig aangepast dat wordt voorkomen dat heel grote clusters worden gedetecteerd. FlexScan maakt net als SatScan gebruik van het Poisson-model, waarbij het aantal events in een gebied een Poisson-verdeling volgt gegeven een bekende achterliggende populatie 'at risk' (de algemene bevolking). Daarnaast is het, net als bij SatScan, mogelijk om de data te analyseren gebaseerd op een binomiale verdeling. Die kan worden gebruikt om een groep te vergelijken met een andere (referentie)groep en niet noodzakelijk de gehele populatie. Binnen FlexScan is het vooralsnog niet mogelijk om ook trends in de tijd te detecteren of 'space-time'-clustering uit te voeren. Al met al kan FlexScan net als SatScan worden gebruikt om:

- gerapporteerde ruimtelijke ziekteclusters te beoordelen op statistische significantie;
- te testen of een aandoening random geografisch is verdeeld is;
- geografische monitoring uit te voeren, om eventuele gebieden met een significant verhoogd optreden van een aandoening te detecteren.

#### *Ruimte-tijdanalyse met SatScan*

In aanvulling op de ruimtelijke *scananalyse* is eveneens een zogenaamde 'time-space'-clusteranalyse uitgevoerd. Daarbij wordt zowel gescreend op clustering in ruimte als tijd. Deze analyse met een tijdcomponent kan alleen worden uitgevoerd met SatScan. De analyse in tijd is op twee manieren uitgevoerd. In eerste instantie zijn alle onderzoeksjaren (periode 1995-2010) als aparte variabelen ( $n=16$ ) opgenomen in de analyse. In aanvulling daarop is ook een analyse uitgevoerd waarbij de gegevens op jaarniveau zijn geaggregeerd tot vier perioden van elk vier jaar. Dit laatste is gedaan om vergelijkbaarheid te creëren met de *disease mapping*, die vanwege de rekencapaciteit gedraaid moesten worden in vierjaarsperioden.

Daardoor wordt in de *scan-analyses* een extra tijdsdimensie ingevoerd, waardoor ook de duur van eventuele verhoging van incidenties in beschouwing wordt genomen (zie Figuur 4).



Figuur 4. Voorbeeld van het toevoegen van de tijdsdimensie in SatScan (bron: Kulldorff, 1997)

Deze aanpak heeft tot doel om te onderzoeken of er binnen de onderzoeksperiode en onderzoekspopulatie sprake is van een gemeenschappelijk patroon van clustering in ruimte en tijd van de gevallen van leukemie in vergelijking tot de referentiegroep.

De uitkomsten van de *scan-analyses* zijn beschreven in hoofdstuk 5.

De vragen die met deze analyses worden beantwoord, zijn (vraag 1c en 1d uit de inleiding):

- Kan er aanvullend op het ruimtelijke aspect clustering in tijd worden gedetecteerd?
- Zijn er verschillen in de uitkomsten tussen de gehanteerde analysemethoden?





## 8 Weergave en visualisatie van de resultaten (achtergrond)

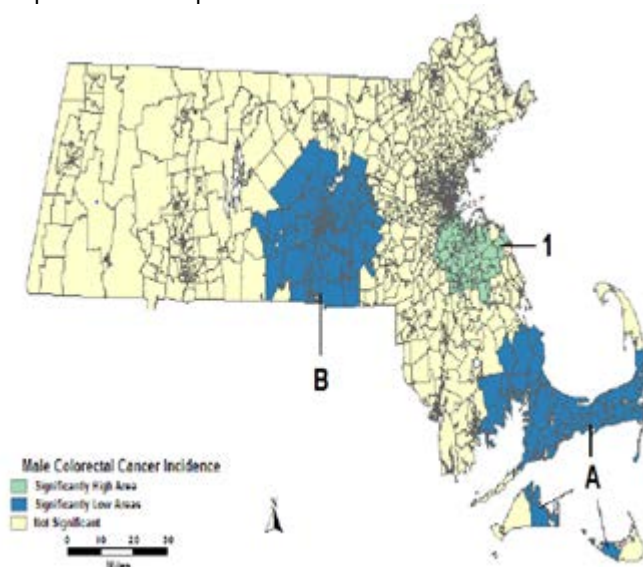
Dit hoofdstuk beschrijft de mogelijkheden om de onderzoeksresultaten zodanig te visualiseren zonder te veel detaillering waardoor privacyaspecten in het geding kunnen zijn. Daarbij is een onderscheid gemaakt tussen de visualisatie van de uitkomsten van de uitgevoerde ruimtelijke analyses enerzijds en een visualisatie van de oorspronkelijke data, geaggregeerd op gemeentelijk niveau, anderzijds.

