

TER DISCUSSIE

De rol van epidemiologisch bewijs in de zorg voor individuele patiënten

Yvo M. Smulders, Marcel Levi, Coen D.A. Stehouwer, Mark H.H. Kramer en Abel Thijs

Vanuit de maatschappij, overheid en zorgverzekeraars klinkt in toenemende mate de roep om normering en standaardisatie van het geneeskundig handelen. Hierdoor ‘verprotocolleert’ de zorg en neemt keuzevrijheid van zorgverleners af.

Het overheersende normeringinstrument voor regulatie en standaardisatie is epidemiologisch bewijs, oftewel ‘evidence’. Evidence bestaat uit empirische gegevens verkregen in groepen personen, bijvoorbeeld in cohortstudies of gerandomiseerde klinische trials. Veelal wordt ‘evidence-based medicine’ (EBM) uitgelegd als geneeskundig handelen waarbij besluitvorming bij een individuele patiënt is gebaseerd op het zo direct mogelijk vertalen van epidemiologisch bewijs. De epidemioloog Sackett, die aan de wieg stond van de EBM-beweging, hanteert echter een bredere omschrijving: ‘integratie van individuele klinische expertise met het best beschikbare bewijs uit epidemiologisch onderzoek.’¹ Deze definitie van Sackett doet meer recht aan de praktijk. Rechtlijnige vertaling van epidemiologisch bewijs naar de klinische praktijk is vaak onmogelijk, en waar wel mogelijk doorgaans riskant.

In dit artikel beargumenteren wij dat epidemiologisch bewijs als normeringsinstrument voor individuele zorg ernstige beperkingen heeft. Tevens bespreken wij hoe epidemiologisch bewijs de zorg beter ten goede kan komen.

BEWIJS

Onze bezwaren tegen het heilig verklaren van epidemiologisch bewijs bij individuele zorgverlening berust op enkele peilers: we weten nog een heleboel niet en het bewijs is niet altijd betrouwbaar en toepasbaar.

HOEVEEL WETEN WE?

Sommigen veronderstellen dat het merendeel van ons geneeskundig handelen is gebaseerd op evidence. Het tegendeel is waar, zoals de figuur laat zien. Dit geldt ook voor deeldomeinen. Bij hart- en vaatziekten bijvoorbeeld bestaat een sterke traditie van epidemiologisch onderzoek. Een analyse van Amerikaanse richtlijnen wees echter uit dat van alle aanbevelingen slechts 11% was gebaseerd op krachtig epidemiologisch bewijs. De helft werd zelfs nauwelijks door epidemiologisch onderzoek ondersteund.² Deze analyse doet vermoeden dat epidemiologisch bewijs schaars is in andere domeinen van de geneeskunde met een minder sterke onderzoekstraditie. Deze bewijsschaarste heeft vele achtergronden, waaronder commerciële belangen en de lage prevalentie van veel ziektebeelden. Een verdere selectie ontstaat doordat epidemiologisch onderzoek zich meestal richt op therapeutisch handelen, waarbij schier eindeloze vergelijkingen met placebo noodzakelijke directe vergelijking van interventies vertroebelen.³ De aandacht voor farmacologische interventies verdringt ook onderzoek naar andere aspecten van therapeutisch handelen, zoals therapietrouw. Zo is er veel onderzoek naar het beste antihypertensivum, maar de kleine verschillen in effectiviteit vallen in het niet wanneer men beseft dat in de praktijk na 2 jaar ongeveer de helft van de patiënten is gestopt met het gebruik van ervan.⁴

Tenslotte is onderzoek naar diagnostiek schaars, wat temeer zorgelijk is omdat fouten in dit domein verantwoordelijk zijn voor meer gezondheidsschade dan fouten in therapeutische besluitvorming.⁵

DE BETROUWBAARHEID VAN EPIDEMIOLOGISCH BEWIJS

‘Al weten we dan niet zo veel, wat wel onderzocht is, weten we tenminste zeker.’ Of klopt die veronderstelling ook niet? Er zijn allerlei problemen met de veronderstelling dat we ons op uitkomsten van onderzoek kunnen

VU Medisch Centrum,

afd. Inwendige Geneeskunde, Amsterdam.

Prof.dr. Y.M. Smulders, prof.dr M.H.H. Kramer en

dr. A. Thijs, internisten.

Academisch Medisch Centrum,

afd. Interne Geneeskunde, Amsterdam.

Prof.dr. M. Levi, internist.

