

EVALUATIE VAN DE TOEPASSING VAN RICHTLIJNEN

F.G. Schellevis, J.C. Wille en W.J.J. Assendelft

KERNBOODSCHAPPEN

- De toepassing van richtlijnen kan worden geëvalueerd met behulp van gerichte evaluatiestudies of met behulp van routinematig beschikbare gegevens.
- Gerichte evaluatiestudies zijn vooral geschikt voor het meten van scores op interne indicatoren (bedoeld voor verbetering van het eigen zorgproces) en evaluatiestudies met routinematig beschikbare gegevens voor externe indicatoren (bedoeld voor het afleggen van verantwoording aan de buitenwereld).
- De kwaliteit van handelen bij patiënten is gekoppeld aan de hulpverlener die deze handelingen uitvoert. Daarom moeten de gegevens met zogenoemde multi-level-analyse per *hulpverlener*, en niet per *patiënt* worden geanalyseerd. Als hiermee geen rekening wordt gehouden, dan worden de indicatorscores overschat en wordt de variatiebreedte tussen hulpverleners onderschat.

Inleiding

De evaluatie van de toepassing van evidence-based richtlijnen in de gezondheidszorg is een essentiële stap in het bevorderen en handhaven van de kwaliteit van de zorg die met de toepassing van richtlijnen wordt nagestreefd. De evaluatie van de toepassing ervan is een van de stappen in de cyclus van kwaliteitsverbetering (zie hoofdstukken 4 en 18).

De evaluatie van de toepassing van richtlijnen zal veelal plaatsvinden aan de hand van indicatoren die kernaspecten van de richtlijnen weergeven (zie hoofdstuk 25). Het gebruik van indicatoren gaat uit van de veronderstelling dat indien zorgverleners handelen conform de richtlijnen, zij zich ook aan de andere onderdelen van de richtlijnen zullen houden. Ook wordt wel verondersteld dat niet alle onderdelen van de richtlijnen in gelijke mate bijdragen aan het uiteindelijke doel dat met de toepassing van de richtlijnen wordt nagestreefd (meestal verbetering, of vertraging van verslechtering van de gezondheidstoestand van de patiënt), maar dat die (aspecten van de) richtlijnen als indicatoren worden geselecteerd, die worden geacht de belangrijkste bijdrage aan het beoogde doel te leveren. Het gebruik van indicatoren voor de evaluatie van de toepassing van richtlijnen is ook efficiënt: het is meestal ondoenlijk om de toepassing van alle (aspecten van alle) richtlijnen te evalueren. De uitkomsten van de evaluatie kunnen vervolgens richting geven aan een eventueel uit te voeren meer gedetailleerde evaluatie. In

overeenstemming met de gebruikelijke theorieën over kwaliteit van zorg wordt ook bij kwaliteitsindicatoren onderscheid gemaakt tussen structuur-, proces- en uitkomstindicatoren, die de kernaspecten van de structuur, het proces en de uitkomst van de zorg weergeven. Daarnaast wordt onderscheid gemaakt tussen zogenoemde interne en externe indicatoren. Dit onderscheid verwijst naar het doel waarvoor de indicator is ontwikkeld en wordt toegepast. Interne indicatoren zijn bedoeld om informatie te verkrijgen voor verbetering van het eigen zorgproces. Het doel van externe indicatoren is verantwoording af te leggen aan de buitenwereld en de mogelijkheid te scheppen tot onderlinge vergelijking.

Om na te gaan in hoeverre zorgverleners de richtlijnen toepassen, moeten scores op dergelijke indicatoren worden berekend. Deze scores hebben over het algemeen de vorm van een breuk, respectievelijk worden uitgedrukt in een proportie (zie kader).

