

Het Geïntegreerd Informatie Netwerk Ouderenzorg als basis voor data-exploratie en complexe (beleidsgeoriënteerde) analyses over patiëntkenmerken en gebruik van zorg[#]

P. Moorer, E.W. Wolffensperger en W.O. Zijlstra**

Samenvatting

Het Geïntegreerd Informatie Netwerk Ouderenzorg (GINO) is een registratie systeem voor het vastleggen van gegevens van cliënten van de Thuiszorg, verpleeg- en verzorgingshuizen. Het GINO kent een groot aantal mogelijkheden om analyses uit te voeren. Omdat in het registratiesysteem een zeer grote verscheidenheid aan gegevens vastlegd, zijn er ook interessante mogelijkheden om data mining op deze gegevens toe te passen. De ontwikkelaars van het GINO zijn echter bang dat deze "verborgen rijkdom" nooit geëxploreerd zal worden, omdat maar weinig mensen de mogelijkheden zien. De auteurs zien ook een mogelijkheid tot data mining die met de huidige technieken niet mogelijk is.

**RU Groningen, GINO b.v., A. Deusinglaan 1, 9713 AV, Groningen
tel: 050-3636271, email: e.w.wolffensperger@med.rug.nl

#

Een eerdere versie van dit artikel is verschenen in "SSS '97, Verborgen Rijkdom, CBS, Voorburg, 1997, pp 169-189.

1. Inleiding

Op veel plaatsen in Nederland rusten grote databases die in het kader van concrete onderzoeksprogramma's tot stand zijn gebracht. Het besef leeft onder beheerders en gebruikers dat een groot deel van deze gegevens tot dusverre onvoldoende is benut. De toegankelijkheid van de datasets en de specifieke historische samenstelling van de data kunnen een probleem vormen. Bovendien zou het wenselijk zijn verschillende datasets zodanig met elkaar te kunnen verbinden dat met een geïntegreerd gebruik van verschillende bestanden specifieke analyses mogelijk worden.

In de periode 1990 tot 1997 is het Geïntegreerd Informatie Netwerk Ouderenzorg (GINO) tot ontwikkeling gebracht, in eerste aanleg als project binnen de RUG, sedert 1994 binnen een verzelfstandigde onderneming. GINO b.v. is specifiek gericht op de ontwikkeling van en distributie van cliënt-volgsystemen in de zorg. De ontwikkeling van kennissystemen, van kwaliteitsinstrumenten voor procesbeheersing en -evaluatie en voor vergroting van de effectiviteit en efficiency van zorgorganisaties staan daarbij op de voorgrond.

Het GINO vormt een rijk informatiesysteem dat rechtstreeks vanuit de werkprocessen van indicatiestelling en zorguitvoering wordt gevoed. De opgeslagen gegevens dienen ook primair de uitvoering van werkprocessen, maar zijn door formulering en historische dossier-opslag bij uitstek geschikt voor systematisch wetenschappelijk onderzoek.

De groeiende spreiding van het gebruik van het GINO onder indicatie-organen en aanbieders van zorg in alle delen van Nederland maakt het geschikt voor grootschalige analyse rond patiënt-kenmerken van oudere hulpvragers en chronisch zieken, en naar het verloop van de hulpverlening aan deze groepen.

Toch bestaat ook bij dit systeem weer het gevaar van een beperkt, en zeker suboptimaal gebruik voor wetenschappelijk onderzoek. Twee voor de hand liggende oorzaken daarvoor zijn dat het GINO buiten de betrokken zorgsectoren

zelf relatief weinig bekendheid geniet. Die bekendheid is er binnen de zorgsector wel, maar daar ontbreekt, paradoxaal genoeg, het inzicht in en de oriëntatie op het gebruik van een machtig onderzoeksframe als GINO voor de evaluatie van de kwaliteit van het eigen functioneren en van lopende ontwikkelingen in de zorgsector.

Het doel van dit paper is een schets te geven van de achtergronden van het GINO als proces- en zorgondersteunend systeem voor indicatiestelling en zorguitvoering en als evaluatie-instrument in zorg en beleid. Verder wordt aangegeven op welke manier van GINO in combinatie met andere databases gebruik kan worden gemaakt.