### 8.1 Visualisatie disease mapping

De uitkomsten van de *disease mapping*-analyses zijn in eerste instantie ook in kaart gebracht. Daarbij krijgt elke postcoderegio een waarde toegekend, zodat een patroon op PC4-niveau voor heel Nederland wordt geboden. Hoewel bij de analyse 'smoothing' is toegepast, wordt daardoor toch op PC4-niveau een patroon zichtbaar. In overleg met IKNL is afgezien van opname daarvan in de rapportage, aangezien dat onvoldoende in overeenstemming is met de privacyvereisten die IKNL op voorhand heeft verbonden aan beschikbaar stellen van de data.

### 8.2 Visualisatie SatScan-analyse

Zoals beschreven levert SatScan als resultante enkel een geodatabestand dat in een GIS-pakket kan worden ingelezen om de clusters te kunnen tonen op de kaart. Het voordeel van de weergave van geclusterde regio's boven de weergave op individueel postcodeniveau (al dan niet na smoothing) is dat de resultaten van de *scan-analyses* (meestal) niet op individueel postcodeniveau liggen, maar op een meer geaggregeerd niveau (regio met het meest waarschijnlijke cluster). Het gaat daarbij vaker om grotere gebieden met grotere aantallen cases, waardoor deze weergave uit privacyoverwegingen de voorkeur heeft boven weergave op individueel postcodeniveau.

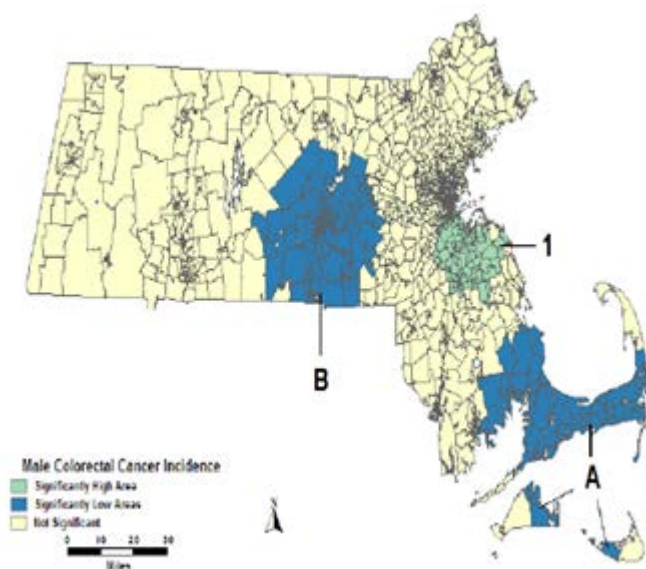


Figuur 5 laat zien hoe de uitkomsten van een SatScan cluster-analyse kunnen worden gevisualiseerd met een regulier GIS pakket, zoals ArcGIS of het open source pakket QGIS. Het betreft in dit geval een clusteronderzoek naar de ruimtelijke verdeling van darmkanker bij mannen, waarbij zowel de statistisch

significant verhoogde als verlaagde clusters in kaart zijn gebracht (respectievelijk de licht- en donkerblauwe gebieden; DeChello en Sheehan, 2007).

### 8.3 Visual Inquiry Toolkit

Een alternatieve visualisatiemethode voor SatScan betreft de zogenaamde Visual Inquiry Toolkit (VIT). Dit betreft een pakket waarbij een programmeeromgeving wordt aangeboden om ruimtelijke gegevens te verwerken, zowel analytisch, visueel als cartografisch. Het beoogt daarmee een bijdrage te leveren aan het exploreren en de beoordeling van ruimtelijke patronen in data. Een onderdeel van het pakket is specifiek gericht op verantwoorde c.q. betrouwbare visualisatie van ruimtelijke clusteranalyse. Het is daarbij mogelijk om de output van Spatial Scan-software als SatScan te importeren ten behoeve van visualisatie. Het programma beoogt de interpretatie van ruimtelijke clusterdetectiemethoden te faciliteren door cartografische representatie van de SatScan-resultaten. Tevens worden visualisatiemethoden aangeboden die de selectie van SatScan-parameters kunnen ondersteunen (Chen, 2008).