<i>Indicator:</i>	het voorschrijven van een smal-spectrumantibioticum bij vermoeden van een door bèta-hemolytische streptokokken veroorzaakte tonsillitis
<i>Score:</i>	het aantal patiënten met een vermoeden van een door bèta-hemolytische streptokokken veroorzaakte tonsillitis aan wie een smal-spectrumantibioticum wordt voorgeschreven gedeeld door het totaal aantal patiënten met een vermoeden van een door bèta-hemolytische streptokokken veroorzaakte tonsillitis

Om een indicatorscore te kunnen berekenen, zijn dus gegevens nodig over (de frequentie van vóórkomen van) bepaalde fenomenen. Dergelijke gegevens kunnen ruwweg op twee verschillende manieren worden verkregen:

- uit speciaal daarvoor ontworpen registraties, al of niet in het kader van gerichte evaluatiestudies;
- uit routinematig en/of voor een ander doel geregistreerde gegevens.

Beide methoden van gegevensverzameling hebben voor- en nadelen. Hier gaan we nadrukkelijk niet in op de evaluatie van de ontwikkeling en/of de implementatie van de richtlijnen zelf (zie de hoofdstukken 29 t/m 32). Hiervoor zijn in het algemeen meer arbeidsintensieve haalbaarheids- en interventiestudies nodig, die gedetailleerde informatie moeten opleveren over zo veel mogelijk aspecten van de richtlijnen. Daarbij zal in het algemeen geen gebruik worden gemaakt van indicatoren.

In dit hoofdstuk worden achtereenvolgens de gerichte evaluatiestudies van de toepassing van richtlijnen en de evaluatiestudies door gebruik van routinematig beschikbare gegevens kort uiteengezet, waarna wij de voor- en nadelen ervan bespreken. Ten slotte maken wij enkele kanttekeningen bij de evaluatie van de toepassing van richtlijnen.

28.1 Evaluatie door gerichte studies

Gerichte evaluatiestudies worden uitgevoerd indien er behoefte bestaat aan een gedetailleerde evaluatie van de toepassing van een of enkele richtlijnen. Een dergelijke studie maakt het namelijk mogelijk om de mate van toepassing van de richtlijnen in detail vast te stellen en, bijvoorbeeld, relaties te onderzoeken tussen de toepassing van richtlijnen (veelal de structuur en/of het proces van zorgverlening) en de uitkomst van de zorg.

28.1.1 KENMERKEN

In de voorbereidingsfase van een gericht onderzoek kan uitvoerig worden nagedacht over welke (aspecten van de) richtlijnen in het onderzoek betrokken worden, naast de al of niet reeds opgestelde indicatoren. De voor het onderzoek benodigde gegevens kunnen door de onderzoekers zelf nauwkeurig worden afgewogen, geselecteerd en gedefinieerd en op de gegevensverzameling zelf wordt zorgvuldig toegezien. Een gericht onderzoek vraagt om de verzameling van relatief veel gegevens op gedetailleerd niveau en zal dan ook alleen op relatief kleine schaal kunnen worden uitgevoerd onder daartoe gemotiveerde zorgverleners. Dergelijke gerichte studies zijn zeer geschikt om interne indicatoren te bestuderen.

28.1.2 METHODOLOGISCHE KANTTEKENINGEN

Deelname aan een dergelijke arbeidsintensieve gegevensverzameling die veelal beperkt zal zijn tot een of enkele onderwerpen draagt het risico in zich van het zogeheten Hawthorne-effect: alleen al door deelname aan het betreffende onderzoek, door het vastleggen van gedetailleerde gegevens over de dagelijkse zorgverlening, wordt de dagelijkse routine doorbroken. Er is sprake van meer aandacht voor de zorg voor een specifieke groep patiënten waardoor er – bewust of onbewust – al veranderingen in die zorg optreden, meestal in de gewenste richting. De te registreren gegevens zijn immers gebaseerd op de te evalueren richtlijnen en door het vastleggen van de verleende zorg wordt de zorgverlener zelf al een spiegel voorgehouden. De intensieve betrokkenheid van de deelnemers is slechts een beperkte periode vol te houden, waardoor de resultaten van de evaluatie een momentopname zullen weergeven; herhaling van de studie na enige tijd in – bij voorkeur dezelfde onderzoekspopulatie – is de enige mogelijkheid om een verandering of een trend te kunnen waarnemen. De motivatie van de deelnemers aan een gerichte studie en de relatief kleine omvang ervan betekenen dat de generaliseerbaarheid van de resultaten noodgedwongen beperkt zal zijn. Dit maakt deze gerichte studies minder geschikt voor externe indicatoren. Een voorbeeld van een gerichte studie wordt in het volgende kader gegeven.