2. Het Geïntegreerd Informatie Netwerk Ouderenzorg

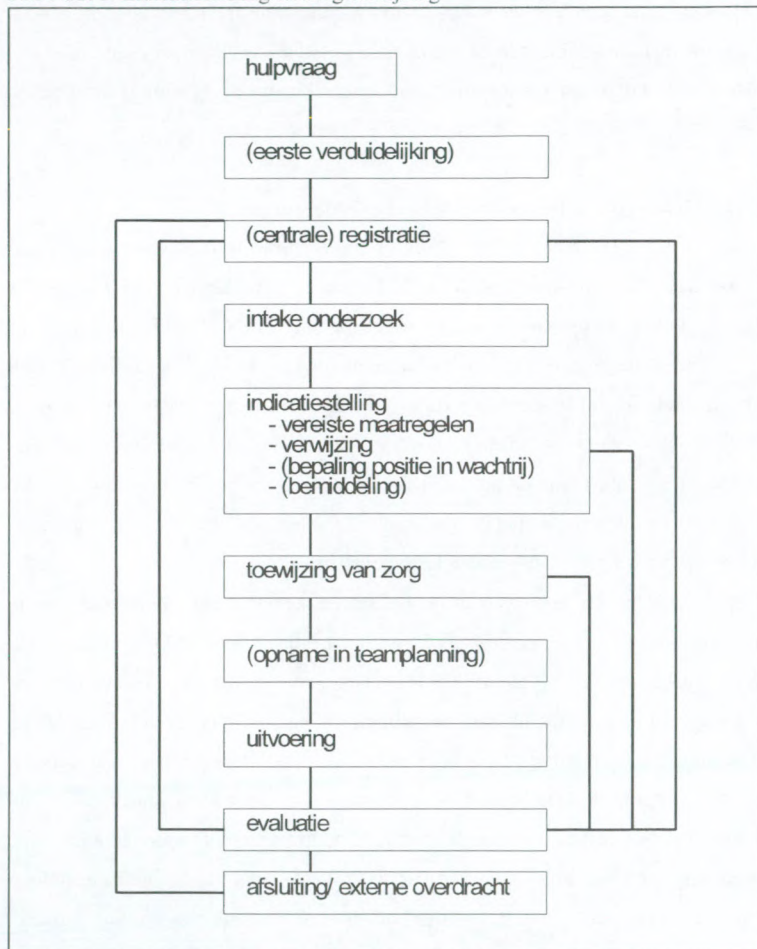
Sinds 1990 is het GINO vanuit het Noordelijk Centrum voor Gezondheidsvraagstukken van de RU Groningen het Geïntegreerd Informatie Netwerk Ouderenzorg (GINO) tot ontwikkeling gebracht. GINO is ontwikkeld als een kwaliteitsinstrument voor procesondersteuning in de alledaagse praktijk van indicatiestelling en zorguitvoering en de evaluatie ervan. Belangrijk achtergrondcriterium is steeds geweest dat vanuit het informatiesysteem hoogwaardige informatie uit het hulpverleningsproces ter beschikking kon worden gesteld voor wetenschappelijk onderzoek zonder dat daarvoor voortdurend hulpverleners behoeften te worden lastiggevallen.

Uitgangspunt bij de opbouw van de basiselementen van het informatiesysteem zoals de begrippen, vragenlijsten, formulieren binnen het GINO (Zijlstra et al, 1991 en Linschoten et al, 1990) is steeds geweest dat de kenmerken van hulpvragers bij verschillende sectoren (thuiszorg, verzorgings- en verpleeghuizen, experimentele voorzieningen) met één gemeenschappelijk begrippenkader konden worden vastgelegd. Daarbinnen bestaat natuurlijk een nadere specificatie van elementen die slechts op een deel van de hulpvragers of voorzieningen van toepassing zijn. Een apart kernpunt van het GINO is dat het de indicatiestelling voor een brede range van zorgfuncties binnen één informatiesysteem ondersteunt,

waarbij de elektronische overdracht van intake-gegevens naar de uitvoering 'in de backoffice' van de zorginstelling is gegarandeerd.

Juridisch en in termen van uitvoering is er een principieel onderscheid tussen de fase van indicatiestelling en zorguitvoering. In de alledaagse werkelijkheid hangen beide nauw samen.

FIGUUR 1: Indicatiestelling en zorgtoewijzing



Indicatiestelling in de zorgsector heeft, kort gezegd, de volgende functies: (1) het - professioneel adequaat - vaststellen van de aanwezige problematiek waarop de hulpvraag berust, en (2) het toetsen van de rechtmatigheid van de gewenste toegang tot professionele ondersteuning.

Een indicatie wordt bij veel cliënten niet éénmalig gesteld: na en/of naast elkaar wordt een combinatie van verstrekkingen aangeboden en -in beperkte mate gecoördineerd- uitgevoerd.

Tijdens het zorgproces kan op basis van nieuwe ontwikkelingen rond een patiënt de situatie zich zodanig wijzigen dat de hulpverlening aanpassing vraagt: een herindicatie of nieuwe beoordeling.

In figuur 1 is de samenhang van intake en evaluatie tijdens de zorguitvoering schematisch weergegeven. Positieve indicatiestelling wordt gevolgd door toewijzing en uitvoering van zorg. Evaluatie leidt mogelijk tot bijstelling. Bij een externe verwijzing voor herindicatie wordt van de cliënt formeel een nieuwe aanvraag in behandeling genomen, het gehele schema van gegevensverzameling, analyse en toetsing wordt opnieuw van boven naar beneden afgewerkt. Bij een interne herziening kan evaluatie leiden tot bijstelling van de uitvoering binnen een uitvoeringsteam, waarbij ook een planmatige herziening van zorgtoewijzing optreedt.

Er kan overigens ook binnen de eigen instelling sprake zijn van een formele indicatiestelling voor specifieke voorzieningen of categorale geldstromen zoals substitutiegelden en middelen voor intensieve en terminale thuiszorg. Een cliënt kan zo op hetzelfde moment formeel gebruik maken van budgetten binnen één instelling op basis van een meervoudige indicatiestelling, en hij kan tegelijkertijd gebruik maken van zorg die door verschillende instellingen wordt verleend, al dan niet op basis van verschillende indicaties.

