

Steunpunt

MILIEU
&
GEZONDHEID

ONDERZOEKSGROEP
FASENPLAN

November 2007

Het steunpunt Milieu en
Gezondheid
is een samenwerkings-
verband tussen de Vrije
Universiteit Brussel, de
Universiteit Gent, de
Universiteit Antwerpen,
de Katholieke
Universiteit Leuven,
het Limburgs
Universitair Centrum,
het Vlaams Instituut voor
Technologisch Onderzoek,
het Provinciaal Instituut voor
Hygiëne en de Universiteit
Maastricht.

Universiteit
Antwerpen

Verantwoording Multicriteria Methode fase 1

Prioriteren cases voor beleid

Fasenplan Biomonitoring

Rapportage



Hans Keune, Johan Springael

Inhoud

1. INLEIDING
2. WAAROM EEN MULTICRITERIA ANALYSE?
3. UITWERKING IN DE PRAKTIJK
4. EVALUATIE
5. REFERENTIES

1. Inleiding

Dit rapport doet verslag van de methodologie die werd gebruikt en geoperationaliseerd in het kader van het fasenplan van het Steunpunt Milieu & Gezondheid inzake een selectie van de humane biomonitoringsresultaten. Het gaat hier over de eerste fase van dit fasenplan waarin de geselecteerde meetresultaten gerangschikt dienen te worden naar prioriteit voor milieu- en gezondheidsbeleid van de Vlaamse Overheid. Deze methodologie is gebruikt en geoperationaliseerd tijdens de verwerking van de resultaten van de expertronde die plaats in de periode februari/april 2007 op basis van desk research en de jury die plaats vond in september 2007, en waaraan verschillende vormen van input van juryleden vooraf gingen in de periode mei/augustus 2007. Omwille van de complexiteit van zowel de gebruikte methode als de operationalisering daarvan in de praktijk hebben we ervoor gekozen aan apart rapport hierover op te stellen. De methodologie is ontwikkeld en toegepast door wetenschappers van de UA. In deze inleiding introduceren we eerst kort het biomonitoringsprogramma (1.1) en het concept fasenplan (1.2). Daarna schetsen we de stand van zaken in de praktische uitwerking van dit fasenplan, namelijk de geselecteerde meetresultaten (1.3) en de desk research (1.4). Tot slot van deze inleiding blikken we vooruit op de rest van dit rapport (1.5). Voor de rapportering van de andere onderdelen van dit proces verwijzen we naar desbetreffende rapporten: desk research (Morrens et al. 2007), expert ronde (Keune et al. 2007a) en jury (Keune et al. 2007b).

1.1 Het Vlaams Humaan Biomonitoringsprogramma

Tussen 2002 en 2006 werd – in opdracht van de Vlaamse Overheid – een grootschalig meetprogramma uitgevoerd, het Vlaams Humaan Biomonitoringsprogramma. Doelstelling was gegevens verzamelen over de blootstelling van de mens aan milieupolluenten én hun mogelijke relatie met gezondheidseffecten. Er werden acht aandachtsgebieden in Vlaanderen aangeduid met een kenmerkende en verschillende milieubelasting, namelijk gebieden rond verbrandingsoven, de Antwerpse agglomeratie, de Gentse agglomeratie, landelijk Vlaanderen, de fruitstreek, Olen, de Albertkanaalzone en de Havenzones van Antwerpen en Gent. Er werden gedurende de periode 2002-2006 in drie meetcampagnes verschillende leeftijdsgroepen onderzocht: pasgeborenen, adolescenten (14-15 jaar) en oudere volwassenen (50-65 jaar). Per campagne werden bij ongeveer 200 mensen uit elk aandachtsgebied gechloreerde verbindingen, zware metalen en metabolieten van PAK's en benzeen gemeten. De onderzochte gezondheidseffecten betreffen onder andere luchtwegklachten, DNA-schade, puberteits-ontwikkeling en kankerrisico.

Voor meer informatie en resultaten zie: <http://www.milieu-en-gezondheid.be>.

1.2 Het Fasenplan

Helaas geven metingen in de mens geen rechtstreekse aangrijpingspunten voor overheidsbeleid. Ze zeggen immers weinig over de mogelijke oorzaken en bronnen van gevonden afwijkingen. Daarom werd in 2004 in opdracht van de Dienst Milieu & Gezondheid van het Departement Leefmilieu, Natuur en Energie (LNE), en in nauw overleg met het intern verzelfstandigd agentschap (IVA) Zorg en Gezondheid, een fasenplan uitgewerkt (Koppen et al. 2005, Aminal Cel Milieu en Gezondheid 2005, Aerts 2005). Met dit fasenplan wordt getracht om op transparante, geobjectiveerde en systematische wijze eventuele afwijkende meetwaarden te herkennen en te evalueren. Op basis van dit afwegingsproces worden dan prioriteiten voor aanpak voorgesteld.

