

Op weg naar meer betrouwbare
prestatieberekening in
verpleeghuizen, verzorgingshuizen en
thuiszorgorganisaties

Dr. Xander Koolman

Dr. Dika Luijendijk

Dr. Lieke Boonen

SiRM – Strategies in Regulated Markets
Postbus 24355
3007 DJ Rotterdam

Rotterdam, 18 februari 2011

INHOUDSOPGAVE

1	Managementsamenvatting.....	3
2	Inleiding.....	4
3	Methoden	6
3.1	Setting	6
3.2	Studiepopulatie	6
3.3	De indicatoren	7
3.4	Verstorende factoren	7
3.5	Statistische analyse	9
4	Resultaten	11
5	Discussie en conclusie.....	15
6	Referenties	17

1 MANAGEMENTSAMENVATTING

Achtergrond en doel. Sinds 2007 meten verpleeg- en verzorgingshuizen de kwaliteit van zorg met behulp van zorginhoudelijke indicatoren. Echter, de prestaties, uitgedrukt in 1 tot 5 sterren, bleken per instelling sterk te wisselen tussen 2007/08 en 2009. Het is onwaarschijnlijk dat dit veroorzaakt wordt door echte verschillen in kwaliteit van zorg tussen de twee meetjaren. Wij hebben daarom onderzocht of toeval de verschillen kan verklaren.

Methoden. Wij hebben eerst de prestaties in 2007/08 en 2009 herberekend met logistische regressie, de frequentistische techniek die nu in gebruik is voor de prestatieberekeningen. Daarna hebben we de prestaties berekend met empirical Bayes multilevel logistische regressie, een methode die minder gevoelig is voor de invloed van toeval bij kleine studiepopulaties. De samenhang tussen de prestaties is gebaseerd op het percentage verpleeg- en verzorgingshuizen dat per indicator drie of meer sterren verschilde tussen de twee meetjaren.

Resultaten. Het percentage huizen dat drie of meer sterren verschilde tussen de twee metingen daalde van 9% naar 4% na toepassing van empirical Bayes. De effecten bleken het sterkst bij instellingen met een klein aantal waarnemingen en bij indicatoren met een zeer lage of juist zeer hoge prevalentie of incidentie.

Conclusie. De prestaties op basis van empirical Bayes vertoonden minder vaak onwaarschijnlijke wisselingen tussen de twee meetjaren. De afname is het sterkst bij die indicatoren en die instellingen waar de invloed van toeval het grootst was. Wij raden aan in de toekomst empirical Bayes methoden te gebruiken zodat de betrouwbaarheid van de prestatievergelijkingen toeneemt.

2 INLEIDING

In het afgelopen decennium heeft het programma Zichtbare Zorg in opdracht van het ministerie van VWS werk gemaakt van de visie van de Tweede Kamer dat zorginstellingen de kwaliteit van hun zorg inzichtelijk maken. Transparantie over de kwaliteit van de zorg dient meerdere doelen. Ten eerste hebben zorgkantoren en zorgverzekeraars deze informatie nodig om optimaal zorg in te kunnen kopen. Zonder informatie over kwaliteit van zorg zouden zorginstellingen enkel op prijs kunnen concurreren. Ten tweede kunnen cliënten en verwijzers met behulp van de informatie de beste plek voor een bepaalde behandeling kiezen. Ten derde kunnen zorginstellingen de kwaliteit van zorg die zij leveren met behulp van deze informatie beoordelen en waar nodig verbeteren. Daarnaast gebruikt de overheid de kwaliteitsinformatie voor toezicht en beleid. Goede informatie over kwaliteit van zorg is dus van groot belang.

Maar wat is kwaliteit van zorg en hoe meet je die? Om deze vraag te beantwoorden werden voor alle zorgsectoren waaronder de sector Verpleging, Verzorging en Thuiszorg (VVe-T) stuurgroepen ingesteld. In die stuurgroepen nemen vertegenwoordigers van cliënten, professionals, zorginstellingen, verzekeraars, beleidsmakers en Inspectie voor de Gezondheidszorg deel. Verantwoorde zorg wordt nu gedefinieerd als zorg die doeltreffend (effectief), veilig, tijdig en patiëntgericht is. Na het vaststellen van de definitie zijn vervolgens voor iedere sector specifieke indicatoren ontwikkeld. Een indicator is een meetbaar aspect van geleverde zorg dat een aanwijzing geeft over de kwaliteit van die zorg. Een voorbeeld is het voorkomen van decubitus bij cliënten in een verpleeghuis. Het Kwaliteitskader Verantwoorde Zorg VVe-T geeft een overzicht van de indicatoren, en beschrijft de achtergronden en verantwoordelijkheden van de betrokkenen (1).

