

RIJKSINSTITUUT VOOR VOLKSGEZONDHEID EN MILIEU  
BILTHOVEN

Rapport nr. 432500 001

**Effecten van medische zorg; een conceptueel kader**

PPM Harteloh, MGW Dijkgraaf, JC Jager

mei 1996

Dit onderzoek werd verricht in opdracht en ten laste van de Directeur-Generaal van de Volksgezondheid van het Ministerie van Volksgezondheid, Welzijn en Sport in het kader van project nr. 432500 (mijlpaal eerste kwartaal 1996). Rijksinstituut voor Volksgezondheid en Milieu, Postbus 1, 3720 BA Bilthoven. tel. 030-2742417, fax 030-2744450

**VERZENDLIJST**

- 1 Directeur-Generaal van de Volksgezondheid
- 2-3 Hoofdinspecteur voor de Gezondheidszorg
- 4 Dr G Elzinga
- 5 Prof Dr Ir D Kromhout
- 6 Hoofd cVTV
- 7 Hoofd afd Gezondheidsverkenningen (cVTV)
- 8 Hoofd (a.i.) afd Modelleren en Informatievoorziening (cVTV)
- 9 Hoofd (a.i.) afd Zorgonderzoek (cVTV)
- 10 Dr J Schouten, arts (cVTV)
- 11 Ir J Jansen (cVTV)
- 12 L Stokx, arts (cVTV)
- 13 drs AEM de Hollander (CCM)
- 14 Dr L Preller (CCM)
- 15 Drs MR Kuijpers (LGM)
- 16 Dr JCC Borleffs (AZ Utrecht)
- 17 Prof Dr AF Casparie (iBMG/EUR/Rotterdam)
- 18 Drs H Stoevelaar (iBMG/EUR/Rotterdam)
- 19 Prof Dr R Leidl (EGZ/RLi/Maastricht)
- 20 Drs R Baltussen (EGZ/RLi/Maastricht)
- 21 Dr G Boas (EGZ/RLi/Maastricht)
- 22 Prof Dr B Colebunders (iTG/UvA/Antwerpen)
- 23 Dr EJ Beck (St Mary's Hospital London)
- 24-34 Leden expertgroep thema 5 VTV-1997
- 35-43 Afdeling Zorgonderzoek
- 44-55 Auteurs
- 56-57 Bibliotheek RIVM
- 58 Bureau Rapportenregistratie
- 59-69 SBD/Voorlichting & Public Relations

70-89 Bureau Rapportenbeheer

90-91 Depot Nederlandse Publikaties en Nederlandse Bibliografie

91-100 Reserve

**INHOUDSOPGAVE**

	<u>Blz.</u>
Verzendlijst	2
Inhoudsopgave	4
Abstract	6
Samenvatting	7
1. Inleiding	8
2. Begrippenkader	9
2.1. Inleiding	9
2.2. Werking en werkzaamheid	9
2.3. Efficacy	10
2.4. Effectiviteit	12
2.5. Aangrijpingspunten voor beleid	14
3 Analysekader	17
3.1. Inleiding	17
3.2. De wetenschappelijk onderbouwing van medische interventies	18
3.3. De analyse van het effect van medische interventies	20
4 De discrepantie tussen efficacy en effectiviteit	27
4.1. Inleiding	27
4.2. Verklaringen en oorzaken	27
4.3. Onderzoek naar oorzaken	31

---

5	Het effect van zidovudine; een case-studie	33
5.1.	Inleiding	33
5.2.	Gegevens over het effect van zidovudine	34
5.3.	De wetenschappelijke onderbouwing van zidovudine	39
5.4.	Efficacy en effectiviteit van zidovudine	40
5.4.1.	Effect op de progressie naar AIDS	40
5.4.2.	Effect op de sterfte bij personen met AIDS	42
5.4.3.	Lokaal gerealiseerde effecten	43
5.5.	De discrepantie tussen efficacy en effectiviteit van zidovudine	49
5.6.	Conclusie	52
6	Onderzoek naar het effect van medische interventies	55
6.1.	Huidige stand van zaken	55
6.2.	Plaatsbepaling	56
6.3.	Nadere uitwerkingen	57
6.4.	Conclusie	58
	Literatuur	59
	Bijlagen	63
A.	Het gerandomiseerd interventie onderzoek	65
B.	Meta-analyse van experimenteel onderzoek	79
C.	Uitkomst-onderzoek	95

**Abstract**

This report describes a conceptual framework for analyzing the effects of medical care. The central concepts, efficacy and effectiveness, are defined. Efficacy is the ideal or potential effect of medical care. It is measured in an experimental setting, e.g. the randomized controlled clinical trial. Effectiveness is the effect of medical care under ordinary conditions. It is measured in outcome research. The causes of a discrepancy between efficacy and effectiveness are described. Identifying and analyzing such a discrepancy is an important aspect of health policy. It provides valuable information about possibilities to improve health care. The conceptual framework is illustrated by a case study: the effect of zidovudine on HIV infection or AIDS.

## SAMENVATTING

Dit rapport behandelt het conceptuele kader voor de analyse van effecten van medische interventies. Centraal staan de begrippen 'efficacy' en 'effectiviteit'. De efficacy is gedefinieerd als het theoretisch haalbare effect van een medische interventie. Het wordt gemeten in experimenteel onderzoek (gerandomiseerde klinische trials). De effectiviteit is gedefinieerd als het in de praktijk gerealiseerde effect van een medische interventie. Het wordt gemeten in observationeel onderzoek (uitkomst-onderzoek). Een discrepantie tussen efficacy en effectiviteit kan worden veroorzaakt door een verschil in: (i) effectmeting, (ii) patiëntenpopulatie, (iii) uitvoering van de medische interventie, (iv) indicatiestelling of (v) de omstandigheden waaronder de zorg is verleend. Het opsporen en analyseren van een dergelijke discrepantie geeft informatie over aangrijpingspunten voor beleidsmaatregelen. Een bestudering van het effect van zidovudine bij een HIV infectie of AIDS illustreert de toepassing van het conceptuele kader.

## 1. INLEIDING

Om de vier jaar brengt het Rijksinstituut voor Volksgezondheid en Milieuhygiëne (RIVM) een volksgezondheid toekomstverkenning (VTV) uit. Het effect van medische zorg vormt een van de thema's in de tweede VTV die in 1997 zal verschijnen (1). Ter ondersteuning van dit thema heeft het RIVM onderzoek geïnitieerd, waarvan het onderhavige project een onderdeel is.

Het doel van dit onderzoek is het formuleren van methode voor de analyse van effecten van medische zorg. Daarbij is een beleidsrelevante invalshoek gehanteerd door onderscheid te maken tussen een maximaal haalbaar (*efficacy*) en een gerealiseerd effect (*effectiviteit*). Het opsporen en analyseren van een dergelijke discrepantie brengt mogelijkheden voor verbetering van de medische zorg aan het licht. De doelgroep van het rapport wordt gevormd door beleidsgerichte onderzoekers.

Het onderzoek omvat: (i) definiëren van de centrale begrippen (*hoofdstuk 2*); (ii) het formuleren van een stapsgewijze aanpak voor analyse van effecten van medische interventies (*hoofdstuk 3*); (iii) inventarisatie van de oorzaken van een discrepantie tussen efficacy en effectiviteit (*hoofdstuk 4*); (iv) illustratie van de analysemethode door een voorbeeld: het effect van zidovudine (AZT) bij HIV infectie of AIDS (*hoofdstuk 5*). Tot slot volgt een beschouwing over de plaats van onze bevindingen in relatie tot gezondheidszorgonderzoek (*hoofdstuk 6*). In de bijlagen wordt gedetailleerd ingegaan op experimenteel onderzoek in de geneeskunde (*bijlage A*) en meta-analyse van experimenteel onderzoek (*bijlage B*), als bronnen van gegevens over efficacy, en uitkomst-onderzoek (*bijlage C*) als bron van gegevens over effectiviteit.

Het onderzoek beperkt zich tot het effect van *medische* interventies in de context van het gezondheidszorgsysteem, alsmede het effect van één medische interventie op één bepaalde ziekte. Zelf- en mantelzorg blijven buiten beschouwing, net als het effect van eenzelfde interventie bij meerdere ziekten (fenytoïne bij epilepsie of hartritmestoomissen) of meerdere interventies bij eenzelfde ziekte (dotteren of trombolysen bij verstoppen van kransslagaders van het hart). Tevens blijft een afweging van het effect tegen de kosten (kosten-effectiviteit/baten), of het effect van een medische interventie in termen van kosten (cost-of-illness), buiten beschouwing.

We bestuderen het effect van een medische behandeling op het niveau van de arts-patiënt relatie (microniveau). De analyse van effecten op mesoniveau (bijvoorbeeld, invloed van behandeling in ziekenhuis op ziekenhuissterfte), of op macroniveau (bijvoorbeeld, invloed van de introductie van een nieuw geneesmiddel op sterftecijfers) blijft buiten beschouwing.



## 2. BEGRIPPENKADER

### 2.1. Inleiding

In de literatuur hanteert men voor het beschrijven van effecten van medische zorg verschillende begrippen (2) (3). Bovendien worden dezelfde begrippen soms op een verschillende wijze gebruikt. In dit hoofdstuk worden de centrale begrippen gedefinieerd, zoals we die in dit onderzoek hanteren. We beginnen met de definitie van een effect. Onder een effect verstaan we een gevolg, een resultaat, een essentiële component van de oorzaak-gevolg relatie (4). Het effect van een medische interventie (preventie, behandeling) is de waargenomen verandering in het beloop van een ziekte, die deze interventie teweegbrengt (5).

### 2.2. Werking en werkzaamheid

De begrippen 'werking' en 'werkzaamheid' geven ons een mogelijkheid om een effect nader te kwalificeren. De werkzaamheid betreft het (waarneembaar) effect van een medische interventie in termen van genezing, sterfte, beperkingen/handicaps of kwaliteit van leven.. De werking betreft het verklarend mechanisme voor het ontstaan van het effect. Een medische interventie kan over een werkzaamheid beschikken zonder dat er sprake is van een (bedoelde) werking (placebo-effect, sommige homeopathische middelen), of over een werking zonder dat er sprake is van een werkzaamheid.<sup>1 2</sup>

Het onderscheid tussen werking en werkzaamheid wordt fraai geïllustreerd door een controverse uit de 16e eeuw. Het was toen gangbaar op het slagveld een wond dicht te schroeien met kokende olie. De chirurgijn Ambroise Paré kon deze gangbare behandeling op een gegeven moment niet meer toepassen, toen de kokende olie opraakte door het grote aantal gewonden. Hij besloot de wonden te verbinden. Tot zijn verbazing had dit echter een veel geringere mortaliteit tot gevolg dan de gangbare behandeling. Door het dichtbranden van de (vuile) wond, ontstond, naast veel pijn en mutilatie, dikwijls een ontsteking door opeenhoping van vuil en bacteriën, in het slechtste geval gevolgd door sepsis en het overlijden van de patiënt. Het (losjes) verbinden van een wond verminderde de kans op deze complicaties (7). Hoewel het verbinden van de wond, net als de toentertijd gangbare therapie (dichtbranden), gespeend was van iedere antibiotische

---

<sup>1</sup> Bij onbedoelde gevolgen is er sprake van bij- of nevenwerking.

<sup>2</sup> Mogelijk is behandeling met zidovudine (AZT) bij HIV/AIDS een voorbeeld van het laatste. Hoewel AZT over een bewezen werking beschikt (de replicatie van het HIV virus wordt geremd), is het nog niet zeker dat er ook een effect in termen van levensverlenging optreedt (6).

werking, was deze wel meer werkzaam. Ook blijkt dat de werkzaamheid van een medische interventie vergelijkenderwijs aan het licht komt. Een principe dat in het medische experiment, de klinische trial, verdere methodologische uitwerking heeft gevonden.<sup>3</sup>

In een klinische trial wordt het effect van een medische interventie in een groep personen op een nauwkeurig beheerste en gestandaardiseerde wijze ('ideal conditions') vergeleken met het beloop van een ziekte of aandoening in een vergelijkbare groep personen die de medische interventie niet krijgt (de controlegroep) (8). Het geneesmiddelenonderzoek staat voor deze onderzoeksopzet model. De eerste fase van de ontwikkeling van een geneesmiddel betreft de *werking*. De werking van een geneesmiddel wordt allereerst onderzocht in het laboratorium, in vitro, aan de hand van een proefdier of met behulp van gezonde vrijwilligers (eerste fase klinische trial).<sup>4</sup> Vervolgens wordt de *werkzaamheid* van het middel bij patiënten onderzocht: in de tweede fase van een trial bij een geselecteerde groep van (bijvoorbeeld ernstig zieke) patiënten, in de derde fase meer grootschalig, bij een voor de beoogde doelgroep representatieve groep van personen.<sup>5</sup> Daarbij is het niet strikt noodzakelijk over kennis van de werking te beschikken, zoals het voorbeeld van Ambroise Paré toont. Op basis van casuïstiek of toevallige bevindingen kan een vermoeden omtrent de werkzaamheid van een bepaalde medische handeling bestaan. Vervolgens wordt geëxperimenteerd. Dit geschiedt op het niveau van de arts-patiënt relatie, zij het binnen een nauwkeurig gecontroleerde (experimentele) situatie. In principe is de methodologie van het geneesmiddelen onderzoek ook op andersoortige medische- en verpleegkundige handelingen toepasbaar, waarbij de werking uiteraard niet noodzakelijkerwijs in biochemische termen behoeft te worden beschreven. In de *bijlage A* is de methode van de klinische trial nader beschreven.

### 2.3. Efficacy

Onder efficacy verstaan we *'The probability of benefit to individuals in a defined population from a medical technology applied for a medical problem under ideal conditions of use'* (9).<sup>6</sup> Deze definitie wordt geoperationaliseerd door gegevens te verzamelen over:

---

<sup>3</sup> klinische trial = het gerandomiseerd klinisch interventie onderzoek, meestal aangeduid met de Engelse term 'Randomized Controlled Clinical Trial' (RCCT). De term 'klinisch' duidt op iedere interactie tussen zorgverlener en patiënt, niet noodzakelijkerwijs op behandeling in het ziekenhuis.

<sup>4</sup> Het voorafgaande laboratorium onderzoek naar de biochemische eigenschappen van het geneesmiddel wordt soms als fase 0 aangeduid.

<sup>5</sup> Een vierde fase betreft de post-marketing surveillance, voornamelijk gericht op bijwerkingen van het geneesmiddel. In deze fase komen soms ook nieuwe indicaties aan het licht.

<sup>6</sup> Afhankelijk van de onderzoeksopzet kan een trial in onze optiek dus zowel betrekking hebben op de werking als op de werkzaamheid.

(i) *'The probability of benefit...'* - kans op een gunstig resultaat- In een klinische trial geeft een significant verschil tussen de interventie en de controlegroep met betrekking tot een bepaalde responsvariabele aan dat de geteste behandeling werkzaam is.<sup>7</sup> De grootte van het verschil geeft een indruk van de mate waarin de uitgeteste behandeling beter is dan de behandeling waarmee deze wordt vergeleken. Dit laatste vormt de basis voor een conclusie over de klinische relevantie van de medische interventie.

(ii) *'...to individuals in a defined population...'* - kenmerken van de patiëntenpopulatie - Te denken valt aan leeftijd, geslacht, sociaal-economische status, diagnose, ernst, of ziektestadium. In de klinische trial vindt doorgaans een bewuste selectie van de proefpersonen plaats met het doel de conclusies van het onderzoek te kunnen generaliseren naar andere, niet-onderzochte groepen van personen. De eigenschappen van de groep personen in de derde fase van een klinische trial, de beoogde doelgroep, vormt (onderdeel van) de indicatiestelling van de behandeling bij toepassing in de klinische praktijk. Deze is van belang voor de reproductie van het onderzoeksresultaat in de klinische praktijk.

(iii) *'...from a medical technology...'* - kenmerken van de medische interventie- Geneeskunde wordt wel omschreven als een technologie toegepast binnen de context van een relatie (10). Allerlei soorten van medische interventies (van chirurgie, geneesmiddelen, tot psychotherapie) zijn als een technologie op te vatten en komen voor onderzoek door middel van een trial in aanmerking. Een bepaalde soort medische interventie legt soms beperkingen op aan de methodologische gestrengheid waarmee deze kan worden onderzocht. Zo is het dubbelblind onderzoek van psychotherapie of chirurgie bijna onmogelijk (11). Daarnaast zijn ook de kunde en vaardigheid van de zorgverlener van belang voor het welslagen van het experiment. Een valide meting van het effect van een bandage na verzwikken van de enkel vereist bijvoorbeeld een goede beheersing van de bandage techniek. Ook is de uitvoering (bijvoorbeeld soort bandagetechniek) een belangrijk kenmerk van de te onderzoeken medische behandeling. Deze is in de klinische trial nauwkeurig geprotocolleerd.

(iv) *'...applied for a medical problem...'* - indicatie voor de medische interventie - De indicatie voor een medische interventie wordt gevormd door het medisch probleem, de klacht, het symptoom of de diagnose waarmee de patiënt zich aandient. Een welomschreven en eenduidige formulering van de indicatie voor een bepaalde behandeling, vaak gepaard gaande met uitgebreid laboratorium onderzoek, maakt onderdeel uit van het onderzoeksprotocol van een klinische trial. Dit moet worden vertaald naar een

---

<sup>7</sup> Vaak gebaseerd op berekening van een oddsratio, dat wil zeggen, een verhouding tussen de kans op gunstig effect in de interventie groep en de kans op gunstig effect in een vergelijkbare groep personen die de onderzochte behandeling niet onderging.

indicatiestelling in de praktijk, waar diagnostische mogelijkheden, ook om financiële reden, beperkter of moeilijk uitvoerbaar kunnen zijn.

(v) '*... under ideal conditions of use.*' - *omstandigheden waaronder de medische interventie is toegepast* - In een klinische trial wordt een situatie geschapen waarin alle mogelijke randvoorwaarden zodanig worden beheerst, dat zij in dezelfde richting werken om een positief effect van de onderzochte medische interventie te bewerkstelligen. Dit impliceert een optimale verpleging en verzorging. Soms is daartoe opname in het ziekenhuis noodzakelijk voor een therapie die in principe buiten het ziekenhuis kan worden toegepast. Manipulatie van de omstandigheden is onderdeel van een experiment en gerechtvaardigd wanneer overdraagbaarheid van de verworven kennis op de praktijk wordt verwacht.

## 2.4. Effectiviteit

Onder effectiviteit verstaan we '*The probability of benefit to a typical patient in a population from a medical technology applied for a medical problem under ordinary conditions by the average practitioner*' (9). De effectiviteit van medische zorg wordt gemeten door observatie van het zorgproces onder gewone, dat wil zeggen, niet experimentele omstandigheden. Zorgregistraties vormen hiervoor de belangrijkste bron. In de Amerikaanse literatuur duidt men dit soort onderzoek aan als 'uitkomst-onderzoek' (12). De effectiviteit kan op eenzelfde wijze worden geoperationaliseerd als de efficacy. Er treedt echter een aantal kenmerkende verschillen op bij de invulling van de diverse aspecten van de definitie.

(i) '*The probability of benefit...*' - *kans op gunstig resultaat* - Veelal betreft het een veelheid en verscheidenheid van 'spontaan' optredende uitkomsten, dat wil zeggen, (ogenschijnlijke) gevolgen van een medische behandeling. Deze kunnen variëren van veranderingen in de gezondheidstoestand, kwaliteit van leven, sociaal-maatschappelijke situatie of globale tevredenheid van de patiënt tot kosten of zorggebruik die de behandeling met zich meebrengt. In uitkomst-onderzoek streeft men naar het meten van een zo breed mogelijk spectrum van deze gevolgen, vaak breder dan in een klinische trial het geval is (12). Een causaal verband is op grond van dit soort onderzoek echter niet zonder meer vast te stellen. Toeschrijving van een gevolg aan een oorzaak vereist immers experimenteel onderzoek. Standaardisatie van de meet- en onderzoeksmethode is dan belangrijk om de vergelijkbaarheid van uitkomsten te garanderen.

(ii) '*...to a typical patiënt in a population...*' - *kenmerken van de patiëntenpopulatie* - In principe is er geen verschil met de kenmerken van de populatie uit de derde fase van een klinische trial (leeftijd, geslacht, sociaal-economische status, diagnose, ziektestadium). In de klinische praktijk kan echter een onwillekeurige selectie plaatsvinden. De samenstelling van de patiëntenpopulatie (case-mix) verschilt daardoor in de praktijk

van die in de klinische trial. Dit heeft gevolgen voor het gerealiseerde effect. Een therapeutisch advies wordt bijvoorbeeld minder goed opgevolgd naarmate de sociaal-economische situatie slechter is, waardoor het effect van een bepaalde medische interventie bij personen met een lage sociaal economische klasse slechter kan zijn dan voor personen met een hoge sociaal economische klasse (13).

(iii) *'...from a medical technology...'* -kenmerken van een medische interventie- Een gecompliceerde medische interventie zal bij toepassing in de praktijk enige oefening vergen voordat hetzelfde effect wordt bereikt als in het experiment. De introductie in de praktijk moet dan ook worden begeleid door instructie (bij- en nascholing), waardoor de kennis opgedaan in het klinische onderzoek wordt verspreid. De vaardigheid van de zorgverlener, in combinatie met persoonlijke voorkeur en interpretatie speelt een rol bij de snelheid waarmee een nieuwe medische technologie wordt geaccepteerd (14).

(iv) *'...applied for a medical problem...'* -indicatie voor de medische interventie- In de klinische praktijk is de indicatiestelling doorgaans minder geprotocolleerd dan in het klinisch onderzoek. Daardoor kan verbreding van de indicatiestelling plaatsvinden, dat wil zeggen, de onderzochte medische interventie wordt ook toegepast op personen waarvoor deze oorspronkelijk niet was bedoeld. Dit heeft gevolgen voor het gerealiseerde effect. Zo is de efficacy van coronaire bypass chirurgie aangetoond voor een groep personen met een bepaalde vernauwing van de kransslagaders van het hart (drietakslijden). In praktijk wordt deze ingreep echter ook toegepast op personen met andere vormen van vernauwingen van de kransslagaders (15). De effectiviteit verschilt daardoor van de efficacy.

(v) *'...under ordinary conditions by the average practitioner.'* -omstandigheden waaronder de medische interventie wordt toegepast- Meestal wordt volstaan met een beschrijving van de situatie waarin de zorg is verleend (soort ziekenhuis, huisarts, etc.). Daarbij kunnen aanwezigheid van medische behandelmogelijkheden (apparatuur e.d.) in de beschouwing worden opgenomen (16).

### De centrale begrippen

In tabel 1 staan de begrippen die nodig zijn voor het beschrijven van het effect van medische interventies. Het verschil tussen de betekenis van begrippen in het gangbare spraakgebruik en de betekenis in dit rapport is aangegeven, alsmede de definities en de wijze waarop een begrip wordt gemeten.<sup>8</sup> De begrippen 'efficacy' en 'effectiveness' zijn ontleend aan het Engels. 'Efficacy' wordt vertaald met werkzaamheid en 'effectiveness'

---

<sup>8</sup>In dit rapport is ervoor gekozen om de Engelse term 'efficacy' te gebruiken. Hoewel we daardoor vervallen in een anglicisme, is er een aantal voordelen. Het begrip is welomschreven, duidelijk gedefinieerd, en het vereist niet steeds de toevoeging 'onder experimentele omstandigheden'. De begripsinhoud van de term 'effectiviteit' is beperkt tot het 'effect onder gewone, gangbare omstandigheden'. In de (met name) wat oudere literatuur wordt deze term nog vaak voor zowel 'efficacy' als 'effectiveness' gebruikt.

met doeltreffendheid (18). In het Nederlands kunnen werkzaamheid en doeltreffendheid beide effectiviteit betekenen (4). Een onderscheid is echter wenselijk. Aanscherping van de terminologie is dus vereist.

Tabel 1. Het begrippenkader voor het effect van medische interventies

in dit rapport	gedefinieerd als	meting in	in het spraakgebruik	
			Engels	Nederlands
effect	verandering in beloop van ziekte		effect	effect, gevolg, resultaat
werking	(beoogde) effect op patho-fysiologisch mechanisme	fundamenteel onderzoek, fase 0/I van klinische trial		
werkzaamheid	waargenomen (uiteindelijk) effect	vergelijkend onderzoek, fase II/III van klinische trial		
efficacy	potentieel effect	klinische trial, het experiment	efficacy	werkzaamheid, effectiviteit
effectiviteit	gerealiseerd effect	uitkomst/observatoneel onderzoek, ofwel de klinische praktijk	effectiveness	doeltreffendheid, effectiviteit

## 2.5. Aangrijpingspunten voor beleid

De *efficacy* geeft de mate aan waarin een bepaald doel van een medische interventie onder experimentele omstandigheden kan worden bereikt. We spreken van een potentieel- of maximaal haalbaar effect. De *effectiviteit* geeft de mate aan waarin een bepaald doel van een medische interventie in de klinische praktijk wordt bereikt, het gerealiseerde effect (12). Als er gegevens zijn over de *efficacy* en de *effectiviteit*, kan men beide met elkaar vergelijken, ervan uitgaande dat de metingen valide en betrouwbaar zijn. Daarbij kan een discrepantie naar voren komen.<sup>9</sup> Een voorbeeld kan dit verduidelijken.

Stel dat men het optreden van AIDS bij een asymptomatische HIV infectie tracht te voorkomen (doelstelling). Uit klinisch onderzoek blijkt het geneesmiddel zidovudine (AZT) dit effect in 96,3% van de

<sup>9</sup>Met 'discrepantie' bedoelen we (niet meer) dan dat de *efficacy* groter of kleiner is dan de *effectiviteit*. We vermijden het woord 'verschil', omdat er anders verwarring optreedt met het inherent verschil tussen *efficacy* en *effectiviteit*.

gevallen te bewerkstelligen (17). Dit is het beoogde effect dat met het geneesmiddel AZT onder optimale omstandigheden is te behalen (efficacy). Gegeven de doelstelling blijft er nog 3,7% 'niet haalbaar effect' over. Om de oorspronkelijke doelstelling te behalen is nader onderzoek naar additionele therapie, of naar een nieuw geneesmiddel nodig. In de klinische praktijk van alledag blijkt AZT slechts in 62,4% van de gevallen het beoogde effect te ressorteren (effectiviteit). Er is sprake van een discrepantie tussen efficacy en effectiviteit (zie hoofdstuk 5).

Een verklaring voor de discrepantie tussen efficacy en effectiviteit is beleidsrelevant. Een dergelijke discrepantie kan aangrijpingspunten voor beleid aan het licht brengen (tabel 2). Deze vormen onderwerpen voor kwaliteitsverbetering, consensusvorming over de juiste behandelwijze, of nader onderzoek naar de omstandigheden van zorgverlening initiëren (bijvoorbeeld, instellen op insuline in thuissituatie of in ziekenhuis). Ook kan een verschil tussen efficacy en effectiviteit een rol spelen bij de allocatie van voorzieningen en middelen. Als een werkzaam geneesmiddel in de praktijk veel minder effect heeft dan mag worden verwacht, is het immers niet raadzaam een dergelijk middel zonder meer aan de praktijk toe te vertrouwen, of uit publieke fondsen te vergoeden. Daarom is het zinvol om te bezien hoe een discrepantie tussen efficacy en effectiviteit kan worden vastgesteld (hoofdstuk 3) en waardoor deze discrepantie wordt veroorzaakt (hoofdstuk 4 en 5).

*Tabel 2. Discrepantie tussen efficacy en effectiviteit als indicator voor beleid*

discrepantie toegeschreven aan:	beleidsmaatregelen:
kans op uitkomst	standaardisering van de meting, nader onderzoek naar efficacy stimuleren, lokaal kwaliteitsbeleid stimuleren
patiëntenpopulatie	verscherping indicatiestelling, risicoselectie
medische interventie	protocollering, budgettering, volumebeperking
indicatiestelling	consensus bevorderen, nader onderzoek naar bepaalde subpopulaties stimuleren
omstandigheden	financiële of structurele maatregelen initiëren





### 3. ANALYSEKADER

#### 3.1. Inleiding

In dit hoofdstuk beschrijven we een stapsgewijze aanpak voor de analyse van effecten van medische interventies. Op individueel niveau is er soms (ogenschijnlijk) een duidelijk effect van een medische interventie. Zo was bijvoorbeeld de toepassing van penicilline op een patiënt met een levensbedreigende infectie zo treffend dat de waargenomen gevolgen zonder meer aan de gift van het middel werden toegeschreven. Hetzelfde geldt voor geneesmiddelen als insuline en sulfonamiden (19). Toepassing van deze behandelingen werd ondersteund door een goede theorie over de werking van het middel. Bovendien was er een groot (behandel)contrast. Dit is niet altijd het geval.

Wanneer een groep patiënten (bijvoorbeeld) klaagt over hoofdpijn en daarvoor een pijnstiller krijgt, kunnen we waarnemen dat de klachten bij een bepaald percentage verdwijnen. We kunnen echter allerminst concluderen dat het verdwijnen van de klachten is toe te schrijven aan het geneesmiddel. Het is immers bekend dat de pijn ook vanzelf had kunnen verdwijnen (natuurlijk beloop). Daarnaast kan het innemen of voorschrijven van een geneesmiddel als zodanig effect hebben gehad (placebo effect). Ook kan er sprake zijn van een waarnemingsfout (selectieve perceptie aan de kant van de arts of de patiënt). Het is derhalve niet gemakkelijk op basis van observatie alleen tot een conclusie over het effect van een medische interventie te komen.

De opzet van de klinische trial (RCCT) is erop gericht de diverse alternatieve verklaringen voor een waargenomen verandering te ondervangen. De opzet van een RCCT maakt een (methodologisch) zuivere vergelijking mogelijk, zodat een conclusie over het effect in termen van een oorzaak-gevolg relatie kan worden getrokken (zie bijlage A). Volgens schattingen is een groot deel (60-80%) van de momenteel in gebruik zijnde medische interventies niet door middel van een RCCT onderzocht (20).<sup>10</sup> Medische handelingen berusten vaak op traditie, intuïtie, en casuïstiek (21). Op sociaal-ethische gronden is het ook veelal niet (meer) mogelijk om een behandeling experimenteel te onderzoeken als deze inmiddels op brede schaal wordt toegepast (22). Het zou bovendien een onevenredig grote en dure onderzoeksinspanning vergen om iedere diagnostische test, medische behandeling of therapeutisch programma met een RCCT te onderbouwen (23). Daarom moeten ook

---

<sup>10</sup> Deze schattingen lopen nogal uiteen. Zij berusten op een registratie van zorghandelingen gedurende een bepaalde periode (noemer) om vervolgens de wetenschappelijke onderbouwing van deze handelingen te bekijken (teller).

andere vormen van effect-onderzoek in de beschouwing worden opgenomen en worden gewogen op bewijskracht om tot een conclusie over de relevantie van een medische interventie te komen.

