



Ziekenhuiskenners Keuren Zichzelf

De Elsevier Onderzoeksaanpak Doorgelicht (een MOA Audit - Maart 2009)

Inleiding:

Gezondheid is van groot belang. Dat wordt voor ieder vooral weer duidelijk als het even minder goed gaat en er een ziekenhuisopname noodzakelijk is. Maar in toenemende mate is er discussie over de kwaliteit van de zorg. Hoe weet je dat je in goede handen bent?

Sinds 1996 publiceert Elsevier in samenwerking met Peter Lagendijk een jaarlijks onderzoek over de kwaliteit van ziekenhuizen in Nederland. Hierbij worden deskundigen als “medisch specialisten, verpleegkader, bestuur/management en huisartsen” benaderd om een oordeel te geven over hen bekende ziekenhuizen, inclusief het ziekenhuis waar zij zelf aan verbonden zijn dan wel veel mee te maken hebben. Het AD publiceert inmiddels ook een ranglijst met als basis kwaliteitscriteria opgesteld door onder meer de “Inspectie van de Gezondheidszorg” en de mening van patiënten. Daarnaast verschijnt een jaarlijkse studie van Roland Berger met als invalshoek financiële, operationele en medische prestaties middels indicatoren van de “Inspectie voor de Volksgezondheid”.

Bij nogal wat betrokken ziekenhuizen en de Nederlandse Vereniging van Ziekenhuizen (NVZ) zijn vragen gerezen inzake de onderzoeksaanpak van Elsevier/Lagendijk en de wijze waarop de uitslag van het onderzoek tot stand komt. Er is derhalve contact gelegd met de Nederlandse Marktonderzoek Associatie (MOA). Deze ziet het als haar taak de reputatie van het markt-en opinieonderzoek in Nederland te bewaken en waar gewenst de gehanteerde standaards van kwaliteit aan de orde te stellen. De MOA heeft vervolgens twee onderzoekspecialisten, Lex Olivier en Mario van Hamersveld gevraagd een “audit” uit te voeren uitgaande van de professionele maatstaven.

Doel daarvan was vast te stellen of het onderzoek en de rapportage adequaat inzicht geven in de prestaties van de beoordeelde ziekenhuizen.
En ook na te gaan of er verbeteringen dienen te worden aangebracht; zo ja welke?

De heer Lagendijk heeft in de onderzoeksperiode herhaalde malen principieel geweigerd om mee te werken aan de audit. Dit ondanks het feit dat hij in de Elsevier publicatie wordt genoemd als onderzoeker-adviseur en de betrokkenen meermalen te kennen hebben gegeven transparant te willen rapporteren. Een deel van de noodzakelijke achtergrondinformatie werd uiteindelijk via Elsevier verkregen.

Het gepubliceerde materiaal is geanalyseerd en de, na hevig aandringen, verkregen achtergrondinformatie over 2008 is bestudeerd. In deze notitie wordt verslag gedaan van de audit. Ook worden de voornaamste bevindingen samengevat.

De Audit

De steekproefstratificatie en de uitnodigingsprocedure

Bij de start van het jaarlijkse onderzoek wordt een uitnodiging gestuurd ter rekrutering van "informanten". De uitnodiging tot deelname is, zo blijkt uit de uitnodigingsbrief, gericht aan de "leden van medische staf". Omdat de identiteit van deze staf, op persoonsniveau, natuurlijk niet bekend is, richt men zich tot een algemeen e-mail adres van het ziekenhuis of tot de e-mail van de afdeling communicatie of PR van het ziekenhuis in kwestie. Het verzoek is om de e-mail door te sturen naar "alle in het ziekenhuis werkzame artsen, in het bijzonder naar artsen met patiënten contact". Hiermee geeft de onderzoeker de controle over de representativiteit uit handen; het initiatief komt bij de beoordelaars die óók zichzelf moeten beoordelen.