Er worden twee soorten indicatoren onderscheiden: cliëntervaringsindicatoren en zorginhoudelijke indicatoren. Een onafhankelijk onderzoeksbureau meet voor iedere zorginstelling éénmaal per twee jaar de cliëntervaring met behulp van de Consumer Quality Index (CQ-index) en doet daarvan verslag (2). De zorginstellingen zelf dienen jaarlijks de metingen voor de zorginhoudelijke indicatoren te verrichten. Een deel van de zorginhoudelijke indicatoren wordt op organisatieniveau gemeten, zoals de beschikbaarheid van een arts of de aanwezigheid van een hitteprotocol, en een deel op cliëntniveau, zoals de aanwezigheid van decubitus en recente valincidenten. Dit artikel richt zich op het onderzoek met de cliëntgebonden zorginhoudelijke indicatoren.

Inmiddels hebben vrijwel alle verpleeg- en verzorgingshuizen vanaf 2007 metingen met de cliëntgebonden zorginhoudelijke indicatoren verricht. De prestaties in 2007/08 en 2009, uitgedrukt in 1 tot 5 sterren, zijn openbaar gemaakt en gepubliceerd in Jaardocumenten Maatschappelijke Verantwoording en op de website www.kiesBeter.nl. De prestatiemeting zou de komende jaren een ingrijpende invloed kunnen krijgen op het dagelijks handelen van de specialist ouderengeneeskunde, zeker wanneer zorgkantoren steeds meer prestatiecontracten gaan afsluiten. Aangezien de specialisten ouderengeneeskunde uiteindelijk verantwoordelijk zijn voor de medische kwaliteit van zorg zullen zij zeer waarschijnlijk worden aangesproken op de zorgprestaties. Tegelijkertijd is zorginhoudelijke inbreng van artsen noodzakelijk om de betrouwbaarheid en validiteit van de prestatieberekening te verbeteren.

Een onderzoek naar trends over de meetjaren 2007/08 en 2009 heeft inmiddels uitgewezen dat in de verpleging en verzorging de gemiddelde prestaties op 3 van de 15 indicatoren verbeterd was (3). Echter, voor individuele zorginstellingen vertoonden de resultaten van 2007/08 weinig samenhang met die van 2009 (3). In andere woorden, de prestaties uit 2007/08 bleken voor een belangrijk deel van de indicatoren vrijwel geen voorspellende waarde te hebben voor die van 2009. Bovendien was er weinig samenhang tussen de verschillende dimensies van kwaliteit, niet tussen de zorginhoudelijke indicatoren onderling, noch tussen cliëntervarings- en zorginhoudelijke indicatoren (zie <http://verpleeghuizen.volkskrant.nl/>). Het is onwaarschijnlijk dat de verschillen tussen de twee meetjaren toe te schrijven zijn aan sterke wisselingen in de kwaliteit van zorg in verpleeg- en verzorgingshuizen. Veel eerder lijken de verschillen het gevolg te zijn van toeval doordat de prestatieberekeningen soms gebaseerd zijn op een beperkt aantal metingen per indicator.

Het doel van onze studie was om te onderzoeken of een statistische methode anders dan de nu gebruikte methode de gevoeligheid voor toeval kon reduceren. Hiertoe hebben wij eerst met de huidige en daarna met de alternatieve analysetechniek, die minder gevoelig is voor toeval, de samenhang tussen de prestaties in 2007/08 en 2009 berekend. Wij hebben de prestaties van thuiszorginstellingen buiten beschouwing gelaten omdat slechts een beperkt deel van de indicatoren ($n = 2$) ontvankelijk is voor toeval. Wij hebben ook niet onderzocht welk effect bias zou kunnen hebben gehad op de prestatieberekeningen (validiteit), en ons beperkt tot het onderzoek naar de rol van toeval (statistische betrouwbaarheid).

3 METHODEN

3.1 SETTING

Zorginstellingen in de VVE-T zijn sinds 2007 verplicht tot het jaarlijks indienen van het Jaardocument Maatschappelijke Verantwoording (JMV). Dit is vastgelegd in de Regeling Verslaggeving van de Wet Toelating Zorginstellingen. Een beschrijving van de prestaties zoals gemeten met de zorginhoudelijke indicatoren is onderdeel van de jaardocumenten. De metingen dienen te worden verricht volgens de instructies uit het handboek van de stuurgroep VVE-T (4). Cliënten die meedoen aan de metingen hebben informed consent gegeven.

Er zijn metingen verricht in 2007, 2008 en 2009. Omdat 2007 en 2008 gezien werden als pilot jaren, mochten de zorginstellingen voor de externe verslaglegging over 2008 gebruik maken van de gegevens uit 2007 of 2008. Daarom worden de metingen uit 2007 en 2008 beschouwd als metingen uit één jaar. Indien een zorginstelling beide jaren aan de meting heeft deelgenomen, dan gebruiken wij enkel de laatste meting in dit onderzoek.