### 3.2. De wetenschappelijke onderbouwing van medische interventies

De wetenschappelijke onderbouwing van een medische interventie kan op verschillende manieren worden gekwalificeerd. Naast *het aantal* onderzoeken dat het effect van medische interventies onderbouwt, zijn er *verschillende vormen* van onderzoek te onderscheiden. Te onderscheiden zijn het experimenteel onderzoek<sup>11</sup> (waaronder de klinische trial), het observationeel onderzoek<sup>12</sup> (waaronder uitkomst-onderzoek), de gevalstudie<sup>13</sup> en casuïstiek<sup>14</sup>. De conclusies die uit de verschillende vormen van onderzoek kunnen worden getrokken, verschillen in bewijskracht.

Van oudsher neemt de casuïstiek in de geneeskunde een belangrijke plaats in. Het gaat hierbij om de rapportage van eenmalig waargenomen gebeurtenissen in de alledaagse medische praktijk. Uit een analyse van dergelijke gebeurtenissen kan een aanwijzing voor de patho-fysiologie van een ziekte, de werkzaamheid van een medische interventie, of een bijwerking van een geneesmiddel voortkomen. Deze aanwijzingen worden vervolgens meer systematisch en op grotere schaal onderzocht in observationeel of experimenteel onderzoek.

In Canada heeft een nationale werkgroep drie verschillende niveaus onderscheiden om het wetenschappelijk bewijs voor het effect van een medische interventie te kwalificeren (24). Het is de bedoeling daarmee (overheids)aanbevelingen voor het medisch beleid te onderbouwen. Naast wetenschappelijk bewijskracht wordt daarbij een onderzoek ook gewogen op praktische waarde. Het wetenschappelijk bewijs wordt als volgt geordend (van hoog naar laag):

- I Ten minste één RCCT beschikbaar;
- II-A Ten minste een cohort of case-control studie beschikbaar, bij voorkeur van verschillende onderzoeksgroepen/centra;

---

<sup>11</sup> experiment: onderzoeksstrategie waarbij verschijnselen op een voorgeschreven (gestandaardiseerde) manier worden waargenomen onder voorwaarden die door de onderzoeker zijn vastgesteld, en waarbij een of meer variabelen - de onafhankelijke variabele(n) of factor(en) - worden gemanipuleerd om de effecten op de reacties van de proefpersonen na te gaan, de afhankelijke variabelen.

<sup>12</sup> observatie: gestandaardiseerde vorm van waarneming van gebeurtenissen of menselijk gedrag.

<sup>13</sup> geval-studie: synchrone of diachrone analyse van een gebeurtenis, bijvoorbeeld hoofdstuk 5, casus met interpretatie, analyse gericht op interne validiteit.

<sup>14</sup> casuïstiek: beschrijving (reconstructie) van een gebeurtenis, bijvoorbeeld melding bijwerkingen geneesmiddelen of klinische lessen in het Tijdschrift voor Geneeskunde.

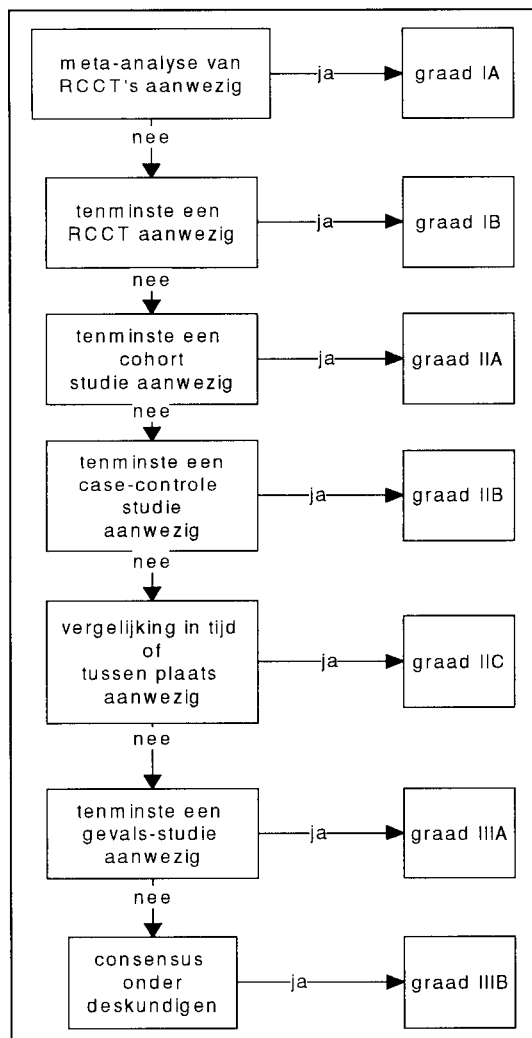
- II-B Vergelijking in tijd en/of plaats met en zonder interventie beschikbaar (historische/regionale verschillen);
- III Meninge van deskundigen met klinische ervaring/autoriteit, beschrijvingen of gevals-studies.

De Canadese indeling is vrij grof. Ze kent slechts drie niveaus en deze niveaus zijn vrij heterogeen van aard. Bovendien ontbreekt de meta-analyse, een vorm van onderzoek die met de grote toename van het aantal onderzoeken meer en meer aan belang wint. Een meta-analyse is een (kwantitatieve) review van meer dan één onderzoek naar eenzelfde onderwerp. De 'National Library of Science' definieert meta-analyse als: '*A quantitative method of combining the results of independent studies (usually drawn from the published literature) and synthesizing summaries and conclusions which may be used to evaluate therapeutic effectiveness, plan new studies, etc., with application chiefly in the areas of research and medicine*' (25). Als er dus meerdere trials met betrekking tot eenzelfde medische interventie voorhanden zijn, kan meta-analyse worden aangewend om tot een schatting van de efficacy te komen (zie bijlage B). Een meta-analyse die een significant effect van een medische interventie aantoont, lijkt een beter bewijs te leveren voor het effect van een medische interventie dan een enkele RCCT of observationeel onderzoek.

Olkin geeft een gedetailleerde hiërarchie van medisch onderzoek, die min of meer correspondeert met de gang van zaken in het wetenschappelijk onderzoek: van aanwijzingen in de vorm van casuïstiek tot experiment (26). Naar gelang de wetenschappelijke bewijskracht worden van laag naar hoog gerangschikt: (i) casuïstiek of het anekdotisch materiaal, (ii) de historische gevalsstudie: series van gevallen, zonder dat men over een controlegroep beschikt, (iii) de analytische gevalsstudie: series van gevallen, afgezet tegen een controlegroep uit de literatuur, (iv) modelbouw en simulatie: analyse uitgaande van een gefingeerd (computer) databestand, (v) observationeel onderzoek (case/controle, cohort), (vi) series van gevallen beschikkende over een historische controle (cross-over design), experiment: (vii) gerandomiseerde klinische trials, uitgevoerd in één onderzoekscentrum, (viii) klinische trials, uitgevoerd in verschillende onderzoekscentra (intrinsieke meta-analyse), (ix) meta-analyse van onafhankelijk van elkaar verrichte klinische trials.

In het *diagram 1* zijn de beide benaderingen samengebracht. Binnen de drie niveaus van de Canadese indeling hebben we een nadere detaillering aangebracht op basis van de indeling van Olkin. Het eerste niveau

*Diagram 1. Wetenschappelijke onderbouwing van medische interventies*



betreft (gegevens uit) het experimenteel onderzoek, het tweede niveau (gegevens uit) het observationeel onderzoek, en het derde niveau de min of meer gesystematiseerde opinies van terzake kundigen. Modelbouw is door ons niet in het diagram opgenomen. Het kan op ieder niveau worden gebruikt en toont vooral de mogelijkheden in termen van verschillende scenario's. Wanneer toepassing van een medische behandeling berust op traditie, intuïtie of individuele opvattingen, is het (vermeende) effect van deze behandeling onvoldoende onderbouwd. Dit neemt uiteraard niet weg dat er geen effect van een dergelijke behandeling aanwezig kan zijn, zoals de geschiedenis van de geneeskunde ons leert.

Het diagram stelt ons in staat de relatie tussen een medische interventie (oorzaak) en het gevolg te kwalificeren op basis van wetenschappelijke onderbouwing. Dit is een belangrijk onderdeel van de analyse van het effect van een medische interventie (zie *diagram 2b*).

### 3.3. De analyse van het effect van medische interventies

De verschillende vormen van wetenschappelijk bewijs moeten een plaats krijgen binnen een analyse die een beleidsmaker, onderzoeker of behandelaar in staat stelt tot een afweging te komen over de relevantie van een medische interventie. De analyse begint met de vraag naar het effect van een medische interventie (X) op een bepaald aspect van de gezondheidstoestand (Y). Vervolgens worden verschillende stappen genomen: definitie van de gehanteerde concepten, verzamelen van gegevens over het effect van een medische interventie, typeren

van de bijeengebrachte gegevens, schatten van de grootte van het effect, en interpretatie. De relevantie van de medische interventie wordt afgelezen aan (i) de grootte van het effect, gekwalificeerd naar efficacy en effectiviteit, alsmede werking en werkzaamheid, en (ii) een typering van de wetenschappelijke onderbouwing van de (causale) relatie tussen een medische interventie en het waargenomen effect.

Het diagram twee geeft het gehele analyseproces weer. Allereerst vindt selectie voor de in aanmerking komende medische interventies plaats (*diagram 2a*). In VTV 1997 geschiedt dit aan de hand van ziekte en aandoeningen. De bijbehorende medische interventies worden opgezocht en zo volledig mogelijk beschreven (27). De medische interventie is te typeren aan de hand van de frequentie, (verwachte) baten, bij- en nevenwerkingen en kosten. Door het kiezen van passende uitkomstmaten is aansluiting op het VTV-model te bewerkstelligen. Een voorbeeld kan dit verduidelijken.

Stel dat men het effect van zidovudine (AZT) bij HIV/AIDS wil analyseren.<sup>15</sup> De behandeling wordt momenteel vaak toegepast bij HIV besmetting. De verwachte baten bestaan uit het verminderen van het optreden van AIDS bij HIV geïnfekteerden of het verbeteren van de overleving bij AIDS. Er zijn aanzienlijke bijwerkingen (28). Het middel is momenteel goed beschikbaar en relatief goedkoop.<sup>16</sup>

Het geneesmiddel AZT wordt in verschillende doseringen voorgeschreven, kent verschillende indicaties en verschillende behandelingsdoelen. Er zal dus moeten worden gedefinieerd wat men nu precies onder 'behandelen met AZT verstaat'. Dit kan door middel van een descriptieve (beschrijvende) definitie ('alle gevallen van AZT gebruik in de beschouwing opnemen'), of door middel van een prescriptieve (conventioneel voorschrijvende) definitie ('beperken tot welomschreven AZT gebruik in bepaalde situatie').

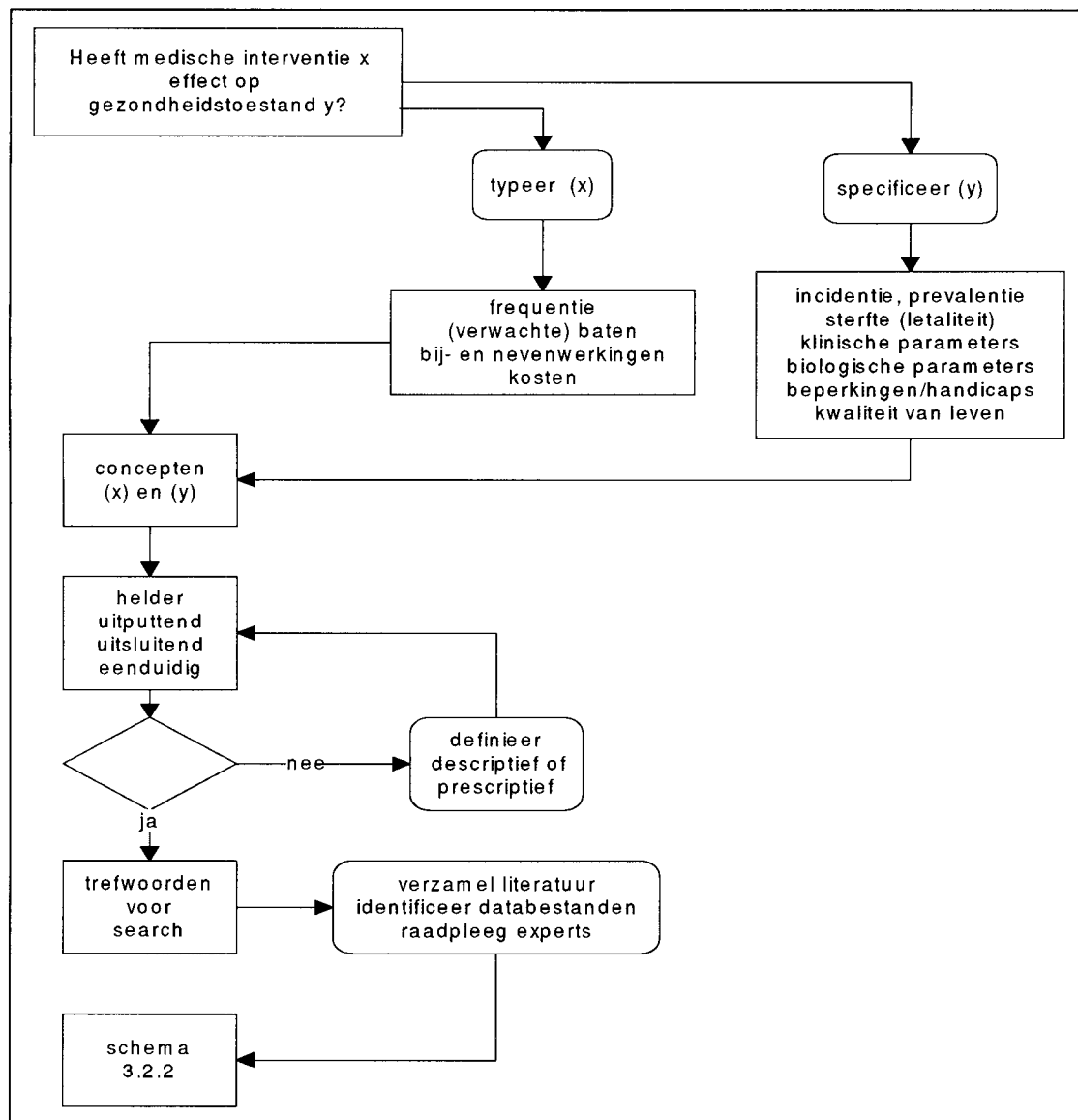
Een descriptieve definitie wordt beoordeeld op overeenstemming van de beschrijving met de werkelijkheid (correspondentie), een prescriptieve definitie op eenduidigheid (coherentie). Tevens moet worden gelet op helderheid van de termen (begrijpelijk, gegeven de doelgroep?), uitputtendheid (sensitiviteit, slaat een term op alle verschijnselen die eronder vallen?), uitsluitendheid (specificiteit,

---

<sup>15</sup> In hoofdstuk 5 wordt dit voorbeeld uitgewerkt om toepassing van de analyse methode te illustreren.

<sup>16</sup> De kosten worden geraamd op \$16,000 per gewonnen levensjaar. Ter vergelijking. Dezelfde berekeningsmethode hanterend, worden de kosten per gewonnen levensjaar van een levertransplantatie geraamd op \$43,000; die van een harttransplantatie op \$26,900; pancreastransplantatie \$16,300; niertransplantatie \$16,300; \$13,500 voor behandeling van hypertensie en \$6,900 voor behandeling met nicotine kauwgum (29).

Diagram 2a. Verzamelen van gegevens over het effect van medische interventies



Legenda

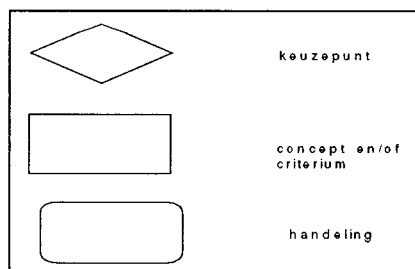
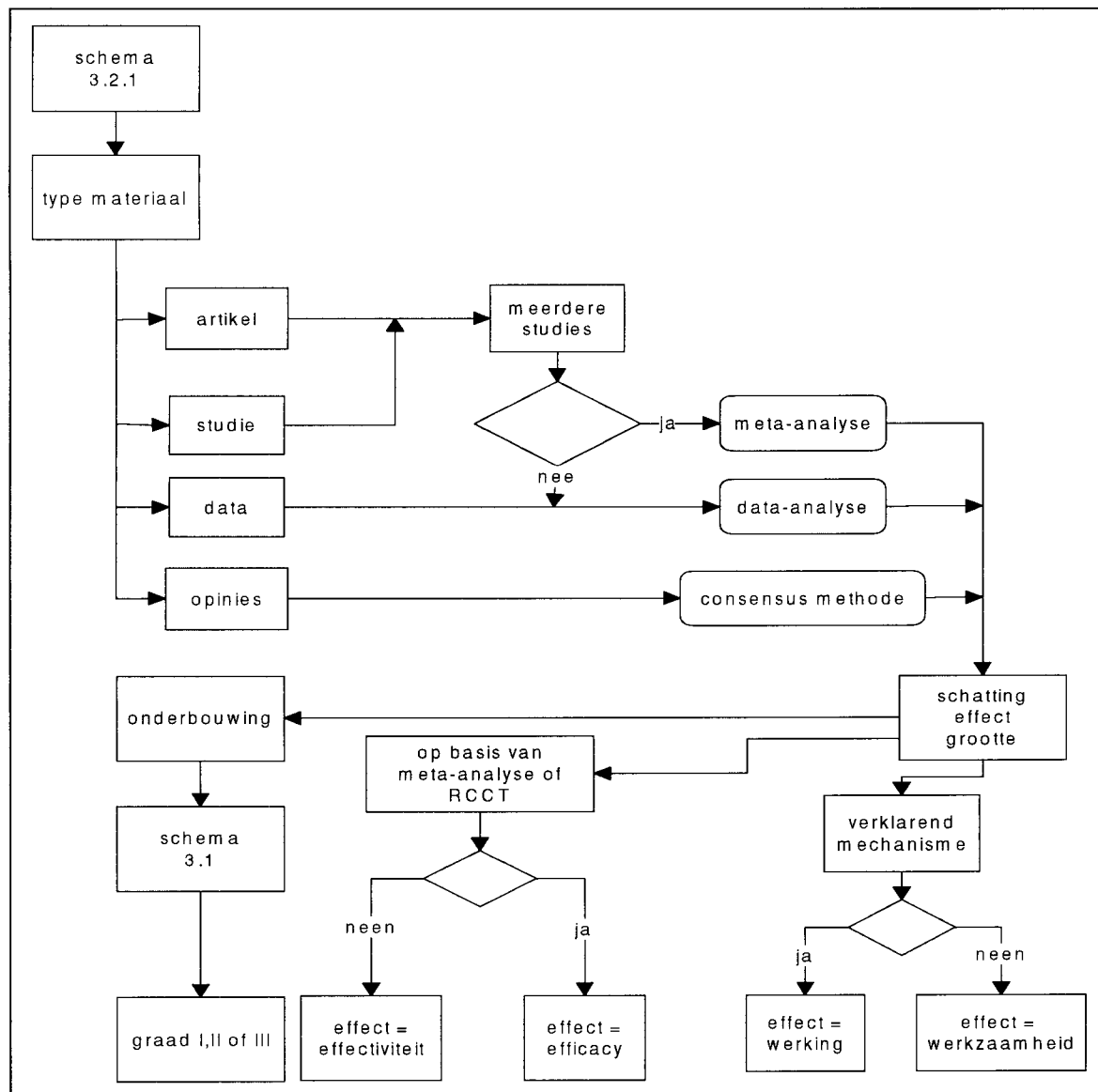
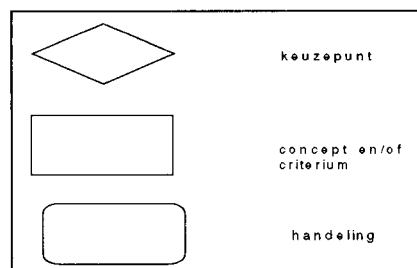


Diagram 2b. Interpreteren van het effect van medische interventies



Legenda



slaat een term niet op verschijnselen die er niet onder vallen?). Zo ook voor de uitkomstmaten die men kiest (Y). Een indicator voor de uitkomst van een medische interventie kan worden gedefinieerd in termen van sterfte, klinische parameters biologische parameters, beperkingen/handicaps, of kwaliteit van leven. Het definiëren gaat door totdat aan de genoemde eisen is voldaan (zie *diagram 2a*).

Bij de gedefinieerde termen worden, indien het literatuurbestand dit vereist, passende zoektermen geformuleerd. Het resultaat is een verzameling zoektermen die het bijeenbrengen van gegevens over het effect van medische interventies aanstuurt. Vervolgens wordt literatuur opgezocht, informatie ingewonnen bij experts op een bepaald gebied, of worden gegevensbronnen geïnventariseerd.

Toepassing van *diagram 2a* levert een heterogene verzameling gegevens op over effecten van medische interventies. In *diagram 2b* worden de verzamelde gegevens geïnterpreteerd. Het kan gaan om het verslag van een onderzoek (een studie/rapport), of artikelen met een uiteenlopende inhoud (verslag van een of meer studies, beschouwing, case-report, meta-analyse, etc.). Daarnaast worden primaire gegevensbronnen (registraties) aangetroffen of zijn opinies van deskundigen ingezameld. Op basis van het soort materiaal wordt actie ondernomen, te weten meta-analyse als er verschillende onderzoeken over hetzelfde onderwerp aanwezig zijn, analyse van primaire of secundaire gegevensbronnen, consensusvorming ingeval van uiteenlopende opinies van deskundigen. Deze verschillende technieken beogen alle informatie over *de grootte* van het effect van een medische interventie op te leveren, zodanig dat de baten van een medische interventie kunnen worden berekend. Lubsen en Tijssen geven hiervoor het volgende model<sup>17</sup>:

$$\text{Net benefit} = \sum v_i(1 - \text{relrisico}_i)\text{risico}_i$$

Het model geeft aan hoe de verwachte reductie in het relatieve risico, dat de wetenschappelijke waarde van een medische interventie aangeeft, kan worden omgerekend naar een reductie in het absolute risico, dat de verwachte opbrengst van een medische interventie voor een bepaalde populatie in de praktijk aangeeft (30).

De interpretatie spitst zich toe op de wetenschappelijke onderbouwing van de (causale) relatie tussen de medische interventie en het waargenomen effect. Hiervoor wordt *diagram 1* gebruikt. Ook

<sup>17</sup> Het relatieve risico wordt afgeleid uit meta-analyse van klinische trials. Het (basis)risico voor een bepaalde populatie op een bepaalde uitkomst uit cohort of case-controlle studies over de betreffende populaties. Het produkt van beide wordt vermenigvuldigd met de waarde (v) die een patiënt toekent aan een bepaalde uitkomst. De sommatie over de (i) mogelijke uitkomsten betreft zowel gewenste als ongewenste effecten van



wordt het effect van de geanalyseerde medische interventie gekwalificeerd in termen van *efficacy* of *effectiviteit*, naar gelang de wijze waarop het is aangetoond (RCCT of uitkomstonderzoek). Als bovendien het verklarende mechanisme waardoor een effect tot stand komt, bekend is, wordt gesproken van *werking*; is dit niet het geval dan wordt gesproken van *werkzaamheid* (zie diagram 2b).

---

een medische interventie (bijwerkingen e.d.), de laatste uiteraard met een negatief teken, en levert uiteindelijk een schatting van de netto opbrengst, als alle mogelijke gevolgen in de berekening zijn opgenomen (30).



## 4 DISCREPANTIE TUSSEN EFFICACY EN EFFECTIVITEIT

### 4.1. Inleiding

Het effect van een medische interventie in de experimentele situatie kan buiten deze situatie niet altijd (volledig) worden gereproduceerd. Er is dan sprake van een discrepantie tussen efficacy en effectiviteit (12) (19) (31) Het opsporen en analyseren van een dergelijke discrepantie is beleidsrelevant. Het geeft een aanwijzing over aangrijpingspunten voor beleid (32). In dit hoofdstuk wordt ingegaan op de oorzaken van een discrepantie tussen efficacy en effectiviteit. In het volgende hoofdstuk komt de analyse ervan aan bod.

### 4.2. Verklaringen en oorzaken

Een discrepantie tussen de efficacy en de effectiviteit van een medische interventie kan ontstaan gedurende het zorgproces. Denk bijvoorbeeld aan het voorschrijven van een geneesmiddel in de alledaagse praktijk. Bij een dergelijke handeling kiest de arts op persoonlijke en medisch wetenschappelijke gronden voor een bepaald middel, passende bij een bepaalde indicatie en een bepaalde patiënt. Hij speelt daarbij in op normen, waarden en verwachtingspatronen van patiënten, maar heeft zelf ook verwachtingen over het verloop van de ziekte bij toediening van een bepaald geneesmiddel. Ook heeft iedere zorgverlener een specifieke manier waarop hij het geneesmiddel aan de patiënt overhandigt, een werkstijl. De een geeft veel informatie. De ander schrijft gehaast een recept uit. De genoemde aspecten van het zorgproces kunnen afwijken van de situatie in de klinische trial en het effect van het geneesmiddel in de praktijk doen verschillen van het effect in de trial.

De verklaring voor een discrepantie tussen efficacy en effectiviteit van een medische interventie kan vanuit een *zorginhoudelijk perspectief* worden gezocht in de ‘factoren’ patiënt, zorgverlener, zorgverlening en de aard van de medische interventie.<sup>1</sup> Tevens spelen de omstandigheden waarin de zorg is verleend een rol (aanwezigheid van apparatuur, voorzieningen, e.d.). Deze kunnen een materiële beperking vormen voor de realisatie van het effect van een medische interventie in de praktijk. Vanuit een *methodologische invalshoek* wordt gelet op de wijze waarop het effect is vastgesteld, ofwel de validiteit van de klinische trial of van het uitkomst-onderzoek. De richting of de mate waarin deze

---

<sup>1</sup> Vanuit een economische invalshoek worden genoemd: externaliteiten, de prijs van een medische interventie, de tijdsdimensie, utilisatie en innovatiekarakteristieken (33). Bespreking van deze aspecten valt buiten de scope van dit rapport (zie *inleiding*).

factoren een eventuele discrepantie beïnvloeden verschilt. In de literatuur worden de volgende oorzaken voor een discrepantie tussen potentieel en gerealiseerd effect aangetroffen.

(i) *meting van het effect*. De validiteit van de gegevens waarop een uitspraak over de efficacy of effectiviteit is gebaseerd, moet in ogenschouw worden genomen. Naast de validiteit en betrouwbaarheid van de trial (*bijlage A*). Daarnaast speelt het type trial een rol. Schwartz en Lellouch onderscheiden verschillende typen trials: *verklarend of pragmatisch* (34). Een trial met een verklarende opzet heeft een theoretische relevantie. Er wordt een bewijs geleverd voor *een causaal verband* tussen een medische interventie en een waargenomen gevolg. Een trial met pragmatische opzet is gericht op een afweging van behandelstrategieën ter ondersteuning van het nemen van een *beslissing in de medische praktijk* (35). Een voorbeeld moge dit onderscheid verduidelijken.

Het vergelijken van bestraling bij kanker met een combinatietherapie bestaande uit bestraling, voorafgegaan door een medicament om de tumor gevoeliger te maken voor de straling. Om het effect van toegevoegde medicament te bepalen kan de opzet van de trial voorschrijven pas met bestraling te beginnen als het medicament enige tijd is toegediend aan de interventiegroep, gelijktijdig met een placebo aan de controlegroep. Een situatie die afwijkt van de praktijk waarin men gestart zou zijn met bestraling op het moment dat de indicatie daarvoor aanwezig is. In een pragmatische trial worden de combinatietherapie en de bestraling op hetzelfde moment (de aanvang van de trial) gestart. Eigenlijk wordt dan een oude handelwijze met een nieuwe vergeleken. Een eventueel verschil in effect kan dan niet zuiver aan de voorbehandeling met het medicament worden toegeschreven, omdat de interventie en controlegroep ook in andere (niet gecontroleerde) factoren, zoals het tijdstip van de bestraling, van elkaar verschillen (34).

Effectiviteit wordt (per definitie) gemeten in de klinische praktijk en zou dus moeten worden vergeleken met het effect in een trial met een pragmatische opzet (35). Naast de typering van Schwartz en Lellouch zijn andere denkbaar. Een verder uitwerking, waarbij het type trial wordt gerelateerd aan de discrepantie tussen efficacy en effectiviteit, lijkt dan ook aangewezen.

(ii) *patiëntenpopulatie*. De patiëntenpopulatie in de trial kan verschillen van de populatie die in de dagelijkse praktijk de medische interventie ondergaat, bijvoorbeeld in leeftijd, geslacht, of sociaal-economische status. Deze laatste is mogelijk minder homogeen, minder ernstig ziek, vertoont een hogere graad van spontane remissie (natuurlijk beloop), en heeft vaker te maken met co-morbiditeit (36).

Ook gedraagt een patiëntenpopulatie in een trial zich anders dan in de klinische praktijk. Er is bijvoorbeeld een verschil in de therapietrouw (compliance). Deze kan in de praktijk slecht zijn, omdat het geneesmiddel slecht smaakt of omdat er te vaak een tablet moet worden ingenomen. In een trial wordt mede dankzij intensieve follow-up en geprotocolleerde uitvoering van de behandeling veelal een grotere compliance dan in de klinische praktijk bereikt (8).

De gevoeligheid van de patiënt voor een therapie is van belang voor het gerealiseerde effect (37). Een verschil in uitgangswaarde maakt het in dezen moeilijker in verschillende situaties eenzelfde effect te realiseren. Additionele therapie speelt hierbij een rol. Deze kan een bepaalde patiënt meer of juist minder gevoelig maken voor een bepaalde medische interventies. Geneesmiddelen interacties vormen hiervan een voorbeeld.