Bij een onderzoek onder 413 patiënten met astma of chronic obstructive pulmonary disease (COPD) uit de praktijken van zestien huisartsen werd onder andere nagegaan in hoeverre de huisartsen zich hielden aan acht verschillende aspecten met betrekking tot de follow-up van deze patiënten (frequentie van controles, uitvoeren van longfunctiemetingen, geven van een stoppen-met-rokenadvies enzovoort).

28.2 Evaluatie door gebruik van routinematig beschikbare gegevens

In Nederland worden in de gezondheidszorg op grote schaal gegevens geregistreerd voor uiteenlopende doeleinden, variërend van abstracte informatie voor de voorbereiding of evaluatie van het beleid in de gezondheidszorg tot exploitatiegegevens van bepaalde sectoren in de zorg. De aard van de gegevens die – vaak op continue basis – worden vastgelegd, loopt uiteen van financieel-administratieve gegevens (bijvoorbeeld voor en door ziektekostenverzekeraars), gegevens over het vóórkomen van aangifteplichtige ziekten, complicatieregistraties of gegevens over de aard en omvang van de verleende zorg. Veel van dergelijke registratiesystemen, respectievelijk gegevensbestanden zijn voor derden – onder voorwaarden – toegankelijk en kunnen in beginsel ook bruikbaar zijn voor de evaluatie van de toepassing van richtlijnen.

28.2.1 KENMERKEN

Studies ter evaluatie van de toepassing van richtlijnen met behulp van reeds beschikbare gegevens worden uitgevoerd indien er behoefte bestaat aan meer globale informatie over de toepassing van richtlijnen, informatie op grotere schaal en/of informatie over de mate van toepassing van meerdere richtlijnen. Indien hiervoor gerichte studies zouden worden ondernomen, zou dit veel inspanning vergen en hoge kosten met zich meebrengen. Gebruik van reeds bestaande gegevens is dus goedkoop en efficiënt. Evaluatiestudies waarin gebruik wordt gemaakt van reeds beschikbare gegevens zijn vooral geschikt voor het meten van scores op externe indicatoren. De belangrijkste nadelen van het gebruik van bestaande gegevensbronnen zijn dat de onderzoekers afhankelijk zijn van de eerder gemaakte keuzes over de aard van de verzamelde gegevens en de definities daarvan. De bestaande gegevens zullen namelijk vaak (mede) voor een ander doel zijn geregistreerd dan voor het berekenen van scores op kwaliteitsindicatoren.

28.2.2 METHODOLOGISCHE KANTTEKENINGEN

Door gebruik van gegevens die veelal voor een ander doel zijn verzameld, bijvoorbeeld in het kader van algemene registraties, is er – door het meestal grootschalige karakter ervan – slechts beperkt inzicht in de betrouwbaarheid

en de validiteit van de gegevens. Tegelijkertijd beïnvloedt deelname aan de gegevensverzameling of registratie zelf de inhoud van de geregistreeerde gegevens minder dan bij een gerichte gegevensverzameling het geval is. Omdat aan de gegevensbron vaak een langdurige registratie ten grondslag ligt, is het bovendien vaak mogelijk om gegevens over langere perioden in het evaluatieonderzoek te betrekken, respectievelijk trends te bestuderen.

Voor een voorbeeld van gebruik van gegevens die voor een ander doel zijn verzameld, wordt verwezen naar het volgende kader.