Figuur 5. Weergave van ruimtelijke clusterpatronen in de incidentie van darmkanker bij mannen in de Verenigde Staten, SatScan-analysemethode (Bron: DeChello, 2007)

### 8.4 Visualisatie FlexScan-analyse

In aanvulling op SatScan biedt ook FlexScan mogelijkheden voor de visualisatie van de uitkomsten van de *scan-analyse*. In tegenstelling tot SatScan worden de resultaten van de FlexScan-analyse direct verwerkt tot een grafische weergave. Daarbij wordt gebruikgemaakt van twee geografische bestanden die nodig zijn als input voor de analyse. Dat betreft in eerste instantie een bestand met de coördinaten van de zwaartepunten van de regio's die worden onderzocht (in het huidige onderzoek PC4-gebieden). In aanvulling daarop wordt bij de visualisatie ook gebruikgemaakt van de burenmatrix ('adjacency tabel') die als input voor FlexScan wordt gebruikt. In die burenmatrix is vastgelegd welke gebieden direct aan elkaar grenzen. Bij de visualisatie van de uitkomsten wordt in kleuren aangegeven wat de meest waarschijnlijke clusters zijn. Elk cluster wordt getoond in een aparte kleur. Daarbij wordt niet de regio ingekleurd, maar de

zwaartepunten met de verbindingen met naburige regio's die behoren tot hetzelfde cluster. Hierbij kan zelf bepaald worden wat als afkappunt wordt gehanteerd voor het visualiseren van de gevonden clusters.

Aangezien de FlexScan-analyse er meer op is gericht om een zo nauwkeurig mogelijke *scan-analyse* uit te voeren, impliceert dit dat de weergave van de clusters nauwkeuriger kan zijn dan de resultaten die zijn afgeleid uit SatScan-analyse, met name indien daarbij de circulaire scantechniek is toegepast. Dit impliceert dat, afhankelijk van het onderwerp en privacygevoeligheid, visualisatie van de resultaten van scananalyses niet altijd mogelijk is in openbare rapportages. Het kan dan wel een hulpmiddel zijn voor de direct bij het onderzoek betrokken organisaties.

In aanvulling op de statistische clusteranalyse ter beoordeling van eventuele clustering in ruimte en/of tijd is eveneens bekeken welke methoden er beschikbaar zijn voor de visualisatie van de onderzoeksbevindingen. Voor beide *scananalysemethoden* zijn mogelijkheden beschikbaar om de resultaten te visualiseren, net als voor *disease mapping*. De zogenoemde 'geovisual analytics'-benadering die is geïmplementeerd binnen het programma VIT biedt visualisatiemogelijkheden van SatScan-analyses, waarbij ook ondersteuning wordt geboden voor selectie van blootstellingsparameters en andere covariaten. Dat is zinvol, omdat de resultaten in zekere mate sensitief zijn voor de selectie van de parameters die worden gehanteerd voor de analyses. De keuze van de parameters kunnen van invloed zijn op de bevindingen (Chen, 2008). Het is daarbij wel van belang om erop te wijzen dat geovisuele analysemethoden als VIT ook beperkingen hebben. Visuele methodieken lijken aantrekkelijker dan puur statistische technieken, maar de interpretatie ervan kan moeilijker zijn dan die van traditionele statistische parameters, zeker bij gebruik door hiermee onervaren personen. Vandaar dat de ontwikkelaar van de VIT-software (Pennsylvania University) aanraadt om uitkomsten met epidemiologen en domeinexperts te bespreken voordat er implicaties aan worden verbonden, zoals bepaalde maatregelen of gezondheidsbeleid. Dit impliceert dat verantwoorde visualisatie van onderzoeksbevindingen een aandachtspunt is voor toepassing in de GGD-context.