3.2 STUDIEPOPULATIE

Voor de verpleeg- en verzorgingshuizen geldt dat in principe alle cliënten in de studiestudiepopulatie opgenomen werden, met uitzondering van cliënten die korter dan 30 dagen in zorg zijn, cliënten die revalidatiezorg, enkel dagactiviteiten, of respijtzorg ontvingen, en cliënten die specifiek voor een bepaalde indicator uitgesloten moesten worden. Voor enkele indicatoren gold bovendien dat cliënten die te ziek waren of die niet aan het onderzoek wensten deel te nemen geëxcludeerd mochten worden. In 2007/08 waren er 1675 verpleeg- en verzorgingshuizen die metingen verrichtten bij 24682 cliënten (gemiddeld 78 cliënten per huis). In 2009 waren er 1633 huizen die metingen verrichtten bij 23987 cliënten (gemiddeld 74 cliënten per huis).

3.3 DE INDICATOREN

De zorginhoudelijke indicatoren die in 2007/08 en 2009 gebruikt werden waren zo gekozen dat zij gezamenlijk de verschillende aspecten van kwaliteit van zorg zo goed mogelijk beschrijven (5). Deze aspecten betroffen de kwaliteit van leven van cliënten, het functioneren van de zorgorganisatie als geheel, het professioneel handelen van de zorgverleners, en de zorginhoudelijke kwaliteit en veiligheid. De kwaliteit van leven van cliënten omvatte vier levensdomeinen: lichamelijk welbevinden en gezondheid, woon- en leefsituatie, participatie, en geestelijk welbevinden. Voor ieder van deze thema's is één of een aantal indicatoren ontwikkeld. Voor iedere indicator is een vragenlijst of ander meetinstrument gekozen. Tabel 1 geeft een overzicht van de zorginhoudelijke cliëntgebonden indicatoren, en bijbehorende meetinstrumenten.

3.4 VERSTORENDE FACTOREN

Verstorende factoren, dat wil zeggen confounders of verschillen in zorgzwaarte (case-mix) tussen cliënten, kunnen prestatievergelijkingen tussen zorgorganisaties vertekenen (6,7). Daarom dient in de data-analyse het effect van de verstorende factoren te worden uitgeschakeld. Op instigatie van de Stuurgroep VVe-T heeft een raad van deskundigen per indicator aangegeven of en zo ja voor welke factoren diende te worden gecorrigeerd (8). De raad moest zich beperken tot de factoren die per cliënt zijn gemeten: geslacht, leeftijd in jaren, en de *Care Dependency Scale* (CDS) (9). De CDS bevat 15 items over zelfstandigheid zoals: kunnen voldoen aan de behoefte te eten en te drinken, kunnen voortbewegen, en zelfzorg kunnen verrichten. De items worden op een schaal van 1 (volledig afhankelijk) tot 5 (geheel zelfstandig) gescoord. In theorie, kan een factor de berekeningen vertekenen, wanneer (1) deze factoren de indicatoren beïnvloeden, (2) deze factoren ongelijk verdeeld zijn over zorgaanbieders, en (3) deze factoren door de zorgaanbieder niet beïnvloedbaar zijn (10). Per indicator zijn de potentieel verstorende factoren terug te vinden in tabel 1. Met statistische tests is vervolgens vastgesteld of de variabelen in de dataset voldoen aan de theoretische voorwaarden. Indien dat het geval was, dan werd voor deze factoren gecorrigeerd in de uiteindelijke analyse (vetgedrukte factoren in tabel 1).

Tabel 1 Overzicht van cliëntgebonden zorginhoudelijke indicatoren

Indicator	Meetinstrument	Verstorende patiëntfactoren*
4.1 Decubitus	5-item vragenlijst over vochtletsel en decubitus	j h m g l c
4.2a Voedingstoestand (wegen)	Cliënten met onbedoeld gewichtsverlies gemeten door instelling	g l c
4.2b Voedingstoestand (navraag)	Cliënten met onbedoeld gewichtsverlies aangegeven door client	g l c
4.2c Voedingstoestand (combinatie)	Combinatie van indicatoren 4.2a en 4.2b	g l c
4.3 Valincidenten	Cliënten met valincidenten	j h m g l c
4.4 Medicijnincidenten	Cliënten met medicijn-incidenten	NVT
4.5a Psychofarmaca	Vragen over gebruik van antipsychotica, anxiolytica, hypnotica en antidepressiva	s
4.5b Antidepressiva	Vragen over gebruik van antidepressiva	NVT
4.6a Vaccinatiegraad- cliënten	Cliënten die gevaccineerd zijn	NVT
4.7a Incontinentie – prevalentie	Cliënten gediagnosticeerd met incontinentie	j h m g l c p
4.7b Incontinentie – diagnose	Wie de diagnose incontinentie gesteld heeft.	NVT
4.8 Verblijfskatheter	Cliënten met verblijfskatheter	j h m g l c p
4.9 Probleemgedrag**	5-item vragenlijst over probleemgedrag	j h m g l c
4.10 Fixatie***	Cliënten waarbij fixatie is toegepast	j h m g l c
7.2 Depressie	7-item vragenlijst over aanwezigheid depressieve verschijnselen.	j s h m g l c

j = jaren oud; s = sexe; h = houding; m = mobiliteit; g = gevaar ontwijken; c = contact vaardigheden; l = leercapaciteit; p = psychofarmaca; NVT = niet van toepassing

* vet = factor is significant gerelateerd aan zowel zorginstelling als aan de indicator De vetgedrukte factoren zijn als case-mixfactoren opgenomen in de analyse wanneer statistische tests (likelihood ratio tests op multinomial en logistische regressie modellen) uitwezen dat zij voldeden aan de theoretische voorwaarden voor verstorende factoren.