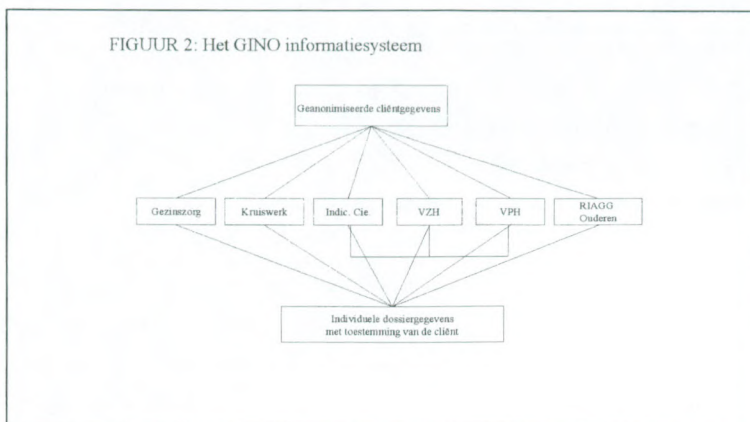
De nieuw opgezette geïntegreerde indicatiestelling berust op een vergelijkbaar

model, waarbij in beginsel alle aanvragen voor verstrekkingen van een hulpvrager binnen één centraal kader worden afgewogen, waarbij alle bronnen uit het regionale zorgcircuit kunnen worden benut. Dat kan overigens binnen uiteenlopende juridische kaders gebeuren.

De afweging van alle hulpvragen binnen één kader vraagt om een gemeenschappelijk en uniform begrippenkader waarin de gegevens worden vastgelegd. Een tussen sectoren goed vergelijkbare vastlegging van gegevens vormt de basis van het GINO: intersectorale uitwisseling van zorginformatie over individuele gebruikers bij overdracht van individuele patiëntengegevens en gedeelde analyses over groepen cliënten tussen zorginstellingen. Een dossier kan in volledige vorm dan wel in verschillende vormen van samenvattende rapportage, met toestemming van de patiënt, elektronisch worden overgedragen tussen verschillende GINO-softwarepakketten die functioneren in de thuiszorg, bij indicatiecommissies en bij verzorgingstehuis en verpleeghuis, zie

De mogelijkheid bestaat om de gegevens vanuit verschillende instellingen weg te schrijven naar een externe database die bijvoorbeeld op regio- of landelijk niveau wordt onderhouden, zie figuur 2. Deze externe database maakt een integratie van samenhangende processen in de zorguitvoering tussen instellingen mogelijk. Tevens zou het onder voorwaarden rechtstreeks als database voor wetenschappelijk onderzoek kunnen fungeren.

FIGUUR 2: Het GINO informatiesysteem



Een dergelijke database vormt een persoonsregistratie. Dat betekent dat deze met sterke waarborgen ter bescherming van de privacy van cliënten en een duidelijke regulering van het gebruik en beveiliging wordt omgeven.

3. Analyse mogelijkheden binnen het GINO

In de voorafgaande paragraaf is reeds ingegaan op de gegevens die in het GINO worden opgeslagen. In deze paragraaf zal nader worden ingegaan op de analyse mogelijkheden, maar voordat daarvan een overzicht wordt gegeven, wordt eerst dieper worden ingegaan op de gegevens die daadwerkelijk worden vastgelegd. Voor een voorbeeld van de gegevensset die in het GINO worden vastgelegd zie Wolffensperger et al, 1996. Geïnteresseerden kunnen -kosteloos- een informatiepakket met een volledige gegevensset verkrijgen bij GINO b.v.

In eerste instantie worden persoonsgegevens en administratieve gegevens vastgelegd: geslacht, leeftijd, woonplaats (postcode), en leefsituatie, maar ook gegevens zoals naam en adres van de huisarts en medische diagnoses. Daarnaast wordt een aantal andere hoofdgroepen van informatie opgenomen:

1. de hulpvraag van de patiënt
2. de aanwezige problematiek naar professionele maatstaven,
3. de mogelijkheden bestaande informele netwerken in de zorg te betrekken,
4. de geïndiceerde hulpverlening en
5. de verstrekte hulpverlening.
6. hoe deze wordt uitgevoerd en zo nodig bijgesteld.

De afgelopen jaren is de ontwikkeling van Regionale Indicatie Organen in een versnelling gekomen. Waar eerst onder GINO-programmatuur de samenwerking tussen aanbieders de stimulans vormde voor gegevensafstemming en –uitwisseling is en nu een noodzaak tot het gebruik van één gemeenschappelijk analysekader voor indicatie-onderzoek voor de brede sectoren van wonen, zorg en welzijn.

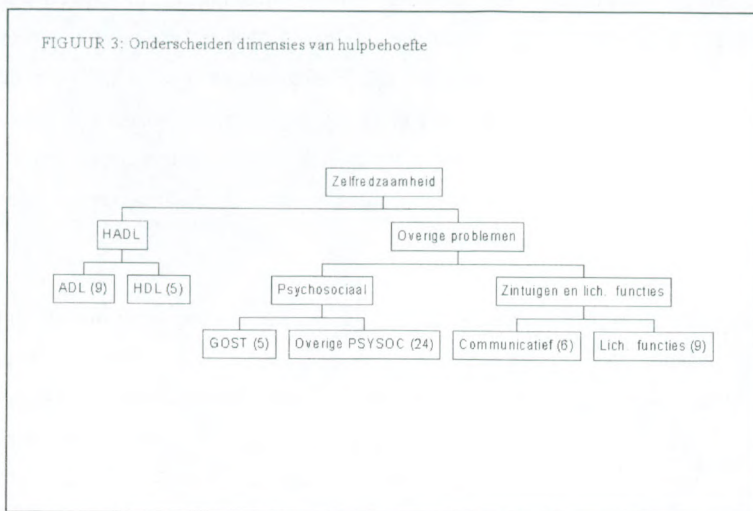
Deze ontwikkeling is landelijk gecoördineerd door het Breed Indicatie Overleg (BIO) waarin alle betrokken partijen deelnemen. Vanuit het BIO zijn proces-protocollen en concept-formulieren gepresenteerd die de praktijk van het nieuwe brede indicatie-onderzoek zouden moeten ondersteunen.