Niet alleen de tastbare kenmerken van de patiënt veroorzaken een discrepantie tussen efficacy en effectiviteit. Ook de normen en waarden van patiënten vertonen grote verschillen. In de praktijk worden therapeutische beslissingen genomen naar aanleiding van de kans op gewenst resultaat en de waarde die een patiënt aan een dergelijk resultaat toekent (21). Er bestaat de mogelijkheid dat deze waarden in de praktijk verschillen van de uitgangspunten die (mogelijk op basis van veronderstellingen) in de klinische trial zijn gehanteerd. Dit verschil in waarden zal doorwerken in de gevolgde behandelstrategie en het uiteindelijke effect van de medische interventie (38).

(iii) *zorgverlener*. De zorgverlener wordt gekenmerkt door gespecialiseerde kennis, verworven door langdurige opleiding en oefening, een grote autonomie in de uitoefening van zijn beroep, een neutrale (dienstverlenende) houding ten opzichte van patiënten, en een relatie met de patiënt gebaseerd op een wederzijds vertrouwen. (32). Deze kenmerken worden gevat onder de noemer competentie en/of werkstijl van de zorgverlener (36). Verschillen in competentie en werkstijl kunnen een discrepantie tussen efficacy en effectiviteit in de hand werken. Deelname aan een trial betekent vaak behandeling door een (hoog) gespecialiseerde zorgverlener in een academisch centrum. Deze bezit meer kennis en ervaring met betrekking tot het onderhavige ziektebeeld dan de zorgverlener die uiteindelijk de behandeling in de alledaagse klinische praktijk gaat toepassen. Bovendien vindt toepassing van de medische interventie in de klinische praktijk plaats in het kader van een langdurige arts-patiënt relatie waarvan ook andere overwegingen dan de rationeel vastgestelde indicatiestelling deel uitmaken (bijvoorbeeld, 'therapeutische diagnostiek' ter geruststelling, of voorkomen van somatische fixatie, etc.). Ook het simultaan verzamelen van gegevens en leveren van de behandeling verstoort in de trial de gebruikelijk arts-patiënt situatie en kan uitval of selectieve perceptie in de hand werken (36).

In de praktijk komt de zorgverlener vaak op ideeën met betrekking tot nieuwe toepassingsmogelijkheden van een medische technologie, hetgeen leidt tot ontdekkingen, zoals het voorbeeld van Paré toonde (*zie hoofdstuk 2*), maar ook tot nodeloze verbreding van de indicatiestelling. Ook problemen met de accuratesse van de diagnostiek en de vaardigheid waarmee een bepaalde medische technologie wordt aangewend, kunnen de discrepantie tussen efficacy en effectiviteit beïnvloeden (23).

(iv) *zorgverlening*. De zorgverlening is te beschouwen als een proces van diagnostiek, indicatiestelling, therapie en evaluatie (follow-up) naar aanleiding van een bepaald medisch probleem (21). Ook de verschillende aspecten van de zorgverlening vormen aangrijpingspunt voor het ontstaan van een discrepantie tussen efficacy en effectiviteit.

In de praktijk kunnen er bijvoorbeeld minder (uitgebreide) diagnostische mogelijkheden zijn, waardoor de medische interventie op een populatie met (iets) andere samenstelling wordt toegepast. Het niet op het juiste tijdstip toegepaste of inadequaat niveau van zorg worden veroorzaakt door verschillen in indicatiestelling (32). In een trial wordt volgens een protocol gewerkt en de medische interventie uitgevoerd door speciaal daartoe geïnstrueerde clinici. Dit draagt bij aan een afwijkende situatie waarin de therapie wordt uitgevoerd (36). Ook wanneer een trial niet wordt uitgevoerd op de gebruikelijke plaats van behandeling, bijvoorbeeld in het academisch ziekenhuis in plaats van in de huisartspraktijk, vallen non-specifieke kenmerken van de behandeling in de laatst genoemde situatie weg.

In de dagelijkse praktijk worden patiënten vaak niet lang genoeg gevolgd om het effect van een medische interventie betrouwbaar vast te stellen. Daarvoor zijn twee verschillende oorzaken. Patiënten stoppen met een behandeling als deze niet snel genoeg werkt, of onttrekken zich aan follow-up als de behandeling juist snel effect heeft gehad (36). Patiënten in een trial zijn doorgaans gemotiveerd om de voorgeschreven behandeling nauwkeurig op te volgen, inclusief de daarbij behorende controles in het kader van follow-up. Bovendien is de omvang en de reden van de uitval in een trial vaak bekend.

Een gecompliceerde behandeling zal (waarschijnlijk) een beter effect opleveren wanneer uitgevoerd onder nauwkeurig geprotocolleerde omstandigheden. Wanneer een trial geen voordeel laat zien van een bepaald, eenvoudig toe te dienen medicament, is niet uitgesloten dat in de klinische praktijk een dergelijk middel toch beter werkzaam is dan meer gecompliceerde behandelvormen, omdat het gerealiseerde effect minder afhankelijk is van vaardigheden van de zorgverlener of compliance van de patiënt (36).

(v) *Omstandigheden waaronder de zorgverlening plaatsvindt*. Te denken valt aan de toegang van de zorg, verzekeringsvoorwaarden en financiering van een bepaalde behandeling, beschikbaarheid van voorzieningen (apparatuur, gebouwen en technieken/specialisten). Het al of niet aanwezig zijn van deze factoren beïnvloedt de verdeling van patiënten met verschillende ziektestadia over de aangeboden medische technologie (33). Bijvoorbeeld, screening op cervixcarcinoom. Personen met een cervixcarcinoom worden actief opgespoord en vervolgens behandeld. Hiertoe is een opsporing met ondersteunende administratie nodig. De efficacy van een screeningsprogramma is hoog, indien toegepast op patiënten met een vroeg stadium van kanker. Hysterectomie bij een stadium I cervixcarcinoom is levensreddend. Dezelfde ingreep bij een stadium IV cervixcarcinoom is slechts palliatief. Introductie in de populatie brengt echter (ook) toepassing op patiënten in verder gevorderde stadia met zich mee. Deelname aan het programma wordt hun om sociaal-ethische redenen niet

geweigerd. De technologie presteert daardoor minder goed dan wordt verwacht. Wel is te verwachten dat het gerealiseerde effect in de loop van de tijd meer in de richting van de efficacy zal naderen.

Ook is er een relatie tussen de uitkomst en de omvang van een gezondheidszorgprogramma beschreven (volume-uitkomst relatie), met name voor chirurgische interventies. Deze is mogelijk te wijten aan: meer behandelen geeft betere uitkomsten door meer oefening en grotere vaardigheid, of beter behandelen geeft een groter aanbod van patiënten en daardoor een grote volume van behandelingen (39) (40). In ieder geval kan een verschil in omvang van de zorgverlening een discrepantie tussen efficacy en effectiviteit in de hand werken.

Bij preventie wordt een discrepantie tussen efficacy en effectiviteit bovenal bepaald door de bereikbaarheid van de doelgroep, deelname aan het preventieprogramma en uitvoering van de preventieve maatregel (41).

#### 4.3. Onderzoek naar oorzaken

De oorzaken van een discrepantie tussen efficacy en effectiviteit zijn in *tabel 3* weergegeven. Zij zijn geordend naar de vijf aspecten van de definities van efficacy en effectiviteit (*zie hoofdstuk 2*). In *hoofdstuk 5* fungeren zij als onafhankelijk variabelen in een onderzoek, waarin wordt getracht een (eventuele) discrepantie tussen efficacy en effectiviteit nader te analyseren. We vergelijken efficacy en effectiviteit *niet* met betrekking tot de onderbouwing van een oorzaak-gevolg relatie. Dan zouden we immers appels met peren vergelijken. De efficacy wordt gemeten in de klinische trial en de onderbouwing ervan bevindt zich dus (per definitie) op niveau I. De effectiviteit wordt (per definitie) gemeten in observationeel onderzoek en de onderbouwing ervan bevindt zich op niveau II of III (*zie diagram 1*). Een vergelijking tussen beide vindt plaats op de gemeenschappelijke noemer, *de grootte van het effect*. We vergelijken de grootte van het effect in de experimentele situatie met de grootte van het effect in de alledaagse situatie, en we trachten een eventueel verschil tussen beide te verklaren vanuit een verschil in situatie.

Tabel 3. Oorzaken van een discrepantie tussen efficacy en effectiviteit

**1. Kans op gunstige uitkomst/verschil in effectmeting:**

in trial (type: verklarend, pragmatisch)  
 of registratie  
 longitudinaal of transversaal  
 persoonsgebonden of groepsgewijs  
 geselecteerde proefpersonen of  
 aselechte steekproef  
 ziekte(stadia)gericht (naar diagnose)  
 interventie gericht (naar verrichting)  
 zorggericht (naar:  
 input/output/outcome)  
 verschil in tijdstip

**2. Patiëntenpopulatie****verschil in:**

sociaal/demografische kenmerken  
 symptomatologie  
 (ernst) ziektestadium  
 co-morbiditeit  
 ziekte duur  
 gevoeligheid voor therapie  
 compliance/therapietrouw  
 termijn van observatie  
 waarden en normen  
 (zorg)consumptiegedrag  
 (risico)gedrag van de  
 patiëntenpopulatie  
 relatie tot tijd (cohort effect)

**3. Kenmerken van de medische interventie****verschil in: (a) aard**

soort (chirurgie, geneesmiddel, etc.)  
 additionele therapie  
 follow-up (aard, intensiteit, duur)  
 gecompliceerdheid van behandeling  
 (eenvoudig/moeilijk)  
 geprotocolleerde uitvoering van de  
 handeling  
 medical audit (niet) aanwezig  
 omvang/volume van zorg  
 simultaan verzamelen van gegevens en  
 uitvoering  
 uitval of selectie

**verschil in (b) uitvoering**

kennis, opleiding en ervaring van  
 zorgverlener  
 vaardigheid van zorgverlener  
 competentie  
 accuratesse van de diagnostiek  
 voorschrijfgedrag  
 professionele prestatie  
 werkstijl

autonomie/werkverband  
 houding ten opzichte van patiënt

**4. Indicatiestelling****Verschil in:**

(juiste) tijdstip/intensiteit  
 diagnose (stadium)  
 ernst, duur, ongerustheid, hinder van  
 klacht/symptoom  
 als prognostische factor gehanteerde  
 klinische parameter

**5. Omstandigheden****Verschil in:**

plaat van behandeling (ziekenhuis,  
 verpleeghuis etc.)  
 toegang  
 verzekeringsvoorwaarden/financiering  
 beschikbaarheid/bereikbaarheid:  
 apparatuur, gebouwen,  
 technieken/specialisten  
 behandelcapaciteit/  
 benutting behandelcapaciteit



## 5 HET EFFECT VAN ZIDOVUDINE; EEN CASE-STUDIE

### 5.1 Inleiding

In dit hoofdstuk wordt het conceptuele kader toegepast op de bestudering van het effect van zidovudine (AZT). AZT is momenteel een van de gangbare medische interventies bij een infectie met het humane immunodeficiëntie virus (HIV). Het HIV is een retrovirus dat bij de mens een vermindering van de immunologische afweer veroorzaakt, en na tien à twaalf jaar aanleiding geeft tot een fulminant ziektebeeld, het Acquired Immunodeficiency Syndrome (AIDS), met een dodelijke afloop binnen één à twee jaar (42). Remming van het ziekteproces en verbetering van de kwaliteit van leven zijn vooralsnog de belangrijkste doelen van medische interventies bij deze ziekte (*zie tabel 4*). Een middel dat het virus (permanent) onschadelijk maakt is nog niet gevonden.

Het doel van dit hoofdstuk is illustratie van (a) de analyse van de grootte van het effect van een medische interventie in termen van efficacy en effectiviteit en (b) de toeschrijving van een verschil in effectgrootte aan factoren die een discrepantie tussen efficacy en effectiviteit kunnen veroorzaken. Op deze manier kunnen we de bruikbaarheid van het conceptuele kader voor de analyse van effecten van medische zorg toetsen en de relevantie ervan voor het gezondheidszorgbeleid aangeven.

Er is gekozen voor een bestudering van het effect van het geneesmiddel AZT bij een HIV infectie. Het geringe aantal behandelingsmogelijkheden dat er vooralsnog voor deze ziekte aanwezig is, maakt de analyse minder ingewikkeld. Therapeutische interacties spelen geen (grote) rol (*zie inleiding*). Tevens is er de directe beschikbaarheid van een klinisch databestand, waarin de effectiviteit kan worden bepaald. Dit maakt het mogelijk een (lokaal) gerealiseerde effect te vergelijken met het in de literatuur gerapporteerde effect. Daarnaast is HIV infectie en/of AIDS een voorbeeld van een ongeneeslijke, langdurige ziekte met een blijvend zorgbeslag, hetgeen het mogelijk maakt de specifieke problemen die het bepalen van een effect van een medische interventie bij chronische ziekten met zich meebrengt, te illustreren.

Tabel 4. Medische interventies bij een HIV infectie (43)

CD4 telling <sup>2</sup>	toegenomen risico op	medische interventie
>500	niet-specifieke ziekte/aandoening	routine controle, vaccinatie (algemeen), voorlichting op gebied van veilige sex, of drugsgebruik (sputomruil)
<500	tuberculose, non-Hodgkin's lymfoom, bacteriële infecties	antiretrovirale therapie (w.o. AZT)
<200	<i>Pneumocystis Carinii</i> pneumonie (PCP), toxoplasmose	PCP profylaxe (Bactrimel), toxoplasmose profylaxe
<100	Cryptococcosis, histoplasmosis, Mycobacterium Avium complex (MAc)	MAc profylaxe
<50	Cytomegalovirus retinitis, non-Hodgkin lymfoom in CZS	profylaxe, voeding, aandacht voor kwaliteit van leven

## 5.2 Gegevens over het effect van zidovudine

Het stramien van *diagram 2a/b* is gebruikt om gegevens over het effect van AZT te verzamelen en te interpreteren. Allereerst is een definiëring van de medische interventie en de uitkomstmaten vereist. Behandeling met AZT is een relatief welomschreven entiteit. De omschrijving van het middel is eenduidig en de indicatiestelling helder (44). Het aantal CD4 cellen in het perifere bloed wordt gebruikt als leidraad voor het voorschrijven van AZT. Over de hoogte van het aantal CD4 cellen waarbij AZT wordt voorgeschreven, bestaat (internationaal) geen overeenstemming. In de Verenigde Staten wordt bij minder dan 500 CD4 cellen per mm<sup>3</sup> AZT voorgeschreven (45). In Nederland bestaat consensus over het hanteren van een grens van 300 cellen mm<sup>3</sup> (46). Het medisch beleid is in Nederland dus iets terughoudender dan in de Verenigde Staten.

De keuze van een passende uitkomstmaat is gerelateerd aan het ziektestadium. De HIV infectie kent vier stadia. Het eerste stadium van de HIV infectie (CDC<sup>3</sup> I) betreft de acute ontstekingsreactie volgend op de besmetting met het virus. Deze is kortdurend (één tot drie weken) en wordt gekenmerkt door griep-achtige symptomen. In het tweede stadium (CDC II) is de geïnfecteerde persoon asymptomatisch. In het derde stadium (CDC III) van de infectie heeft de persoon last van pijnlijke,

<sup>2</sup> Het aantal T-lymphocyten met het CD4 oppervlakte eiwit (CD4-cellen) dat bij telling in een bloedmonster wordt aangetroffen, geeft aan wanneer een bepaalde medische interventie moet worden toegepast (volgens Amerikaanse indicatiestelling).

<sup>3</sup> CDC: Centers for Disease Control

gezwollen lymfeklieren. In het vierde stadium (CDC IV) wordt gesproken van AIDS. Er treedt algehele uitputting, neurologische aandoeningen, kanker of ernstige (opportunistische) infecties op (47). In de literatuur wordt onderscheid gemaakt tussen (a) het effect van AZT bij een asymptomatische HIV infectie<sup>4</sup> en (b) het effect van AZT bij personen met verschijnselen van AIDS. Beide indicaties vereisen een andere uitkomstmaat, respectievelijk, AIDS en sterfte.

De definitie van AIDS is in de loop van de tijd veranderd. Aanvankelijk werd een AIDS-related complex beschreven (ARC)<sup>5</sup> als een stadium dat voorafgaat aan AIDS. In de huidige CDC classificatie kunnen personen met ARC vallen onder CDC III of CDC IV-a. Een gedeelte van de personen met ARC zou dus momenteel als personen met AIDS worden aangemerkt en bij onderzoek naar het effect van AZT op de progressie naar AIDS niet meer voor de interventie(groep) in aanmerking komen. Sterfte is een relatief betrouwbare uitkomstmaat. Het overlijden van een persoon kan vrij eenduidig aan de HIV infectie worden toegeschreven.

Literatuur is opgezocht met behulp van het MEDLINE bestand. Daarbij is een combinatie van de trefwoorden 'AZT' met 'HIV', 'AIDS', en 'Randomized Trial' gehanteerd. Onder invloed van de Cochrane collaboration zijn klinische trials in MEDLINE van een geïndexeerd trefwoord voorzien. We mogen er dus van uitgaan dat alle gepubliceerde trials zijn gevonden. Ook wordt cohort onderzoek aangetroffen. We beschikken niet over rapportage van studies, opinies van deskundigen, of (gepubliceerde) meta-analyses (zie *diagram 2*).

De bijengebrachte gegevens zijn in *tabel 5* weergegeven, gerangschikt naar de vijf factoren die een discrepantie tussen efficacy en effectiviteit kunnen veroorzaken (zie *tabel 3*).<sup>6</sup> De aangetroffen klinische trials en cohort-onderzoekingen betreffen (a) het effect van AZT op *de progressie* van asymptomatische HIV infectie naar AIDS (grijstint in de tabel) en (b) *de sterfte* onder personen met AIDS. Het betreft gepubliceerd onderzoek. Een onderzoek is benoemd naar de eerste auteur van de publikatie en het jaartal waarin publikatie plaatsvond. Ten eerste wordt de belangrijkste uitkomstmaat vermeld met de gevonden waarde in de interventie- en (indien aanwezig) in de controlegroep. Meerdere uitkomsten per interventiegroep duiden op afzonderlijke strata in de interventiegroep. Het (overall) effect met het bijbehorende 95% betrouwbaarheidsinterval is berekend per 100 persoonsjaren op basis

---

<sup>4</sup> Meestal worden personen met CDC III in trials met betrekking tot progressie van asymptomatische HIV infectie naar AIDS opgenomen. Een persoon kan dit stadium overslaan en meteen vanuit CDC II naar CDC IV overgaan.

<sup>5</sup> AIDS-Related Complex: patiënten met koorts, gewichtsverlies, diarree, nachtzweeten, lymfadenopathie, en immunologische stoornissen (48). Momenteel wordt CDC IV-c2/e niet als AIDS beschouwd.

<sup>6</sup> Trials met betrekking tot HIV infectie bij hemophilie patiënten zijn buiten beschouwing gelaten.

van de gerapporteerde gemiddelde observatieduur (gem. od., *tabel 5*.<sup>7</sup> Hierdoor is de effectgrootte van de verschillende onderzoeken te vergelijken.

Vervolgens zijn kenmerken van de gehele onderzoekspopulatie (pat. pop., *tabel 5* vermeld. De vermelde therapietrouw (non-compliance) betreft de interventiegroep bij een klinische trial. Als het gaat om cohort onderzoek is er geselecteerd op AZT gebruik. Het ziektestadium geeft de uitgangspositie van de onderzoekspopulatie weer bij aanvang van de behandeling met AZT.

De medische interventie (med. iv., *tabel 5*) is gekenmerkt door de dosis AZT en de gemiddelde observatieduur van de personen die aan het onderzoek deelnemen. Verschillende doseringen AZT corresponderen met verschillende strata in de interventiegroep of het cohort, dan wel met een verschillend behandschema. De aanduiding 'vroeg versus laat' betreft een onderzoeksopzet waarbij personen met een asymptomatische HIV infectie onmiddellijk AZT toegediend krijgen bij aanvang van de observatieperiode (vroeg behandeling) en de controlegroep pas AZT krijgt wanneer verschijnselen van AIDS optreden (late behandeling). Tot het moment van het optreden van AIDS gebruikt de controlegroep een placebo.

#### *Legenda bij tabel 5*

R	= randomized,
DB	= double blind,
PC	= placebo controlled,
OL	= open label,
BI	= 95%-betrouwbaarheidsinterval
pat.pop	= patiëntenpopulatie
gem. od.	= gemiddelde observatie duur (in maanden).
med.iv	= medische interventie
ind.	= indicatie
omst.	= omstandigheden
-	= gegeven niet vermeld in publikatie

<sup>7</sup> De betrouwbaarheidsintervallen zijn berekend conform *bijlage A, pagina 75-76*.

Tabel 5. Het effect van AZT behandeling op progressie naar AIDS en sterfte

I effect	Fischl et al (28) 1987	Fischl et al (55) 1989	Fischl et al (50) 1990	Volberding et al (17) 1990	Fischl et al (53) 1990	Swanson, Cooper (56) 1990	Hamilton et al (45) 1992	AZ Utrecht (60)	Buira et al (58) 1992	Vella et al (57) 1992	Graham et al (54) 1992	Concorde trial (52) 1993	Concorde trial (52) 1993	Cooper et al (51) 1993
uitkomst maat	sterfte	sterfte	AIDS	AIDS/ARC	sterfte	sterfte	AIDS	AIDS (CDC IV) sterfte	sterfte	sterfte	sterfte	AIDS/ARC	AIDS/ARC	CDC IV
interventie groep	1/145	83/229	13/360	17/453 (a) 19/457 (b)	188/262 169/262	77/308	28/170	23/50 35/65	98/195	113/159	306/2.162	132/507	34/358	52/495
controle groep	19/137	geen	36/351	38/428	geen	historisch	48/168	geen	historisch	historisch	geen	154/508	50/352	94/489
per 100 ppj (95% CI)	2,1 (0,2-11,6)	28,2 (22,4-34,0)	3,9 (2,1-6,8)	3,7 (2,5-5,1)	31,4 (29,6-33,2)	16,7 (14,6-18,8)	7,3 (4,9-10,6)	37,6 (23,8-56,2) 44,1 (30,7-61,6)	26,2 (23,0-29,4)	38,8 (35,0-42,6)	7,1 (6,5-7,7)	8,7 (4,8-12,6)	3,2 (2,2-4,4)	5,5 (4,4-6,6)
II pat pop														
aantal	282	229	711	1.338	524	308	338	50/65	195	159	2.162	1.749	1.749	993
%mannen	95,4	-	94,5	92	88	100	99	93	87	86	100	85	85	88
%homosex	-	-	69	70	-	-	63	68	34,5	18	n.v.	-	-	68
non compli	4/152	-	23%	16/21%	32%	-	10%	-	-	-	-	-	-	10%
ziekte stadium	AIDS (PCP) ARC	AIDS (PCP) ARC	ARC	CDC/III	AIDS (PCP)	AIDS	ARC 200<CD4 <500	CDC/II/III AIDS	AIDS	AIDS	CDC II/III	asymptomatische HIV infectie	asymptomatische HIV infectie	CDC II/III
III med IV AZT dosis mg/dag	1.500 v placebo	1.500/1.200 (vroeg/laat)	1.200 v placebo	500 (a) 1.500 (b) v placebo	1.500 v 1.200 / 600	1.200	1.500 vroeg v late beh	1.000	750-1.200	1.000	wel/niet	1.000 vroeg v late beh.	1.000 vroeg v late beh.	1.000
gem OD. IV ind. aant. CD4 cellen	4	15	11	13	26	11	27	15/15	23	22	24	36	36	23
V omst. land	USA	USA	USA	USA	USA	Australië	USA	NL	Spanje	Italië	USA	UK, Ir, Fr	UK, Ir, Fr	Australië Europa 1989-1991
jaar van entree	1986	1986-1988	1987-1989	1987-1989	1987-1990	1987-1988	1988-1989	1987-1992	1981-1987	1987-1990	1986-1987	1988-1991	1988-1991	1989-1991
soort onderzoek.	R,DB,PC theoretisch analyse	cohort; follow-up uit 1986	R,DB,PC	R,DB,PC	R,OL	cohort	R,DB,PC switch to OL	cohort	cohort	cohort	cohort	R,DB,PC	R,DB,PC	R,DB,PC C, switch to OL
soort effect	efficacy	effectiviteit	efficacy	efficacy	effectiviteit	effectiviteit	efficacy	lokaal effect	effectiviteit	effectiviteit	effectiviteit	efficacy	efficacy	efficacy

De indicatiestelling (ind., tabel 5) is het aantal CD4 cellen dat bij telling in het perifere bloed wordt aangetroffen. In cohortstudies wordt het aantal CD4 cellen vaak niet als indicatie, maar voor een stratificatie bij de analyse gebruikt. Doorgaans bevat een publikatie veel gegevens over laboratorium- en klinische waarden die de patiënt bij indicatiestelling kenmerken. Dit stelt de lezer van het artikel in staat zich een voorstelling te maken van de patiënten die voor behandeling in aanmerking komen.

De omstandigheden (omst., tabel 5) waaronder het onderzoek plaatsvond, worden gekenmerkt door het land, het jaar waarin de onderzoekspopulatie is geworven, en het soort onderzoek. Het land indiceert een bepaalde cultuur en de daarmee samenhangende attitude ten opzichte van de behandeling. Het jaar waarin de onderzoekspopulatie is geworven kan van belang zijn in verband met de gedragsveranderingen die zich in betrekkelijk korte tijd onder risicogroepen hebben voorgedaan. Daardoor verschilt een populatie binnen een paar jaar naar aard en samenstelling, hetgeen gevolgen kan hebben voor het waargenomen effect van AZT.

Naast literatuur is er een bestand beschikbaar met gegevens van 285 personen met HIV/AIDS die tussen 1 januari 1987 en 31 december 1992 in het Academisch ziekenhuis Utrecht (AZU) zijn behandeld. In deze persoonsgebonden, longitudinale registratie bevonden zich 115 AZT-gebruikers. Dit maakt het mogelijk een schets van het effect van AZT in de tijd te geven en een vergelijking te maken tussen verschillende subgroepen in de tijd.<sup>8 9</sup> Het gegevensbestand van het AZU kan worden gebruikt om de vergelijking van de effectiviteit met de efficacy te illustreren.

Op basis van het verzamelde materiaal kunnen we nu de effectgrootte schatten en kwalificeren als efficacy of effectiviteit. Allereerst is echter een specificering van de onderbouwing van het effect van AZT nodig (zie diagram 2).

---

<sup>8</sup> Voor de univariate analyse van het effect van AZT op de progressie van HIV naar AIDS en het overlijden van AIDS patiënten is gebruikt gemaakt van de methode van Kaplan en Meier. Verschillen tussen subgroepen werden onderzocht met de (non-parametrische) log-rangorde toets van Wilcoxon. Het multipele regressie model van Cox is gebruikt om factoren te vinden die de onafhankelijke variabelen, progressie naar AIDS en sterfte, beïnvloeden. De analyse is stapsgewijs opgebouwd, met als drempel  $p < 0,05$  voor introductie van een variabele in het model. De statistische analyses zijn uitgevoerd met het Statistical Analysis System (SAS), versie 6.01.

<sup>9</sup> Aangezien AZT niet geneest, maar het ziekteproces slechts afremt, is de hoogte van het relatief risico afhankelijk van de observatieduur. Dit illustreert een belangrijk aspect van het effect van een medische interventie bij chronische ziekten. De tijd speelt een rol. Het stelt eisen aan de meting van de effectiviteit, wanneer men deze wil vergelijken met de efficacy. De meting moet longitudinaal van aard zijn, zodat het effect in relatie tot de tijd bekend is. Dit stelt weer eisen aan de gegevensverzameling. Het tijdstip van aanvang van de therapie moet nauwkeurig worden geregistreerd, zodat de therapeutische episode is vast te stellen.

### 5.3 De wetenschappelijke onderbouwing van zidovudine

Een verklaring voor het effect van AZT wordt beschreven met behulp van de termen 'werking' en 'werkzaamheid' (zie *diagram 2*). Uit laboratoriumonderzoek blijkt dat AZT het enzym 'reverse transcriptase' inhibeert. Dit enzym heeft het HIV nodig om zich in menselijke cellen met het CD4 oppervlakte-eiwit (T lymphocyten, macrophagen en zenuwcellen) te vermenigvuldigen. Op deze wijze remt AZT de replicatie van het virus in de menselijke cel, zonder deze daarbij schade toe te brengen. De *werking* van AZT is derhalve bewezen (49).

Ook toepassing van het middel op personen met AIDS heeft effect (29). Uit de trial van Fischl et al. blijkt dat de immuunfunctie verbetert. Er vindt een (initiële) stijging van het aantal T-lymphocyten na AZT toediening plaats en een vermindering van de HIV-antigeen concentratie in het serum. Ook is er sprake een vermindering van aantal en ernst van opportunistische infecties, het optreden van Kaposi sarcoma en stabilisering van lichaamsgewicht (28). Het bewijs voor de *werkzaamheid* van AZT is op basis van deze trial te kwalificeren als graad IB. Er is tenminste één gerandomiseerd, dubbelblind, placebo gecontroleerd onderzoek aanwezig dat het effect van AZT op sterfte bij personen met AIDS onderbouwt (zie *diagram 1*). De trial had echter een beperkte duur (gemiddelde observatieduur 4 maanden) en kende een vrij hoge uitval onder de deelnemende personen. De trial van Fischl et al. in 1986 is het enige gerandomiseerde, placebo gecontroleerde onderzoek dat een indruk van het (netto) effect van AZT op sterfte kan geven.

Mede op basis van het onderzoek van Fischl et al. werd AZT voorgeschreven aan personen met AIDS. Nader onderzoek diende uit te wijzen of AZT ook in een eerder stadium van de HIV infectie effect ressorteerde. In 1990 wezen de onderzoeken van Volberding et al. en Fischl et al. op een effect van AZT bij een asymptomatische HIV-infectie (17) (50). Het gebruik van AZT bleek het optreden van AIDS uit te stellen ofwel de AIDS incidentie (per tijdseenheid) te verminderen. Ook hier gaat het om een bewijs voor de *werkzaamheid* van graad IB. Er is tenminste één gerandomiseerd, dubbelblind, placebo gecontroleerd onderzoek aanwezig dat het effect van AZT onderbouwt (zie *diagram 1*). Het voorschrijven van AZT bij HIV of AIDS is dus wetenschappelijk onderbouwd. Derhalve is er sprake van een potentieel effect, een effect onder ideale omstandigheden, dat kan worden vergeleken met het gerealiseerde effect onder lokale omstandigheden.

