

De kwaliteit van wetenschap

Hans van Maanen

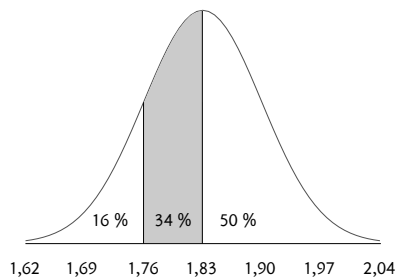
De Volkskrant, 2 april 2011

MENSEN vragen ons wel eens, deugt er dan helemaal niets van het wetenschappelijk onderzoek? Kunnen eenvoudige burgers zich echt niet meer veilig laven aan wetenschappelijke bronnen? Of, als ze beter hebben opgelet: hoeveel procent van het onderzoek deugt er niet?

Het gemakkelijke antwoord op die vraag is natuurlijk: 16 procent. We mogen aannemen dat kwaliteit door talloze kleine factoren wordt beïnvloed, hetgeen in de statistiek betekent dat kwaliteit normaal verdeeld is — net als bijvoorbeeld lichaamslengte en intelligentiequotiënt. Nederlandse jongemannen zijn gemiddeld 1,83 meter lang, met een ‘standaardafwijking’ van 7 centimeter. Dankzij de normale verdeling van lengte weten we dan dat de bulk van de jongens, 68 procent om precies te zijn, 7 centimeter aan weerszijden van het gemiddelde zit, dus tussen de 1,76 en 1,90 meter is. Dat zijn ‘normale’ jongens. Van de overige 32 procent is de helft, 16 procent, aan de korte kant en 16 procent aan de lange kant — zie figuur 1 op de volgende pagina.

Zo is het ook met de kwaliteit van wetenschappelijke artikelen. De bulk, 68 procent, is van gemiddeld niveau, 16 procent is niet zo best en 16 procent is bovengemiddeld goed. Dus 1 op elke 6 artikelen in een wetenschappelijk tijdschrift is onder de maat, is het simpele antwoord.

Wat erg simpel — zeker zo lang we geen cijfer aan dat ‘gemiddelde’ kunnen hangen. Maar dankzij wetenschappelijk onderzoek valt er misschien toch nog wel meer over te zeggen — in ieder geval over medisch onderzoek. Daarvoor wenden wij ons tot de evidence-based medicine, in het bijzonder tot de erfenis van de Schotse arts Archie Cochrane. In een invloedrijk boek uit 1972 wees hij erop dat de geneeskunde veel te weinig gebruik maakte van wetenschappelijke kennis, en dan in het bijzonder van de *randomized controlled trials* — dat zijn zorgvuldig opgezette experimenten waarbij, kortweg, de ene helft van de proefpersonen de te onderzoeken behandeling krijgt, de andere helft een placebo (of de op dat moment beste behandeling). Hij pleitte voor het opzetten van een databank van dergelijke trials en, sterker nog, voor het regelmatig bijeenvegen van trials in grotere ‘reviews’. Beter bewijsmateriaal dan zo’n ‘meta-analyse’ is bijna niet te bedenken.



Figuur 1. De normale verdeling als berg. Hoe hoger de berg, hoe meer jongens eronder passen: 16 procent van de jongens zit meer dan 1 standaardafwijking (7 cm) onder het gemiddelde.

In 1993 werd de Cochrane Collaboration opgezet, een groep onderzoekers – inmiddels zijn het er zo’n dertigduizend over de hele wereld – die zich met Cochrane Reviews bezighouden: kritische, systematische overzichten van trials op allerlei terreinen.

Tegenwoordig verschijnen er tientallen Cochrane Reviews per maand, en hiervan kunnen wij goed gebruikmaken om te kijken wat de kwaliteit van medisch-wetenschappelijk onderzoek eigenlijk is. Het is niet de enige plek waar meta-analyses te vinden zijn, maar in de Cochrane Library staan ze handig bij elkaar, en ze zijn, in principe, volgens een vast stramien opgebouwd. Al die Cochrane Reviews moeten namelijk ook een uitspraak doen over de kwaliteit van de bekeken randomized controlled trials. Was de trial goed opgezet, wist echt niemand wie medicijn en wie placebo kreeg, zijn de uitvallers goed bijgehouden, is er niet selectief gerapporteerd, dat soort dingen. Van de reviewers wordt verwacht dat ze in ieder geval op de zes belangrijkste bronnen van vertekening scoren, met redenen omkleed uiteraard, en die samenvatten in een soort tabel: + is voldoende, ? is niet vast te stellen, en – is onvoldoende. Een heel goed onderzoek scoort zes plussen, een heel slecht onderzoek zes minnen. Mede op grond van dit oordeel bepaalt de reviewer of de totale last van het bewijs uiteindelijk voor of tegen de medische ingreep pleit.