De procesgang voor de brede indicatiestelling die binnen de GINO software wordt ondersteund volgt volledig de criteria van het BIO-protocol van voorjaar 1997. Verschillende specifieke systemen als het RAI (Resident Assessment Instrument) zijn in het GINO voor gebruik in specifieke instellingstypen geïntegreerd.

Een cruciaal onderdeel van het systeem vormt de hulpvraag van de patiënt. Op basis van de hulpvraag moet namelijk besloten worden welke zorg de patiënt zou moeten krijgen. Op basis van wetenschappelijk onderzoek zijn voor het in kaart brengen van de hulpvraag een aantal instrumenten (schalen) ontwikkeld (Zijlstra et al, 1991, Wolfensperger en Zijlstra). Het gaat hierbij om het meten van (1) algemene dagelijkse levensverrichtingen (ADL), (2) huishoudelijke dagelijkse levensverrichtingen (HDL), (3) geheugen- en oriëntatiestoornissen (GOST) en (4) psychosociale problematiek (PSYSOC). De psychometrische eigenschappen van deze schalen zijn goed: goede schaalbaarheid, goede betrouwbaarheid en geen aantoonbare DIF (differential item functioning) (Moorer et al. (submitted)). Naast

de items die nodig zijn voor deze schalen, zijn nog een aantal specifiekere items opgenomen (zie figuur 3).

FIGUUR 3: Onderscheiden dimensies van hulpbehoefte



De hulpbehoefte ((zelf)redzaamheid) is mede bepalend voor de zorg die moet worden gegeven. In het GINO wordt de te geven en verkregen zorg vastgelegd op handelingsniveau. D.w.z vastgelegd wordt of een patiënt moet worden geholpen in bij het strijken of het bereiden van maaltijden (HDL-hulp) of bij het wassen of het tandenpoetsen (ADL-hulp) of dat handelingen bij de patiënt moeten worden verricht, zoals het geven van injecties of het bijhouden van de bloeddruk (verpleegtechnische handelingen) of het aanbrengen van een katheter (medisch-technische handeling) of moet worden geholpen bij het bijhouden van de financiën of het wegbrengen en ophalen van kinderen (begeleiding). Sommige van deze handelingen zijn terug te voeren op de aanwezigheid van een specifieke medische diagnose die bepaalde vormen van zorg (injecties etc) noodzakelijk maken.

ICIDH

Vanaf 1996 is het GINO is ertoe overgegaan de hulpbehoefte te omschrijven binnen de kaders van de International Classification of Impairments, Disabilities

and Handicaps (ICIDH) . De ICIDH is als internationale classificatie door Nederland aanvaard en wordt in toenemende mate als referentiekader gebruikt. Daarmee verandert de indeling van hulpbehoefte in specifieke 'beperkingsterreinen' in de zelfredzaamheidssfeer. Bij deze herziening is ervoor zorg gedragen dat de dimensies van hulpbehoefte opnieuw en vergelijkbaar in de registratie van gegevens aanwezig zijn. In specifiek instrument-onderzoek zal de mate van overeenstemming tussen de in de versie 1997 en oudere versies van het GINO gebruikte begrippenkader op item-niveau en op het niveau van meetschalen nader worden onderzocht.

De gegevens worden vastgelegd in een relationeel database systeem (ontwikkeld met Delphi (Borland), bruikbaar met een reeks van databases, werkend onder Windows 3.xx, 95 en NT). Alle verschillende versies van GINO '97 (Thuiszorg, verzorgingshuizen, verpleeghuizen en indicatieorganen) bevatten relevante geselecteerde data-tabellen uit één grote moederdatabase. Daardoor zijn er tussen de verschillende versies van GINO '97 geen data-definitie problemen. Deze problemen waren wel aanwezig in de eerdere versies van GINO.

Om bepaalde datadefinitie problemen tegen te gaan heeft een verregaande opdeling van datavelden plaatsgevonden. Zo wordt de naam van een persoon in drie afzonderlijke velden opgeslagen: geboortenaam, naam van de partner, voornaam en tussenvoegsels. Daardoor bestaat de moederdatabase uit meer dan 800 database-tabellen.

Alle gegevens die ooit voor een patiënt zijn ingevoerd worden in een historisch record vastgehouden en zijn in hun oorspronkelijke samenhang en voor longitudinale analyse beschikbaar.

De gegevens worden gewijzigd als daar aanleiding toe is, bijvoorbeeld een gewijzigde hulpbehoefte of wijzigingen in de toewijzing van zorg, of als gevolg van herindicatie-procedures. Een instelling kan als regel hebben dat alle patiënten tenminste eens in het half jaar opnieuw beoordeeld moeten worden. Impliciet

wordt verondersteld dat gegevens van de laatste (her)indicatie onveranderd geldig zijn zolang ze niet worden bijgesteld.

De gegevens en de wijze van opslaan in het GINO maken analyses in de zorgsector mogelijk die tot dusverre hebben ontbroken. Enkele ervan worden hieronder besproken.

