

# Methode van bijschatten van StatLinetabel Ziekenhuispatiënten naar diagnose

09

*Agnes de Bruin, Adelaide Ariel, Gerard Verweij, Abby Israëls*

Publicatiedatum CBS-website: 22 juni 2009



## Verklaring van tekens

.	= gegevens ontbreken
*	= voorlopig cijfer
x	= geheim
–	= nihil
–	= (indien voorkomend tussen twee getallen) tot en met
0 (0,0)	= het getal is kleiner dan de helft van de gekozen eenheid
niets (blank)	= een cijfer kan op logische gronden niet voorkomen
2007–2008	= 2007 tot en met 2008
2007/2008	= het gemiddelde over de jaren 2007 tot en met 2008
2007/'08	= oogstjaar, boekjaar, schooljaar enz., beginnend in 2007 en eindigend in 2008
2005/'06–2007/'08	= oogstjaar, boekjaar enz., 2005/'06 tot en met 2007/'08

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

## Colofon

### *Uitgever*

Centraal Bureau voor de Statistiek  
Henri Faasdreef 312  
2492 JP Den Haag

### *Prepress*

Centraal Bureau voor de Statistiek - Grafimedia

### *Omslag*

TelDesign, Rotterdam

### *Inlichtingen*

Tel. (088) 570 70 70  
Fax (070) 337 59 94  
Via contactformulier: [www.cbs.nl/infoservice](http://www.cbs.nl/infoservice)

### *Bestellingen*

E-mail: [verkoop@cbs.nl](mailto:verkoop@cbs.nl)  
Fax (045) 570 62 68

### *Internet*

[www.cbs.nl](http://www.cbs.nl)

## ***Samenvatting***

Van 1995 tot en met 2004 was de nonrespons bij de Landelijke Medische Registratie (LMR) van ziekenhuisopnamen zeer laag. Daarna is deze sterk gestegen, hetgeen het noodzakelijk maakt om de op deze registratie gebaseerde persoonsstatistiek ‘Ziekenhuispatiënten naar diagnose’ bij te schatten voor de ontbrekende ziekenhuisopnamen. De hiervoor ontwikkelde methode wordt in deze nota beschreven. De methode is toegepast vanaf verslagjaar 2005. Als referentiejaar is 2004 gebruikt; de weinige ontbrekende gegevens in dit jaar zijn hiervoor eerst aangevuld, met als gevolg dat ook de cijfers van verslagjaar 2004 (heel licht) zijn bijgeschat. Omdat voor de bijschatting van de latere jaren noodzakelijkerwijs gebruik gemaakt wordt van oude gegevens (2004) is de methode beperkt houdbaar. Op grond van simulatiestudies is de verwachting dat de methode 4 tot 5 jaren toegepast kan worden.

*Trefwoorden: ziekenhuisopnamen, LMR, GBA, persoonsstatistieken, klinische prevalenties en incidenties, bijschatting ontbrekende opnamen, Statlinetabel ‘Ziekenhuispatiënten naar diagnose’.*

## ***Inhoudsopgave***

1.	Inleiding .....	5
2.	Gevolgen van de bijschatting voor de onderschatting in de statistiek .....	6
3.	Methoden van bijschatting van de klinische prevalenties .....	8
3.1	Notatie .....	8
3.2	Mogelijke bijschattingsmethoden .....	8
4.	Gekozen bijschattingsmethode voor de klinische prevalenties .....	11
4.1	Bijschattingsmethode .....	11
4.2	Non-respons in referentiejaar .....	12
4.3	Gedeeltelijk responderende ziekenhuizen in verslagjaar .....	12
4.4	Fusie van ziekenhuizen in verslagjaar .....	13
4.5	Sluiting van ziekenhuis in verslagjaar .....	13
4.6	Nieuw ziekenhuis in verslagjaar .....	14
5.	Nauwkeurigheid van bijschattingsmethode klinische prevalenties .....	15
6.	Schattingsmethode voor de klinische incidenties .....	16
6.1	Aannames .....	16
6.2	Berekening incidenties in verslagjaar .....	18
6.3	Bepalen van de incidenties in referentiejaar .....	19
7.	Nauwkeurigheid van de schattingsmethode klinische incidenties .....	20
8.	Referenties .....	21

## 1. Inleiding

De Statlinetabel ‘Ziekenhuispatiënten naar diagnose’ is gebaseerd op gegevens van de Landelijke Medische Registratie (LMR) van ziekenhuisopnamen, gekoppeld met de Gemeentelijke Basisadministratie (GBA). Op basis van deze koppeling worden voor deze statistiek twee uitkomstmaten afgeleid op persoonsniveau:

- (1) het aantal personen met minstens één ziekenhuisopname voor een bepaalde diagnose in een jaar (klinische prevalentie); en
- (2) het aantal personen met een eerste ziekenhuisopname voor een bepaalde diagnose in een jaar én geen opnamen voor dezelfde diagnose in de vijf voorafgaande jaren (klinische incidentie).

Bij de afleiding van de statistiek wordt er van uitgegaan dat de LMR een volledige registratie is van ziekenhuisopnamen en dat alle opnamen (voor zover deze GBA-ingezetenen betreffen) koppelen met de GBA (Van Sijl en De Ree, 2005). Dit is echter niet helemaal het geval: de LMR heeft van 1995-2004 een kleine nonrespons gehad (0,5-1,5% missende records) en er zijn ook altijd LMR-records die niet koppelen met de GBA (2-4%), bv. als gevolg van fouten in de administratieve gegevens in de LMR en/of GBA. Als gevolg hiervan wordt het aantal personen met ziekenhuisopnamen in deze statistiek licht onderschat. Deze onderschatting kan niet precies gekwantificeerd worden, maar bedraagt in ieder geval minder dan het totale percentage van ontbrekende en niet gekoppelde LMR-records. De onderschatting is minder dan dit percentage, omdat de patiënt de teleenheid is en deze meerdere keren kan zijn opgenomen. Het percentage ontbrekende en niet gekoppelde LMR-records is van 1995 tot 2003 geleidelijk gedaald van 5,6% tot 3,5%. Daarna is het percentage missende LMR-records echter flink gestegen, van 1,1% in 2004 tot 12,0% in 2007. Het totaal percentage missende en niet koppelende records steeg daardoor van 3,9% in 2004 tot 14,3% in 2007 (zie tabel 1).