Met de begeleidende opmerking dat men het eigen ziekenhuis tekort doet door niet mee te werken wordt tevens gesuggereerd dat men de onderzoeksresultaten kan beïnvloeden. Hierdoor sluipt reeds bij de aanvang van het onderzoek de kans op systematische vertekeningen binnen in de onderzoeksopzet en wordt de representativiteit van het onderzoek nog verdere onder druk gezet

Als men dan vervolgens zelf ook nog kan beslissen welke andere ziekenhuizen men beoordeelt, neemt de kans op systematische vertekeningen binnen het onderzoek aanmerkelijk toe. In de onderzoekswereld is het dan ook vrij ongebruikelijk ("not done") om respondenten de mogelijkheid te geven om zelf te bepalen wat men beoordeelt. Nog minder gebruikelijk is het om een respondent (in deze situatie het ziekenhuis) de gelegenheid te geven om zelf te bepalen hoeveel maal men aan een onderzoek deelneemt. Er moet immers worden gevreesd dat hierdoor de geldigheid van de onderzoeks-uitkomsten aanzienlijk wordt gereduceerd

Zelfselectie

Daarnaast verdient het verschijnsel van zelfselectie de aandacht. Zelfselectie impliceert dat men zelf bepaalt of men al dan niet aan het onderzoek deelneemt en dat men zelf bepaalt aan welke onderdelen men meedoet (in dit geval "andere ziekenhuizen"). Immers; het eigen ziekenhuis zal er veelal beter uitkomen dan de andere ziekenhuizen

Het eigenbelang speelt een rol en er is niets bekend over de reden waarom men andere ziekenhuizen in de beoordeling betreft. Wellicht frustraties vanuit het verleden, wellicht slechts rapporterend wat men van anderen heeft gehoord.

Steekproefomvang

Een steekproefomvang van 4968 in 2008 lijkt indrukwekkend, maar is dit niet wanneer men een gedetailleerd oordeel over alle ziekenhuizen wil vellen en hen op basis van dit gedetailleerde oordeel wil "ranken" en hiermee de maat wil nemen.

Een waardeoordeel over een ziekenhuis op basis van 27 waarnemingen is onvoldoende nauwkeurig. Hierbij gaat het niet alleen om statistische betrouwbaarheid, maar ook, of eigenlijk juist om de validiteit van het waardeoordeel. Men beoordeelt niet alleen het ziekenhuis waaraan men zelf verbonden is, maar ook collega/concurrerende ziekenhuizen.

Uit wetenschappelijke studies blijkt steeds weer dat “het bekende” (dus eigen ziekenhuis) altijd hoger gewaardeerd wordt dan “het onbekende” (het andere/ concurrerende ziekenhuis). Daarnaast is de kans levensgroot aanwezig dat de resultaten van het onderzoek worden gemanipuleerd.

Statistische significanties

In de Elsevier publicatie wordt een aantal malen een hoger/lager oordeel gegeven terwijl hier geen stevige statistische basis voor is. Natuurlijk is het Elsevier tijdschrift geen wetenschappelijke publicatie, maar gelet op de gewenste impact van de Elsevier “redactie” in medische kringen alsmede het belang van het onderhavige onderwerp kan een zorgvuldigere toetsing niet meer achterwege blijven. Ook het vertrouwen van de lezer in de betrouwbaarheid van het redactionele niveau van Elsevier schept verplichtingen.

Vraagstelling en vertekeningen

De informant wordt gevraagd een aantal oordelen te geven c.q. cijfers te verstrekken. Niet duidelijk is hoe de inhoud van de vragenlijst tot stand is gekomen respectievelijk of de relevantie van de vragen en verwoording is gevalideerd.

Op diverse plaatsen wordt gevraagd een vergelijkend ja/nee oordeel te geven op kwaliteitsaspecten. Het nee- oordeel kan ontstaan doordat de respondent dit aangeeft, maar ook door het feit dat hij niet “op de hoogte is van” hetzij “geen mening heeft over” het niveau van het kwaliteitsaspect. Voor een zuivere beoordeling van het kwaliteitsniveau van het desbetreffende onderdeel in het desbetreffende ziekenhuis dient het bekendheidseffect en het kwaliteitsoordeel nadrukkelijker te worden gescheiden.

