

# Kanttekeningen bij het gebruik van ‘levels of evidence’ in richtlijnen

Tjerk Wiersma, Jaco Burgers

## Inleiding

Gedurende het afgelopen decennium heeft de ontwikkeling van richtlijnen voor het medisch handelen een grote vlucht genomen. In Nederland nemen de CBO-richtlijnen en de NHG-standaarden een prominente positie in. Inmiddels zijn de richtlijnen – zeker internationaal gezien – zo talrijk geworden dat er meerdere eenzelfde onderwerp behandelen. De onderlinge verschillen gaan nu in het oog springen<sup>1-3</sup> en als reactie ontstaat er aandacht voor het formuleren van kwaliteitscriteria voor richtlijnen of zelfs richtlijnen voor richtlijnontwikkelaars. Het doel van dit alles is waardevolle richtlijnen van minder waardevolle te onderscheiden en de richtlijnontwikkeling op een hoger niveau te brengen.<sup>4-6</sup> Te midden van de vele adviezen en geboden voor goed gedrag worden er twee veel genoemd: de specificatie van de methode die gebruikt is om relevante wetenschappelijke gegevens te verzamelen en de methode die gebruikt is voor de weging van die gegevens.<sup>7</sup> Er zijn inmiddels diverse met elkaar verwante systemen om wetenschappelijke literatuur naar niveau van bewijsvoering (*level of evidence*) te classificeren.<sup>8-11</sup> De methodologie van richtlijnontwikkeling is hierdoor overal sterk beïnvloed. Ook in de CBO-richtlijnen wordt een formeel systeem gehanteerd en diverse niveaus van bewijskracht worden gekoppeld aan de verschillende aanbevelingen.<sup>12</sup> Bij de toepassing van niveaus van bewijsvoering wordt verondersteld dat alle potentiële bewijzen in de literatuur zijn nagezocht en dat een rangschikking van de kracht van het bewijs de interpretatie van de onderbouwing verbetert. Ook wordt met het gebruik van een classificatiesysteem voor wetenschappelijke bewijsvoering en het graderen van bijbehorende aanbevelingen gesuggereerd dat aanbevelingen met een hoger niveau van bewijsvoering juist zijn en met meer kracht dienen te worden uitgevoerd dan aanbevelingen met een lager niveau van bewijsvoering.<sup>9</sup> We gaan in deze beschouwing na of deze suggestie wel terecht is.

## De zoektocht naar het beste bewijs

Een voorwaarde voor het hanteren van een formeel wegingssys-

### Auteursgegevens

Afdeling Richtlijnontwikkeling en Wetenschapsbeleid, Nederlands Huisartsen Genootschap, Postbus 3231, 3502 GE Utrecht:  
dr. Tj. Wiersma, huisarts-filosoof;  
J.S. Burgers, huisarts te Gorinchem  
Correspondentie: dr. Tj. Wiersma, e-mail: T.Wiersma@nhg-nl.org  
Mogelijke belangenverstrengeling: niets aangegeven.

teem is dat de literatuur systematisch is verzameld. Alleen indien getracht is alle relevante literatuur op het spoor te komen, kunnen uitspraken gedaan worden over het beste bewijs of de afwezigheid van bewijs. De hulp van de computer en elektronische databases is daarbij onontbeerlijk.<sup>13</sup> Door een zoekstrategie, veelal een combinatie van trefwoorden, los te laten op een literatuurdatabase kunnen onderzoeksgegevens over een bepaald onderwerp in korte tijd worden opgespoord. Door expliciet melding te maken van de wijze waarop de literatuur gevonden is, wordt het mogelijk medisch-wetenschappelijk onderzoek over uiteenlopende onderwerpen op een voor eenieder controleerbare en reproduceerbare wijze samen te vatten.

Toch is de waarde van deze wijze van werken niet in alle opzichten even duidelijk. Zo geeft een beschrijving van de zoekstrategie als zodanig geen informatie over de hoeveelheid relevante literatuur die over het hoofd is gezien. En ook is een gouden standaard voor een search over het onderwerp in kwestie meestal niet voorhanden.<sup>14</sup> Nogal eens laat de eerste stortvloed aan literatuurreferenties zich alleen tot hanteerbare proporties terugbrengen door concessies te doen aan de sensitiviteit van de search. Vooral als het gaat om minder welomschreven vraagstellingen ontstaan er gemakkelijk serieuze problemen.