Na het doorlopen van een *testcase* voor de biomonitoringsresultaten van DDE werd het concept fasenplan lichtjes bijgesteld. Het bestaat nu uit een voorfase en vier opeenvolgende fasen:

- voorfase: vaststellen en selecteren van overschrijdende biomerkerwaarden
- fase I: evalueren van geselecteerde overschrijdingen naar ernst en prioriteiten
- fase II: opzoeken van oorzaak en bron van de prioritaire overschrijdingen en opstellen plan van aanpak
- fase III: uitvoeren plan van aanpak
- fase IV: evaluatie plan van aanpak

In de voorfase beoordelen experts van het Steunpunt Milieu & Gezondheid elke biomerkermeting van vervuilende stoffen of gezondheidseffecten in de aandachtsgebieden in het licht van: internationale advieswaarden (voor zover beschikbaar), internationale meetwaarden en een Berekende Referentiewaarde (referentiegemiddelde & referentie-90ste percentiel) van alle onderzochte gebieden in de biomonitoring.

Indien uit fase I blijkt dat het probleem als ernstig en beleidsmatig en maatschappelijk relevant wordt ingeschat door experts en ook prioritair beoordeeld wordt door een jury (hierbij wordt ook rekening gehouden met beleidsmatige en maatschappelijke aspecten) om aan te pakken dan zal – indien de bevoegde Minister(s) daartoe besluit(en) – worden overgegaan naar fase II. Deze ministers hebben immers opdracht gegeven aan het Steunpunt Milieu en Gezondheid om de blootstelling en mogelijke effecten daarvan bij de Vlaamse bevolking te meten.

In fase II wordt onderzocht in hoeverre de overschrijding te wijten is aan milieuverontreiniging dan wel aan levensstijlfactoren (aard voeding, werk, hobby's, rookgewoonten) die eventueel systematisch verschillen tussen de aandachtsgebieden. Indien geoordeeld wordt dat de factor milieu van belang is, wordt nagegaan of er een lokale bron is die in verband kan worden gebracht met de geobserveerde afwijkende meetwaarden in het aandachtsgebied. Ook hier wordt gekeken naar beleidsmatige en maatschappelijke aspecten

Elke fase doorloopt achtereenvolgens een aantal stappen. Eerst wordt door een onderzoeksteam van het Steunpunt Milieu & Gezondheid desk research gedaan met betrekking tot de vraag die in betreffende fase voorligt. Op basis hiervan wordt aan experts een aantal vragen voorgelegd samen met documentatie over het vraagstuk. Naast medisch-milieukundige aspecten, wordt hierbij ook steeds naar beleidsmatige en maatschappelijke aspecten gekeken. Vervolgens komen in een jury-discussie maatschappelijke actoren aan het woord om hun inzichten en meningen over het vraagstuk te geven. Het onderzoeksteam maakt van 1) de desk research, 2) de expertronde en 3) de jury-discussie een synthese op basis waarvan de bevoegde Minister(s) op een onderbouwde en afgewogen wijze kunnen beslissen en communiceren of er, en zo ja welke, verdere stappen gezet moeten worden.

1.3 De geselecteerde meetresultaten

De voorfase vond plaats in januari 2007 en resulteerde in de selectie van de zes meest belangwekkende meetresultaten, vanaf nu 'cases' genoemd, van het biomonitoringsprogramma 2002-2006. Deze cases zijn:

1. Dioxine, PCB's en HCB in landelijke gemeenten
2. Dioxineachtige stoffen in Antwerpen (agglomeratie + haven)
3. Benzeen in Antwerpen (agglomeratie), Roeselare en Sint-Niklaas (verbrandingsovens)
4. Cadmium in landelijke gemeenten
5. PCB's in Gent (incl. haven)
6. Astma in Gent en Antwerpen (incl. havens)

Deze zes cases gingen vervolgens door naar fase I waarin ze geëvalueerd en gerangschikt moeten worden prioriteit voor milieu- en gezondheidsbeleid van de Vlaamse Overheid.