De respondent heeft de neiging om het extreem positieve dan wel het extreem negatieve te laten prevaleren. Dit zijn immers de zaken die men zich het beste herinnert. Indien de respondent zich een extreem negatieve ervaring herinnert en op basis hiervan besluit dat hij hierover wil rapporteren, dan moet hij de volledige set vragen beantwoorden. Ook bij de aspecten van het ziekenhuis waarover hij zich geen oordeel heeft kunnen vormen. Immers, het antwoord “weet niet” is niet opgenomen in de vraagstelling.

Het gevaar bestaat dat deelnemers het eigen ziekenhuis (intern) systematisch hoger beoordelen dan de concurrerende ziekenhuizen: het zogenaamde “halo-effect”. Er wordt in de nadere uitleg bij de resultaten van 2008 onderzoek aangetoond dat dit effect zich inderdaad voordoet. Het eigen ziekenhuis scoorde in het verleden 0.6

procentpunten hoger dan de concurrerende ziekenhuizen als het ging om medisch functioneren. Dit stijgt in het laatste onderzoek, op onverklaarbare wijze, tot een verschil van 1.0 procentpunt.

Voor het bedrijfsmatig functioneren van ziekenhuizen ligt het verschil op 0.5 procentpunt. Voor de verpleegverantwoordelijken ligt het verschil v.w.b. het medisch functioneren eveneens op een niveau van 0.5 procentpunt. Uitgaande van een gemiddeld aantal beoordeelde ziekenhuizen van 2.6 zal de verhouding tussen intern en externe beoordelingen op 4:6 moeten liggen. Het terugwegen naar deze verhoudingen is op zichzelf genomen juist maar kan niet gebruikt worden als correctiefactor voor het halo effect- zoals door Lagendijk toegepast- als blijkt dat het halo-effect zelf aan veranderingen onderhevig is.

De validiteit van de Schoolcijfers

In het onderzoek wordt een uitgebreid rekenmodel gehanteerd alvorens te komen tot een gepubliceerde ranking. Hier wordt duidelijkheid over geboden. Blijft het feit dat er een subjectieve factor in de onderzoeksresultaten sluipt. Immers een ander rekenmodel leidt tot andere uitkomsten. Een voorbeeld is de vertaling van schoolcijfers in een puntenwaardering.

Schoolcijfer 7.1 levert 2 punten op
Schoolcijfer 7.3 levert 4 punten op

Een stijging van 2.8% (van 7.1 naar 7.3) in schoolcijfers leidt in het onderzoek tot een verdubbeling van het aantal toegekende punten. Feitelijk worden er relatief kleine verschillen opgeblazen tot enorme verschillen. In situaties waar het gevaar van manipulerende beantwoording van vragen dreigt is een dergelijke schaalombouw niet aanbevelenswaardig. Ook wanneer er ontwikkelingen in de tijd worden gerapporteerd kunnen dergelijke herberekeningen verschuivingen gaan registreren, die er feitelijk helemaal niet zijn.

De validiteit van de puntentelling

In het onderzoek worden gemeten percentages per attribuut onmiddellijk vertaald in punten en de punten worden op hun beurt weer vertaald in gevisualiseerde bolletjes. Het verdient de voorkeur de percentages langduriger als centrale meeteenheid in de analyse te handhaven. Nu is het zo dat een significante verschuivingen binnen een arbitrair gevormde categorie niet resulteert in een verandering in het aantal punten. Het omgekeerde kan natuurlijk ook het geval zijn. Een niet-significante verschuiving in een gemeten percentage leidt tot een verschuiving in het puntenaantal.

De wijze waarop in het onderzoek percentages in een vroegtijdig stadium vertaald worden in punten verlaagt het meetniveau van een ratioschaal naar een nominale schaal en minimaliseert de mogelijkheid om uitspraken te doen over statistische

significantie-niveau's. Dit geldt zeker in situaties waarin gewerkt moet worden met kleine steekproefcellen.

