

Het Bewijs

Achtergrondstudie bij het advies "Zonder context geen bewijs. Over de illusie van evidence-based practice in de zorg."

Auteur: Leo Ottes

Den Haag, juni 2017

Inhoud

1	Inleiding	5
1.1	Leeswijzer	6
2	Evidence-based medicine	7
2.1	Het begrip evidence-based medicine	7
2.2	De bewijspyramide binnen EBM	7
3	De verschillende niveaus van bewijskracht binnen EBM nader geanalyseerd	9
3.1	Inleiding	9
3.2	In vitro en proefdieronderzoek	9
3.3	Mening van experts, ideeën en theorieën	10
3.4	Casusbeschrijvingen: case reports, case series en case-control studies	12
3.5	Sterke en zwakke punten van de verschillende onderzoeksmethoden	16
3.6	Systematische reviews en meta-analyses	17
4	Causaliteit	19
4.1	Inleiding	19
4.2	De cholesterolhypothese: <i>hèt</i> bewijs?	20
4.3	De cholesterolhypothese: de losse eindjes	21
4.4	Oorzaak, gevolg of een samenloop van omstandigheden?	22
4.5	Conclusie	23
5	Is er een alternatief voor de EBM-piramide?	25
5.1	Inleiding	25
5.2	Het GRADE-systeem	25
5.3	The revised evidence pyramid	26
5.4	No pyramid at all: het legpuzzelmodel	27
	Afkortingen	29

1 Inleiding

Evidence-based protocollen en richtlijnen vormen in toenemende mate een richtsnoer voor het handelen van zorgverleners. Toezichhouders en financiers eisen meer en meer dat ze gevolgd worden. Naarmate evidence-based protocollen en richtlijnen een grotere invloed krijgen op de behandelingsopties voor patiënten, wordt het ook belangrijker om te kijken naar de soliditeit van het fundament waarop deze gebaseerd zijn: de evidence. In deze notitie wordt dit fundament, met name de hiërarchie in bewijskracht zoals die binnen EBM wordt gehanteerd, nader beschouwd.

Evidence-based practice is, zoals het begrip reeds aangeeft, gebaseerd op evidence cq. bewijs. Maar wat is bewijs? Is er één soort bewijs of zijn er meerdere vormen. Bestaat er zoiets als *het* bewijs?

In een column in de NRC somt Robert Dijkgraaf een aantal vormen van 'bewijs' op, zoals bewijs per intimidatie: "Het is waar omdat ik het zeg dat het waar is". Of bewijs per voorbeeld: een conclusie trekken uit één, voor de gelegenheid passende, casus. Of een bewering net zolang herhalen dat men het als waar aan gaat nemen (bewijs *ad nauseam* of *zum Kotzen*).¹

Uiteraard is zijn het voorbeelden van het tegengestelde, maar ze werden en worden in de dagelijkse praktijk nog steeds met succes gebruikt. Millennia lang was de het handelen in de geneeskunde gebaseerd op het oordeel van een autoriteit - de dokter. De medische 'kennis' werd van meester op leerling overgedragen en deze kennis stond niet ter discussie. Als iemand in de middeleeuwen ziek was, dan werd er ader gelaten, gepurgeerd etc. Of deze behandelingen ook daadwerkelijk effectief waren stond niet ter discussie.

Een poging om daar verandering in te brengen was de 'experimentele geneeskunde' in het midden van de negentiende eeuw. Pionier op dit terrein was Claude Bernard (1813-1878). Zijn ideaal was *wetenschappelijke geneeskunde* gebaseerd op empirisch onderzoek, zoals hij in *Introduction à l'étude de la médecine expérimentale* (1865) beschreef. Deze geneeskunde was gebaseerd op exacte kennis van de fysiologische mechanismen die ten grondslag liggen aan een specifieke ziekte of aandoening of het therapeutisch effect van een geneesmiddel.

Interessant is dat Bernard statistiek zag als een instrument maar niet als een basis voor de medische wetenschap:

*"Natuurlijk, statistiek kan de prognose van de dokter leiden; tot zover is statistiek nuttig. Ik verwerp het gebruik ervan in de geneeskunde dan ook niet, maar ik heb er wel bezwaar tegen als men niet een stap verder gaat en gelooft in statistiek als het fundament van de medische wetenschap."*²

Noten

¹ Dijkgraaf, R. (2009) Waterdicht bewijs. <http://www.nrc.nl/handelsblad/2009/04/25/waterdicht-bewijs-11718159>

² Bernard, C. (1865) *Introduction à l'étude de la médecine expérimentale*. p. 196, Flammarion, Parijs
Het Bewijs

Ondanks de opkomst van de medische basiswetenschappen, bleef de geneeskunde in de praktijk nog gebaseerd op autoriteit. Pas in de jaren negentig van de vorige eeuw kwam daar verandering in met de opkomst van de evidence-based medicine (EBM).

Inmiddels is de toepassing van de binnen EBM gehanteerde methoden niet beperkt gebleven tot de geneeskunde. Op allerlei terreinen, zowel in de zorg, maar ook daarbuiten, zoals het onderwijs, worden ze toegepast. Een passende term voor deze brede toepassing is evidence-based practice (EBP).

1.1 Leeswijzer

In deze achtergrondstudie wordt specifiek gekeken naar de zgn. evidentie-pyramide die binnen EBM wordt gehanteerd. Is deze bewijshiërarchie houdbaar en zo nee, zijn er alternatieven? Daartoe wordt in het volgende hoofdstuk het begrip evidence-based medicine en de gehanteerde bewijspyramide kort beschreven. Vervolgens worden de verschillende niveaus van bewijskracht nader geanalyseerd. Het blijkt dat elke onderzoeksmethode cq. elk soort bewijs haar sterke en zwakke kanten heeft en dat er geen universeel geldige hiërarchie bestaat. In het hoofdstuk over causaliteit wordt aan de hand van de cholesterolhypothese, waarbij aannames gedaan worden over oorzaak en gevolg geïllustreerd dat het belangrijkste probleem niet zozeer (de hiërarchie van) bewijs is, maar de grote invloed van ideeën over oorzaak en gevolg cq. hypothesen, waarbij er een gebrek is aan falsificatie. Veel onderzoek binnen de gezondheidszorg is gericht op het bevestigen van hypothesen in plaats van het trachten te falsificeren. In het laatste hoofdstuk wordt een alternatief voor de evidentiepyramide geschetst.

2 Evidence-based medicine

2.1 Het begrip evidence-based medicine

Het begrip evidence-based medicine is in 1992 door de Canadese arts Gordon Guyatt geïntroduceerd in een artikel in het tijdschrift JAMA: "Evidence Based Medicine. A new approach to teaching the practice of medicine." Zoals de titel aangaf ging het om een nieuwe onderwijsstroming, in dit geval binnen de Canadese McMaster-universiteit, die zich afzette tegen 'het oude denken' dat vooral gebaseerd was op de meningen van experts en pleitte voor het toepassen van wetenschappelijk bewijs bij de behandeling van de individuele patiënt naast de praktijkervaring van de arts en de voorkeuren van de patiënt: een balans tussen evidence-, practice- en demand-based medicine.

Het beoordelen van wetenschappelijk bewijs bleek evenwel te hoog gegrepen voor individuele artsen en het werd steeds meer een collectieve activiteit van groepen experts, zoals richtlijncommissies. Dit resulteerde in evidence-based protocollen en richtlijnen voor het handelen van de individuele arts. Basis voor de *evidence* vormen de uitkomsten van wetenschappelijk, met name statistisch onderzoek. Nu bestaan er vele soorten van (statistisch) onderzoek, die qua bewijskracht verschillend worden gewaardeerd binnen EBM.

2.2 De bewijspyramide binnen EBM

De verschillen in bewijskracht wordt binnen EBM veelal aangegeven in de vorm van een hiërarchische piramide. Aan de basis van de piramide staat de onderzoeksmethode met de laagste bewijskracht, aan de top die met de hoogste bewijskracht. Er zijn verschillende hiërarchieën in omloop voor verschillende doeleinden, bijvoorbeeld therapeutische studies, en ook zijn er verschillen in indelingen van soorten studies. In figuur 1 is als voorbeeld de hiërarchie zoals die door het Centre for Evidence-based Medicine (www.cebm.net) voor therapeutische studies wordt aangehouden.