1.4 Desk research, expertronde en jury

De eerste stap van fase I was het verzamelen van bijkomende onderzoeksgegevens. Dit gebeurde in de periode januari/maart 2007. Wetenschappers van de UA en VITO deden desk research over de zes geselecteerde cases met betrekking tot drie criteria (Morrens, Colles, Keune, 2007):

1. ernst van het gezondheidsrisico
2. beleidshaalbaarheid
3. maatschappelijke haalbaarheid

Op basis van de verzamelde informatie werd per criterium een vragenlijst opgesteld ter rangschikking van de cases naar prioriteit. In de tweede stap van fase I worden deze vragenlijsten, samen met de desk research, voorgelegd aan experts. De resultaten van de expertronde (Keune et al 2007a) worden vervolgens door een jury (samengesteld uit maatschappelijke belangengroepen) beoordeeld op maatschappelijke en politiek prioriteit op basis waarvan zij de bevoegde Ministers een advies (Keune et al 2007b) voor beleidsprioriteiten geven.

1.5 Opzet rapport

In het volgende hoofdstuk (2) wordt de multicriteria methode geïntroduceerd. We beschrijven waarom voor deze methode gekozen is en welke de theoretische achtergrond is. Hierna duiden we de uitwerking in de praktijk (hoofdstuk 3) en de evaluatie daarvan (hoofdstuk 4).

2. Waarom een multicriteria analyse?

In dit hoofdstuk gaan we in op het gebruik van de multicriteria methode. We behandelen de vraag waarom we gekozen hebben voor deze methode. Daarna schetsen we kort wat de theoretische achtergrond van de methode is. Vervolgens beschrijven we hoe de methode gekaderd is in de conceptuele opzet van het fasenplan (Koppen et al. 2005). In het volgende hoofdstuk (3) zullen we de operationalisering van de specifieke door ons gebruikte methode binnen de vraagstelling van fase 1 van het fasenplan toelichten.

2.1 Vraagstelling fase1

De vraag die centraal staat bij de eerste fase van het fasenplan dat zich buigt over de resultaten van de biomonitoring is: welke meetresultaten van de biomonitoring¹, moeten in eerste instantie prioriteit hebben (of in welke rangorde zaken aan bod dienen te komen) voor verder onderzoek in het kader van het fasenplan en overheidbeleid?

Deze vraagstelling wordt in het kader van de voor het fasenplan uitgewerkte praktijkcyclus (Keune & Goorden 2004) aan een jury voorgelegd. Hierbij staan drie hoofdcriteria bij de beoordeling centraal: ernst van het gezondheidsrisico, beleidshaalbaarheid en maatschappelijke haalbaarheid. De informatie voor deze criteria wordt verzameld in desk research (Morrens et al. 2007). Dit wordt ter beoordeling vervolgens voorgelegd aan experts (Keune et al. 2007a). Dit alles wordt vervolgens aan een jury van maatschappelijke belangengroepen voorgelegd om een advies tot priorisering voor de overheid te komen (Keune et al. 2007b). Het is hierbij niet de bedoeling dat het advies enkel bestaat uit een numerieke rangschikking van de geselecteerde meetresultaten van de biomonitoring op basis van de gebruikte multicriteria methode. Het is de bedoeling dat er een groepsdiscussie met uitwisseling van voorkeuren en argumenten plaatsvindt. De multicriteria methode functioneert hierbij als middel om de discussie mee te structureren en zorgt voor reflectie op de inhoud van de discussie te zorgen.

Complexiteit

Veel complexe vraagstukken waarbij gekozen moet worden tussen verschillende opties op basis van meerdere criteria stellen hoge eisen aan besluitvormers. Het menselijke brein is, zo blijkt uit studies (Miller 1956; Kagan 1988), niet in staat om een zo complexe mix van gegevens gelijktijdig in de overweging te betrekken. Als zich hierbij bovendien de complexiteit voegt van een groepsproces waarin afweging door verschillende actoren van belang is, is er nood aan structurering en visualisering ter ondersteuning van het afwegingsproces. Multicriteria analyse (MCA) biedt hiertoe handreikingen (Keune & Goorden 2005).

Geen miracle tool, geen objectieve scheidsrechter

Om de hierboven geschetste belofte van MCA maar meteen te nuanceren: het is een hulpmiddel, geen technologie die voor u uitrekenet wat de beste keuze is. Een term die MCA-experts ook wel gebruiken is die van een klankbord: MCA structureert en visualiseert wat actoren zelf aan overwegingen inbrengen, en biedt daardoor gelegenheid op goed geïnformeerde en transparante wijze hierop te reflecteren, te leren, en te discussiëren. Waarom kent multicriteria analyse dan toch beperkingen? Eenvoudigweg omdat inschattingen van mensen eraan ten grondslag liggen: multicriteria analyse moet het doen met de kwaliteit aan informatie die erin gebracht wordt (zie bvb.: Dodgson et al. 2000).