## 8.5 Visualisatie van de data op geaggregeerd niveau

In aanvulling op de visualisatie van uitkomsten van de analyses is het in principe ook mogelijk om inzichtelijk te maken hoe de oorspronkelijke data variëren in ruimte en tijd. Om dat mogelijk te maken is gebruikgemaakt van hiervoor ontwikkelde software (Statplanet). Daarbij zijn de van IKNL verkregen gegevens op postcodeniveau geaggregeerd tot gemeentenniveau, waarna deze in een viewer kunnen worden bekeken, zowel statisch als dynamisch. In de dynamische weergave is de variatie door te tijd goed te zien.

Uit bespreking in de interne en externe begeleidingscommissie van het onderzoek is gebleken dat deze wijze van visualisatie verhelderend is, onder meer om de natuurlijke variatie van zeldzame aandoeningen als leukemie inzichtelijk te maken. In overleg met IKNL zal daarom als resultante van dit onderzoek ook een filmpje ter beschikking worden gesteld aan de GGD waarin de opgetreden variatie inzichtelijk wordt gemaakt.



## 9 Literatuur

- Aamodt, G., S.O. Samuelson en A. Skrondal (2006). A simulation study of three methods for detecting disease clusters. *International Journal of Health Geographics*, 5: 15 <http://www.ij-healthgeographics.com/contents/5/15>
- Abellan, J.J., S. Richardson en N. Best (2008). Use of Space-Time models to investigate the stability of Patterns of Disease. *Environmental Health Perspectives* 116, 8, 1111-1119.
- ACS (2014). <http://www.cancer.org/cancer/cancercauses/othercarcinogens/generalinformationaboutcarcinogens/cancer-clusters>, geraadpleegd op 24 februari 2014.
- Alexander, F. (1999). Clustering of Childhood Leukaemia: the Euroclus project. *Radiat Environ Biophys* 71-4.
- Amin, R., A. Bohnert, L. Holmes, A. Rajasekaran en C. Assanasen (2010). Epidemiologic mapping of Florida childhood cancer clusters. *Pediatric Blood Cancer* 54: 511-518.
- ARCGIS (2013). <http://www.arcgis.com/home/item.html?id=63d6023447324429a3f474b23bf46dfc>, geraadpleegd 2013.
- Beale, L. (2012). *Area Weighting Models: three different methods for moving data between different geographies*. ArcGIS.
- Belson, M., B. Kingsley en A. Holmes (2007). Risk Factors for Acute Leukaemia in Children: a review. *Environmental Health Perspectives* 115, 1, 138-145.
- Besag, J.E., J.C. York en A. Mollié (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics* 43: 1-59.
- Bogers, R.P., D. Schram-Bijkerk, J.L.A. Devilee, A.B. Knol en O.R.P. Breugelmans (2014). *Verkenning van mogelijkheden voor onderzoek naar blootstelling aan gewasbeschermingsmiddelen bij omwonenden*. RIVM rapport 630030002.
- Botella-Rocamora, P., A. López-Quílez en M.A. Martínez-Beneito (2013). Spatial moving average risk smoothing. *Statistics in Medicine* 32, 15: 2595-2612.
- Buntinx, F., H. Geys, D. Lousbergh *et al.* (2003). Geographical differences in cancer incidence in the Belgian province of Limburg. *European Journal of Cancer* 39: 2058-72.
- CBS (2012). *Statistische gegevens per vierkant*. Den Haag/Heerlen.
- Chen J, et al (2008), Supporting the Process of Exploring and Interpreting Space–Time Multivariate Patterns: The Visual Inquiry Toolkit. *Cartogr Geogr Inf Sci* 35(1); 33-50.