Figuur 1.
Levels of Evidence for Therapeutic Studies*

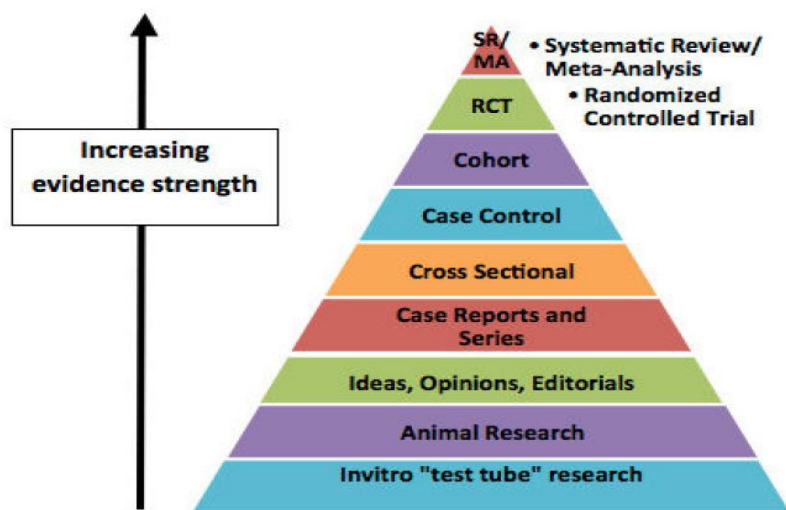
Level	Type of evidence
1A	Systematic review (with homogeneity) of RCTs
1B	Individual RCT (with narrow confidence intervals)
1C	All or none study
2A	Systematic review (with homogeneity) of cohort studies
2B	Individual Cohort study (including low quality RCT, e.g. <80% follow-up)
2C	"Outcomes" research; Ecological studies
3A	Systematic review (with homogeneity) of case-control studies
3B	Individual Case-control study
4	Case series (and poor quality cohort and case-control study)
5	Expert opinion without explicit critical appraisal or based on physiology bench research or "first principles"

* From the Centre for Evidence-Based Medicine. <http://www.cebm.net>.

In het algemeen zijn de verschillen tussen de diverse indelingen klein.

Figuur 2 geeft een meer algemene EBM-piramide weer. In alle indelingen staan op de hoogste trede van wetenschappelijk bewijs de systematische reviews van met name dubbelblind gerandomiseerde onderzoeken, RCT's (randomized controlled trials). Dit is onlosmakelijk verbonden de zgn. Cochrane reviews, genoemd naar Archie Cochrane (1909-1988), een Schotse arts. Hierbij worden via literatuuronderzoek alle wetenschappelijke publicaties over een bepaald onderwerp verzameld en volgens een vast stramien geanalyseerd: de kwaliteit van de onderzoeksopzet, statistische bewerkingen etc. Dit wordt gedaan door een internationaal netwerk van meer dan 28.000 wetenschapper in meer dan 100 landen, de Cochrane collaboration, opgericht in 1993.

Figuur 2: De evidentie-pyramide



Adaptation of the Evidence Pyramid Diagram developed by the Medical Research Library of Brooklyn, SUNY Downstate Medical Center.

3 De verschillende niveaus van bewijskracht binnen EBM nader geanalyseerd

3.1 Inleiding

In dit hoofdstuk worden de verschillende niveaus van bewijskracht binnen de bewijspiramide nader gezien. Elk type onderzoek blijkt zijn sterke en zwakke kanten te hebben, waardoor duidelijk wordt dat de ideale onderzoeksopzet niet bestaat. Ook komt de paradox van de lage waardering van de mening van experts binnen EBM aan de orde. Echter uiteindelijk moeten onderzoeksresultaten – verschillende vormen van bewijs – gewogen worden door experts, waardoor ze aan de top van de hiërarchie staan. Aan het eind van het hoofdstuk wordt in een tabel een overzicht gegeven van de sterke en zwakke kanten van de verschillende onderzoeksopzetten.

3.2 In vitro en proefdieronderzoek

Op de laagste treden van de bewijshiërarchie in de EPB-pyramide staat laboratoriumonderzoek: het onderzoek *in vitro* - in de reageerbuis - en onderzoek gebaseerd op proefdieren. Reden hiervoor is dat het extrapoleren van onderzoeksresultaten naar de mens vaak problematisch is. Een stof die bij proefdieren heel effectief is, blijkt bij mensen vaak niet of nauwelijks te werken. Ook kan een stof kan bij de mens veel giftiger blijken te zijn dan bij een dier. Het omgekeerde komt ook voor. Daar komt bij dat de ene proefdiersoort de andere niet is. Een bekend voorbeeld is de thalidomide (Softenon) tragedie. Deze stof werd eind jaren vijftig van de vorige eeuw als 'volkomen veilig slaapmiddel' op de markt gebracht, eerst in Duitsland en daarna in geheel Europa. Het was een effectief slaapmiddel dat ook aan zwangeren werd voorgeschreven. Er was door de Duitse fabrikant Chemi Grünenthal onderzoek gedaan op muizen naar de giftigheid van het middel en mogelijk schadelijke effecten op de ongeboren vrucht bij gebruik tijdens de zwangerschap. Daar waren geen aanwijzingen voor gevonden. Het Softenondrama is welbekend: meer dan 10.000 kinderen kwamen met ernstige geboortedefecten ter wereld in Europa en Afrika.

Overigens bleek thalidomide bij muizen geen, maar bij konijnen wel geboortedefecten te geven. De fabrikant had het middel op 'het verkeerde proefdier' getest. Naarmate een proefdier meer verwant is met de mens, lijkt het metabolisme meer op dat van de mens. Daarnaast zijn er (infectie)ziekten die alleen bij de mens voorkomen en waarvoor geen adequaat proefdiermodel bestaat.

De basis (bio)wetenschappen kunnen echter niet bestaan zonder deze vormen van onderzoek. In het laboratorium kan met cellen en proefdieren in relatief korte tijd en tegen relatief lage kosten onder gecontroleerde, gestandaardiseerde, praktische omstandigheden geëxperimenteerd worden. Zo kan bijvoorbeeld een gen in een muis uitgezet worden (knock-out muis) en kan gekeken worden wat er gebeurt. Zo kan de functie van een gen opgehelderd worden. Laboratoriumonderzoek is onmisbaar voor het ontrafelen van de oorzaken van ziekten en aandoeningen alsmede het ophelderen van werkingsmechanismen van behandelingen. In het verleden is de geneeskrachtige werking van stoffen meestal bij toeval in de praktijk ontdekt, bijvoorbeeld het ontstekings-

remmende effect van salicylzuur en het daarvan afgeleide aspirine reeds. Pas veel later kwam men via laboratoriumonderzoek achter het werkingsmechanisme, namelijk dat het de (verhoogde) aanmaak bij ontsteking van prostaglandines, hormoonachtige stoffen die onder andere ontstekingsverschijnselen veroorzaken, remmen.

Proefdieronderzoek vormt de basis voor hypothesen. Zo vindt de cholesterolhypothese, de veronderstelling dat cholesterol atherosclerose veroorzaakt, zijn oorsprong in proefdieronderzoek. Konijnen die cholesterolrijke voeding kregen ontwikkelen namelijk atherosclerose. (Op de cholesterolhypothese wordt in hoofdstuk 4 nader in gegaan).

Daarnaast wordt proefdieronderzoek door autoriteiten geëist voor bijvoorbeeld potentiële geneesmiddelen voor de mens. Eerst moet een effect in proefdieren aangetoond worden alsmede de mate van giftigheid, waaronder mogelijk schadelijke effecten op het embryo bij gebruik tijdens d zwangerschap.

3-3 Mening van experts, ideeën en theorieën

De mening van experts, ideeën en theorieën staan gezamenlijk op de derde trede van onderen in de hiërarchie. Reden hiervoor is dat meningen subjectief en niet intrinsiek verifieerbaar zijn. Ervaring uit het verleden laten zien dat experts het regelmatig bij het verkeerde eind bleken te hebben.

Het is overigens merkwaardig dat ideeën en theorieën hoger staan dan in vitro en proefdieronderzoek, aangezien deze vaak gebruikt worden om de validiteit van ideeën en theorieën te toetsen.

Op sommige terreinen van de geneeskunde, bijvoorbeeld bij bacteriële infecties, is de juistheid van een hypothese vrij gemakkelijk aan te tonen. Bijvoorbeeld de hypothese dat de oorzaak van cholera een bacterie is. De vier postulaten³ die Robert Koch eind negentiende eeuw zijn hierbij heel behulpzaam. Vaak is het evenwel niet zo gemakkelijk. Bij de huidige complexe chronische aandoeningen is het causaliteitsdenken grotendeels verlaten en wordt gesproken van risicofactoren. Hierbij worden wel causale relaties verondersteld, maar deze zijn vaak impliciet.⁴ Causale mechanismen lijken wat in de vergetelheid geraakt te zijn. Van verschillende kanten wordt gepleit om deze (weer) een prominentere plaats te geven binnen EBM.⁵

In de geschiedenis van de geneeskunde was een medische behandeling vaak gebaseerd op een idee, theorie of overtuiging. Zo was het aderen in de Middeleeuwen gebaseerd op het idee dat ziekte werd veroorzaakt door een onbalans tussen de vier 'lichaamvochten' of *humores*, te weten bloed (*sanguis*), slijm (*phlegma*), zwarte en gele gal (*melainè* en *xanthè cholè*). Veel ziekten zouden door

Noten

³ De vier postulaten van Robert Koch:

- De ziektekiem moet in ongewoon grote hoeveelheden in de patiënt, dier of plant maar niet in het gezonde dier of plant.
- De ziektekiem moet kunnen worden geïsoleerd en verder gekweekt.
- Een proefdier of dezelfde soort plant dat met de gekweekte kiem besmet wordt, moet dezelfde ziekte krijgen.
- De ziektekiem moet uit het proefdier of de proefplant geïsoleerd kunnen worden, en moet gelijk zijn aan de ziektekiem in het lichaam van de patiënt

⁴ Russo, R., Williamson, J. (2011) Epistemic Causality and Evidence-Based Medicine. *Stazione Zoologica Anton Dohrn*

⁵ Clarke, B., Gillies, D. (2014), Illari, P., Russo, F., Williamson, J. Mechanisms and the Evidence Hierarchy. DOI 10.1007/s11245-013-9220-9

een teveel aan sanguis veroorzaakt worden en de 'logische' oplossing was het aanprikken van een ader en bloed aftappen. Tegenwoordig kan men zich moeilijk voorstellen hoe men in zo'n behandeling en theorie kon geloven, maar tot ver in de 18de eeuw waren artsen ervan overtuigd dat het heilzaam was. Deze overtuiging werd van leermeester op leerling overgedragen: authority-based medicine.