** Dit omvat verbaal en fysiek probleemgedrag, sociaal onacceptabel of storend gedrag, en weigeren van zorg zoals medicatie of hulp bij eten.

*** Dit omvat onrustblad, tafelblad en diepe stoel.

3.5 STATISTISCHE ANALYSE

Wij hebben de prestaties van de zorginstellingen voor 2007/08 en 2009 berekend met een parametrische variant van indirecte standaardisatie. Daarbij worden de werkelijke prestaties vergeleken met de prestaties die verwacht worden op basis van de zorgzwaarte van de cliënten. De methoden zijn daardoor meer geschikt voor toezicht houden dan voor selectieve zorginkoop of voor de keuze van zorginstelling door de cliënt. Voor de berekening van de prestaties gebruikten wij eerst logistische regressie, de gebruikelijke frequentistische techniek voor deze berekeningen (6). Daarna pasten wij een nieuwe empirical Bayes multilevel logistische regressie toe (11, 12). Deze methoden worden toegelicht in het kader.

Box 1 Frequentistische versus empirical Bayes methoden

Met frequentistische technieken worden de geschatte prestaties gebaseerd op de waarnemingen die een zorgaanbieder aanlevert, ook als dit aantal erg klein is. In onderzoek naar prestaties van zorginstellingen worden vaak empirical Bayes technieken toegepast. Deze technieken gebruiken niet alleen de informatie die in de aangeleverde gegevens zit, maar ook informatie die al eerder bekend was over de kwaliteit van zorg (de zogenaamde prior informatie) (O'hagan Luce 2003). Oftewel, empirical Bayes schattingen van kwaliteit van zorg zijn gebaseerd op de metingen van een zorginstelling zelf én op de verdeling van de kwaliteit van zorg in alle andere instellingen. Voorwaarde is wel dat er geen vermoeden mag zijn dat de kwaliteit van zorg in de betreffende zorginstelling sterk afwijkt van die van andere zorginstellingen.

Het relatieve belang van de gegevens van de zorgaanbieder wordt groter naarmate meer gegevens verzameld zijn. Bij instellingen met weinig waarnemingen krimpen de verschillen naar het gemiddelde (shrinkage). Dit leidt voor zorginstellingen met (heel) weinig waarnemingen minder snel tot extreme uitkomsten.

Om de empirical Bayes techniek verder toe te lichten, volgt nu een voorbeeld. Stel dat bij de indicator decubitus de verpleeg- en verzorgingshuizen gemiddeld rond de 3% scoren met een minimum van 0% en maximum van 7%. Stel nu dat een verpleeghuis in een nieuw meetjaar over slechts één cliënt gegevens aanlevert voor de indicator decubitus. Als deze cliënt decubitus blijkt te hebben, dan zou de schatting van de nieuw ontstane decubitus in dit huis volgens de frequentistische methode 100% zijn. Die uitkomst is zeer onwaarschijnlijk.

Het is waarschijnlijker dat de prevalentie in het huis ook tussen de 0 en de 7% ligt. Indien de schatting zou zijn gebaseerd op empirical Bayes, dan zou relatief weinig waarde gehecht worden aan de meting van het verpleeghuis zelf en meer waarde aan de gemiddelde

decubitus uitkomst. De geschatte waarde zal net boven de 3% liggen. Stel nu dat onder 200 cliënten een meting gedaan is en dat 14 cliënten decubitus hebben. Het is nu waarschijnlijk dat de feitelijke kans op decubitus rond de 7% ligt. In dit geval zal het eigen gemiddelde zwaarder meewegen dan de gemiddelde kans op decubitus. De empirical Bayes schatting zal nu net onder 7% uitkomen.

Per indicator hebben wij eerst de prestatie berekend als de ratio $\frac{\text{gemiddelde_score}}{\text{gemiddelde_verwachte_score}}$. Dit is de prestatie-index. De gemiddelde verwachte score wordt bepaald door per cliënt op basis van zijn karakteristieken een verwachte score te berekenen. Vervolgens hebben we per organisatorische eenheid de gemiddelde verwachte score bepaald. Een prestatie-index van 1 betekent dat de zorgorganisatie precies presteert zoals verwacht op grond van de kenmerken van haar cliëntengroep. Als de prestatie-index tussen de 0,0 en 1 ligt betekent dat dat de zorgorganisatie minder scoort dan mag worden verwacht op grond van de kenmerken van haar cliëntengroep. Dit is gunstig voor een indicator zoals decubitus, maar ongunstig voor de indicator vaccinaties. Bij decubitus duidt een lagere score namelijk op een betere zorgkwaliteit terwijl bij de indicator vaccinaties een lagere score juist duidt op een lagere zorgkwaliteit. Een prestatie-index boven de 1 geeft aan dat de zorginstelling hoger scoort dan verwacht op grond van de cliëntenkenmerken. Dit is dan juist ongunstig voor een indicator zoals decubitus, maar gunstig voor de indicator vaccinaties.