## 5.4 Efficacy en effectiviteit van zidovudine

### 5.4.1. Het effect op de progressie naar AIDS

Gegevens over de *efficacy* van AZT worden ontleend aan een aantal klinische trials. Fischl et al. vergeleken het effect van 1.200 mg AZT per dag met dat van een placebo bij personen met een asymptomatische HIV infectie of ARC en een CD4 telling van 200 tot 800 cellen per mm<sup>3</sup> (50). Volberding et al. vergeleken een hoge dosis (1.500 mg per dag) en een lage dosis (500 mg per dag) AZT met een placebo bij personen met een asymptomatische HIV infectie en een CD4 telling van minder dan 500 cellen per mm<sup>3</sup> (17). Er is een significant lagere AIDS incidentie onder AZT gebruikers ten opzichte van de controle groep, met name onder de personen met een CD4 telling van minder dan 500 cellen per mm<sup>3</sup> (17) (50). Beide trials hebben een gemiddelde observatieduur van ongeveer een jaar en waren niet opgezet om het effect van AZT op overleving op lange termijn te onderzoeken (6).

Volberding et al. nemen een AIDS incidentie van 3,7 per 100 persoonsjaren (95% BI: 2,5-5,1) waar in de interventiegroep en van 8,2 per 100 persoonsjaren (95% BI: 5,8-11,3) in de controlegroep (17). Fischl et al. vinden een AIDS incidentie van 3,9 per 100 persoonsjaren (95% BI: 2,1-6,8) in de interventiegroep en van 11,2 per 100 persoonsjaren (95% BI: 7,8-15,5) in de controlegroep (50). Deze waarden voor de effectgrootte ontlopen elkaar weinig, ondanks dat de onderzoekspopulaties niet dezelfde samenstelling hebben. Het onderzoek van Fischl et al. betrof personen met AIDS-related complex, een toestand die momenteel gedeeltelijk tot het AIDS stadium (CDC IV) wordt gerekend, en in de studie van Volberding et al. (mogelijk) als uitkomst fungeerde.

Voor personen met een CD4 telling van meer dan 500 cellen per mm<sup>3</sup> is het effect van AZT op de progressie naar AIDS aanvankelijk minder duidelijk aangetoond (50). Het onderzoek van Cooper et al. bij personen met een asymptomatische HIV infectie en een CD4 telling *boven* de 400 cellen per mm<sup>3</sup> brengt hier verandering in (51). In de interventiegroep wordt een significant lagere progressie naar AIDS aangetroffen dan in de controlegroep. Het onderzoek is echter niet dubbelblind uitgevoerd. De participanten waren op de hoogte van de aard van het middel dat zij gebruikten. Een invloed hiervan op de therapietrouw is dan ook niet uitgesloten (*zie bijlage A*). In de interventiegroep wordt een AIDS incidentie van 5,5 per 100 persoonsjaren waargenomen (95% BI 4,4-6,6). Deze waarde verschilt niet significant met de door Fischl en Volberding waargenomen effectgrootte (bij tweezijdige toetsing).

In een aantal Europese landen wordt momenteel de Concorde trial uitgevoerd. Deze is gerandomiseerd, dubbelblind en placebo gecontroleerd van opzet als het gaat om personen met een asymptomatische HIV infectie. Op het moment dat AIDS optreedt, wordt de blindering verbroken en krijgt de desbetreffende persoon AZT. De Concorde trial kent de langste observatieduur van tot op heden gepubliceerde klinische trials (36 maanden). Vooralsnog is er geen significant verschil tussen



interventie en controlegroep waargenomen, ook niet bij correctie voor de initiële CD4 telling. Er wordt een AIDS incidentie van 8,7 per 100 persoonsjaren waargenomen (95% BI 4,8-12,6) (52).

Hamilton et al. vinden dat AZT de progressie van *symptomatische* HIV infectie (ARC) naar AIDS vertraagt (45). De observatieduur bedroeg in dit onderzoek 2 jaar. Het onderzoek was aanvankelijk dubbelblind van opzet. Na de resultaten van de trial van Volberding werd de blinding verbroken, waardoor de validiteit van de bevindingen is verminderd (*zie bijlage A*). De interventiegroep heeft een AIDS incidentie van 7,3 per 100 persoonsjaren (95% BI 4,9-10,6). Er is geen significant verschil met de controlegroep ( $p = 0,02$ ). Deze bevinding is gedeeltelijk in overeenstemming met die van Fischl et al. (50). Er is een effect van AZT bij symptomatische HIV infectie, maar de gevonden waarde voor de effectgrootte verschilt.

Het bijgebrachte materiaal geeft de indruk dat het effect van AZT in de tijd niet constant is. Naarmate de observatieduur van een trial toeneemt, neemt het waargenomen effect af (*zie ook 5.5*). Misschien is dit te verklaren door de resistentie van het HIV tegen het AZT die soms in de literatuur wordt gerapporteerd (53). Ook speelt mogelijk de toegenomen effectiviteit van de PCP profylaxe een rol (54). Het belang van een longitudinale registratie om het verloop van het effect in de tijd nauwkeurig te kunnen beschrijven en te vergelijken met een ander beloop, neemt hierdoor toe. De uitkomst van een transversale vergelijking zal immers sterk afhangen van de periode waarover is geobserveerd.

Ook lijkt het erop dat het effect van AZT verschilt, naarmate de samenstelling van de onderzoekspopulatie verschilt. Bij personen met een symptomatische HIV infectie wordt een geringer effect waargenomen dan bij personen met een asymptomatische HIV infectie. Dit is in overeenstemming met de verwachtingen als men bedenkt dat de symptomen kunnen worden gezien als een uiting van een verhoogde activiteit van het ziekteproces.

We kunnen nu concluderen dat het verwachte (overall) effect van AZT op een progressie naar AIDS bij een asymptomatische HIV infectie is gelegen binnen een interval van 2,1-12,6 per 100 persoonsjaren (tolerantie-interval; *zie bijlage B*). De observatieduur en de toestand van de onderzoekspopulatie beïnvloeden de effectgrootte als het gaat om een progressie van asymptomatische HIV infectie naar AIDS.

Het verwachte effect kan nu worden vergeleken met het gerealiseerde effect. Cohort-onderzoek dat een indruk van het gerealiseerde effect geeft, wordt in de literatuur vooralsnog niet aangetroffen. Daarom vergelijken we de efficacy met een gerealiseerd effect uit een lokale gegevensbron (*zie 5.4.3*).

#### 5.4.2. Het effect op de sterfte bij personen met AIDS

De eerste klinische trial naar het effect van AZT bij personen met AIDS vond plaats in 1986. Deze werd na 6 maanden gestopt, omdat er in een relatief korte tijd een groot verschil tussen interventie en placebogroep werd waargenomen met betrekking tot het optreden van opportunistische infecties en sterfte. De interventiegroep bestond uit personen met AIDS, met een *Pneumocystis carinii* pneumonie als (indicator)diagnose (28). In de interventie groep werden 24 opportunistische infecties waargenomen, in de placebogroep 45. In de interventiegroep overleed slechts 1 patiënt (1/145), in de placebogroep overleden 19 patiënten (19/137). Dit komt neer op een relatief risico van 0,05.

Nadien wordt geen *placebo* gecontroleerd onderzoek naar het effect van AZT op sterfte bij personen met AIDS meer aangetroffen. De *efficacy* moet worden geschat op 2,1 sterfgevallen per 100 personen per jaar (95% BI 0,2-11,6). De relatief korte observatieduur moet daarbij in acht worden genomen. Uiteindelijk lijkt iedereen met een HIV infectie ook aan de gevolgen ervan te overlijden. De bevinding van Fischl et al. impliceert dat het verloop van het ziekteproces door AZT wordt beïnvloed, niet de uiteindelijke afloop ervan.

Gegevens over de *effectiviteit* kunnen worden ontleend aan cohort onderzoek. In een cohort onderzoek van Fischl et al., een follow-up van personen uit het onderzoek in 1986, wordt bij AZT gebruikers met AIDS een sterfte van 28,2 per 100 persoonsjaren waargenomen (55). In een ander onderzoek van Fischl et al. wordt een lage dosis AZT (600 mg/dag) vergeleken met een hoge dosis (1.200 mg/dag) bij personen met AIDS (*Pneumocystis carinii* pneumonie).<sup>10</sup> De lage dosis AZT blijkt even effectief als de hoge dosis. Er wordt een sterfte waargenomen van 31,4 per 100 persoonsjaren (53). In cohort onderzoek wordt doorgaans een veel hogere sterfte gevonden dan in gerandomiseerd interventie onderzoek.

Het onderzoek van Swanson & Cooper is uitgevoerd in Australië (56). Er wordt een significant lagere sterfte aangetroffen, dan in de Amerikaanse onderzoeken van Fischl et al (53) (16,7 versus 31,4 per 100 persoonsjaren). Het gaat echter om een heterogene populatie AIDS patiënten, die niet alleen PCP hebben, maar uit alle verschillende subgroepen van CDC IV stammen.

Ervan uitgaande dat alle sterfte plaatsvindt na het stellen van de diagnose AIDS, wordt in de Concorde trial een sterfte van 55 op 154 personen waargenomen, hetgeen neerkomt op 11,9 per 100 persoonsjaren (52). Ook hier gaat het om een heterogene verzameling AIDS patiënten. De samenstelling van de onderzoekspopulatie lijkt dus een belangrijke invloed te hebben op de waargenomen effectgrootte. Deze zou het grootste kunnen zijn bij patiënten met een PCP, maar neemt af in meer heterogene patiëntenpopulaties.

---

<sup>10</sup>De PCP profylaxe is random verdeeld over interventie en controlegroep. Het effect van deze profylaxe is inmiddels aanmerkelijk verbeterd en interfereert mogelijk met het bruto effect van AZT.

Opvallend is het grote effect van AZT in studies onder i.v. druggebruikers. In een cohort Italiaanse AIDS patiënten nemen Vella et al. AZT behandeling als onafhankelijke voorspeller van sterfte waar ( $RR^{11} = 0,49$ ; 95% BI 0,37-0,68) (57). Het cohort bestaat voor ongeveer 70% uit i.v. drug gebruikers. In een Spaans cohort AIDS patiënten, voor bijna 65% bestaande uit i.v. drug gebruikers, wordt bij behandeling met AZT een relatief risico van 0,48 op sterfte gevonden (58). Mogelijk gaat het hier om selectie. De i.v. drug gebruikers die bereid zijn AZT te gebruiken, verschillen naar alle waarschijnlijkheid sterk in compliance en risicodrag van de overige i.v. druggebruikers.

#### 5.4.3. Lokaal gerealiseerde effecten

Om de vergelijking van een lokaal gerealiseerd effect met de efficacy te illustreren maken we gebruik van een registratie in het Academisch Ziekenhuis Utrecht (AZU), één van de twaalf centra in Nederland met een regiofunctie voor behandeling van HIV/AIDS (59). Voor zover personen in deze registratie niet aan een experiment deelnemen, is er materiaal voor uitkomst-onderzoek. (60) Voor uitkomst-onderzoek is een definitie van de behandelingsperiode in relatie tot een welomschreven uitkomst van groot belang (zie bijlage C). Een persoon is aangemerkt als een AZT gebruiker indien er in de periode 1987-1992 (a) een klinische opname is geregistreerd waarbij sprake was van AZT gebruik, en/of (b) een poliklinisch AZT gebruik is waargenomen. Het tijdstip waarop de therapie is aangevangen, is niet geregistreerd. Informatie over dit tijdstip is echter van belang om de behandelingsperiode te definiëren. Als begin van de therapie is daarom het moment gekozen waarop een CD4 telling van minder dan 300 cellen per  $mm^3$  is waargenomen (de formele indicatie voor AZT gebruik in Nederland).

De tabel 6 geeft de kenmerken van AZT gebruikers in het AZU, alsmede de belangrijkste uitkomstmaten: de incidentie van AIDS onder personen met een asymptomatische HIV infectie<sup>12</sup> en de sterfte onder personen met AIDS. Het betreft relatief kleine aantallen, hetgeen de lokale situatie doorgaans zal typeren. De registratie geeft een gedetailleerd inzicht in de samenstelling van de patiëntenpopulatie, de observatieduur in termen van het aantal persoonsjaren, en de uitkomsten in het verloop van de tijd. Opmerkelijk is de verschuiving in de samenstelling van de patiëntenpopulatie, zowel naar leeftijd, geslacht als risicogroep.

In het AZU is de CD4 telling de beste voorspeller voor AZT gebruik. Daarnaast zijn leeftijd en observatieduur voorspellende variabelen voor AZT gebruik ( $p < 0,05$ ). Als de leeftijd of observatieduur stijgt, neemt de kans op AZT gebruik toe. Een langere overleving vergroot dus de kans op therapie, waardoor het therapeutisch effect kan worden overschat. De therapie moet in het model als een

---

<sup>11</sup> Relatief Risico

<sup>12</sup> Het stadium CDC III is als een asymptomatische HIV infectie beschouwd.

tijdsafhankelijke co-variant worden opgenomen (42). De verzekeringsvorm is in de analyse opgenomen als indicator voor de sociaal-economische status. De analyse is uitgevoerd met en zonder de tijd om verschillen in resultaat te onderzoeken.

Om het *effect van AZT op de progressie naar AIDS* te onderzoeken is in de AZU registratie een selectie gemaakt van personen met een asymptomatische HIV infectie en een CD4 telling van minder dan 300 cellen per  $\text{mm}^3$  ( $n = 79$ ). Van deze personen kregen er 50 AZT toegediend. De 29 personen die geen AZT gebruikten, vormen de groep waarin het basisrisico van de onderhavige populatie op progressie naar AIDS kan worden geschat (30). Deze groep heeft echter een zeer geringe AIDS incidentie, hetgeen verklaard kan worden door de geringe observatieduur. Een schatting van het basisrisico is niet mogelijk. Daarom moeten we voor het verwachte (overall) effect van AZT op een progressie naar AIDS bij een asymptomatische HIV infectie een tolerantie-interval van 2,1-12,6 per 100 persoonsjaren hanteren (zie bijlage B).

Van de 50 personen met een asymptomatische HIV infectie in de AZU registratie hebben er 23 gedurende de observatieperiode een progressie naar AIDS doorgemaakt. Deze getallen representeren een proces in de tijd. *Figuur 1* beschrijft de incidentie van AIDS onder personen met een asymptomatische HIV infectie in het AZU. Er kregen 17 patiënten AIDS gedurende het eerste jaar nadat zij seropositief waren geworden en de therapie met AZT was begonnen. Na het tweede jaar is dit aantal opgelopen tot 21 en na drie jaar tot 22 personen. De gemiddelde observatieduur van het cohort bedraagt 14,7 maanden. Gemiddeld wordt er onder AZT gebruikers met een asymptomatische HIV infectie een AIDS incidentie van 37,6 per 100 persoonsjaren (95% BI 23,8-56,2) waargenomen. Deze waarde voor de effectgrootte is buiten het tolerantie-interval van de klinische trials gelegen. Derhalve constateren we een discrepantie tussen efficacy en effectiviteit.

Meestal wordt in observationeel onderzoek een veel geringer effect van AZT waargenomen dan in trials. Bias in observationeel onderzoek kan ontstaan, omdat patiënten die voor therapie in aanmerking komen zieker zijn (54). Daardoor kan een therapie zelfs schadelijk lijken (42). Dit is ook iets dat we bij de analyse van het AZU bestand waarnemen. Ook als we co-varianten in de analyse opnemen die uitdrukking geven aan de gezondheidstoestand van de patiënt, zoals de CD4 telling, de duur van de periode gedurende welke de patiënt seropositief was, leeftijd, geslacht, of verzekeringsvorm blijft de groep AZT gebruikers een slechtere prognose houden dan niet AZT gebruikers. Waarschijnlijk is er sprake van een selectie door de behandelend arts op basis van niet in de analyse betrokken kenmerken van de patiënt. Het is een van de valkuilen van de analyse van observationeel onderzoek en onderstreept het belang van een (lokale) schatting van het basisrisico, opdat een zinvolle vergelijking van efficacy en effectiviteit plaats kan vinden.

In de AZU populatie is er een significant verschil in AIDS incidentie tussen personen die poliklinisch AZT ontvingen en personen die niet poliklinisch AZT ontvingen ( $p < 0,01$ ). Ook dit wijst op een selectie. Verder is er een significant verschil in progressie naar AIDS voor het geslacht en de risicogroep ( $p < 0,05$ ). Er zijn geen vrouwen met progressie naar AIDS. Homoseksuele personen hebben het grootste risico op een progressie. Als het verloop van het effect in de tijd in beschouwing wordt genomen (cox-regressie model), blijken leeftijd, geslacht, risicogroep en verzekeringsvorm *geen* voorspellers voor de progressie naar AIDS te zijn ( $p > 0,05$ ). Dit illustreert het verschil tussen een transversale en een longitudinale analyse.

Om het *effect van AZT op de sterfte* te onderzoeken is een selectie gemaakt van personen met verschijnselen van AIDS ( $n = 88$ ). Deze vallen uiteen in twee groepen: personen die al AIDS hadden bij aanvang van de observatieperiode ( $n = 65$ ) en personen die op een bepaald moment in de observatieperiode AIDS kregen ( $n = 23$ ). Er is een significant verschil in overleving tussen personen met AIDS die wel en personen die niet met AZT zijn behandeld in de fase dat zij een asymptomatische HIV infectie hadden ( $p < 0,01$ ; log-rank toets). Dit is toe te schrijven aan de duur van de periode waarin AZT wordt gebruikt voordat de diagnose AIDS is gesteld ( $p = 0,01$ ; log-rank toets). Het is dus niet gerechtvaardigd beide groepen samen te voegen. De groep van 65 personen die AZT gebruikten na het stellen van de diagnose AIDS, is gebruikt om het effect van AZT op de sterfte te onderzoeken. Ook hier is het basisrisico op overlijden geschat in een groep personen met AIDS die geen AZT gebruikten ( $n = 45$ ). Deze is geselecteerd op basis van een CD4 telling van minder dan 300 cellen per  $\text{mm}^3$ .

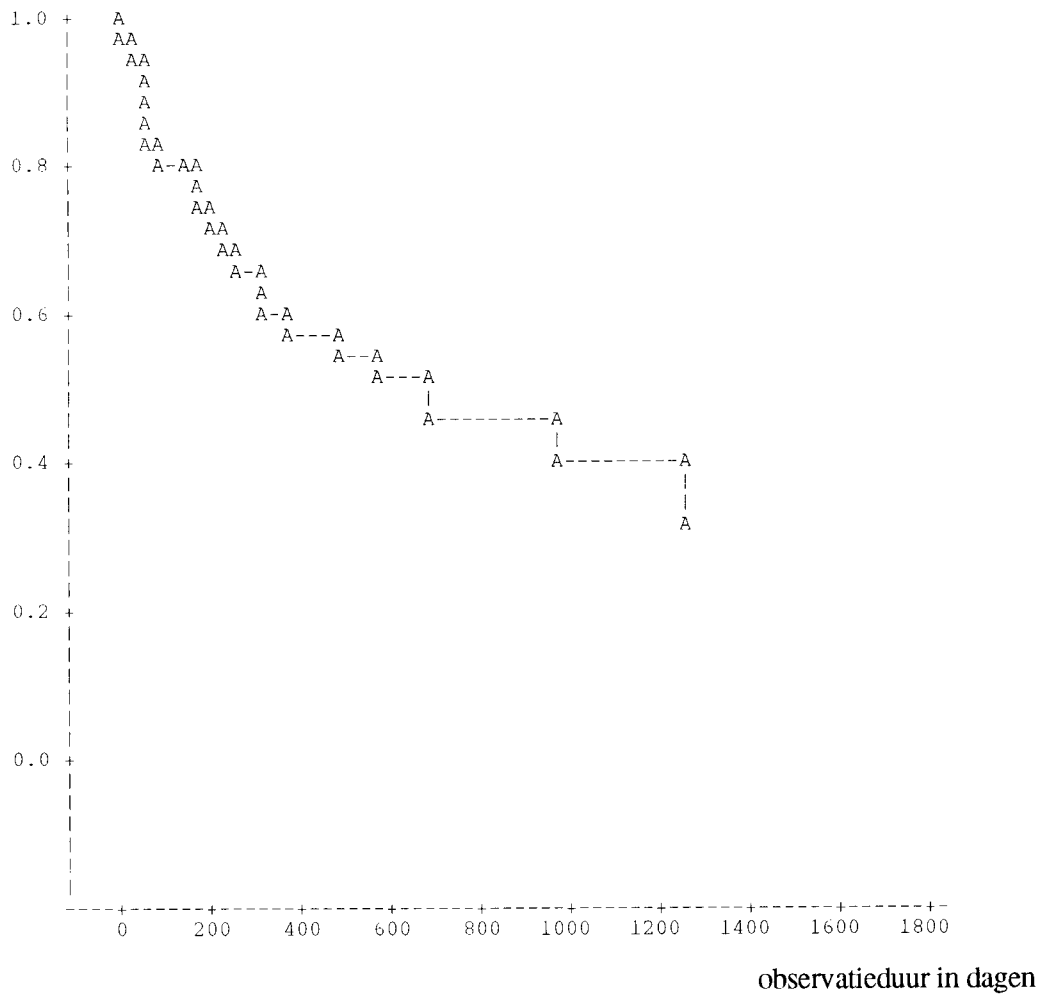
Tabel 6. AZT gebruikers in de Academisch ziekenhuis Utrecht (AZU)

jaar	1987	1988	1989	1990	1991	1992	totaal
aantal (n) <sup>13</sup>	5	18	39	50	65	80	115
aantal persoonsjaren <sup>14</sup>	2,5	12,7	22,4	35,2	45,7	56,0	174,6
gemiddelde leeftijd	46,4	41,0	39,1	39,3	37,5	36,9	38,4
geslacht % mannen	100	94,4	94,9	94,0	90,8	90,0	93,0
verzekeringsvorm % particulier.	40,0	27,8	33,3	34,0	26,2	23,8	27,8
% ziekenfonds	60,0	72,2	66,7	66,0	73,9	76,2	72,2
risicogroep % homoseksueel	100	83,3	79,5	76,0	64,6	62,5	67,8
% heteroseksueel	0	5,6	5,1	8,0	16,9	15,0	13,0
% biseksueel	0	11,1	10,3	8,0	7,7	10,0	8,7
% i.v. druggeb.	0	0	2,6	4,0	1,5	1,3	1,7
% haemofiliepat.	0	0	2,6	4,0	9,2	11,3	8,7
ziektestadium (aantallen)							
CDC II/III <sup>15</sup>	1	8	15	18	25	30	50
CDC IV <sup>16</sup>	4	9	23	27	32	34	65
CDC IV <sup>17</sup>	5	11	28	37	49	53	88
progressie van CDCII/III naar CDC IV (aantallen)	1	1	4	5	9	3	23
sterfte (aantallen)	1	0	6	12	11	14	44
waarvan pmp <sup>18</sup>	0	0	0	1	1	7	9

<sup>13</sup> prevalentie gevallen/jaarprevalentie<sup>14</sup> gegenereerd in het jaar waarin de personen aanwezig zijn (geweest).<sup>15</sup> prevalentie van personen met asymptomatische HIV infectie bij aanvang van de observatieperiode. Noemer voor progressie analyse.<sup>16</sup> prevalentie van personen met CDC IV (AIDS) bij aanvang observatieperiode, exclusief personen met een progressie. Noemer voor sterfte-analyse.<sup>17</sup> prevalentie van personen met CDC IV (AIDS) inclusief personen met een progressie van HIV+ naar AIDS.<sup>18</sup> sterfte onder personen met CDC IV die gedurende de observatieperiode een progressie doormaakten. pmp = personen met een progressie.

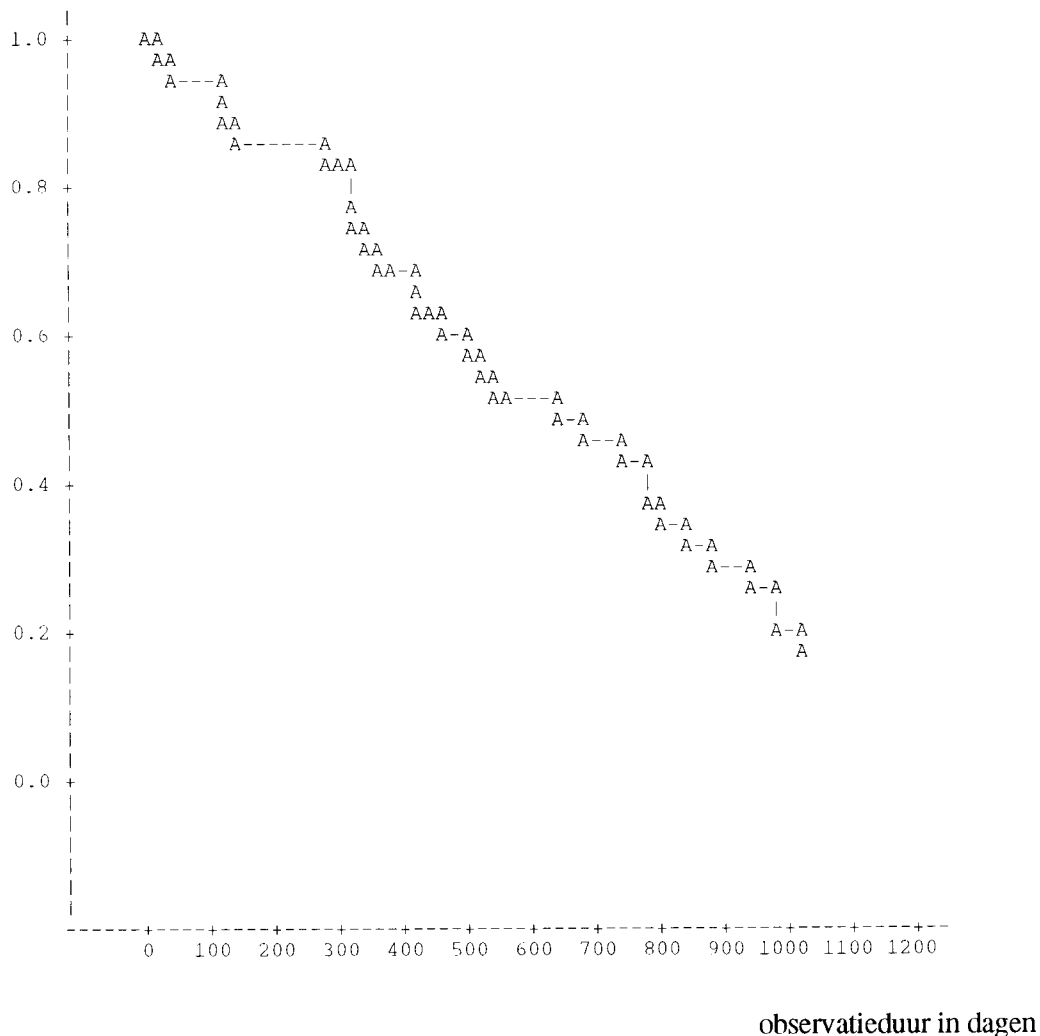
*Figuur 1. Progressie van asymptomatische HIV-infectie naar AIDS bij AZT-gebruik (n = 50)*

% asypt. HIV geïnfecteerden  
(x 100)



Figuur 2. Sterfte onder personen met AIDS bij AZT gebruik (N =65)

% overleving (x 100)



De *figuur 2* beschrijft de overleving onder personen met AIDS in het AZU. Er overleden 16 patiënten (30%) gedurende het eerste jaar nadat de diagnose AIDS was gesteld en de therapie met zidovudine was begonnen. Na twee jaar is dit aantal opgelopen tot 27 personen (55%). In het totaal worden 35 sterfgevallen waargenomen en is de gemiddelde observatieduur 14,6 maanden. Gemiddeld zijn er dus 44,1 sterfgevallen per 100 persoonsjaren (95% BI: 30,7-61,6). Deze cijfers zijn vergelijkbaar met de getallen uit het vervolgonderzoek van Fischl et al. (55). Het lokaal gerealiseerd effect wijkt niet af van het in de literatuur gerapporteerde effect. Over het algemeen liggen de getallen in cohort onderzoek beduidend hoger dan in trials.

Uit het logistische model blijkt verzekeringsvorm een voorspeller voor de sterfte te zijn ( $p < 0,01$ ; RR = 0,12). De sterfte is hoger bij vrijwillig verzekerden dan bij verplicht verzekerden. Als het



verloop van het effect in de tijd in beschouwing wordt genomen (cox-regressie model), blijken leeftijd, geslacht, risicogroep en verzekeringsvorm *geen* voorspeller van de sterfte. Dit illustreert het verschil tussen een transversale en een longitudinale analyse.

Met behulp van de formules van Lubsen en Tijssen is nu het verwachte effect voor de AZU populatie uit te rekenen, zodanig dat het kan worden vergeleken met de efficacy (*zie pagina 24*). Daarvoor maken we gebruik van alle personen met AIDS die een CD4 telling onder de 300 cellen per  $\text{mm}^3$  hebben. In het AZU overlijden 44 van de 88 personen met AIDS (50%) die AZT gebruiken, en 19 van de 43 personen met AIDS (44%) die geen AZT gebruiken. Voor personen met AIDS bedraagt het relatieve risico op sterfte bij AZT gebruik 0,05 (Fischl et al., 1987). In het AZU verwachten we dus een effect van AZT van  $0,05 * 0,44 = 0,02$ . Het waargenomen effect bedraagt 0,50 (95% BI: 0,36-0,67). De verwachte waarde van de effectgrootte ligt buiten het betrouwbaarheidsinterval van de gerealiseerde waarde. We constateren derhalve een discrepantie tussen efficacy en effectiviteit.

## 5.5 De discrepantie tussen efficacy en effectiviteit van zidovudine

Het constateren van een discrepantie tussen efficacy en effectiviteit is aanleiding voor het zoeken naar een verklaring voor deze discrepantie. In de *tabel 5* zijn gegevens over het effect van AZT bijeengebracht. Deze zijn aan een variantie-analyse onderworpen met de effectgrootte als afhankelijke variabele om verschillen in effectgrootte toe te kunnen schrijven aan verschillen in factoren die een discrepantie tussen efficacy en effectiviteit kunnen veroorzaken (*tabel 3*). Een variantie-analyse is gebruikt omdat we te maken hebben met een toets op significante verschillen tussen meerdere klassen (F-toets). De onafhankelijke variabele, de effectgrootte, is van (het vereiste) intervalniveau, namelijk het percentage personen dat na een jaar overlijdt of het percentage seropositieve personen dat na een jaar AIDS krijgt, en kent na logaritmische transformatie bij benadering een normale verdeling. Er is gewogen op basis van de inverse van de variantie. Het aantal studies is echter klein, waardoor een multivariate toetsing niet mogelijk is. Daarom is de analyse telkens univariaat uitgevoerd. Wel moet worden bedacht dat het getal uit een onderzoek een populatie patiënten representeert. Het gaat om een illustratie van het principe, niet om de absolute waarde van de uitkomsten.