Ook in de huidige tijd zijn er allerlei theorieën en ideeën over de oorzaak en remedie van ziekten. Het komt in de geneeskunde regelmatig voor dat men niet goed weet wat oorzaak en wat gevolg is. Zo is het bijvoorbeeld nog steeds niet duidelijk of amyloid plaques die in de hersenen van patiënten met Alzheimer worden gevonden (een) oorzaak van deze ziekte zijn of juist een gevolg. Ondanks deze onduidelijkheid zijn er experts die heilig geloven in de amyloid plaque oorzaak is. Een ander voorbeeld is de cholesterolhypothese waarover later meer.

Als experts zijn het niet met elkaar eens zijn, wat vaak het geval is, zal de bewijskracht op andere manieren verkregen moeten worden. Echter ook als ze het wel met elkaar eens zijn, dan kan er zoals het aderen illustreert, sprake zijn van een collectieve dwaling.

Het moge dan ook niet verwonderen dat binnen de EBP-hiërarchie expert opinions laag scoren. De meningen van een expert is subjectief en niet intrinsiek verifieerbaar.

Binnen de EBM-hiërarchie doet zich evenwel een paradox voor: uiteindelijk moet voor het opstellen van richtlijnen 'het beste bewijs' van de hoogste treden van de EBM-hiërarchie - door een aantal experts - een veel lagere bewijslastcategorie - worden beoordeeld. Daarbij kunnen de interpretaties door de verschillende experts van hetzelfde onderliggende bewijs verschillend zijn en de uiteindelijke 'consensusredeneringen' zijn soms lastig te volgen. Een voorbeeld uit de richtlijn 'Pijn bij kanker'⁶. In deze richtlijn wordt het gebruik van paracetamol als pijnstiller bij kanker onderbouwd met een drietal studies:

"De eerste studie vergeleek het effect binnen 6 uur van achtereenvolgende eenmalige toedieningen van placebo, 650 mg paracetamol, 4 mg butorfanol (een opioïd) of de combinatie van beiden (in gerandomiseerde volgorde) bij 20 patiënten met gemetastaseerde maligniteiten met matig tot ernstige chronische kanker gerelateerde pijn in een dubbelblinde opzet (auteur: Stambaugh). Er was een niet-significante trend voor een beter effect van paracetamol ten opzichte van placebo."

De conclusie luidt evenwel: "De werkgroep is van mening dat het effect van paracetamol bij matig ernstige pijn in de klinische praktijk is aangetoond, ook als stap 1 van de WHO ladder." Als argument geeft zij:

"Het is aannemelijk dat het ontbreken van een significant effect van paracetamol ten opzichte van placebo in het onderzoek van Stambaugh veroorzaakt wordt door een te lage power van het onderzoek om een verschil aan te tonen."

Noten

⁶ Pijn bij kanker (2008). Landelijke richtlijn versie 1.0, Landelijk richtlijnwerkgroep Pijn bij kanker, 10 oktober
Het Bewijs

De tweede studie betrof onderzoek naar het effect van het toevoegen van paracetamol aan morfine. De studie liet geen verschil in effect zien tussen wel of geen toevoeging van paracetamol. Bij de derde studie werd hetzelfde onderzocht en werd wel een verschil gevonden ten voordele van het toevoegen van paracetamol.

De conclusie was echter stellig:

"De werkgroep is van mening dat paracetamol effectief is als eerste stap bij de behandeling van matig ernstige pijn bij patiënten met kanker."

De 'eigenwijze' mening van een individuele expert kan ook levensreddend zijn. In de vorige paragraaf is het Softenondrama beschreven. In de Verenigde Staten werkte in de tijd dat het middel in Europa op de markt kwam Frances Oldham Kelsey als medisch deskundige bij de FDA in de Verenigde Staten. In het British Medical Journal waren inmiddels een paar gevallen beschreven die een mogelijke link legden tussen het middel en een aandoening van de zenuwen bij chronisch gebruik. Wat haar echter het meest ongerust maakte was het ontbreken van gegevens of het middel de placenta al dan niet kon passeren. Immers als een stof niet door de placenta kan, dan kan het ook geen kwaad doen bij de ongeboren vrucht. Door haar toedoen werd Softenon niet tot de markt toegelaten, waardoor de Verenigde Staten deze ramp ontlieden.

3.4 Casusbeschrijvingen: case reports, case series en case-control studies

Een trap hoger in de EBM-hierarchie staan de case reports, case series en case-control studies. Bij case-reports worden individuele ziektegevallen zo objectief mogelijk beschreven. Bij case series gaan het om het beschrijven van een aantal gelijksoortige ziektegevallen. Als hierbij een vergelijking wordt gemaakt met gezonden spreekt men van case-control studies. Case-control studies staan hoger op de EBM-ladder, zelfs hoger dan cross-sectional studies, die later besproken worden.

Case reports

De geneeskunde is van oudsher gebaseerd op de beschrijving van individuele ziektegeschiedenissen. Elk medisch dossier van een patiënt is in feite een casusbeschrijving. De diagnostiek van ziekten en aandoeningen is gebaseerd op casusbeschrijvingen. Een bekend voorbeeld is de ziekte van Alzheimer, genoemd naar degene die het ziektebeeld het eerst beschreven heeft, Alois Alzheimer. In 1906 beschreef hij specifieke afwijkingen in de hersenen van een patiënte met dementie in de vorm van opeenhopingen van materiaal buiten hersencellen, de zgn. amyloïde plaques. Daarnaast vond hij ook afwijkingen in de hersencellen, de zgn. neurofibrillaire tangles.

De kracht van case reports is dat zij nieuwe zeldzame aandoeningen of bijwerkingen aan het licht kunnen brengen, zoals het eerdere voorbeeld van de case reports in het British Medical Journal van een mogelijke link legden tussen het chronisch gebruik van thalidomide en een aandoening van de zenuwen.

De beperking van case reports is, dat ze, geïsoleerd gezien, geen causale oorzaak of effectiviteit van een behandeling kunnen aantonen

Case series

Door ziektebelopen van meerdere patiënten met elkaar te vergelijken, zgn. case series, kunnen patronen onderkend worden. Zo werden amyloïde plaques gevonden bij patiënten die soortgelijke symptomen - dementie – hadden als de patiënt die door Alzheimer was beschreven.

Case series zijn uitermate belangrijk voor de ontdekking van nieuwe ziekten. Als voorbeeld kunnen de beschrijvingen van clusters van toegenomen gevallen van Karposisarcroom en een bepaalde vorm van longontstekingen in 1981 genoemd worden. Deze leidden tot de ontdekking van hiv. Een ander voorbeeld is de epidemie van longontsteking in 1976 die leidde tot de ontdekking van legionella. Zeer actueel zijn case series gepubliceerd in het tijdschrift The Lancet over aangeboren afwijkingen in relatie tot een besmetting met het zikavirus.⁷

Case-control studies

Beschrijving van een enkele of meerdere casus is in onvoldoende om een verband aan te kunnen tonen. Stel dat zich bij iedereen amyloïde plaques vormen (hetgeen op hoge leeftijd het geval lijkt te zijn) en men vindt deze bij patiënten met dementie, dan zegt dat niets over een mogelijk verband. Men zal het altijd moeten vergelijken met mensen die zoveel mogelijk lijken op deze patiënte, bijvoorbeeld in leeftijd en geslacht, maar die de symptomen niet hebben. Gevonden verschillen kunnen dan duiden op een (al dan niet causaal) verband.

Cross-sectional studies

Bij cross-sectioneel, ook wel transversaal onderzoek wordt ieder individu in een groep eenmaal en op hetzelfde tijdstip geobserveerd of gemeten. Er kan dan gezocht worden naar verbanden, bijvoorbeeld een relatie tussen bloeddruk en leeftijd. Cross-sectioneel onderzoek is snel en gemakkelijk uit te voeren en kan een schatting geven van bijvoorbeeld het voorkomen van een bepaalde aandoening, maar geeft geen informatie over een mogelijk causaal verband tussen gevonden correlaties.

Cohortstudies

Cohortstudies kunnen prospectief of retrospectief zijn. Bij retrospectief cohortonderzoek wordt na het vaststellen van een uitkomst terug gegaan in de tijd. Bijvoorbeeld bij een patiënt wordt een mesotheliom, een bepaalde zeldzame vorm van kanker gevonden. Stel dat blijkt dat veel van deze patiënten beroepsmatig veel in aanraking zijn geweest met asbest zijn geweest, dan kan dit wijzen op een verband tussen blootstelling aan asbest en het krijgen van deze vorm van kanker.