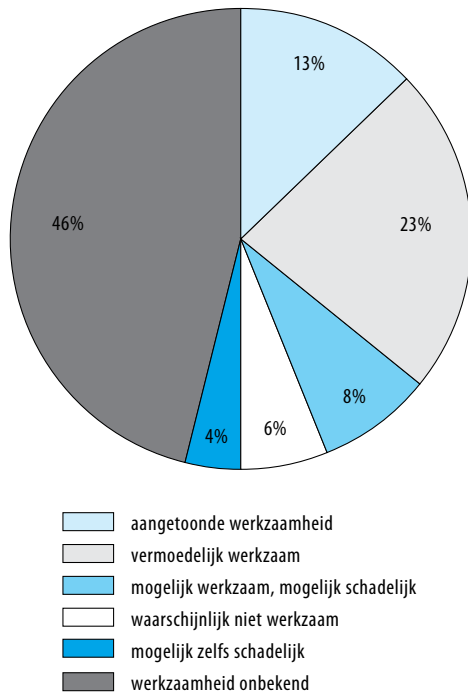
Maastricht Universitair Medisch Centrum,

afd. Interne Geneeskunde, Maastricht.

Prof.dr C.D.A. Stehouwer, internist.

Contactpersoon: prof.dr. Y.M. Smulders

(y.smulders@vumc.nl).



FIGUUR Mate van onderbouwing door epidemiologisch bewijs van veelvoorkomende therapeutische handelingen: bijna de helft is nimmer onderzocht en van slechts 13% is werkzaamheid ondubbelzinnig aangetoond. In de restgroep varieert de ondersteuning uit epidemiologisch onderzoek van 'vermoedelijk nuttig'(23%) tot 'waarschijnlijk nutteloos of zelfs schadelijk'(4%). (Bron: <http://clinicalevidence.bmj.com/ceweb/about/knowledge.jsp>).

verlaten, waaronder beperkingen van het onderzoek zelf, publicatiebias en – misschien wel het belangrijkste – plausibiliteit van de onderzoekshypothese.

Beperkingen van het onderzoek Epidemiologisch vergelijkend onderzoek met een statistisch significant resultaat wordt doorgaans als 'waar' beschouwd. De uitkomst heeft echter betrekking op de data zélf, niet op de wijze waarop data tot stand zijn gekomen. Hierbij kan namelijk vertekening ('bias') zijn opgetreden, die men niet terugziet in de uitkomst, noch in het significantieniveau. Het hele onderzoekstraject is kwetsbaar voor bias, vanaf de formulering van de onderzoeksvraag, selectie van deelnemers en selecties in onderzoeksmethodiek, uitkomstmaat, statistische analysemethoden tot rapportage van uitkomsten.

Selectie door externen Vertekening van de 'epidemiologische waarheid' treedt ook op door publicatiebias: selectie tijdens peer review en publicatie in tijdschriften, veroorzaakt door een voorkeur van referenten en redacties voor prestigieuze auteurs⁶⁷ en voor 'positieve' onderzoeksresultaten.⁸

Plausibiliteit Een zelden onderkende factor in het 'waarheidsgehalte' van epidemiologisch onderzoek betreft plausibiliteit van de onderzoekshypothese. Indien men 2 gelijkwaardige onderzoeken vergelijkt, waarvan de ene een plausibele en de andere een niet-plausibele hypothese bevestigt, dan heeft het eerstgenoemde onderzoek toch een veel grotere kans om waar (hier te lezen als: reproduceerbaar) te zijn. Om dit te begrijpen denke men aan de analogie met de regel van Bayes in de context van diagnostiek.⁹ De juistheid van een diagnose hangt volgens deze regel niet alleen af van de uitslag van een diagnostische test, maar ook van de voorafkansen op deze diagnose. Datzelfde geldt bij epidemiologisch onderzoek. De kans dat een onderzoeksuitkomst juist is, wordt mede bepaald door de vraag of die uitkomst aannemelijk is, onafhankelijk van het significantieniveau van de uitkomst.¹⁰ Voor het bepalen van aannemelijkheid kan men zich baseren op het aanwezige kennisfundament uit eerder onderzoek, op kennis van de fysiologie en pathofysiologie, en op klinische expertise. Het moge duidelijk zijn dat het lastig is over aannemelijkheid van hypothesen consensus te bereiken, waarmee een voedingsbodem is gelegd voor onvermijdelijke – en misschien ook wel gewenste – wisselende interpretatie van gelijke onderzoeksuitkomsten.