Niet alle informatie die op theoretisch gronden relevant is om deze discrepantie te verklaren, is aan de verzamelde literatuur te ontleen. Daarom is een beperkt aantal indicatoren gebruikt om de oorzaken van een discrepantie te operationaliseren. De meting van het effect is gekarakteriseerd met behulp van de omvang (n) van de onderzoekspopulatie en de duur van het onderzoek. Beide indicatoren vertonen een samenhang met de kans op het vinden van een significant resultaat (*zie bijlage A*). De

patiëntenpopulatie is gekarakteriseerd met behulp van de kenmerken geslacht (% mannen), risicogroep (% homoseksuelen) en ziektestadium. De samenstelling lijkt immers invloed te hebben op de effectgrootte (zie paragraaf 5.4.1.). De indicatiestelling betreft het aantal CD4 cellen dat richting geeft aan het voorschrijven van AZT. De interventie is gekenmerkt door de dosering van AZT en de gemiddelde observatieduur van de onderzoekspopulatie als indicatie voor de behandelingsperiode. De omstandigheden zijn geoperationaliseerd aan de hand van het land en het jaar van instroom waarin het onderzoek plaatsvond.

Tabel 7. Discrepantie tussen efficacy en effectiviteit van zidovudine

variabele	indicator	sterfte		progressie	
		F-waarde	p-waarde	F-waarde	p-waarde
kans op uitkomst	omvang	4,80	0,07 ns	0,28	0,62 ns <sup>19</sup>
patiënt populatie	studieduur	7,65	0,03 s	0,54	0,51 ns
medische interventie	% mannen	23,66	0,00 s	3,10	0,15 ns
indicatie stelling	% homoseksueel	.	.	.	ns
omstandigheden	observatieduur	0,58	0,48 ns	30,73	0,01 s
	dosis	0,02	0,91 ns	0,17	0,70 ns
	CD4 telling	0,02	0,90 ns	1,07	0,36 ns
	land	0,90	0,38 ns	4,81	0,09 ns
	jaar v. instroom	0,00	0,99 ns	1,54	0,28 ns

#### Kans op een gunstige uitkomst

Wanneer de omvang van de onderzoekspopulatie als een indicator voor de meting van het effect wordt gebruikt, is er geen significant verschil in effectgrootte voor de progressie van asymptomatische HIV infectie naar AIDS of sterfte bij AIDS ( $p > 0,05$ ). Een relatie tussen omvang van de onderzoekspopulatie en de uitkomst kan wijzen op een publicatiebias (zie bijlage B). Een onderzoek wordt gepubliceerd zodra er een significant resultaat wordt waargenomen, terwijl het bereiken van significantie samenhangt met de omvang van de onderzoekspopulatie. Hetzelfde kan gelden voor de duur van het onderzoek. Voor de onderzoeksduur is er geen significant verschil in effectgrootte voor de progressie van asymptomatische HIV infectie naar AIDS ( $p > 0,05$ ), maar wel voor onderzoek naar de sterfte bij AIDS ( $p < 0,05$ ). In het laatste geval kan dit wijzen op publicatiebias of varieert de prognose in het verloop van de tijd.

<sup>19</sup> sign = significant:  $p < 0,05$ ; F-toets

### patiëntenpopulatie

Er is een opmerkelijk verschil in samenstelling in de patiëntenpopulaties van noord- en zuid-europese landen. In Zuid-Europa bestaat de populatie HIV geïnfecteerde personen voornamelijk uit i.v. drugverslaafden: in een Spaans cohort 65,5% (58), in een Italiaans cohort 69% (57). Invloed hiervan op effectgrootte is niet goed vast te stellen. Te vaak ontbreekt informatie over het percentage homoseksuele personen. Gelet op het percentage mannen in de onderzoekspopulatie is er geen significant verschil in effectgrootte voor de progressie van asymptomatische HIV infectie naar AIDS ( $p > 0,05$ ), maar wel voor de sterfte bij AIDS ( $p < 0,05$ ). De samenstelling van de onderzoekspopulatie heeft dus invloed op de effectgrootte.

### kenmerken van de medische interventie

De dosis van het toegediende AZT en de gemiddelde (observatie)duur gebruiken we als indicatoren voor de aard van de medische interventie. Voor de gemiddelde observatieduur is er een significant verschil in effectgrootte voor de progressie van HIV naar AIDS ( $p < 0,05$ ), niet voor de sterfte onder personen met AIDS.

### indicatiestelling

De CD4 telling is een goede indicator voor de indicatiestelling van de medische interventie (*zie paragraaf 5.1*). Er is geen significant effect van de CD4 telling op het verschil in effectgrootte voor de progressie van een asymptomatische HIV infectie naar AIDS of op de sterfte bij AIDS ( $p > 0,05$ ).

### Omstandigheden

Het land is een indicator voor omstandigheden waaronder de medische interventie wordt toegepast. Zo was AZT in de Verenigde Staten door de hoge prijs niet voor allen toegankelijk. Het werd dan ook aanvankelijk voornamelijk gebruikt door de beter gesitueerden. Ook wordt AZT er vaak gebruikt zonder officieel te zijn voorgeschreven. Wanneer de Verenigde Staten wordt vergeleken met de overige landen is er geen significant verschil in effectgrootte voor de progressie naar AIDS of voor de sterfte onder personen met AIDS ( $p > 0,05$ ).

De omstandigheden kunnen ook worden getypeerd door het tijdstip waarop een therapie is toegepast. Soms is er een verschil in effectgrootte naar jaar van diagnose waargenomen (56). Een analyse van verschil in effectgrootte naar jaar waarin het onderzoek is uitgevoerd, levert geen significant resultaat op ( $p > 0,05$ ).

### Aangrijpingspunten voor beleid

Een verklaring voor een discrepantie tussen efficacy en effectiviteit wordt voor het effect van AZT op de progressie van HIV naar AIDS gevonden in de medische interventie (observatieduur), en voor het effect van AZT op de sterfte bij personen met AIDS in aspecten van de effectmeting (studieduur) en de patiëntenpopulatie (% mannen) (tabel 7). Bijpassende beleidsmaatregelen kunnen vervolgens worden aangewend of geformuleerd om deze discrepantie te beïnvloeden (zie tabel 2). Daarbij valt te denken aan standaardisatie van de effectmeting, protocollering van de behandeling, aanscherping van de indicatiestelling met betrekking tot de te behandelen populatie, of het stimuleren van nader onderzoek naar het effect van AZT bij bepaalde subgroepen.

De analyse van een 'lokale' discrepantie tussen efficacy en effectiviteit vormt een belangrijke aanvulling op de analyse van een discrepantie waarin generaliseerbaar cohort onderzoek als bron van gegevens over effectiviteit wordt gebruikt (een 'globale analyse'). De 'lokale analyse' maakt gebruik van een aan tijd en plaats gebonden gegevensbron, in dit voorbeeld het AZU. Er blijken dan andere variabelen van invloed op de discrepantie dan in de globale analyse. Bovendien blijkt dat de resultaten van een transversale analyse verschillen van die van een longitudinale analyse. Hieruit volgt: (a) een voorbehoud bij de resultaten van de globale analyse. Deze behoeven validering aan de hand van een lokaal gegevensbestand. Discrepantie tussen efficacy en effectiviteit op een globaal niveau is 'slechts' een indicator, een aanwijzing dat er mogelijk iets aan de hand is, en geeft richting aan nader onderzoek op lokaal niveau; (b) een behoefte aan (longitudinale) informatie met een groot detailniveau om de conclusies van een globale analyse te kunnen valideren.

## **5.6 Conclusies**

We kunnen nu een aantal conclusies trekken over de toepasbaarheid van het ontwikkelde conceptuele kader voor de analyse van effecten van een medische interventie.

- (i) De gehanteerde concepten blijken geschikt om het effect van medische interventies in kaart te brengen. De nuancering van de beschrijving neemt toe, waardoor de verschillende interpretatiemogelijkheden beter voor het voetlicht komen.
- (ii) Als een medische interventie over een wetenschappelijke onderbouwing van graad I beschikt kan worden gesproken van efficacy. Deze efficacy kan vervolgens worden vergeleken met effectiviteit. Daarbij moet onderscheid worden gemaakt tussen vergelijking van de efficacy met een *globale* of met

een *lokale* effectiviteit. Beide vergelijkingen vereisen verschillende methoden en stellen andere eisen aan gegevensbronnen.

(iii) Voor analyse van de discrepantie tussen efficacy en een globale effectiviteit dienen de gegevens zo gedetailleerd mogelijk te worden weergegeven. Een synthese door middel van meta-analyse om tot een integrale effectmaat te komen is niet nodig. Voor een analyse van de discrepantie tussen efficacy en een lokale effectiviteit is een geïntegreerde waarde voor de efficacy vereist. Dan is wel meta-analyse nodig als er meerdere onderzoeken over de efficacy aanwezig zijn.

(iv) Voor een vergelijking van een lokaal effect met de efficacy is een schatting van het basisrisico (in de onbehandelde populatie) vereist. Als deze ontbreekt is er geen conclusie over een discrepantie tussen efficacy en een *lokale* effectiviteit te trekken.

(v) De discrepantie tussen een potentieel effect en een gerealiseerd effect is een indicator voor het gezondheidszorgbeleid. De indicator wijst op een mogelijk te behalen gezondheidswinst indien het betrouwbaarheidsinterval van het gerealiseerde effect geen overlap vertoont met dat van het potentieel effect. Nadere analyse is dan aangewezen. Een vergelijking van de efficacy met de effectiviteit op lokaal niveau vormt een aanvulling en validering van een vergelijking van efficacy met de effectiviteit op een globaal niveau. Op basis van de resultaten van de analyse kunnen vervolgens passende beleidsmaatregelen worden genomen.

(vi) De analyse van de discrepantie tussen efficacy en effectiviteit is zodanig uit te voeren dat gebruik wordt gemaakt van de verschillen tussen studies om een verschil in effectgrootte te verklaren. Het veel gehanteerde argument dat uitkomsten van onderzoek niet vergelijkbaar zijn vanwege de verschillen in onderzoeksopzet, wordt daarmee weggenomen. Op deze wijze worden aangrijpingspunten voor een verbetering van de zorgverlening gelokaliseerd. Hierin is de bruikbaarheid van een discrepantie tussen efficacy en effectiviteit als indicator voor beleid gelegen.



## 6 ONDERZOEK NAAR HET EFFECT VAN MEDISCHE INTERVENTIES

### 6.1. Huidige stand van zaken

Het effect van medische zorg kan op verschillende manieren worden onderzocht. Het medisch onderzoek richt zich vooral op het causale verband tussen een medische interventie en een waargenomen verandering in de gezondheidstoestand (klinische trial). Gezondheidszorgonderzoek is in dit verband breder van aard. Het richt zich op het effect van medische interventies in de praktijk, ingebed in een proces van zorgverlening, en neemt daarbij een breed spectrum van (niet alleen medische) gevolgen in ogenschouw.

Informatie over efficacy kan worden ontleend aan (meta-analyse van) klinische trials. De klinische trial is de meest valide vorm van patiëntgebonden onderzoek om een effect vast te stellen. Een beschrijving en weging van de kenmerken van de klinische trial stellen ons in staat de relevantie van medische zorg op waarde te schatten (*bijlage A*). Klinische trials over eenzelfde onderwerp worden onderzocht met behulp van meta-analyse (*bijlage B*). Over de methodiek van meta-analyse van experimenteel onderzoek is inmiddels een (internationale) consensus bereikt; over die van observationeel onderzoek nog niet (61). In 1992 is de 'Cochrane collaboration', een internationaal netwerk voor vergaring van informatie over klinische trials, opgericht. Zowel lopend als afgesloten onderzoek wordt bijeengebracht. Thema- of ziekte/indicatiegewijs wordt meta-analyse verricht. De informatie wordt via internet en diverse publikaties ter beschikking gesteld aan een breed publiek (62).

De informatie die de Cochrane collaboration verzamelt, groeit, maar er zijn nog grote hiaten. Voor veel medische interventies is er geen informatie over efficacy aanwezig. Bovendien is deze informatie niet zaligmakend. Begin tachtiger jaren rapporteerden Kienle en Burkhardt aan de Duitse overheid over het gebruik van klinische onderzoeksgegevens in het kader van het toelatingsbeleid van geneesmiddelen op de markt (63). Zij concludeerden dat klinische trials, mede gelet op de methodologie van dit soort onderzoek, geen waarborg bieden voor de werkzaamheid of veiligheid van geneesmiddelen en adviseerden post-marketing surveillance, ofwel gegevens over effectiviteit, te gebruiken en het middel uit de markt te nemen bij aanwijzingen van schade of gebleken onwerkzaamheid.

De effectiviteit van een medische interventie wordt onderzocht in uitkomst onderzoek ('outcome research'), een vorm van onderzoek die vooral in de Verenigde Staten een grote omvang heeft aangenomen (16). Er wordt gebruik gemaakt van grote gegevensbronnen met zowel administratieve als medisch inhoudelijke gegevens. Het uitkomst onderzoek is gericht op de medische, psychologische en sociaal-economische gevolgen van een medische interventie. (64) (65). Ook op het gebied van uitkomst-onderzoek is sprake van een internationaal netwerk, de pendant van de 'Cochrane collaboration'. Voor Europa is er een

'European Clearing House on Outcome research (ECHO)', gevestigd te Leeds (VK) met vertakkingen in diverse landen. Een 'clearing house' vergaart informatie over het uitkomst onderzoek in een bepaalde regio (66). Ook ondersteunt het onderzoek door het beschikbaar stellen van vragenlijsten voor meting van uitkomsten. Het valideren en standaardiseren van meetinstrumenten, met name voor kwaliteit van leven en patiëntensatisfactie, vormen belangrijke aandachtspunten (67). De methodologie van uitkomst onderzoek is echter nog niet eenduidig en nog volop in ontwikkeling. Ook de interpretatie staat ter discussie. Clinici ervaren informatie over uitkomsten van het zorgproces vaak als een zinvolle terugkoppeling over het effect van hun handelen. Beleidsmakers vinden het moeilijk beleidsmaatregelen op uitkomstgegevens te baseren, omdat de relatie met het zorgproces, doorgaans aangrijpingspunt voor beleid, vaak niet eenduidig of niet voldoende duidelijk is (68).

Als er zowel gegevens over efficacy als effectiviteit aanwezig zijn, rijst de vraag, *of en zo ja hoe*, beide met elkaar kunnen worden vergeleken. Een informatiesynthese waardoor beide onder een gemeenschappelijke noemer (effectgrootte) worden gebracht en waarvan meta-analyse een inherent onderdeel uitmaakt (69), is daartoe onontbeerlijk. Een toeschrijving van verschillen in effectgrootte aan verschillen in patiëntenpopulatie, werkstijl, indicatiestelling, of situatie is vervolgens een methode om een discrepantie tussen efficacy en effectiviteit nader te analyseren. Deze methodiek behoeft verder invulling (*zie hoofdstuk 5*).

Een analyse van de discrepantie tussen efficacy en effectiviteit is een relatief hiaat in gezondheidszorgonderzoek. Het is van belang dat het aantal gegevens hierover toeneemt. Een vergelijking van efficacy en effectiviteit brengt fundamentele vragen met zich mee (o.a. naar causaliteit), is maatschappelijk relevant als het erom gaat welke medische zorg al of niet efficiënt kan worden toegepast, en geeft aanwijzingen voor beleidsmaatregelen op diverse niveaus.

## 6.2. Plaatsbepaling

Binnen de geschetste ontwikkelingen op het gebied van meta-analyse en uitkomst-onderzoek hebben we ons onderzoek een plaats proberen te geven. Het doel van dit onderzoek was het formuleren van een methode voor de analyse van effecten van medische zorg. Daarbij hebben we onderscheid gemaakt tussen efficacy en effectiviteit en beide met elkaar vergeleken. Aanvankelijk lijkt het om een vergelijking van appels en peren te gaan. Dit blijkt uit een controverse die in de literatuur kan worden waargenomen. Sommige personen willen alleen gegevens over efficacy in ogenschouw nemen (70). Anderen wijzen op beperkingen van experimenteel onderzoek, stellen de prestatie van een medische interventie in de praktijk centraal en gebruiken met name gegevens over effectiviteit (71) (72) De controverse betreft de onderbouwing van de causale relatie tussen een medische interventie en het waargenomen gevolg. Om deze controverse een andere wending te geven, exploreerden we een invalshoek waarbij we vanuit het verschil in situatie (experimenteel, praktijk) een



verschil in effectgrootte trachten te voorspellen. De vraag naar het causale verband staat dan niet meer ter discussie. Het gaat om de reproductie van het experimentele resultaat in de klinische praktijk. Deze reproductie is het voorwerp van ons onderzoek geworden.

Met onze case-studie pretenderen we slechts de contouren van een analyse aan te geven. Het is gebleken dat een discrepantie tussen efficacy en effectiviteit een zinvolle indicator voor het gezondheidszorgbeleid kan zijn. Aangrijpingspunten voor beleid komen aan het licht. Het is van belang onderscheid te maken tussen een vergelijking van de efficacy met een globale effectiviteit en vergelijking van efficacy met een lokale effectiviteit. Beide vergelijkingen vereisen een andere methode en stellen andere eisen aan gegevensbronnen (*zie hoofdstuk 5*).

De bruikbaarheid van ons onderzoek voor VTV is tweeërlei. Enerzijds bieden we een kader voor de analyse van het effect van medische interventies. Anderzijds brengt onze methodiek mogelijkheden ter verbetering van de zorgverlening (gezondheidswinst) aan het licht. Het komt daarmee tegemoet aan de wens effecten van medische zorg op een beleidsrelevante wijze te onderzoeken.

### 6.3. Nadere uitwerkingen

Dit onderzoek bevatte een aantal bewust aangebrachte beperkingen om het tot een enigszins hanteerbare omvang terug te brengen (*zie inleiding*). Deze betroffen: complexiteit (effect van één medische interventie op een bepaalde uitkomst), zorgsector (zelf- en mantelzorg buiten beschouwing gelaten), relatie zorg/kosten (buiten beschouwing gelaten), niveau (beperkt tot arts-patiënt relatie).

Een complexere situatie vereist een besliskundige analyse, waarbij informatie over alle mogelijke uitkomsten en behandelstrategieën in een model worden samengebracht en doorgerekend. De analyse van een enkelvoudige behandeling op een bepaalde uitkomst is als één van de bouwstenen van een dergelijk model te beschouwen. Ook een onderscheid tussen medische zorg en preventie vergroot de complexiteit. In principe lijkt er geen verschil in *analysemethodiek*. Als het gaat om primaire of secundaire preventie zijn er echter minder gegevens over efficacy, omdat het experiment moeilijker is uit te voeren en worden andere uitkomstmaten gehanteerd (incidentie, prevalentie, *zie diagram 2b*). Een afweging van medische zorg ten opzichte van preventie is mogelijk een zinvolle aanvulling van onze analyse.

Zelf- en mantelzorg worden niet uitgevoerd op geleide van geneeskundige kennis of een medische indicatie en zijn derhalve strikt genomen niet als een vorm van medische zorg aan te merken. Het is uiteraard niet uitgesloten dat zij geen effect op de gezondheidstoestand kunnen hebben. Dit is echter het terrein van sociaal-wetenschappelijk onderzoek, niet van gezondheidszorgonderzoek.

Ook vanuit economische invalshoek is een discrepantie tussen efficacy en effectiviteit relevant en kan worden getracht deze te verklaren. De kosten van een medische interventie in de experimentele situatie vertonen veelal een verschil met de kosten van een behandeling in de praktijk. Een verschil in methodiek van de kostenberekening is daarvan vaak de oorzaak. Toch kunnen een aantal factoren worden aangewezen als een economische verklaring voor een eventuele discrepantie tussen efficacy en effectiviteit: externaliteiten, prijs, utilisatie, de tijdsdimensie en innovatie-aspecten (33).

Een analyse van de kosten van een medische behandeling kan in principe parallel lopen aan de analyse van het medische effect. Het gaat in beide situaties om toeschrijving (van kosten of een effect) aan een medische interventie, het onderscheiden van verschillende soort effecten of kosten van eenzelfde medische interventie, etc. Er kan bijvoorbeeld een onderscheid tussen kosten/efficacy en een kosten/effectiviteit of kosten/baten analyse worden gemaakt. Door beide analyses met elkaar in verband te brengen ontstaat een beeld van de doelmatigheid van een medische interventie.

Het effect van medische interventies is op verschillende niveaus te onderzoeken. Zo vereist onderzoek naar het effect van een medische interventie op de sterfte in de bevolking een andere methode dan onderzoek naar het effect van een medische interventie op patiënteniveau. Er mag echter een samenhang tussen beide niveaus worden verondersteld. Het effect op individueel niveau culmineert uiteindelijk immers in een effect op populatieniveau. Naar deze samenhang en aanvullende methodologie is nader onderzoek mogelijk. Ontwikkelen van een model dat verschillende niveaus met elkaar in verbinding brengt, maakt hiervan een onderdeel uit. Hiervan zijn voorbeelden in de internationale literatuur aanwezig (73) (74) (75) (76) Een uitwerking op Nederlandse gegevens ontbreekt.

#### 6.4. Conclusie

Een vergelijking van efficacy en effectiviteit wint aan belang naarmate meer gegevens over beide ter beschikking komen. Er is sprake van beleidsrelevantie (mogelijkheden van gezondheidswinst komen aan het licht) en een relatief hiaat in Nederlands gezondheidszorgonderzoek. Ons onderzoek toont hoe een dergelijke vergelijking kan worden uitgevoerd. Hierbij hebben we gebruik gemaakt van een relatief eenvoudig voorbeeld, het effect van zidovudine bij HIV infectie of AIDS. Het gaat om een ziekte met een beperkt aantal therapeutische mogelijkheden en (dus) een beperkt aantal gegevens over het effect van medische interventies. In deze eenvoudige situatie speurden we naar mechanismen die ook in complexere situaties voorkomen. Een uitbreiding van het onderzoek naar andere gebieden (chirurgische interventies), andere niveaus (macro), of andere schaal (meer medische interventies bij eenzelfde ziektebeeld) behoort in principe tot de mogelijkheden en kan bijdragen aan een nadere uitwerking van de door ons gehanteerde analysemethode.

**LITERATUUR**

1. Ruwaard D, Van den Berg Jets A, Jansen J, Kramers PGN, Van der Giessen A, Van Genugten MLL, et al. Definitie voor de opzet van de studie Volksgezondheid Toekomst Verkenning 1997 (rapport nr 431501013). Bilthoven: RIVM, april 1995.
2. Barneveld Binkhuysen FH, Puijlaert CBAJ. Een ineffectief en inefficiënt woordgebruik. Medisch contact 1986; 41: 789-790.
3. Gezondheidsraad (GR). De proef op de som. Over onderzoek en evaluatie in de gezondheidszorg. Rapport no A88/1, GR: 's-Gravenhage, 1988.
4. Dale van. Groot woordenboek van de Nederlandse taal. Utrecht/Antwerpen: Van Dale Lexicografie, 1992 (twaalfde druk).
5. Drummond MF. Principles of economic appraisal in health care. Oxford: Oxford University Press, 1988.
6. Worth LA, Volberding PA. Clinical applications of antiretroviral therapy. In: PT Cohen, Sander MA, Volberding PA (ed.). The AIDS knowledge base: a textbook on HIV disease from the University of California, San Francisco, and the San Francisco General Hospital. Boston: Little Brown Comp., 1994.
7. Nuland SB. Doctors: The biography of medicine. New York: Vintage books, 1988.
8. Lang de R, Lubsen J. (Red.). Klinisch geneesmiddelenonderzoek. Utrecht: Bunge, 1987.
9. Brook RH, Lohr KN. Efficacy, effectiveness, variations and quality. Med Care 1985; 23: 710-722.
10. Pellegrino ED, Thomasma DC. A philosophical basis of medical practice. Oxford: Oxford University Press, 1981.
11. Russell I. Methodological challenges in assessing surgery. In: Scientific Basis of Health Services (abstract book). London, 1995: p31.
12. Lohr KN. Outcome measurement: concepts and questions. Inquiry 1988; 25: 37-50.
13. Williams M. Using outcome information to improve acute care: total hip replacement surgery. In: T Delamothe (ed.). Outcomes into clinical practice. London: BMJ Publishing Group, 1994.
14. Lamas GA, Pfeffer MA, Hamm P, Wertheimer J, Rouleau J, Braunwald E. Do the results of randomized clinical trials of cardiovasculair drugs influence medical practice? N Engl J Med 1992; 327: 241-247.
15. Banta D, Bos M. The relation between quantity and quality with coronary artery bypass (CABG) surgery. Health Policy 1991; 18: 1-10.
16. Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The medical outcome study. An application of methods for monitoring the results of medical care. JAMA 1989; 262: 925-930.
17. Volberding PA, Lagakos SW, Koch MA, Pettinelli C, Myers MW, Booth DK, et al. Zidovudine in asymptomatic human immunodeficiency virus infection. N Engl J Med 1990; 322: 941-9.
18. Bruggenkate ten K. Wolters' Woordenboek. Engels-Nederlands. Groningen: Wolters-Noordhoff, 1992 (twintigste druk, tweede oplage).
19. Cochrane AL. Effectiveness and efficiency. Random reflections on health care services. London: Burgess and Son Ltd, 1971.
20. Antczak-Bouckoms A. Quality and effectiveness issues related to oral health. Med Care 1995; 33 (Supplement): NS123-142.
21. Wulff HR. Principes van klinisch denken en handelen. Utrecht: Bohn, Scheltema & Holkema, 1980.
22. Ferguson JH. Research on the delivery of medical care using hospital firms: Foreword. Med Care 1991; 29; JS1.
23. Greenfield S. The state of outcome research: are we on target? N Engl J Med 1988; 320; 1142-1143.
24. Bonham GH. Health and environment: public health decisions based on evidence, psychology and politics. Environmetrics 1995; 6: 311-318.
25. Houwelingen van JC, Brand R, Hermans J, Keirse MJNC (red.). Meta-analyse. Leiden: Boerhave Commissie, 1995.
26. Olkin I. Statistical and theoretical considerations in meta-analysis. J Clin Epidemiol 1995; 48: 133-146.
27. Ruwaard D, Van den Berg Jeths A, Bol P, Gijzen R, Jansen J, Kramers PGN, et al. Ontwerp van de studie

- Volksgezondheid Toekomst Verkenning 1997 (rapport nr 431501014).  
Bilthoven: RIVM, december 1995.
28. Fischl MA, Richman DD, Grieco MH, Gottlieb MS, Volberding PA, Laskin OL, et al. The efficacy of azidothymidine in the treatment of patients with AIDS and AIDS related complex.  
N Engl J Med 1987; 317: 185-191.
  29. Scitovsky AA, Cline MW, Abrams DI. Effects of the use of AZT on the medical care costs of persons with AIDS in the first 12 months.  
J of AIDS 1990; 3: 904-912.
  30. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment.  
Br Med J 1995; 311: 1356-9.
  31. Diamond GA, Denton TA. Alternative perspectives on the biased foundations of medical technology assessment.  
Ann Intern Med 1993; 118: 455-464.
  32. Williamson JW, Hudson JI, Nevins MM. Principles of quality assurance and cost containment in health care.  
San Francisco/Washington/London: Jossey-Bass Publishers, 1982.
  33. Leidl RM. Some factors to consider when using the results of economic evaluation studies at the population level.  
Int J Technology Assessment in Health Care 1994 (10): 467-478.
  34. Schwartz D, Lellouch J. Explanatory and pragmatic attitudes in therapeutic trials.  
J Chron Dis 1967; 20: 637-648.
  35. Sommer A, Zeger SL. On estimating efficacy from clinical trials.  
Statistics in Medicine 1991; 10: 45-51.
  36. Simon G, Wagner E, Vonkorff M. Cost-effectiveness comparisons using "real world" randomized trials: the case of new antidepressant drugs.  
J Clin Epidemiol 1995; 48: 363-373.
  37. Gerbarg ZB, Horwitz RI. Resolving conflicting clinical trials: guidelines for meta-analysis.  
J Clin Epidemiol 1988; 41: 503-509.
  38. McPherson K. The best and the enemy of the good: randomised controlled trials, uncertainty, and assessing the role of patient choice in medical decision making.  
J Epidemiology Community Health 1994; 48: 6-15.
  39. Flood AB, et al. Does practice makes perfect I: The relation between hospital volume and outcome for selected diagnostic categories.  
Med Care 1984; 22: 98-114.
  40. Flood AB, et al. Does practice makes perfect II: The relation between volume and outcome and other hospital characteristics.  
Med Care 1984; 22: 115-125.
  41. Gorsky RD, Teutsch SM. Assessing the effectiveness of disease and injury prevention programs: costs and consequences.  
CDC: Morbidity and Mortality Weekly Report august 1995; 44 (RR-10): 1-12.
  42. Hoover DR. The effects of long term zidovudine therapy and *Pneumonia carinii* prophylaxis on HIV disease.  
Drugs 1995; 49: 20-36.
  43. Cohen PT, Sander MA, Volberding PA (ed.). The AIDS knowledge base: a textbook on HIV disease from the University of California, San Francisco, and the San Francisco General Hospital.  
Boston: Little Brown Comp., 1994.
  44. Ziekenfondsraad. Pharmacotherapeutisch Kompas 1994.  
Amstelveen: Ziekenfondsraad, 1993.
  45. Hamilton JD, Hartigan PM, Simberkoff MS, et al. A controlled trial of early versus late treatment with zidovudine in symptomatic human immunodeficiency virus infection: results of the Veterans Affairs Cooperative Study.  
N Engl J Med 1992; 326: 437-43.
  46. Everdingen JJE van, Lange JMA. Consensus vroegtijdige interventie met zidovudine.  
Ned Tijdschr Geneesk 1991; 135:1547-1549.
  47. Osmond DH. Classification and staging of HIV disease. In: PT Cohen, Sander MA, Volberding PA (ed.). The AIDS knowledge base: a textbook on HIV disease from the University of California, San Francisco, and the San Francisco General Hospital.  
Boston: Little Brown Comp., 1994.
  48. Harrison TR et al. (eds.). Principles of Internal Medicine (11th edition).  
New York: McGraw-Hill Book Company, 1987.
  49. Mitsuya H, Weinhold KJ, Furman PA, et al. 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro.  
Proc Natl Acad Sci USA 1985; 82: 7096-100.
  50. Fischl MA, Richman DD, Hansen N, Collier AC, Carey JT, Para MF, et al. The safety and efficacy of zidovudine (AZT) in the treatment of subject with mildly symptomatic human immunodeficiency virus type 1 (HIV) infection.