Bij prospectieve cohortstudies volgt men groepen, veelal gezonde, mensen in de tijd en kijkt welke aandoeningen zij krijgen en of er verbanden zijn met bijvoorbeeld leefstijlkenmerken. Een zeer bekende cohortstudie is de Amerikaanse Framingham Heart Study die in paragraaf 4.2 nader aan de orde komt. Nederland heeft overigens een sterke traditie van cohortonderzoek, bijvoorbeeld het ERGO-onderzoek in Rotterdam, LASA, PIAMA, GenerationR en Lifelines. Dergelijke onderzoeken leveren waardevolle informatie over correlaties tussen bijvoorbeeld voeding en ziekte. Het is echter lastig om exacte causale relaties vast te stellen. In hoofdstuk 4 zal aan de hand van de

Noten

⁷ Giovanni V A França et al. (2016) Congenital Zika virus syndrome in Brazil: a case series of the first 1501 livebirths with complete investigation. Lancet Volume 388, No. 10047, p891–897, 27 August

cholesterolhypothese geïllustreerd worden hoe moeilijk het in de geneeskunde veelal is om causale relaties vast te stellen.

Randomized Controlled Trials

Onder RCT's, randomized controlled trials, worden onderzoeken verstaan waarbij proefpersonen willekeurig ('random') over twee of meer groepen verdeeld worden, waarna de groepen worden blootgesteld aan verschillende proefomstandigheden of behandelingen.

Een RCT is idealiter dubbelblind, dat wil zeggen dat noch de behandelaar, noch de patiënt weet welke van de twee behandelingen die worden vergeleken, ontvangt. RCT's worden vooral toegepast in medicijnonderzoek. In veel andere situaties is blinding niet mogelijk of ethisch onacceptabel. Zo zijn er in het verleden schijnoperaties verricht, waarbij de patiënt wel werd geopereerd, maar de betreffende behandeling achterwege werd gelaten. Een bekend voorbeeld is de behandeling van angina pectoris via een operatie waarbij de borstklieslagader werd afgebonden. Dit zou de bloeddoorstroming van het hart vergroten. Driekwart van de patiënten verbeterde na de operatie. Een kwart was zelf geheel genezen. Deze door de Italiaanse chirurg Fieshi in 1939 uitgevonden techniek werd daarna de standaardbehandeling van angina pectoris. Tot dat in 1959 Cobb 17 patiënten opereerde, waarbij hij bij acht de standaardprocedure volgde en bij 9 uitsluitend de insnijdingen in de borst maakte, maar de borstklieslagader ongemoeid liet, zonder dit aan de patiënten te vertellen. Er was geen verschil in resultaat tussen beide groepen.⁸

Dergelijk onderzoek wordt thans onacceptabel geacht. Bij geneesmiddelenonderzoek kan in de regel wel eenvoudig voldaan worden aan de blindingeis bijvoorbeeld door de te vergelijken geneesmiddelen of placebo in identiek uitziende (en smakende) tabletten te verwerken, waarbij alleen de onderzoeker of computer weet wat in welke tablet zit.

De randomisatie zorgt ervoor dat zowel onbewuste als bewuste menselijke invloeden voortvloeiend uit het in de ene of andere groep ingedeeld zijn invloed kunnen hebben op het resultaat. Dubbelblind toewijzen garandeert dat de resultaten en de analyse niet door vooroordelen gekleurd zijn. Dit alles maakt dat de dubbelblinde RCT gezien wordt als de gouden standaard, die boven alle verdenking van potentiële bron van systematische vertekening is verheven.

Bij dit laatste dient opgemerkt te worden dat de RCT vooral op theoretische basis in de jaren 90 van de vorige eeuw is omarmt zonder dat er veel empirisch onderzoek was gedaan of de claims in de praktijk ook daadwerkelijk waargemaakt worden.⁹ Er zijn verschillende onderzoeken gedaan naar de effecten van blinding en randomisatie. In een overzichtsartikel laat Kaptchuk zien dat het beeld complex is. In het algemeen wordt aangenomen dat in niet geblindeerde onderzoeken de effecten van een werkzaam middel zullen worden overschat als dit vergeleken wordt met een placebo. De patiënt weet immers dat hij of zijn een, althans verondersteld, werkzaam middel krijgt of een onwerkzaam. Uit een systematische review die RCT's vergeleek met non-RCT's voor verschillende interventies bleek dit echter niet. Zowel niet als wel gerandomiseerde onderzoeken geven geen

Noten

⁸ Cobb LA, Thomas GI, Dillard DH, Merendino KA, Bruce RA. (1959). "An evaluation of internal-mammary-artery ligation by a double-blind technic.". *New England Journal of Medicine* 260 (22): 1115-8.

⁹ Kaptchuk, T.J. (1998) Intentional ignorance: a history of blind assessment and placebo controls. *Bull. Hist. Med.* 72:389-433

consistent hogere effectresultaten van het behandelingseffect en de variatie tussen wel en niet gerandomiseerd leek niet groter te zijn dan tussen RCT's onderling.¹⁰

In theorie is er geen systematische vertekening bij RCT's. In de praktijk blijkt dit toch het geval te zijn. Alleen al het feit dat een patiënt weet dat hij of zij aan een onderzoek meedoet leidt tot vertekening. Dit is gebleken uit een onderzoek in Frankrijk in de 1990, net voordat het informed consent daar verplicht werd. Een aantal patiënten in het ziekenhuis die leden aan slapeloosheid werd in twee groepen verdeeld. De ene groep werd verteld dat ze aan een RCT meededen en dat ze willekeurig een placebo of een nieuw slaapmiddel kregen. De andere groep werd niets verteld. Iedereen kreeg een placebo. De groep die niets was verteld sliep significant beter dan de RCT-groep.¹¹ Een ander soortgelijk onderzoek, bij patiënten met matige pijn ten gevolge van kanker, waarbij at random een pijnstillers of een placebo werd gegeven, leverde precies het omgekeerde resultaat op. Het effect van zowel de pijnstiller als het placebo bleek in de RCT groep groter te zijn dan in de groep die niet geïnformeerd was.¹² Alleen al het informed consent vertekent de uitkomsten van onderzoek. Ook kan het informed consent leiden tot een selectiebias. Zo blijken hoogopgeleiden vaker deelname te weigeren aan een RCT dan laagopgeleiden.

RCT's zijn zeer kostbaar. De prioritering van onderzoek wordt dan ook onvermijdelijk beïnvloed door sponsors, die hun belangen hebben. De farmaceutische industrie is een belangrijke, zo niet de belangrijkste, sponsor aangezien voor nieuwe geneesmiddelen RCT's door de registratie-autoriteiten worden geëist. Dit levert een bias ten gunste van medicamenteuze behandelingen met geotrooieerde, dure geneesmiddelen.

Een ander belangrijk probleem, dat overigens niet uitsluitend voor RCT's geldt is, dat het onderzoek plaats vindt op groepsniveau. Echter individuen kunnen substantieel afwijken van populatienormen: de gemiddelde patiënt bestaat niet. Daarbij worden bepaalde populatiesegmenten systematisch verwaarloosd: kinderen, vrouwen en patiënten met comorbiditeit. De meeste RCT's vinden plaats bij blanke mannen tussen de 20 en 50 jaar, terwijl ouderen met multimorbiditeit verreweg de grootste patiëntengroep zijn. Overigens geldt dit ook bij laboratoriumonderzoek: proefdieren zijn bijna altijd mannelijk.¹³

Niettegenstaande al de hiervoor geschetste tekortkomingen en problemen met RCT's, zijn ze wel een waardvol instrument voor het ontmaskeren van onwerkzame behandelingen. Een aantal voorbeelden verderop in de tekst illustreert dit.

De N=1 trial

De N=1 trial, een gerandomiseerd experiment bij één enkele patiënt, werd in 1986, 6 jaar voor de introductie van het begrip EBM, door Gordon Guyatt et al geïntroduceerd in een artikel in de New

Noten

¹⁰ McKee et al. (1999) Interpreting the evidence: choosing between randomized and non-randomized studie. BMJ, 319:312-15

¹¹ Dahan, R. et al. (1986) Does informed consent influence therapeutic outcome: A clinical trial of the hypnotic activity of a placebo in patients admitted to hospital. BMJ 293:363-4

¹² Bergmann, J.F. et al.(1994) A Randomized clinical trial of the effect of informed consent on the analgesic activity of placebo and naproxen in cancer pain. Clin. Trials Meta Anal 29:41-7

¹³ Ritskes-Hoitinga, M, Bont, O.de, Lagro-Jansen, T.(2016) Medicijnonderzoek blind voor vrouwen. Medisch Contact 18 februari

England Journal of Medicine.¹⁴ De patiënt krijgt afwisselend het ene middel en een ander middel. De patiënt (en idealiter ook de behandelaar) niet weet welke periode hij of zij het ene of het andere middel krijgt. Om het placebo-effect na te gaan kan met een niet-werkzaam middel afgewisseld worden.