De zaak raakt nog meer gecompliceerd doordat er doorgaans geen algoritmen voorhanden zijn die aangeven hoe de diverse publicaties tot één samenvattende conclusie moeten worden gesmeed. In dit opzicht vormen onderzoeken over behandeling de voorhoede: de meta-analyse van klinische trials heeft een grote vlucht genomen en in meerdere opzichten een hoge graad van perfectie bereikt. Dit betekent echter niet zonder meer dat het vraagstuk van ‘het optellen van clinical trials’ geheel opgelost is. Er is hooguit gedeeltelijke overeenstemming over de wijze waarop gecontroleerd moet worden voor zogenaamde heterogeniteit van de diverse ingesloten onderzoeken.<sup>15</sup> Bovendien is er discussie over de vraag of aan onderzoeken met een betere methodologische kwaliteit meer gewicht moet worden toegekend.<sup>16</sup>

De problemen nemen toe waar het gaat om het samenvoegen van diagnostisch of observationeel onderzoek.<sup>17</sup> Men moet zich bij dit type vraagstukken veelal behelpen met redeneringen waarin de kwaliteit en de relevantie van de verschillende onderzoeken tegen elkaar worden afgewogen. De kans dat subjectieve opvattingen een rol spelen in de discussie moet daarbij niet worden onderschat. Het is dus zeker niet zo dat een systematische literatuurschikking altijd tot eenduidige conclusies leidt.

### De kern

- ▶ Het toekennen van niveaus van bewijsvoering (*levels of evidence*) aan aanbevelingen in klinische richtlijnen, wordt algemeen beschouwd als een teken van kwaliteit van richtlijnen.
- ▶ Niveaus van bewijsvoering suggereren dat aanbevelingen met een hoog niveau van bewijsvoering belangrijker zijn en met meer kracht moeten worden uitgevoerd dan aanbevelingen met een laag niveau van bewijsvoering.
- ▶ Het toekennen van niveaus van bewijsvoering is echter niet eenduidig en vindt plaats op basis van discussies en consensus.
- ▶ Het toekennen van bewijsniveaus is vooral nuttig om de lezer snel inzicht te verschaffen in de aard van de onderbouwing.
- ▶ De kwaliteit van een aanbeveling kan alleen adequaat worden beoordeeld indien de inhoudelijke argumenten die bij de totstandkoming van de aanbeveling een rol hebben gespeeld, worden meegewogen.

### Het gewicht van evidence

Parallel aan de roep om systematisch literatuuronderzoek zouden de aanbevelingen moeten worden onderscheiden naar de mate van bewijs die daarvoor voorhanden is.<sup>18</sup> Aanbevelingen die stoeien op gedegen wetenschappelijk onderzoek, zo luidt de onderliggende gedachte, zouden nauwgezetter ten uitvoer dienen te worden gebracht dan aanbevelingen die berusten op medisch-wetenschappelijke onderzoeken van lager allooi. Door aanbevelingen te labelen wordt de lezer ook in staat gesteld goed onderbouwde van minder goed onderbouwde aanbevelingen te onderscheiden. Wat betreft onderzoek naar de waarde van therapeutische of preventieve interventies wordt het hoogste niveau van bewijs geleverd door de systematische review van enkele trials van goede kwaliteit, waarbij de resultaten van de afzonderlijke onderzoeken consistent zijn. Is er slechts één kwalitatief goede en omvangrijke trial voorhanden, dan wordt de bewijslast iets minder hoog ingeschat. Trials met methodologische mankementen en cohortonderzoek of patiënt-controleonderzoek gelden als bewijs van een duidelijk lagere orde. Van nog minder niveau is niet-vergelijkend onderzoek. De allerminste waarde wordt gehecht aan de mening van deskundigen al dan niet opgetuigd met pathofysiologische redeneringen. Voor verslagen van diagnostische onderzoeken fungeert een analogo classificatiesysteem.<sup>19,20</sup>

Bij het classificeren van medisch-wetenschappelijke literatuur naar bewijskracht kunnen we een aantal kritische kanttekeningen plaatsen.