Deze sterke stijging zou een navenant grotere onderschatting van de prevalentie en incidentie van ziekenhuisopnamen ten gevolge hebben, waardoor de cijfers niet meer goed van jaar op jaar vergelijkbaar zijn. Daarom is voor deze statistiek vanaf verslagjaar 2005 een methode ontwikkeld om de prevalentie- en incidentiecijfers bij te schatten voor de nonrespons (dus voor de missende records). Deze methode is ontwikkeld door Adelaide Ariel en Abby Israëls van de afdeling Methodologie van het CBS (Ariel en Israëls, 2009). De methode van bijschatting wordt in deze nota beschreven. Voor de bijschatting worden de LMR-gegevens uit 2004 gebruikt als ‘compleet’ referentiejaar. Omdat ook 2004 niet helemaal compleet is, is deze eerst volledig gemaakt om als goede referentie te kunnen dienen. Daardoor zijn de prevalentie- en incidentiecijfers voor 2004 ook (heel licht) bijgesteld.

**Tabel 1. Percentage missende en niet koppelende records per jaar in de LMR**

<b>Jaar</b>	<b>Missende records</b>	<b>Niet koppelende records</b>	<b>Totaal missende en niet koppelende records</b>
	%	%	%
1995	1,5	4,1	5,6
1996	1,4	3,8	5,2
1997	1,0	3,7	4,6
1998	0,9	3,6	4,6
1999	0,6	3,5	4,1
2000	0,9	3,4	4,3
2001	0,9	3,3	4,1
2002	0,5	3,1	3,5
2003	0,6	2,9	3,5
2004	1,1	2,8	3,9
2005	3,3	2,7	6,0
2006	10,5	2,4	12,8
2007	12,0	2,3	14,3

## **2. Gevolgen van de bijschatting voor de onderschatting in de statistiek**

In 2004 zijn er 4 ziekenhuizen in de LMR met een gedeeltelijke non-respons, d.w.z. dat bij deze ziekenhuizen van een deel van de opnamen geen microrecords in de LMR zijn geregistreerd. Van de missende opnamen in de LMR is wel bekend welk specialisme deze betreffen en of het om een klinische opname of een dagopname gaat. Op grond van deze gegevens zijn de opnamen van deze ziekenhuizen in 2004 compleet gemaakt door de ontbrekende opnamen te vervangen door de microrecords van vergelijkbare opnamen (zelfde specialisme en soort opname) uit een eerder of later jaar. Op deze wijze is het referentiejaar 2004 ‘compleet’ gemaakt, voor de bijschatting van de latere jaren. Enige uitzondering betreft één kankerkliniek waarvan alle records ontbreken. Omdat deze kliniek ook in alle andere jaren vanaf 1995 in zijn geheel ontbreekt, is besloten om hiervoor ook in 2004 en volgende jaren niet bij te schatten.

Door de bijschatting wordt dus het nonrepons-niveau vanaf 2004 verlaagd (tot ca. 0,4%); de percentages niet koppelende microrecords blijven gelijk. Gevolg hiervan is dat de maximale onderschatting in de prevalenties en incidenties (totaal percentage van missende en niet koppelende records) vanaf 2004 ook substantieel lager wordt (ca. 3%), zie tabel 2. Tabel 2 is equivalent aan tabel 1, de percentages die veranderen door de bijschatting zijn vet gedrukt. Op totaalniveau verandert de maximale onderschatting na bijschatting dus heel geleidelijk, van 5,6% in 1995 tot 2,6% in 2007. De mogelijke vertekening van trends in de prevalentie- en incidentiecijfers zal dus klein zijn; alleen bij licht stijgende trends zou het kunnen zijn dat als gevolg van de afnemende onderschatting er in werkelijkheid geen trend aanwezig is.

Het niveau van de onderschatting kan overigens nog enigszins variëren naar leeftijdsgroep en diagnose. Zo komen niet-gekoppelde LMR-records vaker voor bij jongvolwassenen en zeer ouderen en bij psychische stoornissen en ongevalsletsels. Voor deze diagnoses en leeftijdsgroepen kan het *niveau* van de onderschatting (niet de trend daarin) van het aantal personen met opnamen dus iets groter zijn. Door het ontbreken van één kankerkliniek, kunnen ook de cijfers voor nieuwvormingen een grotere onderschatting vertonen. Omdat van het totaal aantal opnamen voor nieuwvormingen echter maar een relatief klein gedeelte in deze kliniek plaatsvindt (ca. 4%) en een deel van de in deze gespecialiseerde kliniek opgenomen personen al eerder in een ander ziekenhuis zal zijn opgenomen, zal deze onderschatting toch beperkt zijn.

**Tabel 2: Percentage missende en niet koppelende records per jaar in de LMR, na bijschatting vanaf 2004 in de LMR**

Jaar	Missende records	Niet koppelende records	Totaal missende en niet koppelende records
	%	%	%
1995	1,5	4,1	5,6
1996	1,4	3,8	5,2
1997	1,0	3,7	4,6
1998	0,9	3,6	4,6
1999	0,6	3,5	4,1
2000	0,9	3,4	4,3
2001	0,9	3,3	4,1
2002	0,5	3,1	3,5
2003	0,6	2,9	3,5
2004	<b>0,4</b>	2,8	<b>3,2</b>
2005	<b>0,4</b>	2,7	<b>3,0</b>
2006	<b>0,4</b>	2,4	<b>2,7</b>
2007	<b>0,3</b>	2,3	<b>2,6</b>

### 3. Methoden van bijschatting van de klinische prevalenties

#### 3.1 Notatie

We voeren de volgende notatie in voor de LMR in een zeker jaar:

- $O$  = aantal opnamen van uniek gekoppelde patiënten<sup>1</sup>;
- $P$  = aantal patiënten (prevalentie);
- indeling ziekenhuizen: R en NR (respons- en nonrespons-ziekenhuizen);
- indeling opnamen: idem (dus bijv. R-opnamen in R-ziekenhuizen);
- indeling patiënten: r, nr en ov(erlap); dit is een driedeling van de patiënten in degenen die alleen in R-ziekenhuizen zijn opgenomen (r), alleen in NR-ziekenhuizen (nr) en in beide (ov).

Zo is  $O_R$  het aantal (unieke) opnamen in de R-ziekenhuizen, d.w.z. het aantal waargenomen opnamen.  $P_r$  is het aantal patiënten dat alleen in de R-ziekenhuizen is opgenomen, en  $P_{ov}$  het aantal patiënten dat zowel in R- als NR-ziekenhuizen is opgenomen. Er zijn dus  $P_r + P_{ov}$  patiënten waargenomen, en we kunnen deze twee groepen niet van elkaar onderscheiden. Door de nonrespons missen we  $P_{nr}$  patiënten. Dit aantal moet worden geschat om opgeteld bij  $P_r + P_{ov}$  het totaal aantal patiënten te kunnen schatten.