De N=1 trial kan gezien worden als ideale en ultieme onderzoekopzet voor de behandeling van de individuele patiënt. Immers de variatie tussen patiënten wordt uitgeschakeld en er kan een uitspraak gedaan worden over de effectiviteit van een behandeling bij deze patiënt. Nu kent deze vorm van onderzoek wel een fors aantal beperkingen. Zo moet de toestand van de patiënt na de behandeling met het ene middel terugkeren naar de 'begintoestand'. Na bijvoorbeeld staken van de behandeling met een bepaalde bloeddrukverlager, moet de bloeddruk weer stijgen tot exact dezelfde waarde voor de behandeling. Immers om te kunnen vergelijken met een ander middel moet de beginsituatie gelijk zijn. Dit beperkt de mogelijke toepassing van de N=1 trial dan ook tot behandelingen die alleen symptomen bestrijden en geen genezend effect hebben en waarvan ernst en aard van de aandoening in de tijd niet snel verandert. Nu wil het geval dat dit, helaas, bij veel chronische aandoeningen het geval is. Toch wordt de N=1 trial bij chronische aandoeningen weinig toegepast.

3.5 Sterke en zwakke punten van de verschillende onderzoeksmethoden

In de vorige paragrafen zijn voor de verschillende onderzoeksmethoden reeds een aantal sterke en zwakke punten aangegeven. Ho et al hebben deze op een rij gezet, zie tabel 1.¹⁵ Het moge duidelijk zijn dat 'de ideale' onderzoekopzet niet bestaat.

Noten

¹⁴ Guyatt, G et al. (1986) Determining optimal therapy. Randomized trials in individual patients. N. Engl J Med. 314:889-92

¹⁵ Ho, P.M., Peterson, P., Masoudi, F.A. (2008) Evaluating the Evidence: Is There a rigid Hierarchy? Circulation, 118:1675-1684

Tabel 1 Sterke en zwakke kanten van verschillende onderzoeksmethoden (Naar Ho et al)¹⁶

Studieopzet	Sterke kanten	Zwakke kanten
RCT's	Grote interne validiteit Controle over de interventie Controle over de eindpunten	Geringe externe validiteit door strenge selectiecriteria Bias doordat deelnemers vrijwillig participeren Kostbaar en tijdrovend In veel situaties niet toepasbaar
Prospectieve cohortstudies	Kan verschillende eindpunten evalueren Controle over te kiezen onderzoekspopulatie Controle over de meting van variabelen en eindpunten Volgt ontwikkeling in de tijd	Vereist groot aantal deelnemers Kostbaar Tijdrovend Minder geschikt voor infrequent voorkomende uitkomsten Risico op verstorende factoren (confounding)
Retrospectieve cohortstudies	Kan verschillende eindpunten evalueren Volgt ontwikkeling in de tijd Goedkoper en minder tijdrovend dan prospectieve cohortstudies	Vereist groot aantal deelnemers Minder geschikt voor infrequent voorkomende uitkomsten Minder controle over te kiezen onderzoekspopulatie en meetwaarden Risico op verstorende factoren (confounding)
Case-control studies	Geschikt voor zeldzame uitkomsten Kortdurend Kleinschalig en goedkoop	Grote kans op verstorende factoren (confounding) Kans op vertekening (bias) v.w.b. de selectie van controles, overleving en herinnering
Cross-sectional studies	Kan verschillende eindpunten evalueren Kortdurend Kan een schatting van voorkomen van een uitkomst geven	Groot risico op verstorende factoren (confounding) en vertekening (bias) Minder geschikt voor weinig frequent voorkomende zaken Kan geen verband in de tijd aantonen
Case-series	Eenvoudig Goedkoop	Een vergelijkende groep ontbreekt, hetgeen het aantonen van causaliteit sterk beperkt. Niet mogelijk om risico, incidentie en prevalentie te bepalen.
N=1 studie	Ideaal om de effectiviteit van verschillende behandelingen bij de individuele patiënt te bepalen	Slechts in een zeer beperkt aantal situaties toepasbaar: symptomatische behandeling van chronische aandoeningen onder specifieke condities..

3.6 Systematische reviews en meta-analyses

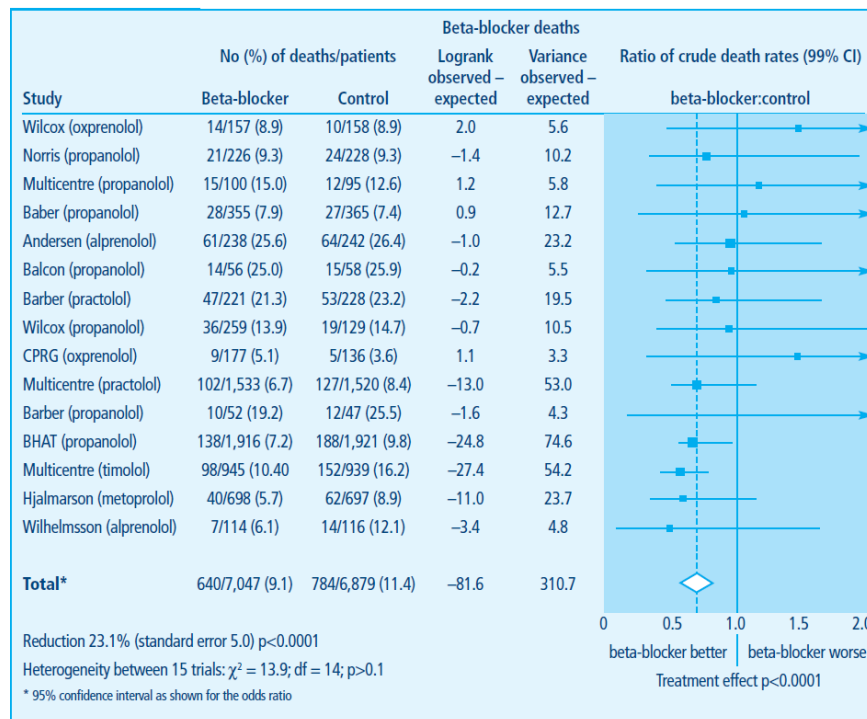
Systematische reviews geven een uitgebreid overzicht van kwalitatief hoogwaardige onderzoeken. Bij meta-analyses worden deze studies statistisch gecombineerd om zo een overall effect te berekenen.

Bekend zijn de Cochrane reviews. Dit zijn systematische reviews en meta-analyses door de Cochrane collaboration. Dit is een internationaal netwerk van meer dan 28.000 wetenschappers in meer dan 100 landen dat is opgericht in 1993. Het is genoemd naar Archie Cochrane (1909-1988), een Schotse arts en pionier op het gebied van evidence-based medicine.

Noten

¹⁶ Ho, P.M., Peterson, P.N., Masoudi, F.A. (2008) Evaluating the Evidence. Is There a Rigid Hierarchy? *Circulation*. 118:1675-1684
Het Bewijs

Bij een Cochrane review worden via literatuuronderzoek alle wetenschappelijke publicaties over een bepaald onderwerp verzameld en volgens een vast stramien geanalyseerd: de kwaliteit van de onderzoeksopzet, statistische bewerkingen etc. Het resultaat van de meta-analyse wordt meestal in de vorm van een zgn. forest plot weergegeven, zoals onderstaand voorbeeld¹⁷.



Bij Cochrane reviews wordt in het strikt de EBM-hiërarchie gehanteerd. Dit betekent dat als het om effectiviteit van geneesmiddelen gaat, bijvoorbeeld betablokkers in het voorbeeld, het bewijsmateriaal vrijwel volledig uit RCT's bestaat.

Noten

¹⁷ Lewis, S., Clarke, M. (2001) Forest Plots: trying to see the wood and the trees. *BMJ* 322:1479-1480
Het Bewijs

4 Causaliteit

"Beyond such discarded fundamentals as 'matter' and 'force' lies still another fetish amidst the inscrutable arcana of modern science, namely the category of cause and effect"

[K. Pearson]

4.1 Inleiding

Uit het voorgaande moge blijken dat de verschillende vormen van bewijs ieder hun sterke en zwakke kanten kennen. Met name de binnen EBM als hoogst gewaardeerd bewijs, de (meta-analyse van) RCT's blijkt in de praktijk nogal wat beperkingen te kennen.

De vraag is evenwel of (al dan niet hèt) bewijs wel het grootste probleem is, zowel binnen als buiten de gezondheidszorg. Volgens John Maynard Keynes, de befaamde econoom uit de eerste helft van de vorige eeuw, wordt de wereld bepaald door ideeën, ideeën van intellectuelen:

"The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed, the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influences, are usually the slaves of some defunct economist. Madmen in authority, who hear voices in the air, are distilling their frenzy from some academic scribbler of a few years back. I am sure that the power of vested interests is vastly exaggerated compared with the gradual encroachment of ideas soon or late, it is ideas, not vested interests, which are dangerous for good or evil."