- COMARE Committee on Medical Aspects of Radiation in the Environment (2011). Further consideration of the incidence of childhood leukemia around nuclear power plants in Great Britain. The distribution of childhood leukaemia and other childhood cancer in Great Britain. Chilton, Dicot, UK: Health Protection Agency; Fourteenth Report.
- DeChello LM, Sheehan TJ (2007). Spatial analysis of colorectal cancer incidence and proportion of late-stage in Massachusetts residents: 1995-1998. *Int J Health Geogr*, 6:20.
- Dusseldorp, A., E.F. Hall en H.P.F.M. van Poll (2013). *Meldingen van milieugerelateerde gezondheidsklachten bij GGD'en. Vierde inventarisatie (2011-2012)*. RIVM rapport 20000004.
- Eenwyk, J. van, L. Bensley, D. McBride *et al.* (1999). *Addressing community health concerns around SeaTac Airport: Second Report*. Washington State Department of Health.
- Elliott, P., J.C. Wakefield, N.G. Best en D.J. Briggs (2000). *Spatial Epidemiology: Methods and Applications*. Oxford, Oxford University Press.
- Fang, Z., M. Kulldorff en D.I. Gregorio (2004). Brain cancer in the United States, 1986-95: A geographic analysis. *Neuro-Oncology* 6: 179-187.
- Gezondheidsraad (2001). *Ongerustheid over lokale milieufactoren: risicocommunicatie, blootstellingsbeoordeling en clusteronderzoek*. Den Haag; publicatienr 2001/10. ISBN 90-5549-370-8.
- Gezondheidsraad (2003). *Gezondheid en milieu: mogelijkheden van monitoring*. Den Haag; publicatie nr 2003/13. ISBN 90-554-492-5.
- Gezondheidsraad (2012) *Childhood leukaemia and environmental factors*. The Hague: Health Council of the Netherlands; publication no. 2012/33. ISBN: 978-90-5549-926-7.
- Gezondheidsraad (2014). *Gewasbescherming en omwonenden*. Den Haag; publicatienr. 2014/02.
- Han, D.W., P.A. Rogerson, J. Nie *et al.* (2004). Geographic clustering of residence in early life and subsequent risk of breast cancer (United States). *Cancer Causes and Control* 15: 921-929.
- Han J, et al (2013). Determining optional cluster reporting sizes for spatial Scan statistics. Manuscript in preparation.
- Hegger, C., M.A. van der Aa, M. Drijver *et al.* (2012). *Kankerclusters: GGD-richtlijn medische milieukunde*. RIVM Rapport 609400002.
- Hegger, C., A.M.J. Reedijk (2013). Kinderleukemie in een wijk met een hoogspanningslijn. Aanpak volgens de GGD Richtlijn "Kankerclusters". *Ned Tijdschr Geneeskunde* 157: 1-5.
- Hjalmar, U., M. Kulldorff, G. Gustafsson en N.Nagarwalla (1996). Childhood leukemia in Sweden: Using GIS and a spatial scan statistic for cluster detection. *Statistics in Medicine* 15: 707-715.

- Hjalmar, U., M. Kulldorff, Y. Wahlquist en B. Lannering (1999). Increased incidence rates but no space-time clustering of childhood malignant brain tumors in Sweden. *Cancer* 85:2077-2090.
- Hund, L., J.T. Chen, N. Krieger en B.A. Coull (2012). A geostatistical approach to large-scale disease mapping with temporal misalignment. *Biometrics* 68, 849-858.
- Imai, J. (1998). Spatial disease clustering in Kochi prefecture in Japan. *National Institute of Public Health Epidemiology and Biostatistics Research* 57-96 (in het Japans).
- Kelfkens G. , Pennders R.M.J., Pruppers M.J.M (2003). Plannen voor nieuwbouwwoningen bij bovengrondse hoogspanningslijnen. RIVM rapport 610150004
- Kelsall, J., J. Wakefield (2002). Modeling Spatial Variation in Disease Risk: A Geostatistical Approach. *Journal of the American Statistical Association* Vol. 97, No. 459, pp. 692-701.
- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* 19, 2555-2567.
- Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics: Theory and Methods* 26, 1481-1496.
- Kulldorff M. (2009). Information Management Services, Inc. SaTScan™ v8.0: Software for the spatial and space-time scan statistics. <http://www.satscan.org>.
- Lawson, A.B. (2009). *Bayesian disease mapping - Hierarchical modeling in spatial epidemiology*. Chapman & Hall/CRC.
- Michelozzi, P., A. Capon, U. Kirchmayer, et al. (2002). Adult and childhood leukemia near a high-power radio station in Rome, Italy. *American Journal of Epidemiology* 155: 1096-1103.
- Mulder, Y.M., M. Drijver en I.A. Kreis (1994). Case-control study on the association between a cluster of childhood haematopoietic malignancies and local environmental factors in Aalsmeer, The Netherlands. *Journal of Epidemiology and Community Health* 48: 161-165.
- Neutra RR. Counterpoint from a cluster buster (Reviews and Commentary). *Am J Epidemiology* 1990; 132(1): 1-8.
- Poll, R. van, M. Drijver (2001). Inventarisatie van milieugerelateerde ziekteclusters (2001). *Tsg/Tijdschrift voor gezondheidswetenschappen* 79: 9-15.
- Plas M. van der , Houthuijs D.J.M., Dusseldorp A., Pennders R.M.J., Pruppers M.J.M. (2001). Magnetische velden van hoogspanningslijnen en leukemie bij kinderen. RIVM rapport 610050007.