Crossectionele analyse

Hierbij kan gedacht worden aan onderzoek naar de relaties tussen hulpbehoefte en geïndiceerde hulpverlening, tussen hulpbehoefte en verstrekte hulpverlening en tussen geïndiceerde hulpverlening en verstrekte hulpverlening. Een vraagstelling zou kunnen zijn in welke mate hoge ADL-behoefte/ADL-beperkingen leidt/leiden tot het uitvoeren van meer lichamelijke en revaliderende taken door een instelling. Of men zou willen weten of een hogere mate van geheugen- en orientatiestoornissen leidt tot het geven van meer begeleiding. Het kan ook zijn dat de sterktes van de relaties mede afhankelijk zijn van andere factoren, zoals leefsituatie, sekse, leeftijd of de aanwezigheid van ziekte of aandoeningen, zoals MS, dementie of reuma.

Onderzoek naar netwerken van zorgrelaties

Naast onderzoek van de bovenstaande relaties is het GINO geschikt om onderzoek te doen naar netwerken van zorgrelaties van patiënten: welk type patiënten maakt veel gebruik van verschillende hulpvormen, welke samenhang in het aanbod in termen van organisatie en producten constitueert zich rond de hulpvrager.

Longitudinaal onderzoek

De grootste kracht van het GINO is echter de mogelijkheid om longitudinale onderzoeken uit te voeren. De opbouw van het GINO en het gegeven dat de gegevens rechtstreeks afkomstig zijn uit het proces van indicatiestelling en zorguitvoering maken het mogelijk om zowel repeated measures designs (Lindsey, 1993) als multiniveau groei-curve-modellen (Bryk en Raudenbush,

1992; hfst 6) te gebruiken.

Via longitudinaal onderzoek kan onderzocht worden welke ziekten of aandoeningen gepaard gaan met een versnelde verslechtering van stoornissen, beperkingen of handicaps in relatie tot patiënten die geen ziekte of aandoening hebben. Ook kan voor verschillende ziekten of aandoeningen nagegaan worden hoe snel of langzaam hun zorggebruik (op handelingsniveau) of qua overige aanwezige hulpvormen in de tijd toe of afneemt. In dit verband kan opgemerkt worden dat algemeen wordt verondersteld dat chronische ziekten en aandoeningen worden verondersteld een steeds groter beroep te doen op De Zorg, maar dat deze aanname nergens onderbouwd is. Uit onderzoek blijkt dat deze aanname nuancering behoeft (Moorer en Suurmeijer (submitted)).

Het GINO kan niet alleen gebruikt worden als hulpmiddel voor analyses, maar ook als steekproefkader voor onderzoek. Het is vaak heel lastig om potentiële respondenten te vinden voor specifieke aandoeningen, zoals dementie, of specifieke combinaties van zorg (fysiotherapie en ergotherapie) (Moorer en Suurmeijer). Het moge duidelijk zijn dat in dergelijke gevallen samenwerking zal moeten worden gezocht met de zorgaanbiedende instellingen.

4. Externe gegevens en de analysemogelijkheden met GINO

In het GINO zelf zijn uitsluitend gegevens opgenomen over de patiënten en hun directe zorg. Over de instellingen zelf en over financiële gegevens van de instellingen zijn geen gegevens opgenomen. Voor onderzoek kunnen deze gegevens ook op cliëntniveau aan de gegevens in het GINO worden gekoppeld. De daarvoor noodzakelijke sleutelvelden zijn aanwezig (b.v. naam intaker, naam van de instelling etc), hoewel niet noodzakelijkerwijs in hetzelfde dataformaat. Wanneer een rechtstreekse koppeling niet mogelijk is, dan kunnen instellings- en/of regionale kenmerken apart geconstrueerd worden en aan de cliëntgegevens worden gehangen met behulp van de hierboven genoemde sleutelvelden.

Met behulp van de instellingsgegevens en het GINO is het bijvoorbeeld mogelijk om economische analyse te doen op de relatie tussen hulpvraag, functioneel hulpaanbod en de relatieve kostenverschillen die daarbij tussen sectoren, in de tijd, en tussen regio's optreden (Wijk P. van der et al, 1995). Het vraagstuk van de vergelijkbaarheid van kostenopbouw van zorgproducten tussen sectoren is daarbij goed onder controle te brengen. De vergelijking van gebruikersgroepen van de verschillende instellingen vindt naar hulpbehoefte-kenmerken op individueel niveau plaats.

Hierboven is alleen de situatie besproken dat de gegevens afkomstig zijn uit een instelling zonder dat koppeling met externe informatie nodig was. Bij het koppelen van gegevens uit het GINO met externe gegevens doen zich twee verschillende omstandigheden voor: (1) koppeling op individueel niveau en (2) koppeling op een geaggregeerd of hoger niveau.

Koppeling op individueel niveau.

Er is sprake van koppeling op individueel niveau als de zorggegevens van de patiënt/cliënt gekoppeld moeten worden aan gegevens van deze patiënt/cliënt die in een andere instelling of organisatie zijn verzameld. Dit kan bijvoorbeeld het geval zijn als medische gegevens van artsen of vanuit ziekenhuizen over patiënten gekoppeld worden aan gegevens uit het GINO. Dit kan bijvoorbeeld nodig zijn als men na wil gaan welk type patiënten (hoeveelheid aanwezige hulpbehoefte) met welke aandoeningen het vaakst naar de huisarts of een ziekenhuis gaan of om na te gaan welk type patiënten welk type medicijnen krijgen voorgeschreven bij welke aandoening (Linschoten et al, 1994).