WAARHEIDSGEHALTE VAN BEWIJS

Om het 'waarheidsgehalte' van uitkomsten van epidemiologisch onderzoek te kennen, moeten we dus niet alleen kijken naar statistische significantie, maar vooral naar bias en plausibiliteit van de uitkomst. De epidemioloog Ioannidis heeft op basis van deze ingrediënten het gemiddelde 'waarheidsgehalte' van diverse types epidemiologisch onderzoek berekend.¹¹ Het ontluisterende resultaat is samengevat in de tabel. Zelfs de vlaggenscheppen van de epidemiologie, de grote klinische trials en de meta-analyses daarvan, met vaak onmetelijk hoge significantieniveaus, halen gemiddeld een 'waarheidsgraad' van slechts 85%. Op dit werk is kritiek gekomen, met name dat de auteur de toegevoegde waarde van replicatie van onderzoeksbevindingen onderschat.¹² Het criterium van replicatie is echter, meer dan welk 'waarheidscriterium' dan ook, kwetsbaar voor rapportage- en publicatiebias.¹³ De kern van het betoog blijft overeind: het significantieniveau zegt lang niet alles over het waarheidsgehalte van de onderzoeksuitkomst.

EPIDEMIOLOGISCH BEWIJS: TOEPASBAARHEID IN DE SPREEKKAMER

Betrouwbare evidence blijkt dus een relatief schaars goed. Resteert de vraag of men, indien het epidemiologisch bewijs voorhanden is én men dit als waar zou aannemen, dit bewijs naar een individuele patiënt kan vertalen.

Een antwoord zou zijn dat het bewijs geldig is voor een

patiënt met een klinisch profiel dat overeenkomt met dat van de gemiddelde deelnemer aan de trial die het bewijs aanleverde. Al lijkt dit redelijk, men realiseert zich dat intensieve selectie aan de inclusie in klinische trials voorafgaat. Het merendeel (60-99,9%) van patiënten met een bepaalde aandoening komt niet in aanmerking voor deelname aan onderzoek.¹⁴ Dit zou niet erg zijn als de redenen voor exclusie geen invloed zouden hebben op de toepasbaarheid van de onderzoeksuitkomst, maar helaas is niets minder waar. Een recente analyse wees uit dat exclusiecriteria van trials veelal omstandigheden betreffen die cruciaal zijn voor extrapolatie van de onderzoeksbevindingen naar de praktijk.¹⁵ Hierdoor zijn klinici aangewezen op onzekere extrapolatie van zowel effectiviteit als bijwerkingenprofielen.

Het risico van laatstgenoemde extrapolaties wordt bijvoorbeeld geïllustreerd door een analyse naar bloedingscomplicaties van vitamine K-antagonisten. De incidentie daarvan nam exponentieel toe met het aantal exclusiecriteria van relevante trials die achteraf op een patiënt van toepassing waren.¹⁶ Een ander voorbeeld is de impact van een studie naar spironolacton bij hartfalen.¹⁷ De aangetoonde effectiviteit werd door artsen vertaald naar hun eigen patiëntenpopulatie, waarbij men zich onvoldoende rekenschap gaf van de verschillen tussen de trialpatiënten en de echte patiënten. Een forse toename van ernstige en soms fatale hyperkaliëmie was het gevolg.¹⁸

Extrapolatie van epidemiologisch bewijs naar de individuele patiënt is dus vaak problematisch. Zelfs indien de patiënt voldoet aan de inclusiecriteria van relevante trials, dan nog is het bewijs dat u juist handelt vele malen zwakker dan het bewijs dat die handeling in algemene zin de

voorkeur heeft. Hier geldt dat kennis en waardering van bijzondere individuele omstandigheden, eerdere ervaringen met soortgelijke patiënten of met dezelfde individuele patiënt en de lastig te vangen component van klinische expertise de doorslag kunnen geven om van een op groepsniveau bewezen onderzocht beleid af te wijken.

IMPLICATIES

De besproken beperkingen in de beschikbaarheid, betrouwbaarheid en individuele toepasbaarheid van epidemiologisch bewijs hebben belangrijke implicaties voor diverse domeinen van de gezondheidszorg, waarvan wij er enkele noemen.

Klinische patiëntenzorg In een tijdperk waarin evidence tot enige norm voor goede geneeskunde dreigt te worden verheven, is het belangrijk dat artsen zich bewust zijn van de beperkingen van epidemiologisch bewijs. Dit zou tenminste weerklink moeten vinden in de zorg voor individuele patiënten. Maar ook in de dialoog tussen zorgverleners is het belangrijk dat de relativiteit van epidemiologisch bewijs geëxpliciteerd wordt. De veelgehoorde opmerking 'daar is geen bewijs voor' is in de meeste gevallen misplaatst of zelfs ongepast, zeker als ermee wordt gesuggereerd dat van handelingen waar geen evidence voor is, zou moeten worden afgezien. Dit zou het overgrote merendeel van ons medisch handelen onmogelijk maken.