Om verwarring te voorkomen krijgt elke zorginstelling op basis van de relatieve prestatie-index voor iedere indicator één tot vijf sterren toegekend. De instellingen die behoren tot de beste 10% krijgen vijf sterren, de volgende 20% vier sterren, de volgende 40% drie sterren, de volgende 20% twee sterren en de laatste 10% één ster. Bij gelijke prestaties werd afgerond in het voordeel van de zorginstelling. Het kan voorkomen dat een instelling gewerkt had aan een betere behandeling van depressie, maar evenveel sterren kreeg als in het voorafgaande meetjaar. Dit komt doordat prestaties uitgedrukt worden op een relatieve schaal en niet op een absolute schaal: het toegekende aantal sterren weerspiegelt de prestatie van een zorginstelling in vergelijking met de gemiddelde prestatie van alle zorginstellingen (mits de prestatie vergelijking betrouwbaar en valide is). Als een verbeterlag in een instelling gelijke tred houdt met de gemiddelde stijging in de sector, dan ontvangt die instelling hetzelfde aantal sterren als het jaar ervoor. De instelling krijgt meer (of minder) sterren dan het jaar ervoor, als de instelling het op de betreffende indicator beter (of slechter) doet dan het sectorgemiddelde.

Tot slot berekenden we de samenhang tussen de meetjaren 2007/08 en 2009 voor de prestaties zoals ze berekend zijn met de gebruikelijke frequentistische en de nieuwe empirical Bayes techniek. Deze samenhang is uitgedrukt in het percentage instellingen dat

tussen 2007/08 en 2009 drie of meer sterren stijgt of daalt. Een dergelijke stijging of daling impliceert een verschuiving van beter respectievelijk slechter dan gemiddeld naar slechter respectievelijk beter dan gemiddeld waarbij de relatieve rang meer dan 60% opschuift. Wij verwachtten niet dat de werkelijke kwaliteit binnen een instelling tussen twee jaren zo sterk zal verschillen omdat kwaliteitsverbetering, of verslechtering een gestaag proces is. Daarom veronderstelden wij dat dergelijke grote verschillen in de sterrentoekenning tussen 2007/08 en 2009 veroorzaakt worden door toeval.

4 RESULTATEN

In totaal 1544 verpleeg- en verzorgingshuizen hebben voor de meetjaren 2007/08 en 2009 informatie aangeleverd over in totaal 49112 waarnemingen (complete informatie voor een indicator voor een cliënt). Tabel 2 toont de prestaties op de individuele indicatoren in 2007/08 en 2009. Vergeleken met 2007/08, laten de prestaties in 2009 gemiddeld genomen een verbetering zien. De indicator 'fixatie' laat de grootste verbetering zien, namelijk een daling van ongeveer 9% naar 5%.

Tabel 2 Gemiddelde prestatie en gemiddeld aantal cliënten per indicator per meetjaar

	Ruwe prestatie 2008	gemiddelde Gemiddeld aantal cliënten 2008	Ruwe prestatie 2009	gemiddelde Gemiddeld aantal cliënten 2009
	Aantal organisatorische eenheden = 1675	Aantal cliënten = 24.945	Aantal organisatorische eenheden = 1633	Aantal cliënten = 24.167
4.1 Decubitus	0.028	121	0.023	84
4.2a Voedingstoestand wegen	0.043	54	0.036	49
4.2b Voedingstoestand navraag	0.027	66	0.029	41
4.2c Voedingstoestand totaal	0.033	109	0.033	85
4.3 Valincidenten	0.106	122	0.101	85
4.4 Medicijnincidenten	0.079	116	0.066	82
4.5a Psychofarma	0.418	113	0.388	78
4.5b Antidepressiva	0.212	110	0.209	80
4.6a Vaccinatiegraad cliënten	0.958	99	0.979	83

4.7a	Incontinentie-prevalentie	0.608	118	0.591	83
4.7b	Incontinentie-diagnose	0.567	67	0.665	47
4.8	Verblijfskatheter	0.043	122	0.042	85
4.9	Probleemgedrag	0.302	121	0.286	85
4.10	Fixatie	0.094	122	0.054	85
7.2	Depressie	0.244	122	0.226	85

De samenhang tussen de prestaties in 2007/08 en 2009 zoals berekend met de huidige frequentistische techniek en de empirical Bayes techniek wordt weergegeven in tabel 3. We zien dat de prestaties op basis van de frequentistische techniek sterk konden wisselen tussen de jaren. Zo zijn er vijf indicatoren waarbij meer dan 10 procent van de aanbieders drie of meer sterren wisselden tussen beide jaren, en voor drie indicatoren ligt dat percentage boven de 20 procent.