- Ann of Internal Med 1990; 112: 727-737.
51. Cooper DA, Gateli JM, Kroon S, Clumeck N, Millard J, Goebel FD, et al. Zidovudine in persons with asymptomatic HIV infection and CD4+ cell counts greater than 400 per cubic millimeter. *N Engl J Med* 1993; 329: 297-303.
  52. Aboulker JP, Swart AM. Preliminary analysis of the Concorde trial. *Lancet* 1993; 341: 889-890.
  53. Fischl MA, Parker CB, Pettinelli C, Wulfsohn M, Hirsch MS, Collier AC. A randomized controlled trial of a reduced daily dose of zidovudine in patients with the acquired immunodeficiency syndrome. *N Engl J Med* 1990; 323: 1009-14.
  54. Graham NMH, Zeger SL, Park LP, Vermund SH, Detels R, Rinaldo CR. The effects on survival of early treatment of human immunodeficiency virus infection. *N Engl J Med* 1992; 326: 1037-1042.
  55. Fischl MA, Richman DD, Causey DM, Grieco MH, Bryson Y, Mildvan D, et al. Prolonged zidovudine therapy in patients with AIDS and advanced AIDS-related complex. *JAMA* 1989; 262: 2405-2410.
  56. Swanson CE, Cooper DA. Factors influencing outcome of treatment with zidovudine of patients with AIDS in Australia. *AIDS* 1990; 4: 749-757.
  57. Vella S, Giuliano M, Pezzotti P, Agresti MG, Tomino C, Florida M, et al. Survival of zidovudine-treated patients with AIDS compared with that of contemporary untreated patients. *JAMA* 1992; 267: 1232-1236.
  58. Buira E, Gatell JM, Batalla J, Zamora L, Mallolas J, Aznar E, et al. Influence of treatment with zidovudine (ZDV) on the long-term survival of AIDS patients. *J of AIDS* 1992; 5: 737-742.
  59. Nationale Commissie AIDS Bestrijding (NCAB). Een kader voor de invulling van de centrumfunctie voor AIDS bij ziekenhuizen. Amsterdam: NCAB, 1991.
  60. Dijkgraaf MWG. Utilization of hospital resources and the costs related to HIV infection. Amsterdam: thesis publishers, 1995.
  61. Cook DJ, Sackett DL, Spitzer WO. Methodologic guidelines for systematic reviews of randomized control trials in health care from the Potsdam Consultation on Meta-analysis. *J Clin Epidemiol* 1995; 48: 167-171.
  62. Kleijnen J, Vet de HCW, Rinkel GJE, Keirse MJNC. De Cochrane Collaboration; systematische overzichten van kennis uit gerandomiseerd onderzoek. *Ned tijdschr Geneesk* 1995; 139: 1478-1482.
  63. Kienle G, Burkhardt R. Der Wirksamkeitsnachweis für Arzneimittel. Stuttgart: Urachhaus, 1983.
  64. Heithoff KA, Lohr KN (ed.). Effectiveness and outcome in health care. Washington: National Academy Press, 1990.
  65. Goldberg HI, Cummings MA (ed.). Conducting medical effectiveness research: A report from the inter PORT work groups. *Medical Care*, July 1994 Supplement (vol 32, nr 7).
  66. Long AF, Dixon P, Hall R, Carr-Hill RA, Sheldon TA. The outcomes agenda: contributions of the UK claring house on health outcomes. *Quality in Health Care* 1993; 2: 49-53.
  67. Ware JE. SF-36 Health Survey. Manual and Interpretation Guide. Boston: The Health Institute, 1993.
  68. Reinertsen JL. Outcomes management and continuous quality improvement: the compass and the rudder. *Quality Review Bulletin/January* 1993: 5-7.
  69. Letzel H. Best-evidence synthesis: an intelligent alternative to meta-analysis. *J Clin Epidemiol* 1995; 48: 19-21.
  70. Sackett DL. The relationship between scientific evidence and clinical practice. In: *Scientific Basis of Health Services* (abstract book). London, 1995: p7.
  71. Black N. Can observational data be used to determine the value of health care interventions? In: *Scientific Basis of Health Services* (abstract book). London, 1995: p14.
  72. Smith G. Can observational data be used to determine the value of health care interventions? In: *Scientific Basis of Health Services* (abstract book). London, 1995: p14.
  73. Cochrane AL. 1931-1971: A critical review with particular reference to the medical profession. In: *Medicines for the year 2000*. London: OHE, 1979.
  74. Teutsch SM. A framework for assessing the effectiveness of disease and injury prevention. *MMWR* 1992; 41: 1-12.
  75. Riggs JE. The nonenvironmental basis for rising mortality from Parkinson's disease.

- Arch Neurol 1993; 50: 653-656.
76. Bunker JP, Frazier HS, Mosteller F. Improving Health: Measuring Effects of Medical Care. *The Millbank Quarterly* 1994; 74: 225-258.
77. Yusuf S, Wittes J, Friedman L. Overview of results of randomized clinical trials in heart disease. *JAMA* 1988; 260: 2088-2093.
78. Cohen AF. *Klinische farmacologie. Tussen middel en kwaal (oratie)*. RU Leiden, 1995.
79. Elferink AJA. De rol van meta-analyse in de registratie procedure van geneesmiddelen. In: JC van Houwelingen, Brand R, Hermans J, Keirse MJNC (red.). *Meta-analyse*. Leiden: Boerhave Commissie, 1995.
80. Vandembroucke JP. De Cochrane Collaboration en 'evidence-based medicine'. *Ned Tijdschr Geneesk* 1995; 139: 1481.
81. Wulff HR, Pedersen SA, Rosenberg R. *Philosophy of Medicine*. Oxford: Blackwell Scientific Publications, 1986.
82. DerSimonian R, Charette LJ, McPeck B, Mosteller F. Reporting on methods in clinical trials. *N Engl J Med* 1982; 306: 1332-7.
83. Pettiti DB. *Meta-analysis, decision analysis, and cost-effectiveness analysis. Methods for quantitative synthesis in medicine*. Oxford: Oxford University Press, 1994.
84. Rümke ChrL. Betrouwbaarheidsintervallen. *Ned Tijdschr Geneesk* 1989; 133: 2013-2015.
85. Marcus SH, Grover PL, Revicki DA. The method of information synthesis and its use in the assessment of health care technology. *Int J Technology Assessment in Health Care* 1987; 3: 497-508.
86. Wortman PM, Yeaton WH. Using research synthesis in medical technology assessment. *Int J Technology Assessment in Health Care* 1987; 3: 509-522.
87. Sacks HS, Berrier J, Reitman D, et al. Meta-analyses of randomized controlled trials. *N Engl J Med* 1987; 316: 450-455.
88. Riet ter G, Bouter LM. Meta-analyse van therapeutische experimenten II. *T Soc Gezondheidsz* 1990; 68: 168-189.
89. L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Int Med* 1987;107: 224-233.
90. Lipsey MW, Wilson DB. The efficacy of psychological, educational, and behavioral treatment. *American Psychologist* 1993; 48: 1181-1209.
91. Veldhuyzen van Zanten SJO, Boers M. *Metanalyse: de kunst van het systematisch overzicht*. *Ned Tijdschr Geneesk* 1993;137: 1594-1599.
92. Cooper HM. *The integrative research review. A systematic approach*. Beverly Hills: Sage Publications, 1984.
93. Jenicek M. *Meta-analysis in medicine*. *J Clin Epidemiol* 1989; 42: 35-44.
94. Brand R, Kragt H. The importance of trends in the interpretation of an overall odds ratio in the meta analysis of clinical trials. In: JC van Houwelingen, Brand R, Hermans J, Keirse MJNC (red.). *Meta-analyse*. Leiden: Boerhave Commissie, 1995.
95. Finney DJ. A statistician looks at meta-analysis. *J Clin Epidemiol* 1995; 48: 87-103.
96. Bouter LM, ter Riet G. Meta-analyse van therapeutische experimenten I. *T Soc Gezondheidsz* 1990; 68: 179-185.
97. Lubsen J. Meta-analyse: een commentaar. *T Soc Gezondheidsz* 1990; 68: 190.
98. Haring HM, Kruit D. *Registraties met beleid*. Rijswijk: Ministerie van WVC, 1994.
99. Delamothe T (ed.). *Outcomes into clinical practice*. London: BMJ Publishing Group, 1994.
100. Ridsdale L. *Evidence-based general practice. A critical reader*. London: Saunders Comp Ltd, 1995.

## **Bijlagen**

### **A Gerandomiseerd interventie onderzoek**

1. Wat is gerandomiseerd interventie onderzoek?
2. Opzet en uitvoering
3. De kwaliteit van gerandomiseerd interventie onderzoek

### **B Meta-analyse van experimenteel onderzoek**

1. Wat is meta-analyse?
2. Opzet en uitvoering
3. De kwaliteit van meta-analyse
4. Formules voor de uitvoering van meta-analyse

### **C Uitkomst-onderzoek**

1. Wat is uitkomst-onderzoek?
2. Opzet en uitvoering
3. De kwaliteit van uitkomst-onderzoek





## **Bijlage A: Het gerandomiseerd interventie onderzoek**

*Gegevens over efficacy komen uit (publicaties van) experimenteel onderzoek: de randomized controlled clinical trial (RCCT). Deze bijlage geeft een beschrijving van de RCCT. Aan de hand van criteria voor de kwaliteit van een RCCT kan worden bepaald of het desbetreffende onderzoek geschikt materiaal vormt voor meta-analyse.*

### **1. Wat is gerandomiseerd interventie onderzoek?**

#### Inleiding

De oudste beschrijving van een vergelijkend experiment wordt in het boek Daniël (1: 1-16) van het Oude Testament aangetroffen. In Babylonische ballingschap namen Daniël en diens metgezellen peulvruchten en water tot zich, terwijl andere dienaren van Nebuchadnezzar het voedsel van de koning en diens wijn bleven gebruiken. Na tien dagen werd het effect van het dieet op de uiterlijk van de deelnemers in beide groepen vergeleken.

Natuurlijk is er vanuit een hedendaags perspectief veel kritiek mogelijk op het experiment van Daniël. De interventie (het dieet) werd niet op gelijke wijze aan de deelnemers toegewezen. Er bestond een verschil in samenstelling tussen de interventie en controlegroep (Joodse ballingen versus heterogene groep hovelingen). De personen in de interventiegroep waren mogelijk meer gemotiveerd om het dieet te volgen. De deelnemers noch de beoordelaars (eunuchen van de koning) waren neutraal ten opzichte van het resultaat van de interventie, hetgeen mogelijk hun waarneming en beoordeling ervan beïnvloedden. Tenslotte was de omvang van het experiment gering (vier personen in de interventiegroep), de observatieperiode was nogal kort, en de rapportage onvolledig (een niet genoemd aantal in de controlegroep) (77). Er is dus alle reden om de uitkomst van dit experiment te wantrouwen.

In de negentiende eeuw werd door de arts Louis en de statisticus Gavaret een formele proefopzet ontworpen om aan de genoemde bezwaren tegemoet te komen, de numerieke methode genaamd (21). Zij werden geïnspireerd door de canons van J.S.Mill en het positivisme van Comte.

Lange tijd bleef de door Louis en Gavaret ontworpen methode in onbruik. In de jaren zestig maakte de gezondheidszorg een explosieve groei door. Op grote schaal kwamen nieuwe medische technieken, geneesmiddelen en voorzieningen ter beschikking, waarvan (vaak op patho-fysiologische gronden) werd aangenomen dat deze een groter effect hadden dan al in gebruik zijnde medische interventies. Dit bracht echter de vraag met zich mee *in welke mate* deze nieuwe technieken en geneesmiddelen een groter effect hadden. De noodzaak tot het ontwikkelen van een betrouwbare methode om een antwoord op deze vraag te vinden nam toe. Optredende calamiteiten als de softenon-

affaire droegen hieraan bij. Het optreden van ernstige aangeboren afwijkingen (softenon) tengevolge van het gebruik van het slaapmiddel thalidomide tijdens de zwangerschap bleek niet voorspelbaar uit het testen van deze stof op proefdieren en een verder formalisering van het onderzoeksdesign te vereisen (63).

De numerieke (experimentele) methode van Louis en Gavaret werd vervolgens uitgewerkt in de vorm van het gerandomiseerd (klinisch) interventie/effectonderzoek (randomized controlled (clinical) trial, RC(C)T). Methodologische aspecten als onbevooroordeelde waarneming, adequate studieomvang en ad random toewijzing van de interventie aan de proefpersonen gingen de bouwstenen van dit soort effectonderzoek vormen. Zorgvuldige opzet en uitvoering van het experiment voorkomt vertekening en maakt een conclusie over de (causale) samenhang tussen interventie en waargenomen effect mogelijk.

### Definitie

Het gerandomiseerd klinisch interventie onderzoek is gedefinieerd als *het verrichten van waarnemingen over de aard en omvang van de werkzaamheid van een interventie op een gezonde of zieke persoon ter beoordeling van de bruikbaarheid als bevordering van de gezondheid of genezing van ziekte* (8).

### Doel

Het doel van gerandomiseerd (klinisch) interventie onderzoek is, (i) het vaststellen van een causaal verband tussen een medische interventie (oorzaak) en een waargenomen verandering in de gezondheidstoestand (gevolg), (ii) een of meerdere voorspellende factoren van het effect van een bepaalde medische interventie identificeren, (iii) een nauwkeuriger kwalificatie van een therapeutisch effect, (iv) nieuwe hypothesen genereren (69).

### Verschillende fasen

Het (klinisch) geneesmiddelonderzoek is een typisch voorbeeld van een gerandomiseerd interventie onderzoek. Het onderzoek naar het effect van een geneesmiddel neemt doorgaans een aanvang in het laboratorium. In het laboratorium worden met behulp van chemische technologie, op basis van farmacologische theorieën, stoffen gemaakt, waarvan een werking wordt vermoed. Een eerste toetsing vindt plaats aan de hand van proefdieren. Het principe van trial and error is in dit stadium vaak een leidraad van de onderzoeker. Hoewel het diemodel berust op een aangenomen verwantschap tussen mens en dier, is het gebruik van een geneesmiddel bij de mens niet volledig op dierproeven te grondvesten. Een geneesmiddel is soms bij het dier niet teratogeen, terwijl het dat bij de mens wel is (de les van de softenon affaire). Dierproeven worden gebruikt om experimenten uit te voeren die bij mensen (soms op ethische gronden) niet mogelijk zijn.

Het diermodel beoogt de *werking* van het geneesmiddel aan te tonen. Het gaat primair om de relatie tussen toediening van het middel en het fysiologische gevolg daarvan. Er wordt getracht een relatie tussen de dosis en het effect vast te stellen. De opzet van de dierproef is doorgaans eveneens vergelijkend, gerandomiseerd, placebo gecontroleerd. Als het dierexperiment een positief resultaat oplevert (dosis-effect respons) gaat men over tot het klinische gedeelte van het onderzoek; klinisch, omdat het onderzoek personen betreft.

Het klinische gedeelte van het onderzoek kan worden ingedeeld in vier fasen. De eerste toediening van een stof is aan gezonde vrijwilligers (*de eerste fase*). In zeer lage dosis die voorzichtig wordt verhoogd, vangt men aan totdat er sprake is van een farmacologisch effect of bijwerkingen optreden. Het doel is in de eerste plaats een indruk te verkrijgen van de toxiciteit van een stof en de farmacologische eigenschappen bij de mens (dosis-effect relatie). De *tweede fase* van het klinisch geneesmiddelenonderzoek is het toedienen van de stof aan een kleine groep van patiënten. In deze fase wordt bepaald in welke dosering het middel effectief kan zijn, zonder al te grote bijwerkingen op te leveren (onder het motto: 'baat het niet, dan schaadt het niet'). De *derde fase* is die van het grootschalig gerandomiseerd, dubbelblind geneesmiddelenonderzoek op basis waarvan een geneesmiddel geregistreerd kan worden en vervolgens op de markt wordt gebracht. In deze fase bestaat de onderzoekspopulatie uit de beoogde doelgroep van patiënten. De *vierde fase* treedt in nadat het middel is geregistreerd en op de markt is verschenen. Voor de geregistreerde indicatie wordt op bijwerkingen gescreend, terwijl er zich tevens nieuwe indicatiemogelijkheden kunnen voordoen (8).

De totale onderzoeksduur kan, afhankelijk van stof en indicatie, oplopen tot tien jaar. Van de tien nieuwe geneesmiddelen die voor fase I in aanmerking komen, komt slechts één uiteindelijk op de markt. Deze tien op één vuistregel gaat ook op voor het onderzoek voorafgaande aan fase I. Het startpunt is dus een honderdtal stoffen, waarvan uiteindelijk slechts één de markt haalt. Met deze ontwikkeling zijn uiteraard veel kosten gemoeid, die ook uiteindelijk in de prijs van het geneesmiddel hun weerslag vinden. Momenteel kost het ontwikkelen van een geneesmiddel ongeveer twaalf jaar en 500 miljoen gulden (78).

Een nieuw geneesmiddel mag in Nederland alleen op de markt worden gebracht als het de eerste drie stadia van het effect onderzoek met succes heeft doorlopen (79). Er wordt verlangd dat in redelijkheid is aangetoond dat het middel in een bepaalde dosis bij een aangegeven indicatie voldoende effectief en veilig is. Hoewel het gerandomiseerd interventie onderzoek is ontwikkeld aan de hand van onderzoek naar de werkzaamheid van een geneesmiddel, is de methode ook toepasbaar bij onderzoek naar de werkzaamheid van andere medische interventies, zoals chirurgische technieken of complete behandelprogramma's. Soms maakt dit uiteraard wel enige modificaties van de methode noodzakelijk (blinding of randomisering niet altijd mogelijk). De veiligheid van het middel wordt geborgd door

post-marketing surveillance, dat wil zeggen, het analyseren van meldingen van (mogelijke) bijwerkingen nadat het middel op de markt is verschenen.

### Gouden standaard

De RCCT is in de geneeskunde zo langzamerhand de gouden standaard voor het aantonen van effect van een medische interventie geworden en geniet in dezen de voorkeur boven case-controle onderzoek, cohort studies of casuïstiek (80). De RCCT is de vrucht van methodologische principes die ons in staat stellen onder nauwkeurig gecontroleerde/beheerste omstandigheden het effect van een medische interventie vast te stellen. Het is een vorm van vergelijkend onderzoek.

De RCCT berust op een aantal principes: (i) kennisverwerving is gebaseerd op een methode (i.t.t. casuïstiek); (ii) deze methode berust op een gestandaardiseerd vergelijken (onderzoeksprotocol); (iii) deze vergelijking beroept zich op de ervaring (gemanipuleerde interventiegroep); (iv) de evaluatie van de verworven kennis vindt plaats met behulp van een formeel systeem (de statistiek, berustende op bepaalde axiomata); (v) de wetenschappelijke bewijskracht van de RCCT is maximaal naarmate de vergelijking beter is uitgevoerd; (vi) het verschil in effect dat dan kan worden waargenomen, is nu met een grotere zekerheid toe te schrijven aan de therapeutische interventie.

De methode en opzet van de RCCT draagt bij aan een minimale (of in ieder geval niet-systematische) vertekening. De verschillen in uitgangssituatie tussen personen in de interventie en controlegroep zijn alleen te wijten aan toevallige variatie, die door randomisatie gelijkelijk is verdeeld over de interventie en de controlegroep. Bij andere vormen van onderzoek is selectieve waarneming en vertekening in de regel minder gemakkelijk uit te sluiten. Het resultaat is dan vaak al snel te ondergraven (69). Daarom is de RCCT als een gouden standaard voor het effect van een medische interventie te beschouwen.

## **2. Opzet en uitvoering**

*Deze paragraaf beschrijft kenmerken van de RCCT. Het moet een lezer van artikelen in staat stellen de kwaliteit van (de beschrijving van) een RCCT vast te stellen.*

De principes van het gerandomiseerd klinisch interventie onderzoek zijn al in de vorige eeuw door Gavaret geformuleerd: (i) keuze van de patiënten (definiëren insluitcriteria), (ii) samenstellen van interventie- en controlegroep, (iii) het minimaliseren van bevoordeling bij de waarneming van het effect en het toepassen van de behandeling (dubbelblinde proefopzet), (iv) het experimentele schema (kruisexperiment of niet; periode voor uitwerking eventuele andere, storende invloeden van

geneesmiddelgebruik), (v) de beoordeling van de behandelingsresultaten (statistische significantie en klinische relevantie), (vi) statistische evaluatie (generaliseerbaarheid, veralgemeniseerbaarheid) (81).

#### Keuze van de patiënten

Om de uiteindelijke toepasbaarheid van het effect van een trial in een populatie te beoordelen, is het nodig te weten welke personen er aan de trial deelnamen en hoe zij voor deelname zijn geselecteerd (insluitcriteria). Als de insluitcriteria niet helder en eenduidig zijn vermeld, is het onduidelijk of de behandeling toepasbaar is op personen die niet aan de trial deelnamen, met andere woorden, de resultaten van de trial zijn dan niet generaliseerbaar en niet interpreteerbaar voor de clinicus die overweegt de desbetreffende behandeling in de praktijk toe te passen. De beschrijving van de insluitcriteria moet derhalve zodanig zijn dat de lezer van het onderzoeksverslag in staat zou zijn een voor de behandeling in aanmerking komende persoon te herkennen, dat wil zeggen, een specificatie naar leeftijd, geslacht, klinische verschijnselen en -parameters.

De insluitcriteria zijn voldoende expliciet als een eensluitend antwoord kan worden gevonden op de volgende vragen. Om welke ziekte of ziekten gaat het en in welk stadium? Hoe wordt de diagnose gesteld? Is de diagnose valide? Is er bij de proefpersonen tevens sprake van andere ziekten die het beloop kunnen beïnvloeden? Welke contra-indicaties worden genoemd? Welke behandelingen zijn al eerder geprobeerd en met welk resultaat?

#### Het samenstellen van groepen

(a) *Aselect*. Bij onderzoek van een populatie beperkt men zich noodzakelijkerwijs vaak tot een deel ervan, bijvoorbeeld, omdat het waarnemen van alle elementen een tijdrovende of kostbare aangelegenheid is, of de waarneming destructief van aard is. Het gedeelte dat uit de populatie wordt genomen, heet de steekproef. Een steekproef moet op een bepaalde manier tot stand komen om conclusies met betrekking tot de gehele populatie toe te laten. Alle elementen van de populatie moeten in principe eenzelfde kans hebben om in de steekproef te worden opgenomen. Indien een steekproef aan deze voorwaarde voldoet spreekt men van een aselechte steekproef. Als het gaat om gerandomiseerd klinisch interventie onderzoek bestaat een steekproef uit een groep van personen uit een populatie die in aanmerking komen voor een behandeling (beoogde doelgroep van patiënten).

(b) *Randomisatie*. De personen die voor het onderzoek in aanmerking komen, worden in verschillende groepen verdeeld en aan verschillende interventies onderworpen. De verdeling van personen in beide groepen geschiedt 'ad random', dat wil zeggen, op basis van het toeval. De kans om in een van beide groepen te geraken voor iedere persoon die voor een medische interventie in aanmerking komt dezelfde.

Randomisatie is een methode om verschillende bronnen van vertekening uit te schakelen (82). Zo is gebleken dat het effect van een medische interventie doorgaans groter is in niet-gerandomiseerd onderzoek dan in RCCT (83). Door middel van randomisatie wordt de variatie in natuurlijk beloop en de invloed van onbekende externe factoren op het waargenomen effect gelijkwaardig over de interventie en de placebogroep verdeeld. Dit is een voorwaarde om beide groepen zo vergelijkbaar mogelijk te doen zijn. Het effect van de interventie tekent zich af in contrast met het natuurlijk ziektebeloop. Een ziekte of aandoening kan immers vanzelf overgaan. De controle groep, die van de interventie verschoond blijft, toont het 'natuurlijk ziektebeloop'. (Hieronder kan zowel het onbehandelde, als het met behulp van een al bekend middel behandelde verloop van een aandoening worden verstaan).

Het moment waarop randomisatie plaatsvindt markeert het beginpunt van de waarnemingsperiode. Voor randomisatie staan verschillende technieken ter beschikking. Deze variëren van het opgooien van een munt tot het gebruik van tabellen. Het laatste is het meest betrouwbaar. Een persoon krijgt bijvoorbeeld op volgorde van binnenkomst een nummer toegekend en wordt vervolgens op basis van randomisatietabellen in de interventie of controlegroep ingedeeld. Het is daarbij verstandig dat de persoon die de randomisatie verricht niet dezelfde is als de persoon die de patiënt voor zich ziet. Dit alles om bevooroordeelde toedeling aan groepen uit te sluiten.

*Samenvattend.* Randomisatie leidt tot: (i) neutralisatie van mogelijke bronnen van toevallige variaties in een behandelingscontrast; (ii) statistische onafhankelijkheid (van respons) van verschillende onderzoekselementen; (iii) betrouwbare vaststelling van de precisie (reproduceerbaarheid) van een behandelingscontrast op grond van respons in het experiment zelf; (iv) markeert de aanvang van het onderzoek.

#### De waarneming van het therapeutisch effect

Het waarneembare effect van een interventie (WE) wordt opgebouwd gedacht uit de volgende componenten:

- (a) het therapiegebonden effect (TGE). Het effect dat werkelijk aan de interventie kan worden toegeschreven (berustende op werking);
- (b) het natuurlijke beloop (NB). De reflectie van het feit dat ziekten, afwijkingen of klachten op onvoorspelbare wijze ook vanzelf kunnen overgaan, verminderen of verergeren;
- (c) waarnemingsfouten (WF);
- (d) het effect van externe invloeden (EI) die samenhangen met het innemen van een geneesmiddel (zoals placebo effecten, maar ook beïnvloeding door leefgewoonten, ect.) of het toepassen van een therapie.
- (e) interactie (IT). Het therapiegebonden effect kan bijvoorbeeld samenhangen met de ernst van het ziektestadium waarin dit wordt toegepast. Dit betekent dat niet voor iedere patiënt met een bepaalde

ziekte of aandoening hetzelfde effect kan worden verwacht. In het model wordt interactie tot uiting gebracht door middel van een produktterm.

(f) random error (ET). Het optreden van toevallige fouten. Aan het model wordt een error term (ET) toegevoegd.

In formule:  $WE = TGE + NB + WF + EI + IT + ET$

Deze formule geldt zowel voor de interventie (WE1), als de controle groep (WE0) (uiteraard voor de laatste op TGE na). In formule:

$WE1 = TGE1 + NB1 + WF1 + EI1 + IT1 + ET1$ ;  $WE0 = NB0 + WF0 + EI0 + IT0 + ET0$

Hieruit volgt:  $TGE = WE1 - WE0$ , indien alle overige componenten tegen elkaar wegvallen (8). De opzet van het gerandomiseerd interventie onderzoek is erop gericht, dat dit laatste het geval zal zijn. Hiervoor is beheersing van de context nodig en een gerandomiseerde toewijzing van een persoon aan de interventie of controlegroep (zie onderdeel 1 en 2).

Om ongelijkheden in waarnemingsfouten als oorzaak van verstoring uit te sluiten moet ervoor worden gezorgd dat de waarnemingsfout in de interventiegroep en in de controlegroep aan elkaar gelijk zijn. Hiertoe dient ten eerste de wijze waarop de waarnemingen in beide groepen worden verricht gelijk te zijn, zonder beïnvloeding van de behandeling, en ten tweede de tijdsperiode gedurende welke waarnemingen en metingen worden uitgevoerd zodanig omschreven te zijn dat er geen vertekening kan worden geïntroduceerd door een invloed van het ziektebeloop op de waarnemingsperiode. Dit zijn voorwaarden voor de zogenaamde *interne validiteit*: betreft het onderzoek ook daadwerkelijk dat wat het beoogt te onderzoeken. Gelijkwaardigheid van uitvoering van waarnemingen wordt bereikt door (a) zowel de betrokkenen waarnemers als de patiënt onkundig te laten van de gebruikte behandeling (*dubbelblinde proefopzet*), (b) het hanteren van een onderzoeksprotocol, volgens welke de waarnemingen en/of metingen op gestandaardiseerde wijze worden verricht.

*Dubbelblinde proefopzet*. Er is ontdekt dat ook (het geven van) een stof zonder de beoogde werking een bepaald effect op de patiënt kan sorteren dat sterk lijkt op het effect van de te onderzoeken stof. Een dergelijk effect wordt *het placebo effect* genoemd. Aangezien men in een RCCT het waargenomen effect aan de stof in het geneesmiddel wil toeschrijven, zal dus ook met het placebo-effect rekening moeten worden gehouden. De controle groep krijgt daarom wel eens een niet werkzaam middel (placebo) toegediend. De Duitse term voor deze proefopzet is '*überlegenheitsnachweis*' (63). De bewijsvoering is relatief. Het effect van een (nieuw) middel wordt gemeten ten opzichte van het effect van een ander middel. Een vergelijking tussen een middel en natuurlijk beloop verschilt in dezen van een vergelijking tussen een middel en een placebo, of tussen een nieuw middel en een gangbaar middel.

Het placebo effect wordt verklaard door te wijzen op het therapeutisch effect van de arts-patiënt relatie als zodanig. Vaak is de komst van de dokter al voldoende om de klacht te doen verdwijnen. Het placebo effect wordt veroorzaakt doordat er bij de patiënt een verlangen bestaat beter te worden, dat wil zeggen, een effect van de behandeling waar te nemen, of er bij de arts een verwachting bestaat dat de behandeling effect zal ressembleren. Deze verwachtingen ten aanzien van het waar te nemen effect beïnvloeden de observatie ervan (82). Het placebo effect verstoort de onbevooroordeelde waarneming en maakt een proefopzet noodzakelijk waarin zowel de arts, als de patiënt onkundig blijven van het feit in welke groep de proefpersoon is onderverdeeld, dat wil zeggen het werkzame middel of het placebo krijgt toegediend. Een dergelijke proefopzet wordt dubbelblind genoemd.