Ook in richtlijnen zijn de beperkingen van epidemiologisch bewijs belangrijk. Doorgaans is de disclaimer ('u kunt in individuele gevallen afwijken van de aanbevelingen' of woorden van gelijke strekking) hiervoor bedoeld. Zaak is dan wel dat de disclaimer serieus wordt genomen, zowel door richtlijngebruikers als door bijvoorbeeld overheid en verzekeraars.

Overheden en verzekeraars Met kracht dient te worden bestreden dat overheden en verzekeraars epidemiologisch bewijs gebruiken om de individuele aanspraak op zorg te reguleren. Bij strikte naleving zou dit het merendeel van het geneeskundig handelen niet-vergoedbaar maken. Hoewel dit uiteraard onrealistisch is, illustreert het dat de selectie van toegestane handelingen op basis van beschikbaarheid van evidence in essentie opportunistisch is, waarbij economische motieven een belangrijke rol zullen spelen.

HET ALTERNATIEF

We willen epidemiologisch onderzoek of EBM niet in een negatief daglicht stellen. Evidence werkt kwaliteitsverhogend als norm voor 'gemiddeld' beleid voor een 'gemiddelde' patiënt. We moeten echter vermijden dat evidence als norm voor individueel beleid gaat gelden of dat de afwezigheid ervan een legitimatie wordt om van handelen af te zien.

TABEL 'Waarheidsgehalte' van diverse types epidemiologisch onderzoek, met diverse graden van aannemelijkheid van de onderzochte hypothese¹¹

type onderzoek*	aannemelijkheid hypothese	kans op 'waar zijn' van significante resultaten
grote gerandomiseerde klinische trial (RCT)	hoog	85%
meta-analyse van grote, eensgezinde RCT's	zeer hoog	85%
meta-analyse van kleinere, tegenstrijdige onderzoeken	gemiddeld	41%
kleine, maar goed uitgevoerde RCT	gemiddeld	23%
klein epidemiologisch 'exploratief' onderzoek	laag	12%

* Gerangschikt op gevoeligheid voor bias: minst bias-gevoelige onderzoeksvormen staan bovenaan, meest bias-gevoelige onderaan.

Balans De moeilijkste vraag is welke mate van keuzevrijheid deze kijk op de geneeskunde toelaat. Louter vertrouwen op pathofysiologisch redeneren heeft in het verleden tot grote fouten geleid, die nota bene door epidemiologisch onderzoek aan het licht zijn gebracht. Klinische expertise is bovendien alles behalve waterdicht. Er dient dus altijd een balans te zijn tussen epidemiologisch bewijs en actuele context. Voor het vinden van deze balans bestaan geen spelregels, maar er is geen principiële reden waarom het niet zou kunnen. Het toepassen van klinische expertise berust immers op, vaak verborgen, opvattingen over causaliteit.¹⁹ De validiteit daarvan hangt af van de vraag hoe goed geprobeerd is deze te weerleggen: een opvatting over causaliteit wordt plausibeler naarmate deze meer intelligente pogingen tot weerlegging heeft overleefd.

Groepsniveau Voor epidemiologisch bewijs als normeringsinstrument zou het uitgangspunt kunnen zijn dat men dit instrument hanteert voor beleid op een hoger integratieniveau dat dan van een individuele patiënt, bijvoorbeeld het niveau van een arts, groepspraktijk of ziekenhuisafdeling.²⁰ Omdat epidemiologisch bewijs zich uitsprekt over een 'gemiddelde patiënt', en dus in feite over een groep patiënten, ligt het niet voor de hand het beleid per individuele patiënt op grond hiervan af te rekenen. Met andere woorden: het integratieniveau waarop zorg de maat wordt genomen, zou in overeenstemming moeten zijn met het niveau waarop epidemiologisch bewijs wordt geleverd: groepsniveau.

Zou een groot percentage van patiënten een praktijk verlaten met een beleid dat niet conform de beschikbare evidence of de algemene aanbeveling is, dan heeft de arts iets uit te leggen. Deze wijze van zorgregulatie op basis van evidence is niet alleen rechtvaardiger voor individuele patiënten, zij is ook eenvoudiger te handhaven. Mits