Veranderingen binnen de onderzoeksaanpak

De onderzoeksomvang is in 2008 verdrievoudigd. Daarnaast is de vraagstelling veranderd. Men heeft de uitslagen van 2007 en 2008 uitslagen tezamen gevoegd; beide jaren wegen even zwaar in de einduitslag. Het zal duidelijk zijn dat het veranderen van opzet en uitvoering gevolgen heeft. Volgtijdelijke vergelijking berust op drijfzand en kan derhalve beter achterwege blijven.

Het Onderzoek Concept Nader Bezien

Het is te betreuren dat Lagendijk geen volledige openheid voorstaat: men zou kunnen stellen dat door dit gebrek aan transparantie en het niet verstrekken van feitelijke informatie hij zich schuldig maakt aan dat wat hij zegt te willen bestrijden.

De validiteit van het concept van Lagendijk is van cruciaal belang: meet hij wat hij denkt te meten? Met andere woorden: geeft hij een solide, basis voor de constructie van de ranglijst van de "beste ziekenhuizen"? Op basis van de beschikbare informatie kunnen de volgende aandachtspunten worden genoemd.

Kernpunt in het onderzoek is het gebruik van opvattingen van (direct) betrokken informanten en het vaststellen van het beste ziekenhuis middels een ranking. Er is geen onafhankelijkheid; de informanten beoordelen zich zelf en zelf gekozen ziekenhuizen. Boswachter en stroper in dezelfde rol?

In deze studie ontbreekt de mening/feitelijke ervaring van de patiënt: de gebruiker van de goederen en diensten. Deze –steeds meer "zelf concluderende consument"- neemt in toenemende mate zelf de regie om de hoge emotionele onzekerheid, die veelal gepaard gaat met vragen rond ziekte en gezondheid, te reduceren.

Kwaliteit wordt niet expliciet gedefinieerd. Wat verwacht men van een ziekenhuis, wat is het belangrijkste? De ranking is gebaseerd op een zevental "pijlers" zoals "medisch functioneren, specialismen, toerusting en kennis, verpleegkundig functioneren, bedrijfsvoering, belang patiënt en belang verwijzer". Lagendijk gaat uit van dezelfde gewichten voor diverse uiteenlopende aspecten en attributenbundels. Dit is zeer discutabel: sluit een en ander aan bij de waardering door bijv. de patiënten?

Daarnaast: ziekenhuizen zijn verschillend, er is in toenemende mate een tendens van specialisatie. Ziekenhuizen kiezen voor enkele richtingen; het verdient de voorkeur vergelijkbare situaties als aandachtspunt te gaan nemen.

De audit toont aan dat de vergelijking van de posities van de ziekenhuizen in de tijd hier problematisch is door gebrek aan continuïteit in onderzoekopzet. Wijzigingen in werving van informanten, vraagstelling, herweging en herberekeningen spelen een

rol- het ontbreekt in zekere zin aan systematiek. In de uiteindelijke rapportage wordt hier onvoldoende rekening mee gehouden.

Voornaamste bevindingen.

Zijn de beste ziekenhuizen bij Elsevier/Lagendijk de beste ziekenhuizen?
Het onderzoek vertoont een aantal tekortkomingen.

De uitslagen geven naar onze mening geen adequaat inzicht in de “performance” van de in het onderzoek beoordeelde ziekenhuizen.

=Wil het Elsevier onderzoek naar de beste ziekenhuis de status krijgen die men zegt na te streven dan zal niet kunnen worden volstaan met de huidige steekproefopzet waarbij zelfselectie prevaleert. Gevolgen:

- Non-response factoren niet onder controle.
- Geen correctie voor Halo factoren mogelijk.
- Het onderzoek wekt de schijn manipuleerbaar te zijn en verliest daarmee de noodzakelijke autoriteit

=De rapportage in Elsevier, waarbij de focus ligt op ranking , is ongewenst :

- De ranking is arbitrair vanwege de subjectiviteit van het de weegfactoren
- De ranking is arbitrair vanwege de afwezigheid van statistische relevanties
- De ranking sluit niet aan bij de wijze waarop consumenten de kwaliteit van de ziekenhuizen/specialismen beleven (Men gaat meestal niet naar Groningen als men in Vlissingen woont).