We hangen hier de index c aan wanneer we duidelijk willen maken dat het gaat om cijfers uit het jaar dat we compleet hebben, en de index i om het incomplete jaar aan te geven waarvoor we  $P_{nr}$  willen schatten. Dus als we het aantal ontbrekende patiënten voor 2006 willen schatten met behulp van 2004 als compleet jaar, dan is  $c=04$  en  $i=06$ . Om de betrouwbaarheid van de methode te onderzoeken zijn er ook simulaties gedaan waarbij we het aantal patiënten in 2004 schatten met behulp van dat van 2000 (dus hier is  $c=00$  en  $i=04$ ). Wanneer het aantallen opnamen of patiënten per diagnose betreft, voegen we een index d toe. We kunnen de notatie zowel voor gewogen als ongewogen cijfers gebruiken. Bij gewogen cijfers worden de gewichten gebruikt zoals beschreven in Van Sijl en De Ree (2005).

#### 3.2 Mogelijke bijschattingsmethoden

Voor de bijschatting van de prevalenties is een aantal verschillende methoden overwogen. In tabel 3 wordt een overzicht van deze methoden gegeven. Eerst wordt per methode de formule gegeven voor de schatter van  $P_{nr}$  (uitgaande van  $c=04$  en  $i=06$ ), en daarnaast de aanname die bij de betreffende methode gemaakt wordt.

---

<sup>1</sup> Vrijwel overal in de nota is sprake van *unieke* opnamen, wanneer we het over opnamen hebben. ‘Uniek’ slaat op het feit dat het opnamen betreft van patiënten die gedurende het gehele jaar koppelbaar zijn met de GBA Waar sprake is van *alle* opnamen geven we dat aan.



**Tabel 3. Verschillende bijschattingsmethoden voor de prevalenties**

Methode	Schatter voor $P_{nr06}$	Formule nr.	Aanname
A	$\hat{P}_{nr06} = O_{NR06} \frac{P_{r06} + P_{ov06}}{O_{R06}}$	(2.1)	$\frac{O_{NR}}{P_{nr}} = \frac{O_R}{P_r + P_{ov}}$
B	$\hat{P}_{nr06} = P_{nr04} \frac{O_{NR06}}{O_{NR04}}$	(2.2)	$\frac{O_{NR}}{P_{nr}}$ constant over jaren
C	$\hat{P}_{nr06} = P_{nr04} \frac{P_{r06}}{P_{r04}} \frac{O_{NR06}}{O_{NR04}} \frac{O_{R04}}{O_{R06}}$	(2.3)	$\frac{O_{NR}/P_{nr}}{O_R/P_r}$ constant over jaren
D	$\hat{P}_{nr06} = P_{nr04} \frac{P_{r06} + P_{ov06}}{P_{r4} + P_{ov04}}$	(2.4)	$\frac{P_r + P_{ov}}{P_{nr}}$ constant over jaren

**Methode A** gaat uit van dezelfde verhouding tussen het aantal opnamen en aantal patiënten voor de NR- en R-ziekenhuizen. Deze ratio's zijn echter niet als gemiddeld aantal opnamen per patiënt te interpreteren, tenzij er geen overlap zou zijn in patiënten tussen NR- en R-ziekenhuizen. In werkelijkheid is er wel sprake van overlap. Om de aanname bij deze methode te toetsen zijn de ratio's tussen het aantal opnamen en patiënten voor 2000-2004 bepaald, er van uitgaande dat in deze jaren dezelfde NR-ziekenhuizen zouden gelden als in 2006. Het resultaat is weergegeven in tabel 4.

**Tabel 4. Verhouding tussen het aantal opnamen en patiënten in 2000-2004 uitgaande van de NR-ziekenhuizen in 2006**

Jaar	Ongewogen		Gewogen	
	NR-ziekenhuizen	R-ziekenhuizen	NR-ziekenhuizen	R-ziekenhuizen
2004	1,611	1,560	1,382	1,335
2003	1,572	1,532	1,345	1,311
2002	1,552	1,511	1,326	1,292
2001	1,510	1,481	1,291	1,267
2000	1,485	1,453	1,271	1,243

Tabel 4 geeft de verhoudingen van de aanname bij formule (2.1) voor de jaren 2000-2004, ongewogen en gewogen. Hieruit blijkt dat ieder jaar de verhouding tussen het aantal opnamen en het aantal patiënten bij de nonrespons hoger is dan bij de respons, d.w.z.  $O_{NR} / P_{nr} > O_R / (P_r + P_{ov})$ . Dit komt voornamelijk doordat in het linkerlid van formule (2.1) de overlap wel in de teller wordt meegenomen, maar niet in de noemer. Dit geeft een overschatting van de uiteindelijke prevalenties. Deze overschatting zal groter zijn naarmate het nonrespons-niveau toeneemt.

Doordat bij methode A geen goede theoretische onderbouwing is te geven (teller en noemer van de aannames hebben betrekking op verschillende deelpopulaties) is deze niet gekozen als bij-schattingsmethode. Wel is nog nagegaan in simulaties hoe groot de overschatting van de prevalenties is bij gebruik van methode A. Hieruit bleek dat methode D nauwkeuriger is dan methode A. Ook om deze reden is afgezien van het gebruik van methode A.

**Methode B** gaat uit van een in de tijd constante verhouding tussen het aantal missende opnamen in NR-ziekenhuizen en het aantal missende nr-patiënten  $O_{NR} / P_{nr}$ . Dit is het linkerlid van formule (2.1); zie tabel 3. De gedachte hierachter is dat de verhouding weliswaar niet goed interpreteerbaar is, maar mogelijk wel stabiel in de tijd.

**Methode C** gaat uit van een constante ratio  $\frac{O_{NR}/P_{nr}}{O_R/P_r}$  tussen de verhouding bij methode B en de hiermee corresponderende verhouding voor de respons, d.i. het aantal opnamen in R-ziekenhuizen gedeeld door het aantal r-patiënten. Ondanks dat de twee verhoudingen niet goed interpreteerbaar zijn, is er wel een symmetrie bereikt tussen respons en nonrespons. Een probleem is echter dat  $P_{r6}$  niet bekend is en dus op één of andere manier zou moeten worden geschat.

**Methode D** gaat uit van een constante verhouding  $(P_r + P_{ov}) / P_{nr}$  tussen (r+ov)-patiënten en nr-patiënten, d.w.z. tussen waargenomen en ontbrekende patiënten. Op het eerste gezicht lijkt het vreemd om het aantal opnamen in R-ziekenhuizen niet mee te nemen. Maar veel opnamen hoeft niet samen te gaan met veel verschillende patiënten. Het kan ook samengaan met de frequentie waarmee patiënten naar het ziekenhuis gaan. Bovendien hangt het aantal opnamen samen met de overlap, en dus niet alleen met patiënten die alleen in R-ziekenhuizen zijn opgenomen. Andere voordelen van methode D boven de andere drie genoemde methoden zijn:

- bij prevalenties per diagnose moet het aantal opnamen per diagnose worden geschat, en die schatting kan onbetrouwbaar zijn;
- de eenvoud.