Allereerst stellen we vast dat het onderscheiden van niveaus van bewijs allerminst vanzelfsprekend is. In de klassieke betekenis in de argumentatietheorie en de wiskunde is bewijzen een alles-of-niets-gebeuren: de redenering is wel of niet een bewijs, een tussenvorm bestaat niet. In deze optiek kan alleen medisch-wetenschappelijke literatuur van niveau I als bewijzend worden beschouwd. Aan onderzoeksgegevens van lager allooi, zoals af-

komstig uit cohortonderzoek en patiënt-controleonderzoek kan in het geheel geen bewijskracht worden verbonden. Ze laten hooguit zien dat van zekere maatregelen een gunstig behandelingseffect niet uitgesloten moet worden geacht of – wat positiever geformuleerd – tot de mogelijkheden behoort. Als het effect op plausible wijze verklaard kan worden met een bekend pathofysiologisch mechanisme lijkt de kans groter dat het behandelingseffect ook daadwerkelijk optreedt. Maar het is het zonneklaar dat in deze gevallen van een onomstotelijk bewijs geen sprake kan zijn.

De lage positie die in de classificatiesystemen wordt toegekend aan de mening van deskundigen lijkt ons niet terecht. Hiermee wordt immers gesuggereerd dat een dergelijke mening uit de lucht komt vallen en geen enkele goede grond heeft. In werkelijkheid zal het veelal om situaties gaan waarin de herkomst van de desbetreffende mening niet goed meer te achterhalen is. Een meer principieel bezwaar verbonden met de introductie van deze categorie betreft het feit dat ze berust op de veronderstelling dat het mogelijk is strikt onderscheid te maken tussen aanbevelingen die berusten op *feiten* en aanbevelingen die berusten op *meningen*. Maar geen aanbeveling kan worden geformuleerd zonder interpretatieve activiteit van haar opstellers. Zo is ook de vertaling van de resultaten van klinische trials of meta-analyses, dat wil zeggen categorie-1-bewijs, in aanbevelingen zelden vanzelfsprekend. Steeds is een beslissing noodzakelijk of een gevonden behandelingseffect van voldoende klinische betekenis is om behandeling te rechtvaardigen. Ook dienen richtlijnmakers te beoordelen of de patiënten in de trials voldoende overeenkomen met de patiënten in de dagelijkse praktijk.<sup>21</sup> Een goed voorbeeld zijn de aanbevelingen voor het voorschrijven van alendronaat bij patiënten met osteoporose. Terwijl van dit middel in een tweetal grote en kwalitatief goede trials een gunstig effect is aangetoond op het aantal nieuwe wervelfracturen, kan er uiteenlopend gedacht worden over de klinische betekenis daarvan.<sup>22,23</sup> Door alleen patiënten met wervelfracturen te includeren zijn de resultaten van de trials niet zonder meer extrapoleerbaar naar patiënten met andere fracturen. Dit voorbeeld toont aan dat zelfs aanbevelingen waarbij gebruik is gemaakt van categorie-1-bewijs, ook ten dele berusten op interpretatie en vatbaar zijn voor discussie.

### Beschouwing

Het indelen van aanbevelingen naar niveau van bewijs geeft geen uitsluitsel over de kracht van de aanbevelingen. Het is veeleer omgekeerd: juist het feit dat aanbevelingen, ongeacht het niveau van de onderbouwing, berusten op een mengsel van feiten en meningen, betekent dat aanbevelingen die onderbouwd kunnen worden met een hoog aangeslagen vorm van bewijs, niet op voorhand correcter zijn of met meer kracht dienen te worden uitgevoerd dan aanbevelingen die slechts kunnen bogen op een bewijssoort van lager allooi. Omdat de randomized clinical trial een betrekkelijk recente uitvinding is en overwegend wordt toegepast bij de vaststelling van de effectiviteit van nieuwe geneesmiddelen, kan een dergelijke stellingname ook gemakkelijk leiden

tot een onwenselijke bevoordeling van recente ontwikkelingen boven al langer bekende therapieën en minder heroïsche benaderingen als leefstijladviezen.

Wellicht is het voornaamste winstpunt van het onderscheiden van niveaus van bewijs dat een dergelijke ordening de oriëntatie in onderbouwing van de aanbevelingen enigszins vergemakkelijkt. De niveaus leveren immers een indicatie voor het type argumenten waarmee de aanbevelingen in kwestie kunnen worden onderbouwd en betwist. Bij cohortonderzoek en patiënt-controleonderzoek betreft het doorgaans de eventuele aanwezigheid van confounders en andere vormen van bias, bij trials gaat het veelal om de klinische relevantie van het gevonden therapeutisch effect, de gerechtvaardigheid van extrapolatie van de resultaten naar andere soorten patiënten en de mogelijkheid van extrapolatie van de effecten over langere tijd. Maar we moeten de oriënterende waarde van de niveaus echter vooral niet overschatten. Zowel bij de onderbouwing van aanbevelingen die enkel berusten op de mening van deskundigen als in de hoger aangeschreven categorieën spelen pathofysiologische inzichten en veronderstellingen een niet te verwaarlozen rol. Deze onderlinge overeenkomst relateert in belangrijke mate de verschillen.