Als er een goede steekproefbasis zou zijn, verdient een rapportage in termen van Top-Middenmoter-Hekensluiter de voorkeur boven een rapportage als ranking die een niet aanwezige nauwkeurigheid suggereert.

=Verandering in beoordeling van de kwaliteit kan alleen maar zuiver vastgesteld worden als de methodologie constant blijft. Door de steeds weer doorgevoerde “verbeteringen” was dit in het afgelopen jaar (2008) niet het geval en gelet op de geplande verbeteringen zullen vergelijkingen in de tijd ook in 2009 beter achterwege kunnen blijven.

Er dienen dus verbeteringen te worden aangebracht

Hoe verder?

Uit het feit dat de adressen van alle ziekenhuizen bekend zijn zou men kunnen afleiden dat er sprake is van een werkbaar steekproefbasis. Omdat er echter per ziekenhuis geen enkele structurele informatie over de identiteit van specialisten, verplegend personeel bekend is, is men genoodzaakt genoeg te nemen met zelf-selecterende respondentengroepen en medewerking van individuele leden van de staf. Zoals eerder aan de orde gesteld is er hierdoor onvoldoende zicht op de effecten

van non-response en het gesignaleerde halo- effect. De oplossing zou kunnen worden gezocht in het beter in kaart brengen van het universum en uit dit universum a-selekt steekproefdeelnemers te selecteren. De deelnemers ontvangen dan, bij niet reageren een aantal reminders. Bij voorkeur middels een beschikbaar email adres. Naarmate de steun binnen de ziekenhuisbranche zou toenemen wordt deze optie haalbaarder.

In het voorgaande is aangegeven dat er ook aandacht zou moeten worden gegeven aan de vragenlijst, de cijfertoekenning, de puntentelling e.d. Daarmee kan vooruitgang worden geboekt, maar de vraag is of aanpassingen dienaangaande op termijn voldoende zijn respectievelijk aansluiten bij de behoeften van de “ziekenhuispatiënt”.

Elsevier heeft pionierswerk verricht in het geven van inzicht op een terrein dat maatschappelijk steeds meer discussie oproept: marktwerking , hogere eisen aan kwaliteit, toenemende kosten, verschuiving naar diensten, vergrijzing, de opkomst van info-media als internet (“ Google als huisarts”),dit zijn enkele van de thema’s die aan de orde komen. Er is echter een sterke behoefte aan betere, betrouwbare duiding en richting.

Zoals in de inleiding is gememoreerd zijn er diverse onderzoeken om “lijstjes” op te stellen, op basis van uiteenlopende informatie. Dat maakt het er voor de geïnteresseerden niet eenvoudiger op. Wij pleiten er dan ook voor die samen te gaan voegen. Idealiter nemen brancheorganisaties in dit soort situaties het initiatief een meer integrale opzet te realiseren: met opinies van “deskundigen”, “patiënten” en officiële instanties, maar ook met een onderbouw van feiten waar mogelijk. Slagvaardiger zelfregulatie – met als vertrekpunt systematisch, onafhankelijk onderzoek- zou dat mogelijk zijn? Wij gaan het zien.

Elsevier/Lagendijk dient de opzet en uitvoering aanzienlijk te verbeteren om betrouwbare uitslagen te kunnen publiceren. Vooral de steekproeftrekking moet worden aangepast. De gehanteerde werkwijze is niet representatief en zet door zelfselectie de deur open voor aanzienlijke vertekening. Ook de manier van ranking en vergelijkbaarheid in de tijd laat te wensen over. Elsevier/Lagendijk heeft momenteel geen solide basis om te meten wat men zegt te meten: De rangorde is niet goed.

Op termijn zouden niet alleen opinies van deskundigen en van patiënten moeten worden meegenomen, maar ook feiten m.b.t. medische prestaties. Goede beoordeling vraagt om een integrale benadering!

Maart 2009

Mario van Hamersveld
Lex Olivier