Voor onderzoek ontstaat er voor dit type onderzoek een lastig probleem: de toelaatbaarheid van het koppelen van gegevens als gevolg van de privacy-wetgeving. In de meeste systemen worden gegevens opgenomen die koppeling geen enkel probleem zouden laten zijn: persoonsgegevens (inclusief postcode en huisnummer) of SOFI-NUMMER zouden dit probleem makkelijk kunnen

omzeilen. Deze gegevens mogen -zonder toestemming van de patiënt- echter niet gebruikt worden. In de discussie zal dit probleem verder besproken worden en zal aangegeven worden hoe dit probleem opgelost kan worden.

Koppeling op geaggregeerd of hoger niveau.

Van koppeling op geaggregeerd niveau is sprake als er geen noodzaak is om gegevens op individueel niveau te koppelen. Dit is het geval als lokale (gemeentelijke), provinciale gegevens of landelijke gegevens worden gebruikt om de analyses uit te voeren. De vraag is welke analyses dan met name interessant zouden kunnen zijn en welke gegevens daarvoor dan noodzakelijk zijn?.

Een belangrijk vraagstuk zou kunnen zijn de kosten-effectiviteit van instellingen in verschillende regio's van Nederland. Wordt deze beïnvloed door de leeftijdsopbouw in de verschillende regio's of spelen specifieke factoren, zoals de bevolkingsdichtheid een rol? Een Thuiszorginstelling in een relatief dunbevolkt gebied kan te maken krijgen met hogere vervoerskosten dan in Thuiszorginstelling in een grote stad.

Ook andere vragen zouden met behulp van het GINO beantwoord kunnen worden:

1. Heeft de inkomensverdeling in verschillende regio's een belangrijke invloed op het aantal patiënten dat instellingen in de verschillende regio's bedienen?
2. Zijn er regionale verschillen in de hulpbehoefte van patiënten en zo ja waardoor worden deze beïnvloed, heeft dat wellicht te maken met de leeftijdsopbouw, de bevolkingsdichtheid of huishoudenssamenstelling, waardoor men minder een beroep kan doen op mantelzorg?
3. In hoeverre hebben provinciale beslissingen met betrekking tot het bouwen of samenvoegen van verzorgingshuizen en/of verpleeghuizen een effect op het gebruik van de Thuiszorg in een regio?
4. Welke regionale effecten heeft een bezuinigingsmaatregel van de rijksoverheid op de mogelijkheden van instellingen om zorg te verlenen en in hoeverre kan dat gerelateerd worden aan de bevolkingsdichtheid, inkomensverdeling of

leeftijdsopbouw in een regio?

5. Wat zijn de effecten van de brede Regionale Indicatie Organen op de distributie van zorg, bijvoorbeeld op de wachtlijsten, zorgzwaarte en kan dat gerelateerd worden aan de bevolkingsdichtheid, inkomensverdeling of leeftijdsopbouw in een regio?

Uit de bovenstaande vraagstellingen blijkt dat gegevens vanuit het GINO afgezet zullen worden tegen bevolkingsgegevens, zoals leeftijdsopbouw en bevolkingsdichtheid, en gegevens over regionale of landelijke inkomensverdeling. Eigenlijk wordt hiermee geïmpliceerd dat de gegevens van GINO afgezet kunnen worden tegen gegevens van een landelijk of regionaal Geografisch Informatie Systeem (GIS).

Analytisch gezien betekent het koppelen van GINO gegevens aan geaggregeerde externe gegevens dat noodzakelijkerwijs gebruik moet worden gemaakt van multiniveau-analyse technieken (Bryk en Raudenbusch, 1992, Longford, 1993). Immers de processen die zich in en tussen zorginstellingen afspelen zijn ingebed in een locale of een provinciale context, waarbij beslissingen op lokaal of regionaal niveau mogelijkheden of onmogelijkheden creëren voor zorginstellingen. Denk bijvoorbeeld aan het afbouwen van het aantal plaatsen voor patiënten in verpleeghuizen of provinciale beslissingen ten aanzien van bouwstops of bouwplannen.

5. De noodzaak en relevantie van data mining voor GINO

In de vorige paragrafen is aangegeven welke analyses binnen of met GINO mogelijk zouden zijn. De mogelijkheden van data mining zijn daarbij niet expliciet aan de orde geweest. In deze paragraaf worden deze mogelijkheden aangegeven.

Totnogtoe is er geen data mining op gegevens van GINO toegepast, omdat daarvoor nog geen opdrachtgever gevonden kon worden.

In de huidige versie van GINO (GINO '97) zijn de dataformaten zodanig aangescherpt dat daardoor geen problemen voor data mining meer zijn. Voor alle verschillende versies (voor Thuiszorg, verpleeghuizen, verzorgingshuizen en indicatieorganen) van het GINO zijn de dataformaten voor gemeenschappelijke databasevelden identiek.

Op de gegevens van de vorige GINO-versies zijn wel analyses uitgevoerd. Daarbij stond met name de relatie tussen hulpbehoefte (ADL, HDL, GOST en PSYSOC) en de aanwezige taken (huishoudelijke, lichamelijk verzorgende, verpleegtechnische, medisch technische en begeleidende taken) centraal. (Zijlstra et al. 1991, Linschoten van et al, 1990). Voor de analyses werd gebruikt gemaakt van "normale" lineaire analyse technieken (regressie- en variantie-analyses).