Dat ideeën, waaronder eveneens 'volkswijsheden' gerekend kunnen worden, binnen een onderzoekszopzet spreken we van hypothesen, ook binnen de gezondheidszorg van oudsher een centrale rol spelen werd reeds geïllustreerd aan het voorbeeld van het aderlaten dat gebaseerd was op het idee van de vier lichaamsvochten, de zgn. *humoraalpathologie*. Er zijn nog legio andere ideeën te noemen die een rol in de geneeskunde hebben gespeeld of nog spelen, zoals de leer der Kritische dagen, de antroposofie, homeopathie etc.

Binnen de natuurwetenschappelijke benadering van de geneeskunde vormen de verschillende onderzoeksmethoden uit de EBM-pyramide de instrumenten om de verschillende ideeën cq. hypothesen te toetsen op hun 'juistheid', in die zin dat ze de juiste causale verbanden weergeven.

Heden ten dage is het als vanzelfsprekend om de humoraalpathologie als 'onzin' terzijde te schuiven, aangezien de daarin veronderstelde causaliteit kant noch wal raakt. De vraag rijst of de hypothesen die heden ten dage voor waar worden aangenomen - mede op basis van de EBM-methodologie - wel juist zijn cq. de juiste causaliteit weergeven. Aan de hand van een voorbeeld, de cholesterolhypothese wordt geïllustreerd dat het heel lastig blijkt 'de waarheid te vinden.'

4.2 De cholesterolhypothese: *hèt* bewijs?

Cholesterolverlagers - statines - worden aan veel verder gezonde mensen voorgeschreven omdat inmiddels overtuigend is aangetoond dat deze geneesmiddelen aderverkalking - atherosclerose - en daarmee hart- en vaatziekten tegengaan. De richtlijn van de Nederlandse huisartsen voor de preventie van hart- en vaatziekten, De NHG standaard Cardiovasculair risicomangement (eerste herziening) beveelt aan om een statine voor te schrijven als het 'slechte cholesterolgehalte', het zgn. LDL-gehalte, in het bloed groter is dan 2,5 mmol/l.¹⁸ Nederland telde in 2013 1,9 miljoen statinegebruikers met een totale kosten van 213 miljoen euro.¹⁹

Zoals gezegd staat het nut van statines in de mainstream gezondheidszorg niet meer ter discussie. Het is interessant te bezien hoe deze discussie is verlopen. Het begon in 1910 toen Windaus ontdekte dat de buikslagaders van patiënten met atherosclerose meer cholesterol bevatte dan van gezonde personen.²⁰ Drie jaar later toonde Anitschkow aan dat cholesterol toegediend aan konijnen het cholesterolgehalte in het bloed verhoogde en atherosclerose veroorzaakte.²¹ In 1939 ontdekte Muller dat er families bestonden bij wie het cholesterol in het bloed sterk verhoogd was. Bij hen kwamen hart- en vaatziekten veel vaker voor.²² In 1950 toonde Gofman et al. aan dat bij deze patiënten het cholesterol zich vooral in bepaalde deeltjes, de zgn. LDL-fracties van het bloed bevond en dat er meer hart- en vaatziekte optrad bij hogere concentraties.²³ Barr et al. ontdekten een jaar later dat de hoeveelheid van andere cholesterolhoudende deeltjes, HDL, juist omgekeerd gerelateerd was aan de kans op hart- en vaatziekte: hoe hoger het HDL hoe minder kans. Het mechanisme werd opgehelderd: cholesterol wordt voor een belangrijk deel in de lever gemaakt en LDL zorgt via het bloed voor de aanvoer naar de weefsels die het nodig hebben. Overtollig cholesterol wordt door HDL afgevoerd naar de lever. De hypothese was is dat als het LDL is verhoogd en het HDL verlaagd, cholesterol wordt afgezet in de bloedvaten. Dit leidt tot een ontstekingsreactie en atherosclerose.

Een zeer grote in 1948 gestarte cohortstudie in de Verenigde Staten, de Framingham Heart Study, genoemd naar een stadje in Massachusetts, liet in 1961 zien dat er een duidelijk verband was tussen een (licht) verhoogd cholesterolgehalte in het bloed en het optreden van hart- en vaatziekte.²⁴

Het onderzoek dat algemeen gezien wordt als doorslaggevend was de 'zeven landenstudie' van Ansell Keys.²⁵ Het liet zien dat hoe lager het gemiddelde cholesterolgehalte van de populatie in een land, hoe lager het aantal doden ten gevolge van hart- en vaatziekten. Een uitzondering vormde Frankrijk: de *French paradox*. Het gemiddelde cholesterolgehalte was daar hoog, terwijl er toch

Noten

¹⁸ NHG-Standaard Cardiovasculair risicomangement (eerste herziening). (2012) Huisarts Wet 55(1)14-28

¹⁹ www.zorgatlas.nl (2017)

²⁰ Windaus, A. (1910) Über den Gehalt normaler und atheromatoser Aorten an Cholesterin und Cholesterinestern. Z. Physiol Chem 67:174-176

²¹ Anitschkow, N.N. (1913) Über experimentelle Cholesterinsteatose und ihre bedeutung für die Entstehung einiger pathologischer Prozesse. Zentralblatt Allgemeiner Pathologie 24:1-9

²² Muller, L. (1939) Angina Pectoris in hereditary xantomatis. Arch Intern Med 64:675-700

²³ Gofman et al. (1950) The role of lipids and lipoproteins in atherosclerosis Science 111:166-171

²⁴ Kannel, W.B. et al. (1961) Factors of risk in the development of coronary heart disease - six year follow up experience. The Framingham Study. Ann. Intern. Med. 55:33-50

²⁵ Keys A. (1980), Seven Countries: A multivariate Analysis of Death and Coronary Heart Disease. Cambridge. Harvard University Press

relatief weinig hart- en vaatziekte voorkam. Dit werd toegeschreven aan het wijngebruik in Frankrijk, dat beschermend zou werken.

Overtuigend bewijs kwam in 1984 toen werd aangetoond dat het geneesmiddel cholestyramine, dat cholesterol in het darmkanaal bindt zodat het niet in het lichaam opgenomen kan worden waardoor het cholesterolgehalte in het bloed met ca. 10% daalt, leidde tot 20% minder kans op hart- en vaatziekte.²⁶

Het verhaal van de statines begint met simvastatine (Zocor®), dat in 1989 door de firma MSD op de Nederlandse markt werd gebracht. Het middel verlaagde het cholesterolgehalte in het bloed met 25% door het remmen van het enzym HMG-CoA dat betrokken is bij de synthese van cholesterol in de lever. De relatieve kans op sterfte door hart- en vaatziekte nam met 42% af.²⁷ Het middel was een uitermate groot commercieel succes. Concurrenten ontwikkelden soortgelijke middelen die nog krachtiger het cholesterol verlaagden, zoals atorvastatine (Lipitor®) en rosuvastatin (Crestor®).

Cholesterol als oorzaak van atherosclerose was hiermee bewezen stelde Steinberg in 2006.²⁸ Een visie die breed door de medische wereld werd en nog steeds wordt gedragen.

4.3 De cholesterolhypothese: de losse eindjes

Zoals gezegd staat de cholesterolhypothese in de mainstream gezondheidszorg niet meer ter discussie. Er zijn echter wel een paar losse eindjes. Hiervoor werd al even de *French Paradox* genoemd die uit het onderzoek van Ansell Keys naar voren kwam. Later is gebleken dat Keys aan cherry-picking had gedaan. Hij had 21 landen bestudeerd en die landen uitgekozen, waarbij hij moeilijk om Frankrijk heen kon, die pasten binnen de cholesterolhypothese.

Op basis van de cholesterolhypothese werden alle cholesterolmiddelen via RCT's getest op hun cholesterolverlagende eigenschap als surrogaatparameter voor de kans op hart- en vaatziekte onder het motto: hoe lager het cholesterol hoe beter. In dit kader werd dan ook een onderzoek gedaan met combinatie van middelen om het cholesterolgehalte van mensen met familiale hypercholesterolemie nog sterker te verlagen. In een RCT waarin een statine al dan niet werd gecombineerd met het middel torcetrapib, bleek, alhoewel het LDL gehalte daardoor met nog eens 35% verminderde en het HDL met 35% steeg, dat de combinatie slechtere uitkomsten had.²⁹ Dit werd verklaard doordat het middel als bijwerking had een verhoging van de bloeddruk, waardoor de kans op o.a. beroerten toenam, hetgeen het positieve effect van de extra cholesterolverlaging meer dan teniet deed.³⁰ Echter ook een combinatie van een statine met een andere stof, ezetimide die

Noten

²⁶ The Lipid Research Clinics Coronary Primary Prevention Trial results II. (1984) The relationship of reduction in incidence of coronary heart disease to cholesterol lowering. *J. Am. Med. Assoc.* 251:365-374

²⁷ Scandinavian Simvastatin Survival Study Group. (1994) Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S) *Lancet* 344:1383-1389

²⁸ Steinberg, D. (2006) An interpretive history of the cholesterol controversy: Part V: The discovery of statins and the end of the controversy. *J. Lipid Res.* 47:1339-1351

²⁹ Kastelein, J.J.P. et al. (2007) For the RADIANCE 1 Investigators: Effect of Torcetrapib on Carotid Atherosclerosis. *N. Engl. J. Med.* 356:1620-1630