#### Het experimentele schema

Wanneer het effect van een therapeutische handeling in de interventiegroep is waargenomen, moet dit in principe overdraagbaar zijn op de controlegroep (veralgemeeniseerbaar). Na een bepaalde tijd worden interventie- en controlegroep soms dan ook omgedraaid. Men spreekt van een kruisexperiment. Een dergelijk opzet heeft voor- en nadelen. Het tijdstip waarop de therapeutische behandeling is uitgewerkt in de interventiegroep, is moeilijk vast te stellen. Daardoor kan deze eigenlijk niet goed meer als controlegroep optreden. Bovendien loopt de blindering het gevaar weg te vallen. Het voordeel is een betere validering van het causaal verband tussen medische interventie en waargenomen verandering in de gezondheidstoestand.

#### Statistische significantie en klinische relevantie

Wanneer met behulp van een gerandomiseerd effect onderzoek de werkzaamheid van een geneesmiddel wordt aangetoond, zegt dit nog niets over de klinische relevantie van het test resultaat. Het kan gaan om een klein, significant effect voor een grote groep personen, of om een groot effect voor een kleine groep personen. Er zal tot een afweging moeten worden gekomen op populatieniveau door de indicatiestelling in ogenschouw te nemen, en vervolgens het effect te wegen naar de potentieel te behandelen populatie. Daarbij moeten ook bij- en nevenwerkingen in ogenschouw worden genomen. Rapportage van uitval is daarbij van belang. Vaak wordt deze namelijk veroorzaakt door bij- of nevenwerkingen, die dan als zodanig niet in het onderzoek worden gezien of vermeld.

Om de klinische relevantie van het waargenomen effect van de interventie op de proefpersonen te verhogen wordt het zogenaamde 'intention to treat' principe gehanteerd. Hieronder wordt verstaan, dat alle relevante ziektegebeurtenissen die optreden tussen het begin en einde van de waarnemingsperiode, worden meegeteld in de groep van *de oorspronkelijke medicatietoewijzing*,



terwijl het einde van de observatieperiode zodanig wordt gekozen dat de interne validiteit is gewaarborgd (8).

In dit verband onderscheidde Schwartz en Lellouch in 1967 al een trial met een verklarende, en een trial met een pragmatische opzet: *'with the explanatory approach, we compare strictly defined treatments on a relative arbitrary class of patients; with the pragmatic approach, loosely defined treatments are compared on patients drawn from a predetermined class, viz. those to which the conclusions of the trial are to be extrapolated. We may say that in the first case the class of patient is defined to fit the predetermined treatments, while in the second the treatments are defined to fit the predetermined class of patients.'* (34)

In een trial met een verklarende opzet wordt het effect tussen interventie en controlegroep vergeleken aan de hand van de patiënten die de therapie feitelijk hebben ondergaan, dat wil zeggen, exclusief uitval. Dit kan betekenen dat de indicatiestelling achteraf moet worden bijgesteld. In een trial met een pragmatische opzet wordt het resultaat gebaseerd op vergelijking van de interventie en controlegroep in hun oorspronkelijke samenstelling (intention to treat) (34).

### Statistische evaluatie

Het sluitstuk van het gerandomiseerd interventie onderzoek wordt gevormd door de statistische analyse. Er moet tot een afweging aangaande de interne en externe validiteit van het onderzoek worden gekomen. Interne validiteit is daarbij een voorwaarde voor *externe validiteit*. De conclusies die op basis van het onderzoek voor de onderzochte populatie geldig mogen worden geacht (interne validiteit), moeten kunnen worden gegeneraliseerd naar personen die niet tot de onderzochte populatie behoren (externe validiteit).

Een klinisch geneesmiddelenonderzoek is extern valide indien de feitelijk onderzochte patiëntengroep in biologische zin representatief is voor een klinisch relevante en herkenbare groep patiënten (8). Op deze manier kan een indicatiestelling worden vastgesteld, ofwel er kan nu op basis van de onderzoeksresultaten criteria worden opgesteld voor het voorschrijven van het onderzochte middel in de alledaagse praktijk. Hier ligt een moment van oordelen. De conclusies worden gegeneraliseerd. Statistische toetsing is een methode om dit oordeel te valideren.

Een statistische toets is een formeel beslissingsschema om aan de hand van de uitkomsten van het onderzoek een kwalitatieve uitspraak te doen over de aanwezigheid van een behandelingseffect. Een statistische toets is een besluitregel die aan de uitkomsten van een steekproef het al of niet verwerpen van een hypothese verbindt. Hiertoe wordt een nulhypothese en een alternatieve hypothese onderscheiden. De nulhypothese is in de situatie van het gerandomiseerd interventie onderzoek

doorgaans de veronderstelling dat het waargenomen effect in de interventie- en de controlegroep niet van elkaar verschilt.

De *p-waarde* is een universele toetsingsgrootheid voor alle statistische toetsen. Van deze toetsingsgrootheid wordt de kansverdeling bepaald onder de vooronderstelling dat de nulhypothese juist is. De wijze waarop de *p-waarde* wordt berekend is afhankelijk van de onderzoeksopzet, het meetniveau van de variabelen en de vooronderstelde verdeling van de meetwaarden. De wijze waarop de *p-waarde* in het beslissingsschema van een statistische toets wordt gehanteerd, is onafhankelijk van de onderzoeksopzet.

Men kiest een waarde  $\alpha$ , die de maximale kans weergeeft, dat de nulhypothese ten onrechte wordt verworpen (fout-positief resultaat). De keuze van  $\alpha$  wordt bepaald door de ernst van de consequenties van een dergelijke fout. Doorgaans kiest men voor  $\alpha$  waarden in de orde van 10%, 5%, of 1%. De waarde van  $\alpha$  markeert een kritieke zone. Valt de waarde van de toetsingsgrootheid in de kritieke zone, dan wordt de nulhypothese verworpen. Een onderzoeksresultaat heet statistisch significant als op basis van de zogenaamde *p-waarde* de nulhypothese (dat er geen effect is) kan worden verworpen. Verwerping van de nulhypothese leidt tot de uitspraak, *dat* er een behandelingseffect is, hetzij in positieve, hetzij in negatieve zin, van de interventie- ten opzichte van de controlegroep.

De *p-waarde* geeft in feite aan in hoeverre het gevonden resultaat al of niet aan het toeval is toe te schrijven. Daarbij is het niet noodzakelijk dat het gevonden effect ook daadwerkelijk aan de interventie kan worden toegeschreven. De proefopzet en het verloop van het onderzoek moeten daartoe ook in de beschouwing worden opgenomen. Iedere vorm van vertekening of selectie moet worden overwogen. Pas dan is de *p-waarde* als maat voor de interne validiteit te beschouwen.

Het *betrouwbaarheidsinterval* geeft de grenzen aan waarbinnen, gegeven een bepaalde kans, de meetwaarde bij herhaling van de steekproef uit eenzelfde populatie kan worden gevonden. Het betrouwbaarheidsinterval is als een maat voor de externe validiteit te beschouwen. Het betrouwbaarheidsinterval wordt meestal benoemd naar de kans. Zo spreekt men van een 95% betrouwbaarheidsinterval als de kans 95% bedraagt dat de meetwaarde bij herhaling van de proefopzet binnen de berekende grenzen zal komen te liggen. Door het vermelden van een interval waarbinnen bij herhaling de gevonden waarde wordt gevonden geeft het betrouwbaarheidsinterval feitelijk meer informatie dan de *p-waarde*. Bovendien is er een innige samenhang tussen *p-waarde* en betrouwbaarheidsinterval. Als het gaat om een verschil tussen de interventie en controlegroep, en het betrouwbaarheidsinterval blijkt de waarde 0 niet te bevatten, is het resultaat ook significant. In de rapportage van klinische onderzoek gaat de voorkeur uit naar het vermelden van het betrouwbaarheidsinterval (84).

Er zijn verschillende manieren om het betrouwbaarheidsinterval te berekenen. Het betrouwbaarheidsinterval hangt af van de veronderstelde verdeling van meetwaarden in de populatie waaruit de steekproef afkomstig is. Wanneer deze verdeling onbekend is, wordt gelet op het waargenomen absolute aantal ( $n > 50$ ) en de (extremiteit van de) gevonden relatieve meetwaarde ( $0,1 < mw < 0,9$ ). Als het waargenomen aantal ( $n$ ) groter is dan 50 en de gevonden meetwaarde ( $p = mw/100$ ) binnen de grenzen van 0,1-0,9 is gelegen, wordt de normaalverdeling verondersteld en voor de berekening van het 95%-betrouwbaarheidsinterval de formule toegepast:

$p \pm 1,96 \sqrt{p(1-p)/n}$ . Het betrouwbaarheidsinterval voor een verschil tussen interventie (treatment) groep en controle (c) groep wordt berekend met behulp van de formule:

$$(p_t - p_c) \pm 1,96 \sqrt{(p_t(1-p_t)/n_t) + (p_c(1-p_c)/n_c)}$$

Als het waargenomen aantal gevallen ( $n$ ) groter is dan 50 en de gevonden meetwaarde ( $p = mw/100$ ) extreem ( $p < 0,1$  of  $p > 0,9$ ), dan wordt eveneens een normaalverdeling verondersteld. Het 95%-betrouwbaarheidsinterval wordt dan echter berekend met de formule:  $p \pm 1,96 \sqrt{p/n}$ . De formule voor het betrouwbaarheidsinterval van het verschil wordt overeenkomstig aangepast.

Als het waargenomen aantal gevallen kleiner is dan 50 dan moet worden aangenomen dat de meetwaarde Poisson is verdeeld. Er wordt gebruik gemaakt van (een gecompliceerdere formule en dus van) tabellen.

Het betrouwbaarheidsinterval speelt een belangrijke rol bij het beoordelen van de klinische relevantie van het onderzoeksresultaat. Er kan nu immers worden vastgesteld in welke orde van grootte bij herhaling van de interventie de winst voor de patiënten zal zijn.

Onder *de power van een onderzoek* wordt de omvang van de steekproef verstaan die nodig is om het minimaal waarneembaar verschil tussen de te vergelijken medische interventies vast te stellen. Voor het berekenen van de power is een inschatting nodig van:

(a) de klinische relevantie van de behandeling (welk verschil in effect is nodig om een nieuwe behandeling boven de oude te prefereren?),

(b) de kans dat de nieuwe behandeling ten onrechte boven de oude wordt verkozen (alfa), fout-positief resultaat.

(c) de kans dat de nieuwe behandeling ten onrechte niet boven de oude wordt verkozen (beta), fout-negatief resultaat.

In een statistische benadering worden twee soorten fouten onderscheiden. Een fout van de eerste soort (alfa) is gedefinieerd als het onterecht verwerpen van de nulhypothese indien deze stelt dat er geen

verschil is tussen twee interventies. Een fout van de tweede soort (beta) is gedefinieerd als het concluderen tot geen verschil tussen twee medische interventies, terwijl dit feitelijk niet het geval is<sup>1</sup>.

De fout van de eerste soort hangt samen met het verwachte en geobserveerde verschil tussen twee medische interventies. Naarmate dit verschil groter is zal de kans op een fout van de eerste soort kleiner zijn. De fout van de tweede soort hangt samen met de steekproefomvang. Als de steekproefomvang toeneemt, neemt een kans op een fout van de tweede soort af.

Het niveau van alfa en beta kan op voorhand worden gekozen. Vervolgens kan de benodigde steekproefomvang worden berekend met behulp van de formule:

$$n \geq [\epsilon_{\alpha} + \epsilon_{2(\beta+\gamma)}]^2 \cdot \frac{2\sigma^2}{\delta^2}$$

De epsilons in deze formule staan voor eenheden standaardafwijking in een normale verdeling corresponderende met waarschijnlijkheden van alfa en 2(beta+gamma). (Bijvoorbeeld. Epsilon 0,05 = 1,96.) Sigma is de veronderstelde standaardafwijking in de beide groepen en delta het verwachte verschil in het gemiddelde van de effectmaat in de beide groepen. Gamma wordt meestal verwaarloosbaar klein geacht.

De power van het onderzoek geeft de relevantie van (met name een negatief) resultaat aan. Onderzoek met een grote power bezit in deze meer bewijskracht. Anders is herhaling van het experiment aangewezen.

### 3. De kwaliteit van gerandomiseerd interventie onderzoek

De beschreven kenmerken van opzet en uitvoering van het gerandomiseerd interventie onderzoek kunnen nu worden gebruikt om de validiteit van de informatie die een RCCT oplevert, te wegen. Om te komen tot een schatting van de kwaliteit van gepubliceerd interventie onderzoek is het nodig de kenmerken te kwantificeren. Daarbij hebben we gekozen voor de meest gedetailleerde indeling die we aantreffen (*tabel 8*). Het lijstje geeft een mogelijkheid de kwaliteit van een studie te schatten en zodoende verschillende studies met elkaar te vergelijken. Het wegen van een RCCT kan aanleiding geven tot het al of niet opnemen van een trial in meta-analyse, of het stratificeren van de onderzoeken naar kwaliteitsscore en analyse per stratum (*zie bijlage B*).

Daarnaast is het wegen van een RCCT naar kwaliteit voor een beschouwing van de discrepantie tussen efficacy en effectiviteit van groot belang. Het stelt ons in staat de onderbouwing van

---

<sup>1</sup> Soms wordt er nog een fout van de derde soort onderscheiden (gamma). Deze ontstaat wanneer wordt geconcludeerd dat interventie A beter is dan interventie B, terwijl feitelijk het omgekeerde het geval is.

een causale relatie tussen medische interventie en waargenomen verandering van de gezondheidstoestand op waarde te schatten.

Met behulp van de *tabel 8* kan per trial een score worden berekend die is gelegen tussen 0 en 100. Deze score kan worden gebruikt om bepaalde trials geheel van meta-analyse uit te sluiten of om een trial naar gelang de score gewogen op te nemen in een gestratificeerde meta-analyse (*bijlage B*) (83).

Tabel 8. Het beoordelen van een RCCT (83)

Studie design		
adequate beschrijving van selectie van proefpersonen (protocol +/- info consent +/-)	3	
beschrijving van onderzochte proefpersonen	3	
uitval beschreven, aantal en reden	3	
interventie gedefinieerd	3	
placebo was niet te onderscheiden van onderzochte stof	1,5	
smaak placebo en onderzochte stof identiek	1,5	
randomisatie geblindeerd	10	
patiënten onkundig van aard behandeling (placebo of interventie)	8	
arts onkundig van aard behandeling (placebo of interventie)	8	
artsen en patiënt geblindeerd ten op zichte van klinisch effect	4	
aantal personen in trial vooraf vastgesteld	3	
adequaatheid randomisatie geëvalueerd (analyse base-line karakteristieken)	3	
adequaatheid blinding geëvalueerd	3	
compliance vastgesteld	3	
biologisch effect gemeten van de actieve stof	3	
Totaal		60
Analyse		
test statistiek en overschrijdingskansen vermeld	3	
indien geen effect gevonden, statistische power vermeld	3	
betrouwbaarheidsinterval gegeven	2	
life-table of tijdreeks-analyse verricht	2	
indien geïndiceerd, regressie analyse verricht	2	
statistische analyse adequaat gegeven de studieopzet	4	
intention to treat principe gehanteerd	4	
bijwerkingen beschreven en (statistisch geanalyseerd)	3	
analyse van subgroepen	2	
statisticus geblindeerd ten opzichte van interventie	2	
meerdere malen gekeken naar preliminaire resultaten	3	
Totaal		30
Presentatie		
startdatum en einddatum gegeven	2	
analyse vergelijkbaarheid groepen bij aanvang studie	2	
alle gebeurtenissen gedefinieerd als eindpunt in de analyse gebruikt	2	
overlevingscurven geconstrueerd, gegevens voor constructie vermeld	4	
Totaal		10
Totaal		100

## **Bijlage B: Meta-analyse van experimenteel onderzoek**

*Een meta-analyse geeft de mogelijkheid tot een schatting van de effectgrootte te komen indien meerdere RCCT's over eenzelfde onderwerp aanwezig zijn. Deze bijlage geeft een overzicht van meta-analyse van experimenteel onderzoek.*

### **1. Wat is meta-analyse?**

#### Inleiding

De meest valide en betrouwbare gegevens over het effect van een medische interventie worden aangetroffen in experimenteel onderzoek (zie bijlage A). In een wetenschappelijk discours waarin het gaat om onderzoeksresultaten van anderen te falsifiëren, verifiëren of te specificeren, worden vaak meerdere vergelijkbare onderzoeken naar eenzelfde medische interventie aangetroffen. Dergelijke onderzoeken hebben zelden een gelijklopend resultaat. Uiteenlopende uitkomsten worden toegeschreven aan toevallige verschillen, verschillen in studieopzet, -uitvoering of omstandigheden, en verschillen in onderzoekspopulatie. Strijdigheden ontstaan onder meer wanneer de interpretatie van de uitkomsten verschilt, of de statistische analyse geen uitsluitsel geeft over de significantie van het onderzoeksresultaat. Er is daarom gezocht naar een methode om op systematische wijze de verschillende onderzoeksresultaten te integreren (37). De informatiesynthese of de meta-analyse biedt hiertoe een mogelijkheid (85) (86) (87) (88).

In een meta-analyse verzamelt men het relevante materiaal, selecteert en ordent dit op basis van inclusie en exclusie criteria, vat de gegevens samen, en verricht een statistische analyse van de gegevens (83). Een meta-analyse neemt in beschouwing: grootte van het effect van verschillende studies, omvang van studies, kwaliteit van de studie, mate van vertekening, en de waarde van de studieopzet. Meta-analyse is bijvoorbeeld toegepast om vooruitgang in chirurgie of anesthesie te bepalen, de effectiviteit van antibiotica profylaxe bij colonchirurgie te onderzoeken, of een vergelijking te maken van het geven van prednison met immunosuppressiva en zonder immunosuppressiva bij een lupus nephritis (89). Ook op het gebied van psychische ziekten is meta-analyse mogelijk, zoals de studie van Lipsey & Wilson laat zien (90).

Systematiek is de grote kracht van een meta-analyse. Een informele review berust teveel op het subjectieve oordeel of anekdotisch bewijsmateriaal. In veel overzichtsartikelen en redactionele commentaren blijft onduidelijk hoe het besproken materiaal is geselecteerd en beoordeeld (91). Meta-analyse tracht hierin verandering aan te brengen. Een protocollaire aanpak is vereist om de

reproduceerbaarheid van de meta-analyse te garanderen en een eventuele bias zo gering mogelijk te doen zijn.

### Definitie

Een meta-analyse is een kwantitatieve beschouwing van (gepubliceerd) onderzoek. De 'National Library of Science' definieert meta-analyse als: '*A quantitative method of combining the results of independent studies (usually drawn from the published literature) and synthesizing summaries and conclusions which may be used to evaluate therapeutic effectiveness, plan new studies, etc., with application chiefly in the areas of research and medicine*' (25).

### Doel

Het doel van meta-analyse is: (i) de statistische power voor betrouwbare bepaling van effecten van medische interventies te doen toenemen; (ii) onzekerheid over strijdige resultaten van verschillende onderzoeken weg te nemen; (iii) effectschattingen te verbeteren; (iv) vragen te beantwoorden die op basis van een individueel onderzoek alleen niet kunnen worden beantwoord (87) (92). Daar kan nog aan worden toegevoegd dat een meta-analyse soms nieuwe hypothesen genereert op basis van het (sterk) afwijken van sommige onderzoeksresultaten (secundaire analyse van het bijeengebrachte materiaal).

## **2. Opzet en uitvoering**

Een meta-analyse kan worden uitgevoerd als de bijeengebrachte studies op een betekenisvolle wijze kunnen worden samengevoegd (25). Belangrijk is dat daarbij de onderzoeken min of meer dezelfde onderzoeksvraag hebben, dezelfde uitkomstmaten hanteren, en voldoende gegevens bevatten (34). De gedachte achter de meta-analyse is het opvatten van (veelal gepubliceerde) onderzoeksverslagen als onderzoeksobjecten. Vervolgens worden dezelfde stappen doorlopen als bij een gewoon onderzoek, tot en met de statistische verwerking toe. Jenicek geeft deze schematisch weer (*zie diagram 4*) (93). Hij onderscheidt twee fasen: een kwalitatief gedeelte en een kwantitatief gedeelte. Belangrijke stappen in een meta-analyse zijn:

- (i) het opzoeken en beoordelen van de artikelen,
- (ii) het selecteren van de gevonden artikelen (RCCT's: *zie bijlage A*),
- (iii) specificatie van het eindpunt (respons, ziektevrije overleving, overleving),
- (iv) definitie van het eindpunt voor meta-analyse<sup>2</sup>,

---

<sup>2</sup> Een onderzoek rapporteert vaak meerdere schattingen van het gekozen klinische eindpunt. Voor meta-analyse moet dan een keuze worden gemaakt. De volgende richtlijnen kunnen worden gehanteerd. (1) Kies de schatting gecorrigeerd voor leeftijd (minstens) en een verzameling variabelen waarvan bekend is dat deze als confounders kunnen worden aangemerkt.



- (v) definitie van de studiekarakteristieken, zoals (omvang studie: aantal patiënten, jaar van onderzoek, land van onderzoek, populatiekarakteristieken, kwaliteit onderzoek, duur follow-up),
- (vi) aanmaken data-base met gegevens uit de verschillende studies.
- (vii) statistische analyse van de gegevens.

Recentelijk is door experts een protocol voor de uitvoering van meta-analyse van experimenteel onderzoek opgesteld. Het is gepubliceerd in een themanummer van de '*Journal of Clinical Epidemiology*' van januari 1995 (61). Een protocol voor de uitvoering van een meta-analyse observationeel onderzoek wordt in het voortzicht gesteld. We bespreken nu een aantal vooronderstellingen en statistische achtergronden van de meta-analyse.

#### Heterogeniteit of homogeniteit

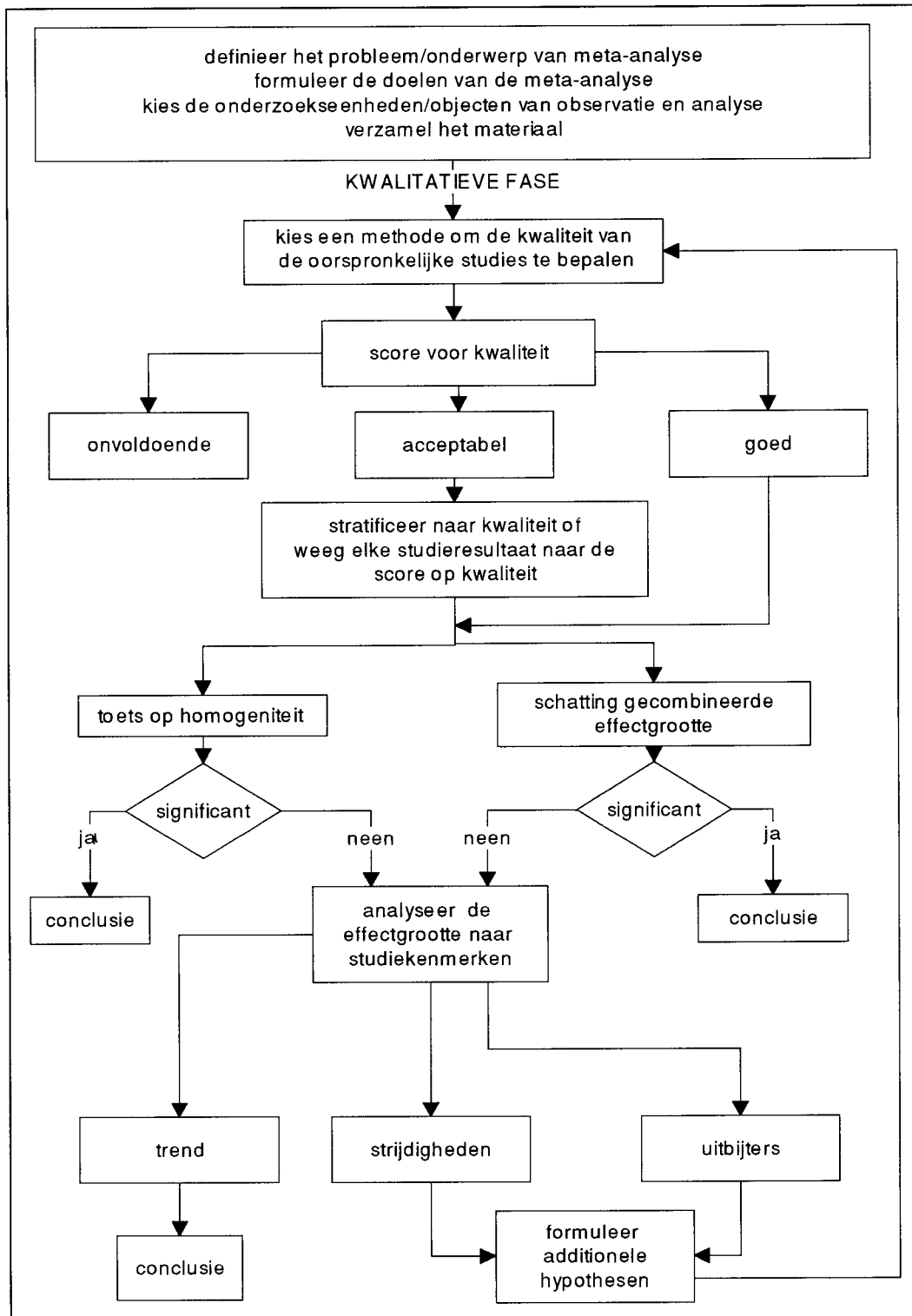
Heterogeniteit betekent dat *het werkelijke verschil tussen de interventie en de controlegroep per onderzoek kan variëren*. Het kan worden veroorzaakt door verschil in onderzoekspopulatie, verschil in (uitvoering van de) behandeling en verschil in behandelaar (zie ook de factoren die een rol spelen bij discrepantie tussen efficacy en effectiviteit). Meestal wordt aangenomen dat er geen sprake is van een heterogeniteit. Deze vooronderstelling kan statistisch worden getoetst. In geval van heterogeniteit is het niet mogelijk de resultaten van de afzonderlijke studies op een zinvolle wijze samen te nemen. Er wordt een tolerantieinterval berekend (random-effect model).

Er wordt gesproken van homogeniteit van het behandelingseffect *als wordt verondersteld dat de verschillen tussen interventie en controlegroep in elke studie even groot zijn*. Verschil betekent in dit geval verschil in het succespercentage tussen de interventie en de controlegroep. Per object/studie wordt verondersteld dat dit verschil hetzelfde is, terwijl de waargenomen verschillen tussen studies berusten op de steekproefgrootte en niet op de aard van de interventie. Het is nu wel mogelijk de afzonderlijke studieresultaten op een zinvolle wijze samen te nemen. Hiervoor staan verschillende benaderingen ter beschikking.

---

(2) Kies de best gecorrigeerde schatting (het meeste variabelen in het model). (3) Kies de schatting uit de samenvatting van een publikatie (83).

Diagram 3. Uitvoering van een Meta-analyse (93)



### Fixed-effect of random-effect model

Wanneer wordt aangenomen dat de medische interventie een bepaald effect bewerkstelligt, en de waargenomen verschillen tussen de verschillende onderzoeken op een toevallige variatie berusten, wordt een fixed-effect model gebruikt. Laat men deze veronderstelling vallen dan wordt een random-effect model gebruikt. Er wordt aangenomen dat de grootte van het effect een bepaalde verdeling vertoont over een verzameling onderzoeken.

De keuze voor een fixed of random effect model is niet gemakkelijk. Indien de toets voor homogeniteit wordt doorstaan leveren beide modellen eenzelfde resultaat. Het probleem is echter dat de statistische power van de toets op homogeniteit vaak niet groot is. Aangezien het aantal (n) gelijk is aan het aantal studies, opgenomen in de meta-analyse. Het is derhalve altijd verstandig een inschatting van de kwaliteit van de individuele studies te maken, of visuele inspectie van de gegevens middels een grafiek toe te passen (89). Soms wordt er zelfs voor gepleit a priori uit te gaan van een random-effect model. In geval van gerandomiseerd interventie onderzoek lijkt dit echter niet rationeel. Waarom immers de resultaten van verschillende studies samenvoegen als wordt verondersteld dat er in werkelijkheid sprake is van een random effect?

Een andere aanwijzing voor een keuze tussen beide modellen wordt geleverd door de puntschattingen zelf. Lopen deze veel uiteen dan ligt een random-effect model voor de hand. Ook inspectie van het betrouwbaarheidsinterval geeft een aanwijzing. Bij een breed interval lijkt een random effect model aangewezen <sup>3</sup> (83).

De keuze van een model is afhankelijk van het doel van de analyse en de vooronderstellingen. Een fixed-effect model gaat uit van de verrichte onderzoeken. In een random-effect model worden de verrichte onderzoeken gezien als een steekproef uit een hypothetische verzameling onderzoeken. Als de vraagstelling van de meta-analyse is gericht op de verrichte studies, is het rationeel het fixed-effect model aan te wenden. Als het erom gaat te weten of toepassing van een interventie een toekomstig effect bewerkstelligt, is het rationeel een random-effect model aan te wenden (83).

De keuze van een model is secundair aan de analyse van een eventueel gebrek aan homogeniteit. Er gaan zelfs stemmen op om bij een gebrek aan homogeniteit de resultaten van afzonderlijke studies niet samen te nemen (83). Als de toets voor de homogeniteit significant is, wijzende op het feit dat het zeer onwaarschijnlijk is dat de verschillen tussen studies op toeval berusten, is het berekenen van een gemeenschappelijke oddsratio te betwijfelen en moet dit ook feitelijk niet worden gedaan (94).

---

<sup>3</sup> Een random-effect model geeft een te klein gewicht aan kleine studies.

### Statistische analyse

In meta-analyse ligt de nadruk op statistische analyse van de uitkomsten van verschillende onderzoeken met als doel een samenvatting van resultaten op overzichtelijke manier en kwantificering van de mate van onzekerheid (89). Er zijn doorgaans gegevens op twee niveaus: ten eerste gegevens over de kwaliteit van de studies; ten tweede gegevens over de uitkomsten van de studies (83). De meta-analyse is verwant aan het klassiek probleem van het combineren van het resultaat van onafhankelijke statistische toetsen dat al door Fisher in 1948 wordt behandeld<sup>4</sup>. De meta-analyse maakt gebruik van (i) de geschatte effectgrootte, inclusief de standaard fout of (ii) betrouwbaarheidsintervallen van de diverse interventiestudies, en geeft als uitkomsten een puntschatting en betrouwbaarheidsinterval voor het bestudeerde effect (25).