van een aantal waarborgen voorzien, zou op deze wijze epidemiologisch bewijs toch normerend kunnen werken, waarbij onuitlegbare praktijkvariatie bestreden kan worden, zonder dat het de keuzevrijheid bij individuele patiënten te sterk reduceert. De belangrijkste voorwaarde is beschikbaarheid van adequate meetinstrumenten en criteria om te beoordelen óf, en zo ja in welke mate, een specifieke patiëntenpopulatie afwijkt van een 'gemiddelde populatie' en hoe deze afwijking een ander gemiddeld geneeskundig handelen rechtvaardigt. Hierbij valt voor de eerste lijn te denken aan verschillen in socio-economische achtergrond en etniciteit van patiëntenpopulaties. In de tweede lijn kunnen verwijspatronen van aanvullend belang zijn. Tertiaire zorginstellingen, die bijvoorbeeld complexe ziektegevallen behandelen, kan moeilijker met normen voor gemiddeld beleid, die uit (in algemene populaties uitgevoerd) epidemiologisch onderzoek verkregen zijn, de maat worden genomen.

CONCLUSIE

Evidence-based medicine dreigt – door een verkeerde interpretatie van het begrip – foutief toegepast te worden om geneeskundig handelen bij een individuele patiënt te formuleren, te normeren en te toetsen. Daarvoor komt gemiddeld geleverde zorg meer in aanmerking, mits aan enkele voorwaarden wordt voldaan.

Belangenconflict: geen gemeld. Financiële ondersteuning: geen gemeld.

Aanvaard op 7 maart 2010

Citeer als: Ned Tijdschr Geneeskd. 2010;154:A1910

[➤ Meer op www.ntvg.nl/opinie](http://www.ntvg.nl/opinie)

LITERATUUR

- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312:71-72.
- Tricoci P, Allen JM, Kramer JM, Califf RM, Smith SC Jr. Scientific Evidence Underlying the ACC/AHA Clinical Practice Guidelines. *JAMA*. 2009;301:831-41.
- Sox HC, Greenfield S. Comparative Effectiveness Research: A Report From the Institute of Medicine. *Ann Intern Med*. 2009;151:203-5.
- Van Wijk BL, Klungel OH, Heerdink ER, de Boer A. Rate and determinants of 10-year persistence with antihypertensive drugs. *J Hypertens*. 2005;23:2101-7.
- Newman-Toker DE, Pronovost PJ. Diagnostic Errors--The Next Frontier for Patient Safety. *JAMA*. 2009;301:1060-2.
- Garfunkel JM, Ulshen MH, Hamrick HJ, Lawson EE. Effect of institutional prestige on reviewers' recommendations and editorial decisions. *JAMA*. 1994;272:137-8.
- Link AM. US and Non-US Submissions: An Analysis of Reviewer Bias. *JAMA*. 1998;280:246-7.
- Song F, Parekh-Bhurke S, Hooper L, et al. Extent of publication bias in different categories of research cohorts: a meta-analysis of empirical studies. *BMC Med Res Methodol*. 2009;9:79.
- Vermeulen M. Dwalingen in de methodologie. XXXVI. Van 'likelihood'-ratio's en de regel van Bayes. *Ned Tijdschr Geneeskd*. 2001;145:2421-4.
- Rawlins M. De testimonio: on the evidence for decisions about the use of therapeutic interventions. *Lancet*. 2008;372:2152-61.
- Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2:e124.
- Moonesinghe R, Khoury MJ, Janssens AC. Most published research findings are false-but a little replication goes a long way. *PLoS Med*. 2007;4:e28.

- 13 Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. *BMJ*. 2009;339:b2680.
- 14 Rothwell PM. External validity of randomised controlled trials: "To whom do the results of this trial apply? *Lancet*. 2005;365:82-93.
- 15 Van Spall HGC, Toren A, Kiss A, Fowler RA. Eligibility Criteria of Randomized Controlled Trials Published in High-Impact General Medical Journals: A Systematic Sampling Review. *JAMA*. 2007;297:1233-40.
- 16 Levi M, Hovingh GK, Cannegieter SC, Vermeulen M, Buller HR, Rosendaal FR. Bleeding in patients receiving vitamin K antagonists who would have been excluded from trials on which the indication for anticoagulation was based. *Blood*. 2008;111:4471-6.
- 17 Pitt B, Zannad F, Remme WJ, et al. The Effect of Spironolactone on Morbidity and Mortality in Patients with Severe Heart Failure. *N Engl J Med*. 1999;341:709-17.
- 18 Juurlink DN, Mamdani MM, Lee DS, et al. Rates of Hyperkalemia after Publication of the Randomized Aldactone Evaluation Study. *N Engl J Med*. 2004;351:543-51.
- 19 Stehouwer CDA. De oorzaak ontmaskerd? Oratie. VU Boekhandel, 2001.
- 20 Smulders YM. Vrijheid om af te wijken. *Med Contact* 2008;43:1763-6.