Tabel 3 Percentage zorgaanbieders met prestaties die tussen 2007/08 en 2009 drie of meer sterren verschilden

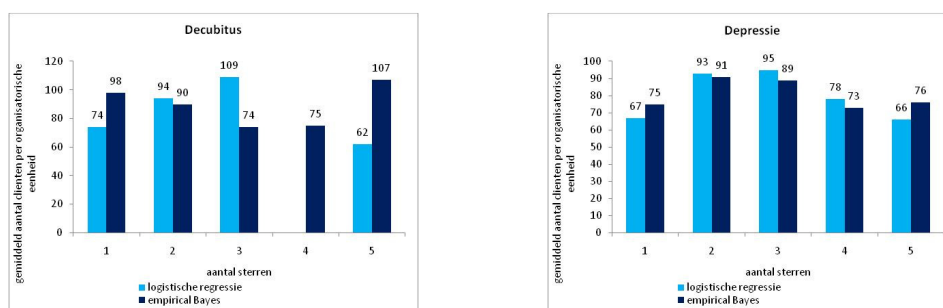
				Logistische regressie	Empirical Bayes
Gehele dataset (alle indicatoren)			9		4
4.1	Decubitus		16.9		6.5
4.2a	Voedingstoestand wegen		21.0		8.8
4.2b	Voedingstoestand navraag		23.8		9.5
4.2c	Voedingstoestand totaal		11.9		6.4
4.3	Valincidenten		7.0		5.9
4.4	Medicijnincidenten		4.5		4.5
4.5a	Psychofarma		3.2		2.9
4.5b	Antidepressiva		1.9		1.1
4.6a	Vaccinatiegraad cliënten		20.8		6.8
4.7a	Incontinentie-prevalentie		4.8		0.9
4.7b	Incontinentie-diagnose		5.5		4.9

4.8 Verblijfskatheter	5.6	2.4
4.9 Probleemgedrag	3.5	3.4
4.10 Fixatie	8.6	3.2
7.2 Depressie	5.2	4.6

De empirical Bayes techniek liet een grotere samenhang zien tussen de prestaties van instellingen op basis van de sterntoekenning. Voor geen enkele indicator geldt dat meer dan 10 procent van de zorginstellingen drie of meer sterren wisselde. Het verschil in prestaties tussen de twee meetperioden bleek het meest af te nemen bij indicatoren waarvoor de variatie in aantal sterren in een periode het grootst was. Dit zijn tevens de indicatoren met een incidentie of prevalentie die dicht bij 0 of dicht bij 1 ligt, waardoor deze indicatoren gevoeliger zijn voor toeval.

Figuur 1 laat voor decubitus zien dat de sterindeling op basis van logistische regressie leidt tot een oververtegenwoordiging van instellingen met weinig waarnemingen in beide extremen. De figuur geeft het gemiddeld aantal cliënten van instellingen met respectievelijk 1, 2, 3, 4 of 5 sterren weer. Bij de schattingen op basis van logistische regressie zien we dat juist de instellingen die door het kleine aantal waarnemingen gevoelig zijn voor toeval meer kans maakten op een zeer goede of een zeer slechte beoordeling. Bij de empirical Bayes schattingen zien we daarentegen dat de instellingen met meer waarnemingen een of vijf sterren scoorden. Het verschil tussen logistische regressie en empirical Bayes hangt vooral af van het aantal cliënten en de proportie cliënten die positief scoren op de betreffende indicator. Zo bleek het effect van de nieuwe methode kleiner bij de indicator depressie die door de hogere prevalentie minder gevoelig is voor toeval (figuur 1). (De voorbeelden zijn gekozen omdat zij illustratief voor alle bevindingen.)

Figuur 1 Gemiddeld aantal cliënten per organisatorische eenheid uitgezet tegen het behaalde aantal sterren: logistische regressie versus empirical Bayes*



* Er was er was geen enkele organisatorische eenheid die 4 sterren kreeg met logistische regressie berekening van de decubitus indicator.

De instellingen leverden voor gemiddeld 72 cliënten per indicator gegevens aan. De variatie tussen instellingen was echter groot en varieerde tussen 1 cliënt en 349 cliënten. In tabel 4 is aangegeven hoeveel procent van de instellingen voor minder dan 25, tussen de 25-50, tussen de 50-75, tussen de 75-100 en meer dan 100 cliënten gegevens heeft aangeleverd. 14% van de instellingen leverde voor minder dan 25 cliënten gegevens aan. 61% van de instellingen levert voor minder dan 75 cliënten gegevens aan. Er is geen minimum of maximum grens voor het aanleveren van gegevens.

Tabel 4 Gewichten van de eigen waarnemingen in de empirical Bayes berekeningen

Gemiddelde aantal cliënten	Gewicht eigen waarnemingen	% instellingen
<25 cliënten	37%	14%
tussen de 25 en 50 cliënten	52%	22%
tussen de 50 en 75 cliënten	60%	25%
tussen de 75 en 100 cliënten	67%	17%
meer dan 100 cliënten	78%	23%

De prestatie van kleinere organisaties wordt meer beïnvloed door het gemiddelde in de populatie dan de prestaties van grotere organisaties. We laten dit zien aan de had van het gewicht dat de eigen score heeft in de empirical Bayes schatting. Tabel 4 laat zien dat bij kleinere organisaties het gewicht van de eigen prestaties minder zwaar woog dan het populatie gemiddelde. Hoe groter de organisatie, hoe zwaarder de eigen prestaties wogen.