³⁰ Bots, M.L. et al. (2007) Torcetrapib and carotid intima-media thickness in mixed dyslipidaemia (RADIANCE 2 study): a randomised, double-blind trial. *Lancet* 370(9582):153-60

deze bijwerking niet heeft, bleek slechtere uitkomsten te geven.³¹ Aangezien torcetramid en ezitimide een ander werkingsmechanisme heeft dan statines, rijst de vraag of het aangetoonde positieve effect dat statines hebben bij secundaire preventie, dus bij patiënten die reeds hart- en vaatziekten hebben, enkel het gevolg is van een verlaging van het cholesterolgehalte in het bloed. Statines lijken namelijk ook het immuunsysteem te beïnvloeden.³² Verder blijkt hormoonvervangende therapie tijdens de menopauze bij vrouwen, waardoor het LDL-gehalte in het bloed met gemiddeld ca. 11% daalt en HDL met 10% stijgt, geen positief effect te hebben op de kans op hart- en vaatziekte.³³

Tot slot bleek er geen significant verschil te bestaan tussen kans op overlijden door een ziekte - *all-cause mortality* - tussen degenen die statines gebruikten en de placebo groep.³⁴

4.4 Oorzaak, gevolg of een samenloop van omstandigheden?

Het cholesterolverhaal illustreert goed een veelvuldig voorkomende probleem in de geneeskunde: "Wat is oorzaak en wat is gevolg?"

In geval van cholesterol lijkt het zo simpel: er wordt een verband gevonden tussen het cholesterolgehalte in het bloed en het vóórkomen van hart- en vaatziekten. Vervolgens blijkt dat het actief verlagen van verhoogd cholesterol met statines de kans op hart- en vaatziekte de kans verlaagt. Dus is cholesterol de, of in elk geval een, oorzaak van hart- en vaatziekte.

Er zijn altijd wel een paar mensen die kritisch zijn en onderzoeksresultaten die niet passen, zoals de 'losse eindjes' die in de vorige paragraaf zijn beschreven, niet negeren, maar nader onderzoeken. Zij zijn degenen die volgens Popper zorgen voor vooruitgang in de wetenschap, namelijk door het trachten hypothesen te falsificeren. Helaas wordt maar al te vaak gezocht naar bevestiging van hypothesen, waarbij men, zoals het voorbeeld van Keys aangeeft, zelfs fraude niet schuwt.

Er wordt dan ook gezocht naar alternatieve hypothesen die alle waargenomen feiten, ook die welke niet passen in de oorspronkelijke hypothese, kunnen verklaren. Een voorbeeld is de 'Mevolonate hypothese' als alternatief voor de cholesterolhypothese.³⁵ Deze gaat uit van ontsteking als oorzaak van hart- en vaatziekten en verklaart het gevonden verband tussen cholesterol en hart- en vaatziekten, terwijl het cholesterol niet de oorzaak is. Als deze hypothese pogingen tot falsificatie doorstaat, dan zou dit mogelijk tot geheel andere therapeutische interventies kunnen leiden. Maar dan moeten er uiteraard wel pogingen tot falsificatie ondernomen worden.

Maar ook gewoon met gezond verstand kan men vraagtekens zetten bij de massale behandeling van patiënten met statines. Het enzym HMG-CoA reductase dat door statines wordt geremd is een

Noten

³¹ Kastelein, J.J.P. For the (2008) ENHANCE Investigators: Simvastatin with or without Ezetimibe in Familial Hypercholesterolemia. *N.Engl. J. Med.* 358: 1431-1443

³² Palinski, W. (2000) Immunomodulation: a new role for statins? *Nature Medicine* 6:1311-1312

³³ Hulley, S et al. (1998) Randomized Trial of Estrogen Plus Progestin for Secondary Prevention of Coronary Heart Disease in Postmenopausal Woman. *JAMA* 280:605-13

³⁴ Diamond D.M., Ravnskov, U. (2015) How statistical deception created the appearance that statins are safe and effective in primary and secondary prevention of cardiovascular disease. *Expert. Rev. Clin. Pharmacol.* 8:189-199

³⁵ Keizer, H.G. (2012) The Mevalonate hypothesis: a cholesterol-independent alternative for the etiology of atherosclerosis. *Lipids in Health and Disease* 11:149

cruciale stap in de aanmaak van vele stoffen in de cel, naast cholesterol o.a. ook vitamine K, coenzym Q10 en alle steroidhormonen. Door behandeling met een statine worden al deze processen beïnvloed. Het is onmogelijk om met aantal kortdurende onderzoeken, RCT's duren slechts enkele jaren, de effecten van een dergelijk geneesmiddel op de lange termijn in te schatten. Behandeling met statines als secundaire preventie, dus bij patiënten die reeds een hart- of vaatziekte hebben is op basis van dergelijke onderzoek goed te verdedigen, de tijdshorizon is hierbij immers kort. De vraag is echter of het verantwoord is een groot deel van de gezonde westerse bevolking aan een dergelijk experiment, dat ook nog eens zeer kostbaar is, bloot te stellen terwijl de uitkomsten op lange termijn ongewis zijn.

4.5 Conclusie

In de aanloop naar de hypothese en 'het bewijs' van de cholesterolhypothese zien we alle vormen van bewijs uit de EBP-piramide voorbij komen: casusbeschrijvingen, experimenteel onderzoek, meningen van experts, cohortonderzoek tot en met RCT's. Bij de cholesterolcasus zagen we dat RCT's met name ingezet werden bij het bepalen van de effectiviteit van geneesmiddelen c.q. statines. Hierbij werd cholesterolverlaging als surrogaat maat genomen voor de uitkomst van de behandeling en op basis van de uitkomsten wordt een groot deel van de westerse wereldbevolking met statines behandeld.

Echter zoals hiervoor is aangegeven, zijn er gerede twijfels omtrent het nut van statinebehandeling op harde eindpunten op de langere termijn: leeft men er uiteindelijk langer en gezonder door. Deze vraag kan niet op basis van RCT's, die noodzakelijkerwijs van relatief korte duur zijn, beantwoord worden. Andere onderzoeksvormen, zoals cohortstudies zijn hier wel geschikt voor.

RCT's zijn wel zeer waardevol als het gaat om het ontmaskeren van onwerkzame therapieën. Zo zagen we dat bij de werking van de wereldwijd meest gebruikte pijnstillers, paracetamol, vraagtekens gezet kunnen worden.

Geconcludeerd kan worden dat elke onderzoeksmethode zijn sterke en zwakke kanten heeft. Daarnaast staat een onderzoek niet op zichzelf. Als bijvoorbeeld in een RCT gebruik wordt gemaakt van een surrogaat maat, dan valt of staat de bewijskracht met de validiteit hiervan. Hetzelfde geldt voor systematische reviews en meta-analyses van RCT's, die binnen de EBM-piramide op de hoogste trede staan. Eén cohortstudie die zich richt op harde uitkomsten levert waardevoller bewijs dan een meta-analyse van 100 RCT's die op een verkeerde surrogaat maat zijn gebaseerd.

5 Is er een alternatief voor de EBM-piramide?

5.1 Inleiding

In de inleiding werd de vraag gesteld: bestaat *hèt* bewijs? In de voorgaande hoofdstukken is aannemelijk gemaakt dat het antwoord ontkennend is en dat de evidence-pyramide zoals die binnen de evidence-based benadering kritisch bezien dient te worden. Elke vorm van bewijs, of dit nu voortvloeit uit de casusbeschrijving, laboratoriumonderzoek of RCT, heeft zijn sterkte en zwakte. We zagen dat een hypothese heel krachtig is, of deze nu juist is of niet en de vormgeving van RCT's beïnvloed. In de EBM-hiërarchie staat de hypothese laag op de ladder. Voor een onbewezen hypothese, cq. een hypothese waar nog weinig moeite is gedaan deze te falsificeren, is dit terecht. Echter een hypothese die vele stormen heeft doorstaan verdient een hoge, misschien wel de hoogste, plaats.

Omdat RCT's zo hoog op de hiërarchische ladder staan, staan behandelingen waarvan het effect met RCT's bepaald kan worden, met name geneesmiddelen, in het algemeen in hoger aanzien dan behandelingen waarbij dit niet kan. Onterecht zoals het voorbeeld van statines laat zien. De effecten van behandelingen en preventie op de lange termijn (all-cause mortality) zullen uiteindelijk toch vooral van cohortstudies moeten komen.

Daarnaast is de toepassing van kennis verkregen uit met name RCT's problematisch, aangezien het uitkomsten op groepsniveau betreft, waarbij de onderzochte groepen vaak sterk verschillen van de groepen uit de dagelijkse praktijk, waarbij bijvoorbeeld vaak sprake is van multimorbiditeit. Daarnaast moet zorgverlener de resultaten op groepsniveau – de gemiddelde patiënt - trachten te vertalen naar de individuele patiënt, want daar gaat het in de gezondheidszorg uiteindelijk om.

In dit hoofdstuk worden een tweetal alternatieven besproken, het GRADE-systeem en de Reversed pyramid. Tot slot wordt nog een derde alternatief besproken, namelijk het legpuzzelmodel.