- Rothman KJ. A sobering start for the cluster busters' conference. *Am J Epidemiology* 1990; 132 Suppl(1): S6-S13.
- Rue, H., S. Martino en N. Chopin (2009). Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). *Journal of the Royal Statistical Society Series B*, 71, 319-392.
- Schmiedel, S., M. Blettner, P. Kaatsch en J. Schutz (2010). Spatial clustering and space-time clusters of leukemia among children in Germany, 1987–2007. *Eur J Epidemiology* 2010 Sep;25(9):627-33
- Sheehan, T.J., L.M. DeChello, M. Kulldorff *et al.* (2004). The geographic distribution of breast cancer incidence in Massachusetts 1988-1997, adjusted for covariates. *International Journal of Health Geographics* 3: 17.
- Song, C., M. Kulldorff (2003). Power evaluation of disease clustering tests. *International Journal of Health Geographics* 2: 9.
- Takahashi, K., T. Yokoyama en T. Tango (2010). *FlexScan: Software for the Flexible Scan Statistic*. National Institute of Public Health, Japan.
- Tango, T. (2000). A test for spatial disease clustering adjusted for multiple testing. *Statistics in Medicine* 19, 191-204.
- Visser, O., J. van Wijnen en F. van Leeuwen (2005). Incidence of cancer in the area around Amsterdam Airport Schiphol in 1988–2003: a population-based ecological study. *BMC Public Health* Volume 5, issue 1, 127-132.
- Wartenberg, D. (2002). Investigating Disease Clusters: Why, When and How? *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 164, Part 1, pp. 13-22.
- Wheeler, DC. (2007). A comparison of spatial clustering and cluster detection techniques for childhood leukaemia incidence in Ohio, 1996-2003. *Int J Health Geogr* 27, 6-13.

## BIJLAGE A. Begeleiding en advies

### **Externe begeleidingscommissie**

De externe begeleidingscommissie is voortgekomen uit de werkgroep van de richtlijn medische milieukunde 'Kankerclusters'. Deze begeleidingscommissie heeft onder andere het projectvoorstel en de conceptrapportage beoordeeld op bruikbaarheid van de resultaten in de GGD praktijk. De commissie bestond uit:

- mw. C. Hegger, arts (GGD Rotterdam-Rijnmond)
- mw. M. van der Aa (IKNL)
- mw. M. Drijver, arts (wetenschappelijke staf Gezondheidsraad)
- dhr. F. Duijm, arts (GGD Groningen)
- mw. drs. P.A.M.J. Esser (GGD Zuid Limburg)
- mw. R.H. Keuken, arts (GGD Kennemerland)
- mw. ir. A.G. Zijlstra (GGD IJsselland)

### **Interne begeleidingscommissie**

Binnen het RIVM is een aantal deskundigen opgetreden als adviseur voor het clusteronderzoek. Hierin zijn diverse disciplines vertegenwoordigd (epidemiologie, statistiek, GIS, communicatie, gezondheidswetenschappen). Deze begeleidingscommissie dacht onder andere mee met de methodische vraagstukken en de implicaties van het onderzoek en bestond uit:

- mw. prof. dr. H.C. Boshuizen
- dhr. ir. D.J.M. Houthuijs
- dhr. dr. ir. J. van de Kasstelee
- dhr. prof. Dr. Ir. E. Lebret
- mw. drs. C.M.A.G. van Wiechen

### **Methodologisch adviseur en data-aanvraagbegeleider vanuit IKNL**

Dr. O. Visser, hoofd kankerregistratie bij IKNL, is geraadpleegd bij de voorbereiding voor het opstellen van het onderzoeksplan, gezien zijn grote ervaring met de beoordeling en analyse van kankercluster meldingen.