Het probleem is echter dat er aanwijzingen zijn dat de relaties wel eens niet-lineair zouden kunnen zijn. Dit geldt onder meer voor de relatie tussen GOST en andere gegevens. Mensen waarbij sprake is van een toenemende geheugen- en orientatiestoornis (GOST) hebben meer zorg nodig, maar vanaf een bepaalde ernst van de geheugen- en orientatiestoornissen neemt de zorg weer af. Met name hoeven er minder begeleidende activiteiten worden uitgevoerd.

De relaties tussen de verschillende gegevens kunnen ook veel complexer van aard zijn. Voor bepaalde groepen van cliënten kan een relatie tussen verschillende gegevens op tijdstip A wel aanwezig zijn, maar deze relatie zou op een tijdstip B wel eens niet aanwezig hoeven te zijn. Voor bepaalde ziekten kan het plotseling optreden van afwijkende situaties, zoals schubs bij personen met MS of het opvlammen van een reumatische aandoening, de relatie tussen hulpbehoefte en de aanwezige zorg ineens anders doen zijn.

In bovenstaande situaties zijn toepassingen van gebruikelijke statistische analyse-technieken minder geschikt. Voor statistische analyses moet in ieder geval duidelijk zijn welke (lineaire) relaties op theoretische gronden verondersteld

mogen en kunnen worden. In de bovenstaande situaties is dat niet zo: data mining technieken zouden hier uitkomst kunnen bieden.

De situatie die hierboven beschreven wordt lijkt veel op de situaties zoals die voor verzekeringsmaatschappijen, banken en marketing bureaus gelden. Het gaat erom om de risicofactoren voor de hoeveelheid benodigde zorg of de duur van de te verlenen zorg te onderzoeken en in kaart te brengen. Het pakket moet op zoek gaan naar de verschillende patronen die er bestaan tussen persoonskenmerken (zoals leeftijd, geslacht en medische aandoening) en de verkregen zorg (lees ook kosten), zorgduur of de overlevingsduur.

Ook het grote aantal databasetabellen in de moederdatabase maakt het wenselijk om gebruik te maken van data mining technieken. Voor één enkele standaard analyse moeten tamelijk veel datamanagement handelingen worden uitgevoerd voordat de eigenlijke analyse kan plaatsvinden. Als er een heel duidelijk idee is over welke analyse er moet worden uitgevoerd dan wegen de kosten wel op tegen de baten. Maar om zo maar even iets na te gaan is dat niet meer het geval.

Momenteel zijn er bij GINO geen concrete plannen om daadwerkelijk data mining op de gegevens in haar pakketten toe te passen. Dit paper is ook bedoeld om beleidsmedewerkers en wetenschappelijk onderzoekers er op te wijzen dat er grote mogelijkheden zijn voor onderzoek als ze data mining-technieken toepassen op GINO-gegevens. Hopelijk kan daarmee voorkomen worden dat deze gegevens sub-optimaal gebruikt zullen worden.

Hoewel er nog geen data mining op GINO-gegevens zijn uitgevoerd, is er wel kennis over mogelijke softwarepakketten die voor de data mining gebruikt zouden kunnen worden en welke niet. Op dit moment zijn er vier pakketten (WizWhy van WizSoft Inc., Datamining Suitetm van Information Discovery Inc., "Neural Connections" van SPSS Inc. en The SAS[®] Data Mining Solution van SAS[®]) die in aanmerking komen om op de GINO-gegevens toe te passen. Deze pakketten

kunnen omgaan met een groot aantal verschillende database-tabellen, met vele gigabytes aan data en beschikken over meerdere methoden.

De huidige methoden voor data mining kunnen één probleem overigens niet oplossen. In de tabellen van GINO is niet alleen de aanwezigheid van een extra record (herziening van gegevens, zoals HADL, GOST of uitgevoerde handelingen) van belang, maar de tijdsduur tussen herzieningen bevat belangrijke informatie. Uit de snelheid waarmee gegevens veranderen kan afgeleid worden of een bepaald probleem, bijvoorbeeld de ontwikkeling van dementie, interventie of aanpassing van beleid noodzakelijk maakt. De huidige data mining technieken zijn echter niet in staat om deze informatie uit de databases te halen. Het gaat hierbij namelijk om de structuur van de database zelf en niet om de informatie in de database.

6. Conclusie en discussie

Met GINO is, ook in combinatie met externe informatiebronnen en data miningtechnieken, een groot aantal vraagstellingen te beantwoorden waarop nu niet of met veel moeite antwoord kan worden gegeven. Enkele daarvan zijn in het bovenstaande kort aangestipt.

Voor een optimaal gebruik van het GINO als breed onderzoeksinstrument is het belangrijkste probleem: het ontbreken van een "Centrale Database" of "Datawarehouse". In de opzet uit 1992 van GINO stond de opbouw van een "Centrale Database" waar alle gegevens van alle gebruikers van het GINO konden worden ondergebracht, centraal (zie figuur 2). Hiermee zou het mogelijk zijn om op een nationaal of regionaal niveau analyses uit te voeren. Voor de opbouw en het onderhoud van een "Centrale Database" is financiering nodig die tot op heden niet is gevonden. De bestaande wijzen van opslaan maken regionale of lokale vergelijkende onderzoeken wel mogelijk, maar dat vereist medewerking van de indicatie-organen en van elke afzonderlijke instelling die als GINO-gebruiker gegevens ter beschikking kan stellen.