Het zonder meer samenvoegen en bijvoorbeeld middelen van de resultaten van verschillende interventiestudies is onjuist. De patiëntenpopulatie, behandelaars, behandelingsstrategieën, definitie van eindpunt kunnen zodanig verschillen dat een samenvoegen van de gegevens leidt tot een onjuiste conclusie. Bovendien treedt er een kanskapitalisatie op door herhaling van het experiment. Het is daarom van belang te komen tot als onafhankelijk te beschouwen studieresultaten, en (de verdeling van) een nieuwe toetsingsgrootte te construeren waarmee een gezamenlijk effect kan worden getoetst.

Daarom wordt gestreefd naar het berekenen van een gemeenschappelijke oddsratio met behulp van de Mantel-Haenszel methodiek. Voordelen van deze aanpak zijn: elke interventiegroep wordt met de eigen controlegroep vergeleken (tegenaan afhankelijkheid), studies worden gewogen naar omvang, er is een toets op homo/heterogeniteit mogelijk, de grootte van het effect kan worden geschat (als verhouding) en getoetst (87).

Een andere manier is het berekenen van de grootte van het effect door het verschil te nemen van het effect in de interventie- en controlegroepen met betrekking tot een bepaalde effectmaat, gedeeld door de standaarddeviatie van de controlegroepen (90)<sup>5</sup>. Daardoor wordt het mogelijk resultaten met verschil in eindpunt samen te voegen, omdat alle resultaten van individuele studies in dezelfde dimensieloze grootte worden uitgedrukt. Dit maakt het mogelijk verschillende interventiestrategieën met elkaar te vergelijken. De conclusies zijn echter niet gemakkelijk klinisch interpreteerbaar (87).

<sup>4</sup> De specifieke p-waarden van de diverse studies ( $p_i$ ) worden gecombineerd in de toetsingsgrootte  $L$ , waarbij  $L = -2 \sum \ln p_i$ , chi-kwadraat verdeeld met  $2S$  vrijheidsgraden ( $S$  is aantal studies).

<sup>5</sup> Grootte van het effect is berekend als  $(M_t - M_c)/s$ , waarbij  $M_t$  het gemiddelde van de interventie groep is,  $M_c$  het gemiddelde van de controlegroep, en  $s$  de gepoolde standaard deviatie van de controlegroepen. Een effect grootte van twee betekent bijvoorbeeld een verschil tussen interventie en controlegroep van twee standaard-deviaties.

Uitkomstmaten

(i) *Odds ratio*. De verhouding tussen het percentage geslaagde interventies en het percentage niet geslaagde interventies in de interventiegroep, en in de controlegroep wordt aangeduid als oddsratio.  $OR = ad/bc$  (zie paragraaf 4). Per studie (n) wordt berekend: de natuurlijk logaritme (logodds ratio):

$$\text{delta}_n = \ln \left( \frac{p_i/(1-p_i)}{p_c/(1-p_c)} \right), \text{ met de bijbehorende standaard fout.}$$

Verwisseling van de ratio's in interventie en controlegroep kan leiden tot een reciproque waarde van de oddsratio. Daarom wordt wel gewerkt met een logaritmische schaal: de log oddsratio. Voor de natuurlijke logaritme van de odds ratio bestaat een eenvoudige formule voor de standaard fout: se

$$(\ln(\text{oddsratio})) = \sqrt{1/a + 1/b + 1/c + 1/d}$$

(ii) *Het risicoverschil tussen interventie en controlegroep*, gedefinieerd als het verschil in succespercentage tussen de interventie en de controlegroep. Voor het samenvoegen van de resultaten wordt de volgende weg bewandeld. Per studie wordt berekend:  $\text{delta}_n = p_i - p_c$   $p_i = x_i/n_i$ ,  $p_c = x_c/n_c$ , met de bijbehorende standaard fout (SE). Als de homogeniteitsveronderstelling opgaat, wordt het effect over de studies gesommeerd met behulp van de formule  $\text{delta}_{\text{tot}} = \sum w_i \text{delta}_i$  met  $w_i = 1/\text{se}_i^2$ . Het wordt vervolgens gedeeld door  $\sum 1 / (\text{se}_i)^2$  (zie paragraaf 4)

(iii) *Effectgrootte*. In een meta-analyse wordt de grootte van het effect, indien de toets op homogeniteit is doorstaan, berekend door per studie het effect ( $\mu$  alfa) in de interventiegroep te verminderen met het effect ( $\mu$  nul) in de controlegroep en dit verschil te delen door de gemiddelde standaarddeviatie ( $\sigma$ ) van beide groepen. Op deze wijze ontstaat per studie een dimensieloze (schaal-onafhankelijke) schatting van de kritische effectgrootte. Deze wordt vervolgens over de studies gesommeerd en gedeeld door het aantal studies. Dit wordt teruggerekend naar het bijbehorende percentage met betrouwbaarheidsinterval. Het te verwachten effect bij toepassing van de medische interventie in de praktijk is binnen dit interval gelegen<sup>6</sup> (92). Indien de test voor homogeniteit niet wordt doorstaan, wordt volstaan met het geven van een tolerantieinterval, ofwel de uitersten van het gerapporteerde effect in de verzamelde studies vormen de grenzen waarbinnen het te verwachten effect zal liggen.

---

<sup>6</sup> Effectgrootte per studie(d):  $d = (\mu_a - \mu_0) / \sigma$       Totaal Effect:  $d_c = \sum d / n$

Berekening van de effectgrootte op deze manier veronderstelt (a) studies onafhankelijk van elkaar (verschillende onderzoekspopulaties), (b) ongeveer gelijke varianties, (c) normale verdeling van studieresultaten.

Het bijbehorende betrouwbaarheidsinterval volgt uit de formule:  $t = d_c \sqrt{df_{\text{error}} / 2}$

### Interpretatie

De oddsratio drukt, indien de toets op significantie wordt doorstaan, uit dat er een effect is van de interventie ten opzichte van de controlegroep. De oddsratio is te interpreteren als de verhouding van de kans op een positief resultaat in de behandelgroep en de kans op een positief resultaat in de controlegroep (niet-behandeld/placebo/oude therapie) met betrekking tot bepaald effect. De oddsratio geeft nog geen indicatie over de grootte van het verschil tussen de onbehandelde en de behandelde groep.

Als de gemiddelde oddsratio ( $w$ ) voor een populatie bekend is, en de populatie homogeen mag worden genoemd, kan de mate van herstel in de controlegroep ( $p$ ) worden gebruikt om het behandel-effect ( $d$ ) te bepalen:  $d = p(1-p)(w-1)/[p+w(1-p)]^2$  (95).

Tabel 9. Het verslag van een meta-analyse (83)

referentie	auteur + jaartal
jaar van onderzoek	
onderzoekspopulatie	leeftijd, geslacht, omvang (n), overige (relevante kenmerken)
soort interventie	
indicatie voor interventie	
klinisch eindpunt	
resultaat	aantallen mbt tot klinische eindpunt in interventie versus aantallen in controlegroep (kruistabel)
OR (95%-bt interval)	
Indien fixed-effect model:	
Toets op homogeniteit	
Mantel-Haenszel	indien typical oddsratio, rate ratio, risk ratio
Peto	indien benaderde oddsratio,
general-variance	indien ratio (alle typen), of verschil
Indien random-effect model	
Dersimonian & Laird	indien ratio (alle typen), of verschil
Conclusie	

Als er meerdere experimentele onderzoeken over de relatie tussen een interventie en een uitkomst voorhanden zijn en de homogeniteitsveronderstelling opgaat, wordt het gezamenlijk effect berekend met behulp van meta-analyse. We kunnen nu spreken van 'efficacy'. De effectgrootte is de

<sup>7</sup> herstel in controlegroep:  $p$   
herstel in interventiegroep:  $p + d$   
oddsratio:  $d$

kans op een bepaald effect in de interventiegroep (puntschatting met het bijbehorende betrouwbaarheidsinterval). Het interval zal de waarde bevatten van het effect indien de onderzochte behandeling wordt toegepast op een gelijkende populatie en op een identieke wijze (verwachte effect). Deze kans wordt verkregen als het gewogen gemiddelde van de kans op positief resultaat in de interventiegroepen.

### 3. De kwaliteit van meta-analyse van experimenteel onderzoek

#### Inleiding

De wetenschappelijke kwaliteit van meta-analyse moet een afspiegeling zijn van de methodologische merites van de klinische trial (37). Bij een experiment worden randomisatie en blinding angewend om vertekening tegen te gaan (*zie bijlage A*). Bij meta-analyse worden insluitcriteria geformuleerd en eventueel niet gepubliceerd materiaal opgespoord. Elke studie krijgt via toetsing aan de hand van een aantal criteria een validiteitsscore toegekend, bijvoorbeeld door middel van beoordeling door verschillende beoordelaars die niet op de hoogte zijn van de uitkomsten van de studies (blinding). Men moet besluiten of: (i) een studie wordt opgenomen, (ii) een studie wordt opgenomen in een bepaald stratum, (iii) alle studies worden opgenomen.

#### Vertekening (bias)

Bouter en Ter Riet onderscheiden naar analogie met het gebruikelijk epidemiologische onderzoek de volgende bronnen van vertekening: selectie-bias, informatie-bias, confounding en effectmodificatie (96).

Om selectiebias te vermijden is het belangrijk een studie op te nemen, ongeacht het resultaat. Een aanwijzing voor selectiebias is een verband tussen de omvang van een studie en de uitkomst. De omvang van een studie is soms een weerspiegeling van een verschil in hanteren van insluitcriteria als het gaat om toelating van personen tot een RCCT. Daarnaast speelt publikatiebias een rol, met name wanneer de grootste trials het kleinste effect melden. Dit kan erop wijzen dat de kleine trials alleen werden gepubliceerd omdat er een grote oddsratio werd gevonden. Het is daarom verstandig de studies naar omvang te ordenen en te bezien, bijvoorbeeld met een eenvoudige regressie analyse, of er een verband tussen de oddsratio's en de omvang van de studies aanwezig is. Als de oddsratio een functie is van de omvang van de trial, is er sprake van effect-modificatie, ofwel er is geen sprake van het effect van de interventie op de uitkomst, maar het effect verschilt naargelang het niveau van bepaalde afhankelijke variabelen (94). Ook kan worden bezien welk percentage van het gepubliceerde onderzoek een significant resultaat rapporteert. Als dit veel hoger is dan het aangegeven significantieniveau, is er een aanwijzing voor publikatiebias (83).

Soms is het goed een schatting te maken van het 'baseline' risico. Het is immers mogelijk dat een effect zich alleen manifesteert omdat de interventiegroep een veel hogere uitgangswaarde van een bepaalde parameter vertoont dan de controlegroep. Het 'baseline' risico kan geschat worden als het effect in de controle/placebogroep (94).

Een regressie naar de inverse van de trial omvang ( $1/n$ ) levert een schatting op van de oddsratio voor de populatie. Kortom, als er een eenvoudige en klinisch relevante variabele is waarmee men de trials in de analyse kan ordenen, en de oddsratios tonen een non-random patroon in deze ordening, dan bestaat er een interactie tussen de oddsratio's en de variabele (94).

Brand en Kragt stellen derhalve voor dat in het protocol van een meta-analyse kan worden opgenomen: (i) een onderzoek naar de omvang van de studie versus oddsratio (als wordt gedacht aan een verandering van ervaring met het verrichten van het aantal handelingen), (ii) een separate analyse van de gepoolde oddsratio in de placebo en interventiegroep (als wordt gedacht dat het uitgangsrisko van belang is voor het waarnemen van effect van de behandeling), (iii) het jaar van publikatie versus oddsratio. Met name als wordt gedacht aan een verandering van ervaring met de interventie in de tijd. Evaluatie naar studieomvang, follow-up duur en jaar van publikatie zijn inmiddels al bijna obligate onderdelen van het vaststellen van de validiteit van de meta-analyse. (94)

### Publikatiebias

Een meta-analyse is zeer gevoelig voor publikatiebias. Het toevoegen of weglaten van een studieresultaat kan de uitkomst van de meta-analyse soms net al of niet significant doen zijn. Het is derhalve van belang een dergelijke vertekening uit te sluiten of de eventuele invloed ervan na te gaan (gevoeligheidsanalyse).

Niet elk onderzoeksresultaat wordt in de literatuur gepubliceerd, met name negatieve bevindingen ontbreken vaak (hypothese kan niet worden bevestigd of er wordt geen significant effect gevonden). De Publikatiebias is mede afhankelijk van het publikatiebeleid van de vaktijdschriften en de bekendheid met het veld van de onderzoeker die de meta-analyse uitvoert. Er wordt getracht publikatiebias tegen te gaan door aan te moedigen dat alle onderzoeken in principe ter publikatie worden aangeboden en publikatie te stimuleren op basis van relevantie van de vraagstelling, kwaliteit van het onderzoek en verslaglegging, en niet op grond van de gevonden significanties. Een andere methode om publikatiebias te bestrijden is het aanleggen van registraties met trials waarop de meta-analyse wordt gebaseerd. De Cochrane data-base is hiervan een voorbeeld.

Er zijn drie manieren om met publikatiebias om te gaan: (i) negeren, bijvoorbeeld indien *de werking* van een medische interventie op biologische gronden plausibel is. In het algemeen zal publikatiebias ertoe leiden dat een interventie te positief wordt beoordeeld. In feite heeft dit dan geen



negatieve gevolgen voor de behandeling van de patiënt, indien de veiligheid van de interventie is gewaarborgd. De patiënt krijgt een gelijkwaardige of betere behandeling (79). (ii) probeer al het gepubliceerde en ongepubliceerde materiaal te verzamelen, bijvoorbeeld door onderzoekers te enquêteren, of advertenties in tijdschriften te plaatsen, (iii) gebruik statistische of kwasi-statistische technieken om de grootte van de publicatiebias in te schatten, bijvoorbeeld de bureauladetest of herhaling van analyse met random weglaten van studies of power analyse (83).

### Een kritische noot

Een meta-analyse is meer en meer het patroon van een statistische analyse van 'gewone' onderzoeksobjecten gaan vertonen: *'if meta-analysis is to be accepted as a scientific discipline, each meta-analysis should be conducted like a scientific experiment, beginning with a clear plan of the question to be answered and the methods to be employed'* (87).

Lubsen heeft hiertegen bezwaar gemaakt. Hij meent dat een gepubliceerd onderzoeksrapport principieel verschilt van een verzameling gegevens, omdat betrokken onderzoeker op een bepaald moment de beslissing neemt om tot publicatie van zijn bevindingen over te gaan en daarmee het onderzoek af te sluiten. Het gaat hier niet om publicatiebias of toevallige variabiliteit. Het is een inherent aspect van het onderzoek dat zich aan de meta-analyse onttrekt. Wanneer de beslissing onzuiver is, is het niet mogelijk deze weer recht te trekken (97).

Het samenvoegen van gegevens geschiedt echter ook binnen klinisch onderzoek (onderzoek op verschillende plaatsen binnen een trial, multi centre trial). De problemen die daarbij spelen, lijken erg op de problemen die spelen bij het samenvoegen van trials, te weten homogeniteit van de patiëntenpopulatie, gelijke werking/performance van de interventie, gelijke omstandigheden van additionele therapie, zelfde wijze van interpretatie van uitkomsten (37). Daarom is een zogenaamde sensitiviteitsanalyse nodig.

Een sensitiviteitsanalyse vormt het sluitstuk van de meta-analyse. Doel is een nader inzicht in de validiteit te verkrijgen. Er wordt bekeken hoe stabiel het resultaat van de meta-analyse is. Een stabiel resultaat verandert niet als een of enkele onderzoeken bij de berekening worden weggelaten. Het is derhalve goed een dergelijke toets in de sensitiviteitsanalyse op te nemen. Er worden ad random enkele onderzoeken weggelaten om te zien wat er met het resultaat gebeurt.

Een andere toets is de zogenaamde bureaulade test. Er wordt berekend hoeveel onderzoeken met een alternatief resultaat nodig zouden zijn om de conclusie van de synthese te veranderen. Dit is ook een manier om inzicht te verkrijgen in een eventuele publicatiebias (91).

Tabel 10. Het beoordelen van de kwaliteit van een meta-analyse (91)

-	Is de onderzoeksvraag precies geformuleerd?
-	Zijn de methoden duidelijk omschreven?
-	Is het literatuuronderzoek voldoende uitgebreid om alle relevante publikaties te identificeren?
-	Zijn de criteria om onderzoeken in de analyse op te nemen voldoende expliciet?
-	Is de validiteit van de opgenomen onderzoeken beoordeeld?
-	Is dit oordeel reproduceerbaar en zonder bias?
-	Is de variatie in de uitkomsten van de onderzoeken geanalyseerd?
-	Zijn de uitkomsten van de onderzoeken op een juiste wijze gecombineerd?
-	Komen de conclusies overeen met de resultaten van de onderzoeken?

#### 4. Statistische formules voor meta-analyse

##### Meta-analyse op Odds-ratio's uit klinische trials

Indien er homogeniteit wordt verondersteld, kan de Cochrane-Mantel & Haenszel procedure worden gebruikt om in een meta-analyse een gepoolde oddsratio en het betrouwbaarheidsinterval te berekenen (93).

De gegevens worden samengevat in een twee bij twee tabel.

	Interventie		
Effect	+	-	totaal
pos	a	b	g
neg	c	d	h
totaal	e	f	m

Oddsratio per stratum/interventiestudie  $i$ :  $OR_i = (a_i \times d_i) / (b_i \times c_i)$

met als gewicht:  $weight_i = 1 / variance_i$

en als variantie:  $variance_i = n_i / (b_i \times c_i)$

95%-betrouwbaarheidsinterval: 95% C.I. =  $e^{\ln OR_i \pm 1.96 \times \sqrt{\text{var } OR_i}}$

**Berekening van verwacht (expected)/geobserveerd (observed) effect**

$$O_i = a_i \quad E_i = \frac{(a_i + c_i)(a_i + b_i)}{m_i}$$

$$V_i = \frac{(a_i + c_i)(b_i + d_i)(a_i + b_i)(c_i + d_i)}{m_i^2 (m_i - 1)}$$

$$OR_i = e^{(O_i - E_i)/V_i} \quad 95\% \text{ CI}_i = e^{(O_i - E_i)/V_i \pm 1.96/\sqrt{V_i}}$$

**Pooling over N strata:**

$$O = \sum_{j=1}^N O_j \quad E = \sum_{j=1}^N E_j \quad V = \sum_{j=1}^N V_j$$

$$OR = e^{(O-E)/V} \quad 95\% \text{ CI} = e^{(O-E)/V \pm 1.96/\sqrt{V}} \quad \chi^2 = (O-E)^2/V$$

**Toets op homogeniteit:**

$$\chi^2 = \sum \{ (O_i - E_i)^2 / V_i \} - \{ \sum (O_i - E_i) \}^2 / \sum V_i$$

met N-1 vrijheidsgraden

O = Observed; E = Expected; V = Variantie; OR = Oddsratio; CI = Confidence Interval/ betrouwbaarheidsinterval.

Meta-analyse op effectverschil in klinische trials (25)

**i Basisinformatie die nodig is voor het uitvoeren van een meta-analyse:**

- 1)  $\Delta$ : schatting van het effectverschil tussen A en B,  
 se: standaard error van deze schatting.

- 2) Een 2x2 tabel per studie:

	+	-	Totaal
A	$X_A$	$n_A - X_A$	$n_A$
B	$X_B$	$n_B - X_B$	$n_B$

- 3) Het uitrekenen van de effectmaat en de standaard error:

In dit geval het verschil in kans op respons:

$$\Delta = p_A - p_B \quad p_A = X_a / N_a \quad p_b = X_b / N_b$$

$$se = \sqrt{\frac{p_A(1-p_A)}{n_A} + \frac{p_B(1-p_B)}{n_B}}$$

Het hierbij behorende 95%-betrouwbaarheidsintervallen:

$$\Delta - 1.96*se < \text{werkelijke verschil} < \Delta + 1.96*se$$

**ii Toets op homogeniteit**

De formele toets voor de nulhypothese, 'er sprake is van homogeniteit', wordt gegeven door de volgende toetsingsgroottheid:

$$\chi^2 = \sum_{i=1}^n \frac{(\Delta_i - \Delta_{hom})^2}{se_i^2}$$

Als  $p < 0,05$  wordt de nulhypothese verworpen. Er is dan sprake van heterogeniteit.

**iii Analyse in het geval van homogeniteit**

Het overall verschil in effect ( $\Delta_{hom}$ ) wordt het best geschat met het gewogen gemiddelde:

$$\Delta_{hom} = \frac{\sum_{i=1}^n w_i \Delta_i}{\sum_{i=1}^n w_i} = \frac{\sum_{i=1}^n \Delta_i / se_i^2}{\sum_{i=1}^n 1 / se_i^2}$$

$\Sigma$ : sommatie over n studies

$$w_i \propto \frac{1}{se_i^2}$$

( $\propto$  betekent evenredig met)

De standaard deviatie van  $\Delta_{hom}$  wordt met de volgende formule berekend:

$$se_{hom} = \frac{1}{\sqrt{\sum 1 / se_i^2}}$$

Het 95%-BI is dan:  $-1,96 * se < \text{werkelijke verschil} < +1,96 * se$

**iv Analyse in het geval van heterogeniteit**

Als er sprake is van heterogeniteit wordt het gemiddelde verschil geschat door het rekenkundig gemiddelde van de individuele verschillen:

$$\Delta_{het} = \bar{\Delta} = \frac{1}{n} \sum \Delta$$

De hierbij behorende standaard deviatie wordt als volgt uitgerekend:

$$sd = \sqrt{\frac{\sum (\Delta_i - \Delta_{het})^2}{n - 1}}$$

De standaard error is de standaard error of the mean:  
 $se_{het} = sd / \sqrt{n}$

Berekening van de tussen-studie standaarddeviatie:

$$sd_{tussen} = \sqrt{sd^2 - \overline{se^2}}$$

$\overline{se^2}$  = gemiddelde  $se^2$

Berekening van het 95%-tolerantie interval:

$$\Delta_{het} \pm 1,96 * sd_{tussen}$$

Als de *tussen-studie spreiding niet groot* is, is het beter om de gemiddelde schatting van het effect te berekenen met een gewicht:

$$w_i \propto \frac{1}{sd_{true}^2 + se_i^2}$$



## Bijlage C: Uitkomst-onderzoek

### 1. Wat is uitkomst-onderzoek?

#### Definitie

Om het effect van een medische interventie op de gezondheid te beoordelen wordt gezocht naar uitkomsten van medisch handelen, waarbij onder een *uitkomst* wordt verstaan: *de uiteindelijk tot stand gekomen verandering in de gezondheidstoestand van een patiënt of populatie, die kan worden toegeschreven aan een geneeskundige interventie* (66). Uit deze definitie volgt:

(a) Het is noodzakelijk een verandering in de gezondheidstoestand van een populatie, of een individu vast te stellen. Dit vereist afhankelijk van het perspectief, doel en situatie, verschillende effectmaten, de keuze voor een bepaalde effectmaat, of de constructie van een index. Een aantal veel gehanteerde uitkomstmaten zijn: (niveau van) klinische parameters, aantal administratieve handelingen, door de patiënt zelf gerapporteerde verschijnselen, (ziektevrije) overleving, morbiditeit, complicaties, fysiek functioneren, kosten, klinische heropnamen, gezondheidstoestand, kwaliteit van leven, patiëntensatisfactie. (16)

(b) Toeschrijving van de verandering aan een geneeskundige interventie, hetgeen het aantonen van een causaal verband vereist en het uitsluiten van andere (autonome) invloeden op de geconstateerde verandering. In principe geschiedt dit op basis van experimenteel onderzoek (*zie bijlage A*). Maar doorgaans is dit niet het uitgangspunt in uitkomstonderzoek. Uitkomstonderzoek is veelal observationeel van aard. Buiten de experimentele situatie vereist het criteria met betrekking tot de wetenschappelijke validiteit en betrouwbaarheid van het onderzoek waarin een effect is aangetoond, alsmede gegevens over de toepassing van een bepaalde handeling in de praktijk.

### 2. Opzet en uitvoering

#### Opzet

Om het effect van een medische interventie in de klinische praktijk vast te stellen, moet men: (i) een behandelingsperiode definiëren, (ii) passende uitkomstmaten kiezen (overleving, ziektevrije overleving, kwaliteit van leven), (iii) methoden ontwikkelen om variabiliteit in patiëntenkenmerken te beheersen (leeftijd, geslacht, ernst, co-morbiditeit), (iv) methode kiezen om de variabiliteit in aanbodfactoren (case-mix) te beheersen (werkstijlverschillen), (v) methode kiezen om effecten van een medische interventie te onderscheiden van effecten op de gezondheid door andere factoren (gedrag, behuizing, opleiding, etc.), (vi) de periode van gegevensverzameling moet lang genoeg zijn om *alle* relevante

effecten in beeld te brengen (follow-up), (vii) de omvang van de steekproef moet groot genoeg zijn om significante verschillen waar te kunnen nemen (power van het onderzoek). (66)

In principe is er een breed scala aan uitkomstmaten beschikbaar. Een uitkomst moet: (i) een functionele impact hebben, (ii) een welomschreven diagnostische categorie betreffen, (iii) een hoge preventie hebben, (iv) beïnvloedbaar zijn door medische zorg, (v) een goed begrepen relatie hebben met voorafgaande preventie, medische zorg, revalidatie, (vi) een goed begrepen relatie met niet medische factoren hebben (32).

### Uitvoering

Voor meting van de effectiviteit van medische zorg verzamelt in Nederland de SIG informatie over de volgende indicatoren: sterfte, verbetering van de gezondheidstoestand (voor- en nameting), patientsatisfactie, recidief (heropname), tijdigheid (van behandeling, wachtlijst), complicaties, wondinfectie, infecties, geneesmiddelenvergiftiging, stralingsgevolgen, FONA-meldingen, PA-diagnose, niet medisch ongeval, uitwendige oorzaken (98).

*De medical outcome study.* In de Verenigde Staten is in 1989 een grote studie opgezet ter bestudering van invloeden op uitkomsten van medische zorg ('MOS: Medical Outcome Study') (16). Het doel was te bepalen welke variaties in uitkomsten van zorg verklaard kunnen worden door variaties in het gezondheidszorgvoorzieningen, werkstijl van de clinicus, klinisch specialisme en het ontwikkelen van bruikbare instrumenten om uitkomsten te volgen en te beoordelen. Als resultaat en klinisch eindpunt (hieronder werden traditioneel door clinici gemeten resultaten verstaan in termen van symptomen en tekens, lab waarden, dood) werden gemeten: lichamelijke- en sociale functie in het dagelijks leven; beleving van de algemene gezondheidstoestand van patiënten; patiëntsatisfactie (64).

Het conceptuele raamwerk van deze studie wordt gevormd door de termen structuur, proces, uitkomst. Het zorgproces is een term, waarmee men de medische interventie en de gebeurtenissen die daarmee samenhangen pleegt aan te duiden. De structuur van de gezondheidszorg is een beschrijving van de verhouding en onderlinge verbanden van de samenstellende elementen van het gezondheidszorg-systeem. Resultaten of uitkomsten van gezondheidszorg zijn de vermeende gevolgen van medisch ingrijpen in de vorm van handelingen en het scheppen van voorzieningen.

De structuurkenmerken worden geoperationaliseerd met behulp van: organisatie, specialisatie (mix), financiële prikkels, werkbelasting, toegankelijkheid/comfort; kenmerken van hulpverlener: leeftijd, geslacht, specialiteit, opleiding, economische motieven, houding/meningen, voorkeur, arbeidssatisfactie; kenmerken van patiënt: leeftijd, geslacht, diagnose/toestand, mate van ernst, comorbiditeit, gezondheidsgewoonten, houding/mening, voorkeuren (16).



Het zorgproces wordt gemeten met behulp van: consulten, medicatie, verwijzingen, labonderzoek, ziekenhuisopnamen, continuïteit van zorg, coördinatie van zorg, bestedingen; interpersoonlijke werkwijze van een arts, participatie van de patiënt, counseling, communicatie niveau (16).

Het uitkomstonderzoek heeft in de Verenigde Staten themagewijs onderzoek op gang gebracht naar meer specifieke onderwerpen, uitgevoerd door multidisciplinaire teams ('Patiënt Outcome Research Teams', PORT's). Doorgaans wordt meta-analyse verricht naar aanwijzingen voor efficacy en vervolgens effectiviteit bepaald met behulp van registratie van het zorgproces (65). Over de resultaten wordt in afzonderlijke rapporten met enige regelmaat verslag gedaan (64).

### Interpretatie

Uitkomsten zijn te beschouwen als (i) bewijs voor efficacy van een medische interventie als het gaat om waarnemingen in een klinische trial, (ii) aanwijzing voor effectiviteit als het gaat om waarnemingen in zorgregistraties, (iii) uitkomsten als een eindstadium van routinematige zorg (geen verdere zorg meer mogelijk, bijvoorbeeld sterfte of beperkingen/handicaps), (iv) uitkomsten als metingen van gezondheid van de bevolking (bijvoorbeeld kwaliteit van leven, ervaren gezondheid) en zogenaamde schildwacht gebeurtenissen (sentinel events) gerelateerd aan gezondheidszorg (bijvoorbeeld rapportage bijwerkingen geneesmiddelen of klachten).

### **3. De kwaliteit van uitkomst-onderzoek**

Het gebruik van uitkomsten om het effect van een medische interventie te bepalen kent een aantal problemen: (i) er is een afhankelijkheid van de kennis van het natuurlijk beloop van een aandoening, (ii) hulpverleners zijn niet voldoende onderwezen in het denken over de gevolgen op het niveau van de groep; ze beperken zich tot het niveau van de individuele patiënt. Een skoop, die gegeven de problematiek (schaarste van zorg en verdeling) van vandaag de dag te smal is, (iii) de gegevens betreffende gevolgen van interventies worden niet routinematig verzameld, (iv) de relatie tussen het zorgproces en de uitkomst is vaak niet gefundeerd. Er is nog een teveel aan storende variabelen en een gebrek aan methoden om dit te ondervangen. (99) (100)

### Conclusie

Het uitkomstonderzoek is een belangrijke bron van informatie over de effectiviteit van medische zorg. Sedert 1989 heeft het in de Verenigde Staten een grote vlucht genomen. Er is veel aandacht voor methodologieontwikkeling en standaardisering van meetinstrument en -methoden. Ook is er sprake van

internationale netwerkvorming. De methodologie is echter nog minder ver ontwikkeld dan die van andere vormen van medisch onderzoek, zoals RCCT/case-controle etc. Over de interpretatie van uitkomstonderzoek lopen de meningen uiteen. Bestudering van de opzet, uitvoering en resultaten van Amerikaans uitkomstonderzoek verdient aanbeveling.

De literatuurverwijzingen in de bijlagen zijn terug te vinden in de literatuurlijst op blz.59-62