## BIJLAGE B. Disease mapping (Bayesiaanse analyse)

Tabel B1. Disease mapping: Tabel incidenties over alle jaren (1995-2010).

Diagnosegroep	Leeftijd	Gem. incidentie per pc4*	SD	Laagste aantal per pc4	Hoogste aantal per pc4	Totaal aantal
Alle kankersoorten	0-19 jaar	0,153	0,428	0	14	9912
	0 jaar	0,011	0,114	0	4	710
	1-4 jaar	0,038	0,202	0	3	2467
	5-9 jaar	0,026	0,165	0	3	1699
	10-14 jaar	0,028	0,171	0	3	1823
	15-19 jaar	0,05	0,237	0	13	3213
G1: Leukemie	0-19 jaar	0,041	0,208	0	3	2673
	0 jaar	0,002	0,045	0	1	132
	1-4 jaar	0,017	0,13	0	2	1069
	5-9 jaar	0,01	0,099	0	2	631
	10-14 jaar	0,007	0,086	0	2	468
	15-19 jaar	0,006	0,077	0	2	373
G2: Hersentumoren	0-19 jaar	0,016	0,127	0	2	1037
	0 jaar	0,001	0,028	0	1	52
	1-4 jaar	0,005	0,068	0	2	296
	5-9 jaar	0,005	0,068	0	2	299
	10-14 jaar	0,003	0,059	0	2	222
	15-19 jaar	0,003	0,051	0	1	168
G3: Overig	0-19 jaar	0,096	0,335	0	14	6202
	0 jaar	0,008	0,1	0	4	526
	1-4 jaar	0,017	0,135	0	3	1102
	5-9 jaar	0,012	0,111	0	2	769
	10-14 jaar	0,018	0,134	0	2	1133
	15-19 jaar	0,041	0,217	0	13	2672

\* n = 64736 (4046 PC4-gebieden maal 16 jaar)



## BIJLAGE C. Omgevingsgerelateerde risicofactoren kinderleukemie

*In onderstaande tabel is een overzicht opgenomen over de waarschijnlijkheid dat de genoemde omgevingsfactoren een rol spelen in het ontstaan van leukemie bij kinderen. Bron: Gezondheidsraad, 2012.*

Potential exposure	Type of leukaemia	Likelihood of causality	Proposed measures	Proposed research
<i>Ionising radiation</i>				
Ionising radiation	Childhood leukaemia	Established	Exposure reduction (ALARA)	Assessment of exposure of unborn children and infants. Biological effects of pre- and postnatal diagnostic X-rays (esp. CT scans)
<i>Electromagnetic fields</i>				
Extremely Low Frequency magnetic fields	Childhood leukaemia	Possible	Exposure reduction	
	ALL	Possible	Exposure reduction	
	AML	Unknown	Exposure reduction	
Radiofrequency radiation	Childhood leukaemia	Unknown		Monitoring actual exposures
<i>Ultrasound</i>				
Diagnostic ultrasound scans	Childhood leukaemia		Optimisation and standardisation	Effects of more intensive or frequent ultrasound scans
- routine/limited		Unlikely	Optimisation and standardisation	
- frequent/intensive		Unknown	Optimisation and standardisation	
<i>Chemicals</i>				
Pesticides	Childhood leukaemia	Possible to likely	Preconceptional counselling to reduce exposure	Monitoring exposures among occupationally exposed women

Potential exposure	Type of leukaemia	Likelihood of causality	Proposed measures	Proposed research
	ALL	Possible to likely	Preconceptional counselling to reduce exposure	Monitoring exposures among occupationally exposed women
	AML	Uncertain to possible	Preconceptional counselling to reduce exposure	Monitoring exposures among occupationally exposed women
Benzene	Childhood leukaemia	Likely	Preconceptional counselling to reduce exposure	Monitoring exposures in specific situations
Formaldehyde	Childhood leukaemia	Possible	Exposure reduction	
Styrene	Childhood leukaemia	Uncertain	Exposure reduction	
Other organic solvents	Childhood leukaemia	Unknown	Exposure reduction	Epidemiological and experimental studies for commonly encountered organic solvents
Inorganic arsenic in drinking water	Childhood leukaemia	Possible	Exposure reduction	
Paternal tobacco smoking	Childhood leukaemia	Likely	Refrain from tobacco smoking	
Maternal tobacco smoking	Childhood leukaemia	Possible	Refrain from tobacco smoking	
Parental marijuana smoking	AML	Uncertain	Refrain from marijuana smoking	Epidemiological studies
	ALL	Unknown	Refrain from marijuana smoking	Epidemiological studies
Parental alcohol consumption	Childhood leukaemia	Possible	Refrain from alcohol consumption	
Maternal cured meat intake	Childhood leukaemia	Uncertain	Exposure reduction	Effects of cured meat ingestion during pregnancy