Het Landelijk Informatie Netwerk Huisartsenzorg (LINH) – www.linh.nl – waaraan ongeveer 180 huisartsen deelnemen, baseert zijn gegevens op elektronische medische dossiers (EMD's) die primair bedoeld zijn voor de zorgverlening aan individuele patiënten. Uit de EMD's kon het opkomstpercentage in het jaar 2000 voor het bevolkingsonderzoek op baarmoederhalskanker worden berekend (66 procent). Gegevens uit het medisch dossier verklaarden bovendien voor 11 procent van de vrouwen waarom zij terecht niet aan het bevolkingsonderzoek hadden deelgenomen (bijvoorbeeld vanwege een status na uterusextirpatie of zwangerschap).

28.3 Gerichte studies en studies op basis van bestaande gegevens vergeleken

In tabel 28.1 zijn de kenmerken van gerichte studies en studies op basis van bestaande gegevens samengevat.

Tabel 28.1 Kenmerken van gerichte studies en studies op basis van bestaande gegevens voor het berekenen van scores op kwaliteitsindicatoren

	Gerichte studies	Studies op basis van bestaande gegevens
Schaalgrootte (aantal zorgverleners)	klein	groot
Invloed van evaluator op de te verzamelen gegevens	groot	klein
Aantal gegevens per indicator	groot	klein
Aantal te evalueren indicatoren	klein	groot
Kosten	hoog	laag
Continuïteit	meestal niet	meestal wel
Risico van Hawthorne-effect	groot	klein
Analyses op 'mechanismen' mogelijk	ja	meestal niet
Vooral geschikt voor	interne indicatoren	externe indicatoren

Dit overzicht kan nuttig zijn indien er keuzes gemaakt moeten worden over de wijze waarop een evaluatie van de toepassing van richtlijnen op basis van indicatoren kan worden uitgevoerd.

Met het oog op de duidelijkheid zijn de twee typen studies hier als tegenpolen beschreven. In werkelijkheid bestaan uiteraard tussenvormen, zoals grootschalige gerichte studies (zie kader).

Sinds 1996 bestaat de registratie PREZIES (PREventie van ZIEkenhuisinfecties door SurveillancE; www.prezies.nl), waaraan ongeveer de helft van de Nederlandse ziekenhuizen deelneemt. Deze registratie is specifiek opgezet voor het meten van een aantal welomschreven infecties in ziekenhuizen gerelateerd aan medische behandelingen. Hoewel de registratie gericht voor dit doel is opgezet, heeft deze ook veel kenmerken van grootschalige evaluatiestudies.

28.4 Enkele kanttekeningen

Wij plaatsen enkele opmerkingen wat betreft het meet- en analyseniveau en de interpretatie van de uitkomsten bij de evaluatie van de toepassing van richtlijnen.

Het meetniveau en het analyseniveau

Een veelvoorkomende valkuil bij de evaluatie van de toepassing van richtlijnen is dat de resultaten van de studie worden gepresenteerd op hetzelfde niveau als waarop de gegevens zijn verzameld. In veel gevallen zullen de gegevens over de toepassing van richtlijnen worden verzameld c.q. geregistreerd op het niveau van patiënten of contacten met patiënten. In de richtlijn wordt echter over het algemeen het handelen van de zorgverlener bij patiënten beschreven. De uitkomsten van de evaluatie dienen daarom te worden beschreven op het niveau van de zorgverlener. De uitkomsten geven dan ook inzicht in een mogelijke clustering van scores van eenzelfde hulpverlener. Hiervoor zijn verschillende statistische analysetechnieken beschikbaar waarbij rekening wordt gehouden met de zogenoemde statistische afhankelijkheid van waarnemingen (patiënten van eenzelfde zorgverlener), zoals multiniveau- ofwel multi-levelanalyse. Daarbij is het ook mogelijk onderscheid te maken tussen de variatie die optreedt door verschillen tussen patiënten en door verschillen tussen hulpverleners. Indien geen rekening wordt gehouden met deze hiërarchie in de gegevens, worden de indicator-scores overschat, dan wel de variatiebreedte onderschat.