5.2 Het GRADE-systeem

Binnen de evidence-based community realiseert men zich dat de EBM-hiërarchie niet absoluut is. Om de kwaliteit van de verschillende vormen van bewijs te wegen is het zgn. GRADE (Grading of Recommendations Assessment, Development and Evaluation)-systeem ontwikkeld. Binnen dit systeem worden vijf niveaus van kwaliteit, A tot en met D, van bewijs onderscheiden.³⁶

— A. Hoge kwaliteit van het bewijs

Deze hoogste categorie van bewijs is gedefinieerd als: 'Het is zeer onwaarschijnlijk dat verder onderzoek ons vertrouwen in de schatting van het effect zal veranderen.' Voorbeelden die

Noten

³⁶ GRADE Working Group (2007) Grading of Recommendations Assessment, Development and Evaluation, modified by the EBM Guidelines Editorial Team
Het Bewijs

daarbij genoemd worden zijn een aantal onderzoeken van hoge kwaliteit die consistente resultaten laten zien. In speciale gevallen kan het één grote, kwalitatief hoogwaardige multicentre onderzoek. Ook goed uitgevoerde observationele studies die grote effecten laten zien kunnen dit bewijs leveren.

— B. Matige kwaliteit van het bewijs

Hierbij 'is het waarschijnlijk dat nader onderzoek een belangrijke invloed heeft op ons vertrouwen in de schatting van het effect en dit de schatting kan veranderen. Voorbeelden zijn één hoogwaardig onderzoek en meerdere studies met enkele tekortkomingen.

— C. Lage kwaliteit van het bewijs

In dit geval is het zeer waarschijnlijk dat nader onderzoek een belangrijke invloed heeft op ons vertrouwen in de schatting van het effect en het is waarschijnlijk dat daardoor de schatting veranderd. Voorbeelden zijn één of meer onderzoeken met ernstige tekortkomingen.

— D. Zeer lage kwaliteit van het bewijs

Dit deze situatie is elke schatting van een effect zeer onzeker. Als voorbeelden worden genoemd: expert opinion, het ontbreken van direct wetenschappelijk bewijs en één of meer onderzoeken met zeer ernstige tekortkomingen.

Binnen de definities staat het begrip 'ons vertrouwen' centraal. Dit sluit naadloos aan bij de betekenis van 'evidence' zoals in de achtergrondstudie *Betekeningen van bewijs* van Gabie de Jong nader is beschreven: "The facts, signs or objects that make you believe that something is true." Dus het uiteindelijke resultaat is de mening van (een groep van) experts. In hoeverre bijvoorbeeld zorgverleners handelen op basis van richtlijnen is weer afhankelijk van het vertrouwen dat zij hebben in de mening van de experts die ze opgesteld hebben.

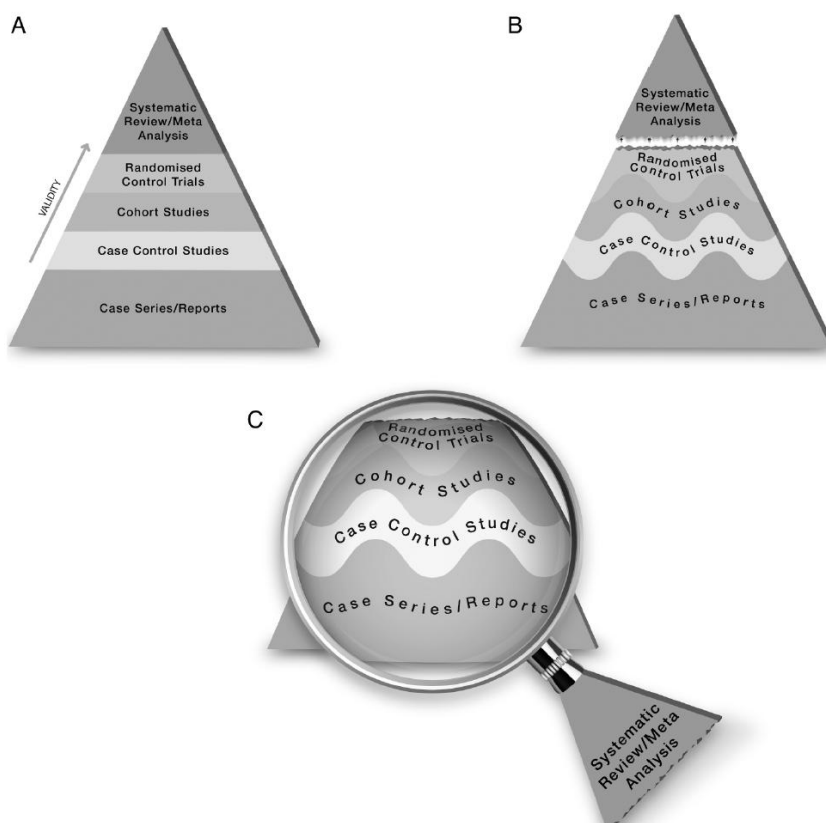
5.3 The revised evidence pyramid

Binnen het GRADE systeem is de strikte hiërarchie van bewijskracht reeds deels verlaten. Zo kan de bewijskracht van een observationele studie die grote effecten laat zien hoger gewaardeerd worden dan een RCT met heel kleine effecten. Recentelijk hebben Murad et al een voorstel voor een herziening van de EBM-pyramide gedaan.³⁷ In figuur 3 is dit schematisch weergegeven. In Figuur 3A is de klassieke pyramide weergegeven. Belangrijk is de 'aftopping' van de pyramide (figuur 3B): de systematische reviews en meta-analyses maken geen deel meer uit van de pyramide, maar vormen het vergrootglas of de bril waarmee naar al het aanwezige bewijs wordt gekeken (figuur 3C). Bij het uiteindelijke oordeel wordt alle beschikbare informatie meegenomen.

Noten

³⁷ Murad, M.H., Asi, N., Alsawas, M. Alahdab, F. (2016) New evidence pyramid. Evidence Based Medicine, Online First, 10.1136/ebmed 2016-110401

Figuur 3 : The revised evidence pyramid (Murad et al.)



5.4 No pyramid at all: het legpuzzelmodel

We zouden nog een stap verder kunnen gaan en de hiërarchie geheel afschaffen. Wat resulteert is een aantal bewijzen in de vorm van resultaten van onderzoek naar een bepaald fenomeen van verschillende aard en omvang, zoals resultaten van case series, proefdieronderzoek, cohortstudies en RCT's die in de loop der tijd zijn verricht. Elk van deze 'puzzelstukjes van bewijs' hebben hun sterke en zwakke kanten. Door ze bij elkaar te leggen ontstaat een beeld. Dit is in eerste instantie heel wazig, maar nieuw onderzoek zal steeds nieuwe puzzelstukjes aanleveren, waardoor het beeld steeds helderder wordt, of juist niet. Zo kan blijken dat bepaalde puzzelstukjes die oorspronkelijk als 'hoekpunten van de puzzel' werden gezien helemaal niet blijken te passen. Vanuit de Popperiaanse gedachte zijn juist niet-passende puzzelstukjes het interessantst. Zij kunnen ertoe leiden dat de puzzel geheel opnieuw opgebouwd moet worden.

Bij een legpuzzel die we in de winkel kopen, zijn, als het goed is, alle puzzelstukjes in de doos aanwezig en passen ze, als we ze goed neerleggen, precies in elkaar en krijgen we een compleet en haarscherp beeld. Binnen de (bio-)wetenschap ligt dit veel gecompliceerder. We moeten zelf door onderzoek de puzzelstukjes ontdekken, waarbij we niet zeker zijn over de vorm en de inhoud van het puzzelstukjes. We weten ook niet uit hoeveel puzzelstukjes de puzzel bestaat. Het beeld zal nooit haarscherp zijn en we zullen dan ook nooit weten wanneer de puzzel compleet is en we het volledige beeld hebben en of we überhaupt wel de juiste puzzel hebben opgelost.



In de inleiding werd de vraag gesteld of er zoiets bestaat als *het* bewijs. Uit het voorgaande moge duidelijk zijn dat dit niet bestaat. Als we *het* bewijs zien als de compleet opgeloste puzzel, dan moge het duidelijk zijn dat wij dat ideaal nooit kunnen bereiken. Maar met een gedeeltelijk opgeloste puzzel kunnen we al ver komen.



Afkortingen

NRC	Nieuwe Rotterdamsche Courant
EBM	Evidence Based Medicine
EBP	Evidence Based Practice
JAMA	Journal of de American Medical Association
RCT	Randomized Controlled Trial
WHO	World Health Organization
FDA	Food and Drug Administration
ERGO	Erasmus Rotterdam Gezondheid Onderzoek
RVS	Raad voor Volksgezondheid en Samenleving
LASA	Longitudinal Aging Study Amsterdam
PIAMA	Preventie en Incidentie van Astma en Mijt Allergie
LDL	Low Density Lipoproteins
HDL	Hight Density Lipoproteins
HMG-CoA	3- H ydroxy-3- M ethylglutaryl- C oenzyme A
Q10	Quinon 10
GRADE	Grading of Recommendations Assesment, Development and Evaluation