### Conclusie

We kunnen dus vaststellen dat het aanbrenge van een hiërarchie in de bewijslast niet simpelweg resulteert in een hiërarchie van aanbevelingen. Op de keper beschouwd betreft het hiërarchisch ordenen vooral de vormgeving van de onderbouwing. Het is echter niet aangetoond dat het gebruik van dit vormprincipe de inhoudelijke kwaliteit van de aanbevelingen doet toenemen. Er bestaat een gevaar dat toename van de aandacht voor de vorm de aandacht afleidt van de inhoudelijke argumentaties. Omtrent aanbevelingen geannoteerd met type-I-bewijs kan men abusievelijk automatisch denken dat het met de aanbeveling in kwestie wel goed zit, terwijl van een aanbeveling waarbij melding wordt gemaakt van de mening van deskundigen gemakkelijk kan worden gedacht dat deze uit de duim gezogen is en inwisselbaar is voor de eigen mening. Bij beide typen onderbouwing is het echter beter het oordeel te baseren op de achterliggende overwegingen.<sup>24</sup> Van richtlijnen mag worden verlangd dat de overwegingen die ten grondslag liggen aan de gemaakte keuzen – en dit geldt in het bijzonder wanneer het schimmige of omstreden kwesties betreft – zo goed mogelijk worden geëxpliciteerd en dat deze explicatie kan worden beschouwd als een uitnodiging erover te discussiëren. Dergelijke discussies worden maar ten dele beslecht door het toekennen van niveaus van bewijsvoering. Uiteindelijk kan de kracht waarmee een aanbeveling moet worden uitgevoerd alleen op inhoudelijke gronden worden beoordeeld.

### Literatuur

- 1 Feder G. Management of mild hypertension: which guidelines to follow? *BMJ* 1994;308:470-1.
- 2 Thompson R, McElroy H, Sudlow M. Guidelines on anticoagulant treatment in atrial fibrillation in Great Britain: variation in content and implications for treatment. *BMJ* 1998;316:509-13.

- 3 Koes BW, Van Tulder MW. Het beleid bij lage rugpijn: een vergelijking van nationale richtlijnen uit vier landen. *Huisarts Wet* 1998;41:57-61.
- 4 Shaneyfelt TM, Mayo-Smith MF, Rothwangl J. Are guidelines following guidelines? The methodological quality of clinical practice guidelines in the peer-reviewed medical literature. *JAMA* 1999;281:1900-5.
- 5 Cluzeau FA, Littlejohns P, Grimshaw JM, Feder G, Moran SE. Development and application of a generic methodology to assess the quality of clinical guidelines. *Int J Quality Health Care* 1999;11:21-8.
- 6 National Health and Medical Research Council. A guide to the development, implementation and evaluation of clinical practice guidelines. Commonwealth of Australia, 1999.
- 7 Cook D, Giacomini M. The trials and tribulations of clinical practice guidelines. *JAMA* 1999;281:1950-1.
- 8 Harris RP, Helfand M, Woolf SH, Lohr KN, Mulrow CD, Teutsch SM, et al. for the Methods Work Group, Third U.S. Preventive task Force. Current methods of the U.S. Preventive Services Task Force. A review of the process. *Am J Prev Med* 2001;20:21-35.
- 9 Harbour R, Miller J, for the Scottish Intercollegiate Guidelines Network Grading Review Group. A new system for grading recommendations in evidence based guidelines. *BMJ* 2001;323:334-6.
- 10 <http://cebmrj2.ox.ac.uk/docs/levels.html>.
- 11 <http://www.ahrq.gov/clinic/evrptfiles.htm#strength>.
- 12 Van Everdingen J. Van consensus naar CBO-richtlijn. *Ned Tijdschr Geneesk* 1999;143:2086-9.
- 13 Assendelft WJJ, Van Tulder WJ, Scholten RJP, Bouter LM. De praktijk van systematische reviews. II. Zoeken en selecteren van literatuur. *Ned Tijdschr Geneesk* 1999;143:656-61.
- 14 Zaat JOM. Worstelend met de literatuur. Zoeken in Medline via Internet – een vergelijking van vijf zoekmachines. *Huisarts Wet* 1999;42:11-7.
- 15 Scholten RJP, Assendelft WJJ, Kostense PJ, Bouter LM. De praktijk van systematische reviews. V. Heterogeniteit tussen onderzoeken en subgroepanalyses. *Ned Tijdschr Geneesk* 1999;143: 843-8.
- 16 Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA* 1999;282:1054-60.
- 17 Zeegers MP, Heisterkamp SH, Kostense PJ, Van der Windt DAWM, Scholten RJP. De praktijk van systematische reviews. VII. Het combineren van de resultaten van observationele onderzoeken. *Ned Tijdschr Geneesk* 2000;144:1393-7.
- 18 Hadorn DC, Baker D, Hodges JS, Hicks N. Rating the quality of evidence for clinical practice guidelines. *J Clin Epidemiol* 1996;49:749-54.
- 19 <http://www.sign.ac.uk/guidelines/fulltext/50>
- 20 Kwaliteitsinstituut voor de Gezondheidszorg CBO. Handleiding voor werkgroepleden. Utrecht: CBO, 2001.
- 21 Rosser WW. Application of evidence from randomized controlled trials to general practice. *Lancet* 1999;353:661-4.
- 22 Liberman UA, Weiss SR, Bröll J, Minne HW, Quan H, Bell NH, et al. Effect of oral alendronate on bone mineral density and the incidence of fractures in postmenopausal osteoporosis. *N Engl J Med* 1995;333:1437-43.
- 23 Black DM, Cummings SR, Karpf DB, Cauley JA, Thompson DE, Nevitt MC, et al. Randomised trial of the effect of alendronate on risk of fracture in women with existing vertebral fractures. *Lancet* 1996;348:1535-41.
- 24 Vandenbroucke JP. Hoe wordt medische kennis gemaakt? Gezondheidsraadlezing, 30 september 1999.