Interpretatie van de uitkomsten

Zoals eerder is aangegeven, bestaat de score op de indicatoren meestal uit een percentage, bijvoorbeeld het gemiddelde percentage patiënten bij wie hulpverleners het betreffende aspect van de richtlijn hebben toegepast. Een belangrijke vraag is welk percentage gehanteerd moet worden om te kunnen spreken van toepassing van de richtlijn c.q. van kwalitatief goede zorg. Dit geldt zowel voor de interne als de externe indicatoren. Veel richtlijnen zijn, noodgedwongen door het ontbreken van 'evidence' voor alle onderdelen, deels gebaseerd op consensus onder experts. Daarnaast wordt, bijvoorbeeld in de NHG-Standaarden, expliciet aangegeven dat – mits gemotiveerd

– in individuele gevallen van richtlijnen kan worden afgeweken. Bijvoorbeeld vanwege specifieke kenmerken van de betreffende patiënt of omdat de patiënt aangeeft een andere keuze te willen maken. Over het algemeen is 100 procent toepassing van richtlijnen ongewenst en onhaalbaar. Een pragmatische keuze voor de maximumscore is om de hoogst waargenomen score in een representatieve onderzoekspopulatie als maximum te beschouwen ('best practice'). Deze hoogste score kan dan worden gebruikt om de scores van anderen tegen af te zetten. Een dergelijke benadering wordt ook wel benchmarking genoemd.

Naast de maximale score is ook de variatiebreedte van de score (bijvoorbeeld de standaarddeviatie, het betrouwbaarheidsinterval van de schatting of de *range*) van belang; deze geeft immers de mate van verschil tussen de zorgverleners weer. In aanvulling hierop kan ook nog worden aangegeven bij welk deel van de zorgverleners de richtlijnen in het geheel niet of voor 100 procent worden toegepast. In het volgende kader wordt hiervan een voorbeeld gegeven.

In een nationaal representatieve groep van 79 huisartspraktijken is op basis van de elektronische medische dossiers (EMD's) nagegaan bij welk gemiddeld percentage patiënten per praktijk de richtlijn is toegepast om eens per halfjaar het serumcreatininegehalte bij patiënten met hartfalen te bepalen. De gemiddelde score voor deze indicator in de 79 praktijken was 78 procent, met een betrouwbaarheidsinterval van 73 tot 83 procent. Daarnaast bleek dat deze laboratoriumtest in 2 van de 79 praktijken nooit wordt aangevraagd.

Slot

Er zijn twee mogelijke gegevensbronnen voor de evaluatie van de toepassing van richtlijnen: gegevens ontleend aan gerichte studies en routinematig beschikbare en/of voor een ander doel verzamelde gegevens. De keuze voor een van beide hangt sterk af van het doel van de evaluatie. Als er behoefte is aan gegevens om het zorgproces zo nodig bij te kunnen sturen, dan ligt een gerichte, kleinschalige studie voor de hand. Is het doel van de evaluatie de kwaliteit van de zorg te vergelijken met die van andere instellingen of zorgverleners, dan ligt gebruik van registraties of anderszins beschikbare gegevens meer voor de hand. In beide gevallen dient de evaluator rekening te houden met belangrijke methodologische implicaties, opdat de validiteit van de uitkomsten zo groot mogelijk is.

Literatuur

Braspenning JCC, Schellevis FG, Grol RPTM (red.). Tweede Nationale Studie naar ziekten en verrichtingen in de huisartspraktijk. Kwaliteit huisartsenzorg be-
licht. Utrecht/Nijmegen, NIVEL/WOK, 2004.
Schellevis FG, Westert GP e.a. Kritisch lezen van informatie uit grote registratiebe-
standen. Huisarts Wet 1999; 42: 591-6, 601.
Website www.tellenenmeten.nl