**Tabel 5. Stabiliteit in de tijd van de verhoudingen in de aannames voor methode B, C en D.**

Jaar	B		C		D	
	ongew.	gew.	ongew.	gew.	ongew.	gew.
2004	1,6114	1,3817	1,0197	1,0221	0,1107	0,1104
2003	1,5715	1,4276	1,0134	1,0138	0,1127	0,1127
2002	1,5520	1,3263	1,0147	1,0143	0,1118	0,1119
2001	1,5104	1,2913	1,0078	1,0076	0,1126	0,1126
2000	1,4849	1,2714	1,0112	1,0118	0,1122	0,1121
Gemiddelde relatieve afwijking t.o.v. het gemiddelde	0,0250	0,0388	0,0030	0,0034	0,0054	0,0056

Tabel 5 toont in hoeverre de verhoudingen in de aannames bij de methoden B-D constant zijn. De verhoudingen zijn berekend voor zowel de ongewogen ('ongew.') als de gewogen ('gew.') prevalenties. Hieruit blijkt dat methode B niet voldoet. Bij methode C en D variëren de verhoudingen weinig over de tijd. Methode C geeft de kleinste gemiddelde relatieve afwijking. Omdat bij methode C het probleem bestaat dat  $P_{r06}$  niet eenvoudig te bepalen is, is gekozen voor methode D als bijschattingsmethode. Methode D is eenvoudig omdat er geen schatting nodig is van het aantal uniek-gekoppelde opnamen (totaal en per diagnose).

#### 4. Gekozen bijschattingsmethode voor de klinische prevalenties

##### 4.1 Bijschattingsmethode

Voor de bijschatting van de prevalenties is gekozen voor de in paragraaf 3 genoemde methode D. Dat wil zeggen dat de ontbrekende patiënten  $\hat{P}_{nr,i}$  geschat worden voor de jaren  $i = 2005, 2006, \text{etc.}$  met de formule:

$$\hat{P}_{nr,i} = P_{nr,04} \frac{P_{rov,i}}{P_{rov,04}}, \quad (4.1)$$

met  $P_{rov} = P_r + P_{ov}$ , en met referentiejaar  $c=2004$ ,

zodat de totale prevalentie  $\hat{P}_{tot,i}$  als volgt geschat wordt:

$$\hat{P}_{tot,i} = P_{rov,i} + \hat{P}_{nr,i} = P_{rov,i} \left(1 + \frac{P_{nr,04}}{P_{rov,04}}\right) \quad (4.2)$$

Als  $P_{rov,04} = 0$ , dan wordt gekozen voor  $\hat{P}_{nr,i} = P_{nr,04}$ , waardoor  $\hat{P}_{tot,i} = P_{rov,i} + \hat{P}_{nr,04}$

Bovengenoemde vereenvoudiging in geval  $P_{rov,04} = 0$  is nodig omdat anders de noemer in de formules nul is.

De prevalenties worden op het diepste niveau van de outputtabel geschat, dat wil zeggen per combinatie van diagnose, leeftijd en geslacht (simultaan). De prevalenties op hogere aggregatieniveaus worden dan afgeleid door sommatie van de schattingen op het diepste niveau. Zo worden de aantallen nonrespons-patiënten per diagnose als volgt geschat:

$$\hat{P}_{nr,04}^d = \sum_g \sum_l \hat{P}_{nr,04}^{d,g,l}$$

met  $d$ =diagnose,  $g$ =geslacht en  $l$ =leeftijd

En bijvoorbeeld de nonrespons-patiënten naar diagnose en geslacht als volgt:

$$\hat{P}_{nr,04}^{dg} = \sum_l \hat{P}_{nr,04}^{d,g,l}$$

In de volgende paragrafen wordt toegelicht hoe bij deze bijschattingsmethode wordt omgegaan met speciale situaties. Met name sluiting van ziekenhuizen en opening van nieuwe ziekenhuizen kunnen een probleem vormen bij deze methode. In 4.5 en 4.6 wordt besproken hoe hier mee omgegaan kan worden.

## **4.2 Non-respons in referentiejaar**

De bijschattingsmethode gaat idealiter uit van een referentiejaar waarin er geen nonrespons is. Gekozen is voor het jaar 2004 als referentiejaar omdat dit het meest recente jaar in de LMR is met een lage nonrespons (zie tabel 1). Er zijn in 2004 vijf ziekenhuizen met nonrepons, waarvan een kankerkliniek die in het geheel geen microrecords aan de LMR heeft aangeleverd. Omdat dit ziekenhuis ook in eerdere jaren ontbrak, wordt deze ook in 2004 (en latere jaren) als non-respons beschouwd. Voor dit ziekenhuis wordt dus niet bijgeschat (zie ook hoofdstuk 2).

Voor de vier ziekenhuizen met gedeeltelijke nonrespons wordt het 2004-bestand compleet gemaakt door de ontbrekende records te vervangen door uniek gekoppelde opnamen uit het zelfde ziekenhuis uit 2003 of 2005. Hiervoor wordt random getrokken uit opnamen met hetzelfde zorgtype (dagopname of klinische opname) en specialisme als de ontbrekende opnamen. Bij het bepalen van het aantal te trekken uniek gekoppelde opnamen wordt rekening gehouden met de verhouding tussen het oorspronkelijke aantal opnamen en het aantal uniek gekoppelde opnamen voor de betreffende instelling. Voor de twee ziekenhuizen waarvan vrijwel uitsluitend opnamen uit de laatste maanden van 2004 ontbreken wordt getrokken uit dezelfde maanden van 2003, zodat er een aaneengesloten periode van opnamen blijft. Bij de andere twee ziekenhuizen wordt getrokken uit opnamen van het jaar 2005. Het persoonsnummer van de getrokken opnamen uit 2003 en 2005 blijft gehandhaafd. De aanname hierbij is dat de kans op heropnamen bij dezelfde persoon niet veel verschilt als je het kalenderjaar 2004 neemt of een periode in 2003-2004 of in 2004-2005. Bij dit kleine aantal aangevulde records is dit een redelijke aanname.

Door het toevoegen van de nonrespons-records worden de (bijgeschatte) prevalenties in 2004 iets hoger dan de oorspronkelijk gepubliceerde, niet bijgeschatte prevalenties. Maar omdat de nonrespons in 2004 laag is, zijn deze verschillen erg klein.