Bij de inrichting van een "Centrale Database" en de analyses van GINO-gegevens dient rekening te worden gehouden met het vraagstuk van privacy-bescherming van cliënten en hulpverleners. Op grond van de wet is het namelijk niet toegestaan om gegevens op te slaan die "te herleiden op de persoon" (Ganzeboom, 1997). Om de gegevens van één cliënt in de tijd, verkregen uit verschillende instellingen, te kunnen koppelen worden bij de export vanuit GINO de identiteitsgegevens geanonimiseerd via een versleuteling. Deze functie kan standaard worden opgenomen, zodat volledig gecontroleerd, maar geheel anoniem, gegevens van personen in een longitudinaal record als kralen aan een snoer kunnen worden geregen. Het GINO is bereid om het versleutelingsalgoritme beschikbaar te stellen aan (wetenschappelijke) instellingen die daar belang bij zouden hebben. Op deze wijze zou het mogelijk zijn om gegevens van verschillende (wetenschappelijke) onderzoeken met elkaar te verbinden. Noodzakelijke voorwaarde is dan wel dat de voor de versleuteling benodigde gegevens beschikbaar zijn. (In gegevensbestanden zou dan opgenomen kunnen worden of de gegevens volgens het algoritme versleuteld zijn.)

Op één belangrijke analysemogelijkheid van het GINO is nog niet ingegaan: de mogelijkheid om survivalanalyses uit te voeren. Deze analysemogelijkheid is namelijk uitsluitend mogelijk binnen een "Centrale GINO Database". Patiënten hebben meestal een "zorgcarrière": ze komen met een zorgvraag binnen bij een Thuiszorginstelling en komen later in een verzorgingshuis en nog later in een verpleeghuis terecht. Als de gegevens worden ondergebracht in een "Centrale GINO Database" dan is het mogelijk om na te gaan wat de zorgcarrière is voor verschillende type patiënten, bijvoorbeeld de zorgcarrière van dementie of MS-patiënten.

7. Tot slot

Het gebruik van de gegevens uit de GINO software die hierboven is beschreven, is afhankelijk van praktische toepassing van deze programmatuur in de alledaagse

praktijk van indicatie-organen en zorgaanbieders. In de RIO-markt worden op dit moment drie producten aangeboden, naast GINO RIO het pakket ZO van Zorgnet B.V. en RIO 2000 van Prevalent. In prijs, voorwaarden en implementatiekosten zijn de verschillen tussen deze drie marginaal. ZO en RIO 2000 zijn echter uitsluitend gericht op procesondersteuning en bevatten geen voor wetenschappelijk onderzoek en/of beleidsanalyse getoetste instrumenten. De andere twee aanbieders hebben de datastructuur niet vastgelegd om optimaal gebruikersgemak te bieden. De consequentie daarvan is dat tussen instellingen feitelijk geen vergelijkende analyses mogelijk zijn. Data mining met ZO en RIO 2000 is derhalve nu geheel uitgesloten.

Literatuur

Bryk AS & Raudenbusch SW. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage Publications, Newbury Park.

Ganzeboom HBG. (1997) Advies van de commissie Kordes over privacy en enquetegegevens. *Facta*, 5, 4, 20-22.

Lindsey JK. (1993) *Models for repeated measurements*. Clarendon Press, Oxford.

Linschoten CP, Zijlstra WO en Wolffensperger EW (1990) *Hulpbehoefte en zorgverlening. Eerste meting onder ouderen in Zuid-Oost Groningen*. Styx Publications, Groningen.

Longford NT (1993) *Random coefficient models*. Princeton University Press, New York.

Moorer P., Wolffensperger EW. & Zijlstra WO. Kengetallen binnen het GINO: onderzoek naar schaalbaarheid, betrouwbaarheid en "differential item functioning". (submitted)

Moorer P. & Suurmeijer ThPBM, Zwanikken CP, Bakker RH en Dassen Th.W.N.
Use of services among persons with multiple sclerosis. (Submitted)

Wolffensperger EW. en Zijstra WO., (1993) Integratie van zorginformatie in de sectoren verpleging en verzorging: het Geïntegreerd Informatie Netwerk Ouderenzorg. *Gezondheidszorg en Management*, 12.

Wolffensperger EW. en Zijstra WO. (zonder datum) *Het gebruik van kengetallen binnen het GINO*. Noordelijk Centrum voor Gezondheidsvraagstukken, RUG, Groningen.

Wolffensperger EW., Ogtrop van J. en Zijstra WO. (1996) *Een ICIDH-gestuurd instrument voor geïntegreerde indicatiestelling van zorg, wonen en WVG: GINO-ZVN*. GINO b.v-ZVN N.V. Groningen.

Wijk P. van der, Wolffensperger EW. en Zijstra WO. (1995) *Economische evaluatie van verpleeghuisvervangende zorg thuis en in het verzorgingstehuis: verpleeghuiszorg in perspectief in Zuid-Oost Groningen*. Noordelijk Centrum voor Gezondheidsvraagstukken, RUG, Groningen.

Zijlstra WO, Wolffensperger EW en Te Velde BP (1991) *Registratie van verschuivingen in de zorg voor ouderen*. Styx Publications, Groningen.

Ontvangen: 07-06-1998
geaccepteerd: 15-01-1999