Maar als we nu eens een min of meer willekeurige greep doen uit alle Cochrane Reviews, en de kwaliteitsbeoordelingen van al die reviews tezamen nemen? Dan hebben we een soort ‘meta-review’ of desnoods een ‘meta-meta-analyse’, en kunnen we iets zeggen over de gemiddelde kwaliteit van de randomized controlled trial – van het hoogst aangeschreven medische bewijsmateriaal.

Zo togen wij begin maart 2011 naar de website van de Cochrane Library, naar de pagina met de meest recente reviews, en haalden daarvan de rapporten op. Of

het een representatieve steekproef is weten we niet, maar willekeurig was het wel. Het leverde een stapel van 29 genummerde reviews.

Daarvan vielen er meteen al ettelijke af. Bijvoorbeeld omdat de auteurs wel een zoektocht door de literatuur hadden ondernomen, maar niets hadden gevonden. Zo bleek er geen onderzoek gedaan te zijn naar het nut van een vetarm dieet voor mensen met hoog cholesterol (volgnummer 7957) en naar het nut van screenen op zaadbalkanker (7853).

Nog een paar andere meta-analyses vielen af omdat ze mogelijke bronnen van bias niet hadden samengevat in een snel toegankelijke grafiek met +, ? en -. Zo resteerden 16 reviews. Sommige betroffen een analyse van slechts twee onderzoeken (welke chirurgie bij parese, 7468), over een ander onderwerp waren er meer dan veertig (welk gonadotrofine bij IVF, 5354). In totaal werden in de 16 reviews 195 randomized controlled trials geanalyseerd.

Om de grafieken te kunnen samenvatten, gaven we elke + het cijfer 10, elk ? een 5, en elke - een 0. Er valt te twisten over de vraag of we zo coulant moeten zijn met een ? (dus uit het oorspronkelijke onderzoek valt niet op te maken of de bias vermeden is), maar we moeten keuzes maken, en heel veel maakt het niet eens uit.

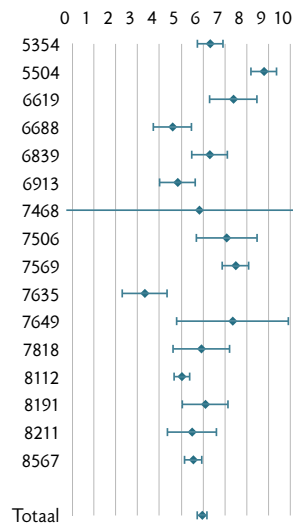
Elke trial in een review kreeg een gemiddeld rapportcijfer over de criteria (een paar hadden er wat meer dan zes, en eentje had er vijf), waarna voor elke review van die rapportcijfers het gemiddelde werd bepaald. Zo kwam bijvoorbeeld de grote meta-analyse, 5354, ten slotte uit op een 6,3 – voldoende, maar het houdt niet over.

Geen enkele meta-analyse kon iets in de buurt van een 10 uitdelen. Het best uit de bus, negen studies van allemaal hoge kwaliteit, kwam 5504, over het nut van levodopa voor het *restless leg syndrome*: rapportcijfer 8,8. Bepaald onvoldoende was onderzoek naar darmoperaties, 7635, met een eindcijfer van 3,3. ‘De kwantiteit en vooral de kwaliteit van de data zijn laag,’ aldus de reviewers.

Nog wat citaten: ‘Alle studies waren van lage kwaliteit’ (6913). ‘Over het geheel genomen was de methodologische kwaliteit laag tot matig’ (7506). ‘Geen studie was geheel vrij van bias’ (7818). ‘De kwaliteit van de onderzochte trials was twijfelachtig’ (8191). ‘De kwaliteit was gering’ (8567). ‘Drie studies, met een hoog risico op bias, hadden een fatale tekortkoming en werden van de meta-analyse uitgesloten’ (8112). En dan betrekken reviewers ook nog maar zelden de commerciële belangen in hun afwegingen.

Figuur 2 vat de rapportcijfers samen; de balkjes aan weerszijden geven het zogeheten 95%-betrouwbaarheidsinterval: hoe korter de balkjes, hoe meer vertrouwen we hebben in het cijfer. Vandaar dat de meta-analyse met maar twee studies, 7468, lange balken heeft: daar valt weinig staat op te maken. Slechts drie reviews zijn vrij zeker van hun voldoende, vier resulteren in een stellige onvoldoende.

Rest nog het eindcijfer te bepalen. Dat komt uit op 5,97, met een 95%-betrouwbaarheidsinterval van 5,7 tot 6,2. Laten we het zo zeggen: als we het



Figuur 2. De rapportcijfers van de 16 meta-analyses met hun 95%-betrouwbaarheidsintervallen en het eindresultaat.

op deze manier uitrekenen en afronden, komt de gemiddelde kwaliteit van *randomized controlled trials* — toch de hoeksteen van de *evidence based medicine* — uit op een 6, en dan zitten we er met een kans van hooguit 1 op 1400 naast. Het stemt tot nadenken.