## **4.3 Gedeeltelijk responderende ziekenhuizen in verslagjaar**

In de bijschattingsmethode worden ziekenhuizen die geheel nonresponderen in het verslagjaar waar de bijschatting voor wordt uitgevoerd (vanaf 2005) toegewezen aan de NR-groep in 2004.

Ziekenhuizen die in het verslagjaar slechts gedeeltelijk responderen krijgen in 2004 een R- en een NR-deel. Dit gebeurt door iedere opname van zo'n ziekenhuis aan het R- of NR-deel toe te wijzen. De bedoeling hiervan is dat voor de LMR 2004 het missingdatapatroon van het verslagjaar zo goed mogelijk wordt nagebootst. Hiervoor wordt het volgende procedé gevolgd:

1. Eerst worden indien mogelijk onderdelen van het ziekenhuis met 100% respons en onderdelen met 100% nonrepons geheel toegewezen aan resp. het R-deel en het NR-deel in 2004. ‘Onderdelen’ zijn hierbij gedefinieerd als combinaties van zorgtype (dagopname of klinische opname), specialisme en/of periode in het jaar (bijv. alle dagopnamen tijdens het laatste halfjaar).
2. Voor de resterende onderdelen met  $x\%$  nonrespons qua opnamen ( $0 < x < 100$ ) wordt het volgende gedaan:
  - a. Als het om een te verwaarlozen aantal nonrespons-opnamen gaat (minder dan 2%) dan worden alle opnamen in 2004 aan het R-deel toegewezen. Door het negeren van de nonrepons worden de bijgeschatte prevalenties in principe licht onderschat, maar bij deze lage nonrespons-niveaus is dat verwaarloosbaar;
  - b. In alle overige gevallen wordt in 2004 random  $x\%$  van het onderdeel toegewezen aan het NR-deel en de niet getrokken opnamen aan het R-deel.

#### **4.4 Fusie van ziekenhuizen in verslagjaar**

Een fusieziekenhuis is doorgaans een R- of een NR-ziekenhuis. Alle onderdelen worden dan in 2004 als zodanig beschouwd. Wanneer een fusieziekenhuis gedeeltelijk heeft gerepondeerd, dan zijn de hierboven genoemde opties voor gedeeltelijke respons toe te passen. Het afsplitsen (zie onderdeel (1) van paragraaf 4.3) zou dan ook kunnen plaatsvinden op grond van de vestiging, wanneer het fusieziekenhuis uit meerdere, gefuseerde vestigingen bestaat.

#### **4.5 Sluiting van ziekenhuis in verslagjaar**

Sluiting van een ziekenhuis of van een bepaald onderdeel van het ziekenhuis zal meestal tot gevolg hebben dat men naar een ander ziekenhuis gaat. Met name ziekenhuizen in dezelfde regio zullen de opnamen van het gesloten ziekenhuis overnemen.

We kunnen ervan uitgaan dat het gesloten ziekenhuis in 2004 volledig heeft gerepondeerd of dat er zo goed mogelijk voor is geïmputeerd. Voor de bijschatting kan dan gekozen worden om dit ziekenhuis (of ziekenhuisonderdeel) uit het LMR-bestand van 2004 te halen en formule (4.2) toe te passen op de overige ziekenhuizen. Dit komt neer op het naar rato verdelen van het aantal patiënten in het bewuste ziekenhuis in 2004 over respons (inclusief overlap) en nonrespons (saldo). Dit kan zowel worden toegepast voor de totale prevalentie als per diagnose.

Heeft men het idee dat de patiënten van het gesloten ziekenhuis geheel worden overgenomen door R-ziekenhuizen of geheel door NR-ziekenhuizen, dan kan gekozen worden om het gesloten ziekenhuis niet weg te halen uit het 2004-bestand, maar om het geheel als R, resp. NR te beschouwen. Op ad-hoc basis kan men kijken hoe de verhouding tussen R- en NR-ziekenhuizen in de regio ligt voor de aantallen patiënten en kan men de verdeling naar R en NR navenant aanpassen. Er is hoe dan

ook sprake van een aanname, die niet verifieerbaar is en voor een methodebreuk kan zorgen.

Voor sluiting van een deel van een ziekenhuis, bijvoorbeeld van een vestiging of een bepaald specialisme, geldt hetzelfde als hierboven. Een probleem is dat men niet altijd op de hoogte kan zijn van sluiting van een deel van het ziekenhuis. Daarom is het goed om als kwaliteitscontrole uitschieters te detecteren in aantallen opnamen per ziekenhuis en in verhoudingen tussen aantallen opnamen en patiënten. Dit zou kunnen leiden tot ad-hoc beslissingen zoals hierboven besproken.

#### 4.6 Nieuw ziekenhuis in verslagjaar

Als er een nieuw ziekenhuis bijkomt in een verslagjaar (niet ontstaan door fusie van bestaande ziekenhuizen) dan wordt aangenomen dat de meeste patiënten die in een nieuw ziekenhuis zijn opgenomen, anders in een ander ziekenhuis zouden zijn opgenomen. Het nieuwe ziekenhuis moet worden meegenomen in de bijschatting, omdat weglaten tot een vermindering van het aantal opnamen leidt.

Wanneer het nieuwe ziekenhuis heeft gerepondeerd, kan formule (4.2) worden toegepast op de in 2004 bestaande ziekenhuizen en vervolgens informatie van het nieuwe ziekenhuis eraan worden toegevoegd. Dit laatste kan door toevoeging van de patiënten in het nieuwe ziekenhuis die niet in de andere respons-ziekenhuizen zijn waargenomen. Strikt genomen moet men ook de overlap met de patiënten uit de NR-ziekenhuizen weglaten, door deze eerst te schatten. Dit lijkt slechts nodig als er een NR-ziekenhuis in de buurt van het nieuwe ziekenhuis ligt. De toe te passen formule wordt dus

$$\hat{P}_{tot,i} = P_{rov,i} + \hat{P}_{nr,i} + \hat{P}_{nieuw,i} = \left[ P_{rov,i} \left( 1 + \frac{P_{nr,04}}{P_{rov,04}} \right) \right]_{oud} + \hat{P}_{nieuw,i} \quad (4.3)$$

Bovenstaande methode gaat ervan uit dat de R- en NR-ziekenhuizen naar rato patiënten naar het nieuwe ziekenhuis zien gaan. Men kan hier op ad-hoc basis van afwijken wanneer in de buurt van het nieuwe ziekenhuis een overschot aan R- of NR-ziekenhuizen is. Zo kan men wanneer alle ziekenhuizen in de regio responderen, ervoor kiezen het nieuwe ziekenhuis bij de toepassing van formule (3.5) meteen als R-ziekenhuis mee te nemen. Men gaat er dan wel vanuit dat het nieuwe ziekenhuis geen patiënten aantrekt, die anders thuis waren gebleven. Als is in te schatten van welke ziekenhuizen de opnamen in het nieuwe ziekenhuis afkomstig zijn, kan naar een ad-hoc oplossing worden gezocht voor de splitsing van de patiënten over respons en saldo.