Potential exposure	Type of leukaemia	Likelihood of causality	Proposed measures	Proposed research
PCBs	Childhood leukaemia	Likely	Preconceptional counselling to reduce exposure	Epidemiological studies
Plasticisers	Childhood leukaemia	Possible	Preconceptional counselling to reduce exposure	Epidemiological studies
Other chemicals	Childhood leukaemia	Unknown	Exposure reduction	Epidemiological studies on agents associated with DNA damage
Biological and other factors				
Infectious exposures	Childhood leukaemia	Uncertain		More research in this area
	ALL	Uncertain		
	AML	Uncertain		
High birth weight	ALL	Possible		
	AML	Possible		
Low birth weight	ALL	Unknown		
	AML	Uncertain		
Socio-economic status	Childhood leukaemia	Unknown		



In onderstaande tabel is een overzicht opgenomen over de waarschijnlijkheid van milieugerelateerde beschermende factoren in het ontstaan van leukemie bij kinderen (in de categorie biologische en andere factoren). Bron: Gezondheidsraad, 2012.

Potential exposure	Type of leukaemia	Likelihood of causality	Proposed measures	Proposed research
Early social contacts	ALL	Likely	Encourage	
	AML	Unknown	Encourage	
Allergies	ALL	Possible		
	AML	Unknown		
Breast-feeding	ALL	Likely	Encourage	
	AML	Likely	Encourage	
Maternal vitamin supplementation during pregnancy	ALL	Uncertain		More research on protective effect of vitamin supplementation during pregnancy
	AML	Unknown		
Maternal folate supplementation	ALL	Uncertain	Encourage	
	AML	Unknown	Encourage	

## BIJLAGE D. Afkortingen en toelichting van gebruikte termen

### Afkortingen

ALL	Acute lymfatische leukemie
AML	Acute myeloïde leukemie
CAR	Conditioneel Auto Regressiemodel (bayesiaanse analyse)
CBS	Centraal Bureau voor de Statistiek
GBA	Gemeentelijke Basisadministratie
IKNL	Integraal Kankercentrum Nederland
NKR	Nederlandse Kanker Registratie
PC4, PC6	Viercijferige of zescijferige postcode
SKION	Stichting Kanker Oncologie Nederland

### Gehanteerde Termen

DIC	Een grootheid waarmee voor een en dezelfde uitkomst kan worden bepaald of uitbreiding van een model met extra verklarende variabelen statistisch gezien een verbetering van het model oplevert. Hierbij geldt in principe: hoe lager, hoe beter. Indien bij het toevoegen van een effect aan een model de DIC niet wordt verlaagd is de toevoeging statistisch gezien niet zinvol
Haematopoiëtische maligniteiten	Kanker van bloed/bloedvormend weefsel
INLA	Pakket ten behoeve van bayesiaanse analyse
CI	Confidence Interval: betrouwbaarheidsinterval rondom risicoschatting.
Correctie voor meervoudig toetsen	Methodiek om te corrigeren voor het verkrijgen van statistisch significante resultaten die (alleen) zijn veroorzaakt door het uitvoeren van een groot aantal statistische toetsen en daardoor een toevalsbevinding kan zijn.
SatScan	Software voor data-analyse waarbij gebruik wordt gemaakt van 'space-time scan statistics'
SIR	Gestandaardiseerde incidentieratio: risicomaat
Scananalyse	In rapport gebruikte term voor ruimtelijke en tijd-ruimteanalyses gebaseerd op scanstatistiek, met de pakketten SatScan en/of FlexScan.

**RIVM**

*De zorg voor morgen begint vandaag*