## Commentaar

Wiersma en Burgers plaatsen een aantal kritische kanttekeningen bij het classificeren van medisch-wetenschappelijke literatuur naar bewijskracht in het kader van richtlijnontwikkeling.

Hun eerste bezwaar is dat het onderscheid van categorieën niet overeenstemt met de klassieke betekenis van bewijs zoals die gebruikt wordt in de argumentatietheorie en de wiskunde. Bewijzen is in deze vakgebieden een alles-of-niets-fenomeen; er

is wel of geen bewijs, een tussenvorm bestaat niet. Tegen die redenering is wel wat in te brengen. Er is niet zoets als hét bewijs in de geneeskunde in algemene (abstracte) zin. Bij de invulling van het begrip bewijs hangt veel af van het doel dat men zich gesteld heeft. In de geneeskunde wordt bewijsvoering vooral gebruikt voor een drietal doelen:

- ▶ het onderbouwen van een oorzaak-gevolgrelatie;
- ▶ het staven van een diagnose;
- ▶ het aantonen van de effectiviteit van een interventie.

Naast het doel heeft men verder te maken met het niveau waarop het te leveren bewijs betrekking heeft: populatieniveau of individueel niveau.

Bij de behandeling van de individuele patiënt vindt er een integratie van gegevens en redeneringen plaats; deze integratie gebeurt via causaal denken en niet via probabilistisch denken.

Bij het doen van een uitspraak op populatieniveau over de waarde van een test of interventie heeft men altijd te maken met de onzekerheden die aan een steekproef kleven. De geneeskundige wetenschap heeft een stochastisch karakter. Om toevalsfactoren zoveel mogelijk uit te sluiten wordt in een experimentele situatie vergelijkend onderzoek gedaan. Bij dergelijk onderzoek is het voor het leveren van bewijs ook van belang dat variabelen die de uitkomst kunnen vertekenen zoveel mogelijk worden uitgesloten, bijvoorbeeld door een randomisatieprocedure en blinding van onderzoekers en onderzochten. Maar ook als alles loopt zoals het zou moeten lopen, is op vrijwel alle trials wel iets aan te merken. Wij zijn het dan ook niet eens met de stelling van Wiersma en Burgers dat, gezien het alles-of-niets-fenomeen, de uitkomsten van clinical trials of meta-analyses (type-I-bewijs) bewijzend zouden zijn. Er zijn altijd interpretatieverschillen. Wat door de ene arts als een onomstotelijk bewijs wordt beschouwd, is voor een ander niet meer dan een aanwijzing in een bepaalde richting, en voor een derde zelfs helemaal geen bewijs.