Ook indien het nieuwe ziekenhuis niet respondeert, kan men de bovenstaande procedure, inclusief formule (4.3), toepassen en achteraf het nieuwe ziekenhuis eraan toevoegen. Schatting van het aantal extra patiënten (saldo) in het nieuwe ziekenhuis moet dan echter geschieden op basis van de verhouding tot het aantal opnamen. Dit zou men kunnen doen door voor een vergelijkbaar R-ziekenhuis (spiegel-ziekenhuis) het saldo te schatten en te relateren aan het aantal opnamen. Wanneer het om cijfers per diagnose gaat, moet ook naar het specialisme, zorgtype

en soort ziekenhuis worden gekeken. Per geval moet worden besloten of inderdaad één ziekenhuis als ‘spiegel’ kan dienen voor alle diagnoses.

## 5. Nauwkeurigheid van bijschattingsmethode klinische prevalenties

Om de nauwkeurigheid van de bijschattingsmethode te schatten zijn simulaties uitgevoerd op oude LMR-jaren waarvan de werkelijke prevalenties bekend zijn. Het missingdatapatroon van 2006, resp. 2007, is hierbij globaal toegepast op de LMR 2004. Dat wil zeggen dat alle instellingen die in 2006, resp. 2007 niet of niet-volledig repondeerden in zijn geheel uit het bestand van 2004 zijn weggehaald. Voor 2006 correspondeert dit met 10,3% van de opnamen en voor 2007 met 14,6% van de opnamen. Vervolgens is voor verslagjaar 2004 de bijschattingsmethode uitgevoerd (zie paragraaf 4.1), met als referentiejaar 2000. De op deze wijze bijgeschatte prevalenties voor 2004 zijn vergeleken met de werkelijke prevalenties van 2004. Op basis van deze vergelijking is de relatieve fout (afwijking in % van de absolute waarde van de werkelijke prevalentie) van de bijgeschatte prevalenties bepaald op totaalniveau, op niveau van hoofdgroepen van diagnoses en op verschillende niveaus van subdiagnosen, leeftijd en geslacht. De resultaten staan vermeld in tabel 6. De cijfers zijn berekend op basis van de gewogen prevalenties, volgens Van Sijl en De Ree (2005).

**Tabel 6. Relatieve fout t.o.v. de absolute waarden van de prevalenties bij bijschatting van LMR 2004 met LMR 2000, met toepassing van het missingdatapatroon van 2006 en 2007.**

Missing data patroon	Publicatieniveau	Relatieve fout (%)
2006	Totaal	0,16
	ICD-hoofdgroepen van diagnoses	0,54
	Subdiagnosen (volgens Beldojo-lijst)	1,10
	Mannen x subdiagnose	1,41
	Vrouwen x subdiagnose	1,34
	Subdiagnose x geslacht x leeftijd	4,5
2007	Totaal	0,16
	ICD-hoofdgroepen van diagnoses	0,72
	Subdiagnosen (volgens Beldojo-lijst)	1,14
	Mannen x subdiagnose	1,43
	Vrouwen x subdiagnose	1,31
	Subdiagnose x geslacht x leeftijd	5,7

De gepresenteerde relatieve fouten op niveau van diagnose/geslacht/leeftijd zijn gemiddelden over alle betreffende categorieën.

Uit tabel 6 blijkt dat de relatieve fout bij de totale prevalentie en de bij prevalenties van ICD-hoofdgroepen van diagnoses kleiner dan 1% is. Dit geldt zowel voor de bijschatting volgens het missingdatapatroon van 2006 als voor die met het missingdatapatroon van 2007. Naarmate het publicatieniveau gedetailleerder is, wordt ook de relatieve fout groter. Dit is ook te verwachten omdat het aantal patiënten in deze subgroepen kleiner is. Bij uitsplitsing naar subdiagnosen volgens de Beldoza-lijst (een lijst met 94 diagnosegroepen) en geslacht is de gemiddelde relatieve fout  $< 1,5\%$ . Op het diepste publicatieniveau volgens de Beldoza-lijst, met uitsplitsing naar geslacht en 20 leeftijdsgroepen, is de gemiddelde relatieve fout  $4,5\%$  voor 2006 en  $5,7\%$  voor 2007. Alleen bij de zeer zeldzaam voorkomende diagnose *wiegedood* is de fout substantieel groter dan  $5\%$  (nl.  $17\%$ , zowel voor missingdatapatroon 2006 als die van 2007).

Op grond van deze resultaten kan geconcludeerd worden dat de bijschattingmethode voldoende nauwkeurig is. Bij uitsplitsingen in de publicatietabel waarbij sprake is van lage aantallen zullen de cijfers vanwege de geringere betrouwbaarheid afgerond of niet weergegeven moeten worden. Aangezien de statistiek al weergegeven wordt in relatieve cijfers (per 10 000 inwoners) met slechts een decimaal, is het de verwachting dat deze presentatie meestal zal voldoen.

De simulaties zijn uitgevoerd met referentiejaar t-4 (2000 als referentiejaar voor 2004). De bijschatting zal uitgevoerd worden vanaf 2005, met als referentiejaar 2004. Op grond hiervan is de verwachting dat de bijschattingmethode bij vergelijkbare nonrespons-niveaus in ieder geval tot en met verslagjaar 2008 zal voldoen. De houdbaarheid van de methode hangt ook af van allerlei externe factoren, zoals veranderingen in de ziekenhuispopulatie en de kwaliteit van de diagnosecodering in de LMR.

## **6. Schattingsmethode voor de klinische incidenties**

### **6.1 Aannames**

Bij de klinische incidenties gaat het om aantallen patiënten die in een bepaald jaar zijn opgenomen (voor een bepaalde diagnose), en niet (voor dezelfde diagnose) in de daaraan voorafgaande vijf kalenderjaren (Van Sijl en De Ree, 2005).

Voor de klinische incidenties zijn dezelfde schattingsmethoden onderzocht als bij de klinische prevalenties. Hieruit kwam naar voren dat ook voor de incidenties methode D voldoet.

We hanteren hier dezelfde notatie als in paragraaf 3.1. Alleen wordt de 'P' van prevalentie nu vervangen door de 'I' van incidentie, en de ' $I_r + I_{ov}$ ' (de respons-incidentie, inclusief de overlap) wordt kortweg aangeduid als ' $I_{rov}$ '.