5 DISCUSSIE EN CONCLUSIE

De inspanning van zorgorganisaties, hulpverleners, en cliënten om data te verzamelen over de kwaliteit van de geleverde zorg is uiteindelijk alleen alle moeite waard als de informatie die gegenereerd wordt een correct beeld geeft van de werkelijke verschillen in kwaliteit van zorg. In andere woorden, de prestatievergelijkingen moeten valide en statistisch betrouwbaar zijn. Wij hebben onderzoek gedaan naar de statistische betrouwbaarheid van de cliëntgebonden zorginhoudelijke indicatoren. Wij vonden dat de prestaties in 2007/08 weinig samenhang lieten zien met die van 2009 als ze gebaseerd zijn op logistische regressie, de totnogtoe gebruikte frequentistische techniek voor de prestatie berekeningen. Dit gold voor een groot deel van de indicatoren.

Het is waarschijnlijk dat toeval het gebrek aan samenhang verklaart in de resultaten berekend met een frequentistische techniek. Betrouwbaarheid vraagt een correcte steekproefgrootte passend bij de gewenste statistische betrouwbaarheid en het gewenste onderscheidend vermogen (13). Toeval heeft relatief een groter effect naarmate het aantal waarnemingen en het werkelijke verschil in kwaliteit van zorg kleiner zijn. Voor alle indicatoren geldt dat minstens één organisatie bij een zeer klein aantal cliënten metingen had verricht (≤ 7). Daardoor is het vermogen van de analyse om verschillen in uitkomsten op te pakken, het zogeheten onderscheidingsvermogen, laag.

Empirical Bayes is minder gevoelig voor toeval dan de frequentistische logistische regressie (11,12). Deze nieuwe methode maakt gebruik van zowel de scores per organisatie als de gemiddelde scores over alle organisaties. Indien een organisatie weinig waarnemingen heeft dan ligt de score van deze organisatie dicht bij de gemiddelde score. Wanneer een organisatie veel waarnemingen heeft dan is de score van die organisatie vooral gebaseerd op de eigen waarnemingen. Dat empirical Bayes minder gevoelig is voor toeval, zien we terug in de hogere samenhang tussen de berekende prestaties van 2007/08 en 2009. Toepassing van empirical Bayes reduceert dus de toevalsvariantie in de berekende prestaties van instellingen met voldoende waarnemingen. De prijs die hiervoor betaald moet worden is dat empirical Bayes resultaten krimpen naar het gemiddelde (shrinkage) en daardoor minder onderscheidend zijn. Of anders gezegd, indien alle zorgaanbieders een groot aantal waarnemingen zouden aanleveren dan zou blijken dat de verschillen in onderlinge kwaliteit groter zijn dan nu blijkt uit de empirical Bayes schattingen. Als gevolg van deze beperking van de methode kunnen kleine instellingen zich op sommige indicatoren niet of moeilijk onderscheiden vanwege te weinig waarnemingen. Zij krijgen voor deze indicatoren vrijwel altijd een gemiddelde beoordeling. Bij toekomstige presentaties van de resultaten zal daarom worden aangegeven of het aantal waarnemingen voldoende was voor een statistisch betrouwbare schatting. Wij adviseren om daarnaast de

toekomstige prestatievergelijking in de zorg verder te ontwikkelen om deze beperking te ondervangen.

De hierboven besproken technieken garanderen niet dat de prestatievergelijkingen valide. Valide vergelijkingen vereisen de afwezigheid van vertekening (14). Vertekening kan optreden door onvoldoende correctie voor verschillen in zorgzwaarte van cliënten (confounding bias). In de terminologie van Zichtbare Zorg heet dit dat de prestaties niet vergelijkbaar zijn. In ons onderzoek zijn de versturende factoren geselecteerd door een raad van inhoudsdeskundigen. Zij konden kiezen uit de items van de CDS, leeftijd, geslacht en de overige indicatoren. Het is waarschijnlijk dat er daardoor versturende factoren niet opgenomen zijn in de berekeningen. Dat kan extra belangrijk zijn bij de beoordeling van zorginstellingen die zijn gespecialiseerd in bepaalde groepen cliënten, zoals patiënten met de ziekte van Korsakow. De ernst van deze zogeheten residuele confounding is zonder aanvullend onderzoek echter niet te bepalen.