Een tweede, meer principieel bezwaar tegen de introductie van categorieën van bewijs zien Wiersma en Burgers in het feit dat die indeling berust op de veronderstelling dat het mogelijk is rigoureus onderscheid te maken tussen richtlijnen die berusten op feiten en richtlijnen die berusten op meningen.

Dat onderscheid is inderdaad lang niet altijd even duidelijk, maar de kunst is juist om dit zo zichtbaar mogelijk te maken. Het CBO hanteert daarom voor elke paragraaf van de richtlijn een vast stramen, waarin meningen en feiten zo goed mogelijk uit elkaar worden getrokken.

De tekst begint met een aantal uitgangsvragen, die zo concreet mogelijk worden gesteld in de vorm van zogenoemde PICO's. PICO staat voor *population, intervention/indextest, control/comparison* en *outcome*. Een PICO houdt dus in: bij welke patiënten heeft men wat onderzocht/behandeld, in vergelijking waarmee en met welk resultaat? Daarna volgt een tekst waarin de uitgangsvraag wordt beantwoord en de voor- en nadelen van diagnostiek en behandeling expliciet worden aangegeven. Door per onderzoek de onderzoeks-

kenmerken te scoren in een tabel maken we duidelijk hoe zwaar de wetenschappelijke bewijskracht is en welk gewicht hieraan kan worden toegekend. Als er geen literatuur van het hoogste niveau voorhanden is, geven we weer wat op een lager niveau is gevonden. Tegenstrijdige of ontbrekende resultaten vermelden we zo expliciet mogelijk.

De uiteindelijke conclusie dient niet verder te gaan dan wat uit de literatuur naar voren komt. Als er geen harde onderbouwing is te geven op basis van de literatuur proberen we toch een conclusie te trekken op basis van ervaringen en meningen binnen de werkgroep en dat zo expliciet mogelijk aan te geven. Overigens kunnen dit ook harde uitspraken zijn indien het om zaken gaat die al op klinische gronden zijn 'bewezen', maar waar nooit een RCT voor is uitgevoerd of zou kunnen worden uitgevoerd, bijvoorbeeld het toedienen van compatibel bloed. In zo'n geval doet een D-classificatie de bewijskracht zeker tekort. In de komende CBO-richtlijnen zullen we hier expliciet aandacht aan besteden.

Na de conclusie volgt een paragraaf met 'overige overwegingen'. Hierin komen zoveel mogelijk contextuele aspecten aan de orde, zoals de beschikbaarheid van de aanwezige voorzieningen, de bereikbaarheid en toegankelijkheid daarvan voor de patiënt. Het is – gezien de globale strekking van zo'n conclusie – niet wenselijk om ook meer individuele factoren, zoals de klachtenperceptie en -presentatie of de angsten en wensen van de patiënt hierin op te nemen. Wel nemen we hierbij ook het kostenaspect mee. Kosteneffectieve richtlijnen geven idealiter aan welke ruimte artsen hebben in de kloof tussen *efficacy* – datgene wat in trialverband is bewezen – en *effectiveness* – datgene wat in de eigen situatie kan worden gerealiseerd. Ten slotte brengen we ethische overwegingen en haalbaarheidsaspecten vanuit het gezichtspunt van patiënten, zorgverleners en ziektekostenverzekeraars in kaart.

Het uiteindelijke advies is veelal de neerslag van wetenschappelijke overwegingen en de normatieve argumenten zoals die binnen de groep worden beleefd. Ook de kostenoverwegingen, organisatorische aspecten of patiëntgebonden voorkeuren zullen het advies kleuren. Zij kunnen ertoe leiden dat het advies en de conclusie niet met elkaar sporen.

De bedoeling van het expliciet uit elkaar trekken van deze onderdelen is dat de tekst transparant wordt en dat de lezer/gebruiker stap-voor-stap de gedachtegang van de richtlijnmakers kan volgen en de zorgvuldigheid kan beoordelen waarmee zij tot bepaalde uitspraken zijn gekomen.

Opstellers van richtlijnen die niet zorgvuldig te werk zijn gegaan in hun beoordeling van de literatuur of die de *levels of evidence* hebben gebruikt of misbruikt om hun eigen stokpaarden te berijden, kunnen op deze wijze beter dan voorheen op hun vingers worden getikt.

J.J.E. van Everdingen, T.A. van Barneveld, CBO

Correspondentie: dr. J.J.E. van Everdingen, CBO, Postbus 20064, 3502 LB Utrecht; e-mail: j.vaneverdingen@cbo.nl  
Mogelijke belangenverstrengeling: niets aangegeven.