Methode D komt voor de incidenties neer op het schatten van  $I_{nr,06}$  voor verslagjaar 2006 met de formule:



$$\hat{I}_{nr,06} = I_{nr,04} \frac{I_{rov,06}}{I_{rov,04}} \quad (6.1)$$

Net zoals bij de prevalenties is hier de aanname dat er een constante verhouding is tussen de waargenomen (incidente) patiënten en de ontbrekende (incidente) patiënten ( $I_{rov}/I_{nr}$ ). Echter, anders dan bij de prevalenties, kunnen de respons-incidenties in 2006 ( $I_{rov,06}$ ) en volgende jaren niet direct berekend worden op grond van de gegevens van dat jaar alleen, omdat voor het bepalen van de incidenties gegevens van de vorige 5 kalenderjaren nodig zijn. Voor 2006 zijn dus volledige gegevens nodig van 2001-2005. Omdat vanaf 2005 de gegevens niet compleet zijn, moeten dus ook de respons-incidenties in 2006 geschat worden. Hiervoor gaan we uit van de aanname dat de incidenties constante subsets zijn van de prevalenties, dus dat  $I_{rov}/P_{rov}$  constant is over de jaren. Bij het berekenen van de incidenties is dit dus een extra aanname die nodig is. Op grond van deze aanname kunnen de respons-incidenties als volgt worden geschat:

$$\hat{I}_{rov,06} = P_{rov,06} \frac{I_{rov,04}}{P_{rov,04}} \quad (6.2)$$

In tabel 7 staan de schatters en bijbehorende aannames voor de schattingsmethode voor de incidenties samengevat.

**Tabel 7. Schattingsmethode voor de incidenties**

Method	Schatters voor $I_{nr,06}$ en $I_{rov,06}$	Formulennr.	Aannames (constant over jaren)
D	$\hat{I}_{nr,06} = I_{nr,04} \frac{I_{rov,06}}{I_{rov,04}}$	(6.1)	$\frac{I_{rov}}{I_{nr}}$
	$\hat{I}_{rov,06} = P_{rov,06} \frac{I_{rov,04}}{P_{rov,04}}$	(6.2)	$\frac{I_{rov}}{P_{rov}}$

De stabiliteit van beide aannames is onderzocht op grond van de gegevens van 2000-2004. De aanname bij formule (6.1) bleek redelijk constant over de tijd, met een relatieve afwijking t.o.v. het gemiddelde van 0,009 (bij de prevalenties was deze 0,006). Het constant zijn van de verhouding tussen de respons-incidenties en respons-prevalenties, de aanname bij formule (6.2), is onderzocht voor de subdiagnoses volgens de Beldojo-lijst (94 diagnosegroepen), voor de jaren 2000, 2003 en 2004. Hieruit kwam naar voren dat ook deze verhouding relatief constant is, gezien de relatief geringe standaarddeviatie per diagnose. Voor zowel het missingdatapatroon van 2006 als van 2007 bleek de gemiddelde standaarddeviatie 0,013 te zijn.

## 6.2 Berekening incidenties in verslagjaar

Op grond van de formules (6.1) en (6.2) kan de totale incidentie in 2006 als volgt geschat worden:

$$\begin{aligned}
 \hat{I}_{tot,06} &= \hat{I}_{rov,06} + \hat{I}_{nr,06} \\
 \Rightarrow \hat{I}_{tot,06} &= P_{rov,06} \left( \frac{I_{rov,04}}{P_{rov,04}} \right) + I_{nr,04} \left( \frac{I_{rov,06}}{I_{rov,04}} \right) \\
 \Rightarrow \hat{I}_{tot,06} &= P_{rov,06} \left( \frac{I_{rov,04}}{P_{rov,04}} \right) + I_{nr,04} \left( \frac{P_{rov,06} (I_{rov,04} / P_{rov,04})}{I_{rov,04}} \right) \\
 \Rightarrow \hat{I}_{tot,06} &= P_{rov,06} \left( \frac{I_{rov,04}}{P_{rov,04}} \right) + I_{nr,04} \left( \frac{P_{rov,06}}{P_{rov,04}} \right) \\
 \Rightarrow \hat{I}_{tot,06} &= \frac{P_{rov,06}}{P_{rov,04}} (I_{rov,04} + I_{nr,04}) \\
 \Rightarrow \hat{I}_{tot,06} &= \frac{P_{rov,06}}{P_{rov,04}} I_{tot,04} .
 \end{aligned}$$

Veralgemeiseerd voor verslagjaar, met jaar  $i = 2005, 2006, \text{etc.}$  wordt de formule:

$$\hat{I}_{tot,i} = \frac{P_{rov,i}}{P_{rov,04}} I_{tot,04} \quad (6.3)$$

Indien  $P_{rov,04} = 0$ , dan wordt gekozen voor  $\hat{I}_{tot,i} = I_{tot,04}$

Omdat als  $P_{o,04} = 0$ , ook  $I_{rov,04} = 0$  (incidenties zijn een deelverzameling van de prevalenties) dan is  $I_{o,04} = I_{rov,04} + I_{nr,04} = I_{nr,04}$

Formule (6.3) kan uitgebreid worden tot lagere publicatieniveaus (naar diagnose, leeftijd en geslacht) door de bijpassende index toe te voegen.

De incidenties in het verslagjaar worden dus berekend door de (totale) incidentie in het referentiejaar (2004) te vermenigvuldigen met de verhouding van de respons-prevalentie in het verslagjaar en de respons-prevalentie (volgens het missingdatapatroon van het verslagjaar) in het referentiejaar.

De respons-prevalenties zijn al berekend voor de bijstelling van de prevalenties, zie formule (4.2). Dat betekent dat voor het berekenen van de bijgeschatte incidenties alleen nog de totale incidenties voor het referentiejaar 2004 bepaald moeten worden. Omdat 2004 niet helemaal compleet is, moet hier ook een schatting voor gemaakt worden. Dit wordt beschreven in de volgende paragraaf.