Verder is het belangrijk dat geen vertekening optreedt door onjuiste gegevens (informatie bias) of onvolledige gegevens (selectie bias). In Zichtbare Zorg-terminologie: de registratiebetrouwbaarheid dient hoog te zijn. Een deel van de zorginstellingen registreerden de gegevens met het Elektronisch Cliënt Dossier, de Registratietool Verantwoorde Zorg, de Landelijke Prevalentiemeting Zorgproblemen, of Rai-view. Sommige van deze systemen dienen meerdere doelen, zoals aansturing, verbetering van de kwaliteit van zorg, en verbeterde uitwisseling van patiëntgegevens tussen hulpverleners. Het is waarschijnlijk dat meetfout (of juistheid) hiermee samenhangt. Daarnaast kunnen er andere oorzaken zijn van verschillen in registraties, waaronder opleidingsniveau van degene die de meting uitvoert of strategisch gedrag.

Voor een deel van de instellingen zijn de aangeleverde gegevens onvolledig. Onvolledigheid kan leiden tot een niet-representatieve steekproef. Vooralsnog kunnen de zorginstellingen zelf invloed uitoefenen op de selectie van cliënten in de 'steekproef'. Zo kunnen zorgaanbieders de keuze van cliënten om af te zien van deelname beïnvloeden. Naar verwachting zal de prikkel tot strategische registratie groter worden indien de financiële consequenties van de prestaties toenemen.

Wij concluderen dat prestatievergelijkingen voor de verpleeg- en verzorgingshuizen minder gevoelig zijn voor toeval indien zij geanalyseerd worden met behulp van empirical Bayes in plaats van een frequentistische techniek. Ons advies is empirical Bayes vanaf 2010 te gebruiken voor de berekening van de prestaties van verpleeg- en verzorgingshuizen. Dat zal de betrouwbaarheid van het prestatievergelijkingen vergroten. Hierdoor is de kans kleiner dat zorginstellingen op onterecht sterk wisselde resultaten worden afgerekend en dat zorginkopers en consumenten de kwaliteitsinformatie zullen negeren.

6 REFERENTIES

- (1) Zichtbare Zorg Verpleging Verzorging Thuiszorg. Kwaliteitskader Verantwoorde Zorg VVe-T. Den Haag: Stuurgroep Kwaliteitskader Verantwoorde zorg VVe-T, april 2010. (pdf te vinden op www.zichtbarezorg.nl)
- (2) www.centrumklantervaringzorg.nl
- (3) SiRM. Standaardisatie zorginhoudelijke indicatoren; Verpleging verzorging en zorg thuis, 3e meetronde 2009. Den Haag: SiRM, 2010.
- (4) Handboek Zichtbare Zorg Verpleging Verzorging Thuiszorg. Handboek bij het Kwaliteitskader Verantwoorde Zorg Verpleging Verzorging en Zorg Thuis (VVT) Langdurige en/of complexe zorg. Den Haag: Stuurgroep Verantwoorde zorg VVe-T, mei 2007.
- (5) Zichtbare Zorg Verpleging Verzorging Thuiszorg. Bijlage Indicatorenset Verantwoorde zorg behorend bij het Kwaliteitskader Verantwoorde Zorg Verpleging Verzorging en Zorg Thuis (VVT) Langdurige en/of complexe zorg. Den Haag: Stuurgroep Verantwoorde zorg VVe-T, mei 2007.
- (6) Bours GJ, Halfens RJ, Berger MP, Huijter Abu-Saad H, Grol RT. Development of a model for case-mix adjustment of pressure ulcer prevalence rates. *Medical care* 2003; 41: 45-55.
- (7) Hernán MA, Hernández-Díaz S, Werler MM, Mitchell AA. Causal knowledge as a prerequisite for confounding evaluation: An application to birth defects epidemiology. *American Journal of Epidemiology* 2002 155:2 (176-184)
- (8) Koolman Advies en Onderzoek. Standaardisatie zorginhoudelijke indicatoren verpleging, verzorging en zorg thuis. Rotterdam: Koolman Advies en Onderzoek, 2008.
- (9) Dijkstra A, Buist GAH, Dassen TWN, Heuvel van den WJA, Het meten van zorgafhankelijkheid met de ZorgAfhankelijkheidsSchaal Care Dependency Scale (CDS) Een handleiding. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken, Rijksuniversiteit Groningen, 1999.
- (10) Zichtbare Zorg Verpleging Verzorging Thuiszorg. De sterren uitgelegd. Naar vergelijkbare zorginhoudelijke kwaliteitsinformatie. Den Haag: Stuurgroep Kwaliteitskader Verantwoorde zorg VVe-T, november 2009.
- (11) O'Hagan A , Luce BR . A Primer on Bayesian Statistics in Health Economics and Outcomes Research. Sheffield: Centre for Bayesian Statistics in Health Economics, 2003. (pdf te vinden op <http://www.shf.ac.uk/~st1ao/pdf/primer.pdf>)

(12) Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation. Chichester: Wiley, 2004.

(13) SiRM. Poweranalyse zorginhoudelijke indicatoren; Verpleging verzorging en zorg thuis. Den Haag: SiRM, 2010.

(14) Bouter LM, Dongen van MCJM, Zielhuis GA. Epidemiologisch onderzoek: opzet en interpretatie. Zesde druk. Houten: Bohn Stafleu van Loghum, 2010