### 6.3 Bepalen van de incidenties in referentiejaar

We gebruiken 2004 als referentiejaar voor de schatting van incidenties in de daaropvolgende jaren. Echter, ook 2004 is niet volledig waargenomen. Voor de prevalenties is 2004 volledig gemaakt door records uit 2003 en 2005 te imputeren voor de ontbrekende records van de gedeeltelijk responderende ziekenhuizen; zie paragraaf 4.2. Hiermee hebben we  $\hat{P}_{nr,04}$  bepaald, een schatting voor de prevalentie van de non-respons in 2004. In vergelijking met de formules in de vorige paragrafen is het van belang op te merken dat het hier gaat om de missing data in 2004 zelf, en *niet* om het toepassen van missingdatapatronen van latere jaren op 2004. We schatten nu de incidentie in 2004 door voor respons en nonrespons dezelfde verhouding ‘incidentie / prevalentie’ te veronderstellen:

$$\hat{I}_{nr,04} = \hat{P}_{nr,04} \frac{I_{rov,04}}{P_{rov,04}} \quad (nr \text{ betreft hier nonrespons van 2004 zelf})$$

De totale incidenties in 2004 worden dan geschat door

$$\begin{aligned} \hat{I}_{tot,04} &= I_{rov,04} + I_{nr,04} \\ \Rightarrow \hat{I}_{tot,04} &= I_{rov,04} + \left( \hat{P}_{nr,04} \frac{I_{rov,04}}{P_{rov,04}} \right) \\ \Rightarrow \hat{I}_{tot,04} &= I_{rov,04} \left( 1 + \frac{\hat{P}_{nr,04}}{P_{rov,04}} \right) \quad (nr \text{ betreft nonrespons van 2004 zelf}) \quad (6.4) \end{aligned}$$

Dit betekent dat de oorspronkelijke, niet bijgeschatte incidenties in 2004 ( $I_{rov,04}$ ) worden gecorrigeerd met behulp van de oorspronkelijke, niet bijgeschatte prevalenties in 2004 ( $P_{rov,04}$ ) en de geïmputeerde prevalenties voor 2004 ( $\hat{P}_{nr,04}$ ). De geïmputeerde, individuele records worden dus gebruikt voor de schatting van  $P_{nr,04}$ , maar voor de schatting van  $I_{nr,04}$  worden de geïmputeerde records verder niet gebruikt. Ook voor de schatting van incidenties in de volgende jaren worden die records niet gebruikt, omdat hiervoor wordt uitgegaan van de verhouding tussen de respons-prevalenties; zie formule (6.3).

$\hat{P}_{nr,04}$  kan ook afgeleid worden uit de totale, wel bijgeschatte, prevalenties van 2004 ( $\hat{P}_{tot,04}$ ) minus de oorspronkelijke, niet bijgeschatte, prevalenties van 2004 ( $P_{rov,04}$ ), want

$$\hat{P}_{tot,04} = P_{rov,04} + \hat{P}_{nr,04}$$

Formule (6.4) gaat dan over in:

$$\hat{I}_{tot,04} = I_{rov,04} \frac{\hat{P}_{tot,04}}{P_{rov,04}} \quad (op \text{ basis van rov en nr van 2004 zelf}) \quad (6.5)$$

Indien  $P_{rov,04} = 0$ , dan is ook  $I_{rov,04} = 0$ . In dit geval wordt gekozen voor  $\hat{I}_{tot,04} = 0$ .

Op deze wijze zijn de totale incidenties in 2004 berekend, en ingevoerd in formule (6.3). Omdat de nonrespons in 2004 erg klein was, verschillen de bijgeschatte incidenties voor 2004 niet veel met de oorspronkelijk gepubliceerde (niet bijgeschatte) incidenties voor 2004. Dit geldt ook voor de bijgeschatte prevalenties voor 2004 (zie paragraaf 4.2).

## 7. Nauwkeurigheid van de schattingsmethode klinische incidenties

Net zoals bij de prevalenties zijn er simulaties uitgevoerd om de nauwkeurigheid van de schattingsmethode voor de incidenties te bepalen. De LMR 2000 is hierbij genomen als compleet jaar, en de LMR 2004 als het jaar waar de incidenties geschat moeten worden, met toepassing van het missingdatapatroon van 2006, resp. 2007. De op deze wijze geschatte incidenties voor 2004 zijn vergeleken met de werkelijke incidenties van 2004. De relatieve fouten van de schatting staan vermeld in tabel 8. Hieruit blijkt dat op hogere aggregatieniveaus de gemiddelde relatieve fouten ongeveer even groot zijn als bij de prevalenties. Op gedetailleerd publicatieniveau zijn de fouten iets groter dan bij de prevalenties.

**Tabel 8. Relatieve fout t.o.v. de absolute waarden van de incidenties bij schatting van LMR 2004 met LMR 2000, met toepassing van het missingdatapatroon van 2006 en 2007.**

Missing data patroon	Publicatieniveau	Relatieve fout (%)
2006	Totaal	0,16
	ICD-hoofdgroepen van diagnoses	0,49
	Subdiagnosen (volgens Beldoza-lijst)	1,22
	Mannen x subdiagnose	1,54
	Vrouwen x subdiagnose	1,46
	Subdiagnose x geslacht x leeftijd	5,04
2007	Totaal	0,10
	ICD-hoofdgroepen van diagnoses	0,63
	Subdiagnosen (volgens Beldoza-lijst)	1,28
	Mannen x subdiagnose	1,52
	Vrouwen x subdiagnose	1,56
	Subdiagnose x geslacht x leeftijd	6,45

De gepresenteerde relatieve fouten op niveau van diagnose/geslacht/leeftijd zijn gemiddelden over alle betreffende categorieën.

Op grond hiervan kan geconcludeerd worden dat de schattingsmethode ook voor de incidenties voldoende nauwkeurig is. Maar ook hier geldt dat de methode beperkt houdbaar is (zie ook hoofdstuk 5), te meer daar bij de incidenties uitgegaan wordt van een extra aanname, nl. dat de incidenties vaste fracties zijn van de prevalenties. De schatting van de incidenties gebruikt alleen de werkelijke incidentiegegevens uit een oud jaar (2004), welke gecorrigeerd worden met het quotiënt van de respons-prevalenties in het verslagjaar en die in het oude jaar. Hoewel de aanname dat incidenties vaste fracties zijn van prevalenties een redelijke veronderstelling bleek voor enkele oude LMR-jaren vanaf 2000, hoeft dit niet altijd het geval te zijn. Als bij bepaalde diagnoses de klinische incidentie sterk afneemt of toeneemt, hoeft de klinische prevalentie hiermee geen gelijke tred te houden. Bij de huidige schattingsmethode wordt een dergelijk verschil dan niet waargenomen. Dit is een beperking waarmee rekening gehouden moet worden bij het interpreteren van de geschatte klinische incidenties.

## **8. Referenties**

Ariel, A. en A. Israëls. Schatten van prevalenties bij de LMR vanaf 2006. Interne CBS-nota, Sector DMV. CBS Den Haag, 2009.

Ariel, A. en A. Israëls. Schatten van incidenties bij de LMR vanaf 2006. Interne CBS-nota, Sector DMV. CBS Den Haag, 2009.

Van Sijl, M. en J. de Ree. Ophogen op persoonsniveau van gegevens van de Landelijke Medische Registratie gekoppeld met de GBA. CBS-nota 0160-05-SOO. CBS Den Haag, 2005 (zie [www.cbs.nl](http://www.cbs.nl)).