2.2 Theoretische achtergrond Multicriteria analyse

Multicriteria analyse (MCA) is een verzamelnaam voor verschillende kwantitatieve en/of kwalitatieve analysemethoden, die gehanteerd worden ter ondersteuning en/of structurering van een beslissing waarin de graad van complexiteit zeer hoog is. Zoals de term "multicriteria" zelf uitdrukt, worden deze methoden gebruikt indien er met verschillende, soms tegenstrijdige, aspecten (criteria) rekening moet gehouden worden in de selectie, rangschikking of groepering van mogelijke keuze-alternatieven. Door de multidimensionale complexiteit van het probleem is het vaak zo dat de mens niet in staat is zonder

¹ Voor zover die niet bestaande normen overschrijden, maar wel boven de Vlaamse referentiewaarde liggen ofwel hoger dan vergelijkbaar onderzoek buitenland.

een dergelijk hulpmiddel de volledige complexiteit van het probleem te overzien. Het gebruik van multicriteria methoden kan hiervoor een hulpmiddel zijn. (Roy, 1996)

Deze methoden kunnen op verschillende manieren gebruikt worden, doch dient de nadruk erop gelegd te worden dat deze methoden niet bedoeld zijn om een beslissing te nemen in plaats van de besluitvoerder(s). Zij dienen enkel om de besluitnemer(s) beter te informeren en de veelheid aan gegevens op een zo gestructureerd mogelijke manier te presenteren en samen te vatten. Tevens hebben ze tot doel de besluitnemer te helpen bij het formuleren van zijn/haar persoonlijke preferenties, zodat de uiteindelijke beslissing op een veel transparantere en meer geïnformeerde basis kan geschieden.

Al de multicriteria methoden voldoen ongeveer aan dezelfde structuur. Verschillende alternatieven moeten vergeleken worden met elkaar op basis van verschillende criteria. Alternatieven hoeven niet gelijk te scoren op een criterium. Het kan hierbij bvb. zo zijn dat een alternatief goed scoort op één criterium, maar slecht op een ander criterium, terwijl dat bij een ander alternatief precies andersom is. In de theorie van de MCA gaat men er dikwijls van uit dat de scores van al de alternatieven op al de criteria gegeven zijn of toegeleverd kunnen worden, meestal via het gebruik van databanken, bestaande interne kennis of via raadpleging van experts. In het geval van expertinschattingen kunnen we spreken over scores als inschattingen, beoordelingen of evaluaties.

Het doel van multicriteria methoden is o.a. op basis van extra informatie een oplossing te zoeken voor een inherente wiskundige probleem wat zich stelt indien men met ongelijksoortige criteria werkt. Deze extra informatie kan samengevat worden onder de noemer "besluitnemer-specifieke informatie". Een besluitnemer is, anders dan een expert, een actor die zich buigt over de vraag welke criteria meer en minder belangrijk zijn. Dit omvat, afhankelijk van de gehanteerde methode, o.m. gewichten of belangrijkheden van de criteria, preferenties van de besluitnemer betreffende de verschillende scores, enz. Het feit dat er met de preferentiële structuur van de besluitnemer rekening wordt gehouden, impliceert dat men niet meer zal kunnen spreken in termen van de "optimale" of "beste" oplossing, daar deze wiskundig gesproken niet meer bestaat. Wel kan men het hebben over de "meest geschikte" oplossing voor een specifieke besluitnemer, i.e. best aansluitend bij diens voorkeuren.

Als laatste stap in de methoden worden de scores gecombineerd met de voorkeuren (gewichten) van de besluitnemer, ofwel de besluitnemer-specifieke informatie, om uiteindelijk tot een globaal advies te komen voor een welbepaalde besluitnemer. Het is in deze laatste aggregatiestap dat de meeste multicriteria methoden methodologisch van elkaar verschillen. In een MCA-basischema ziet dit er als volgt uit (tabel 1):

Tabel 1

	C1	C2	C3
	w1	w2	w3
A1	s11	s12	s13
A2	s21	s22	s23
A3	s31	...	
A4	...		
A5			
A6			

(A=alternatief, C=criterium, w=gewicht, s=score)

Binnen het domein van de MCA zijn er hoofdzakelijk twee grote paradigma's, nl. de methoden gebaseerd op meerdere utiliteitsfuncties en de zgn. "outranking"-methoden. Het eerste paradigma waarin verschillende utiliteits- of nutsfuncties met elkaar gecombineerd worden, is een rechtstreekse uitbreiding van de nutsfunctietheorie uit de economie en probeert een antwoord te bieden op het wiskundige probleem inzake ongelijksoortige criteria. Door de verschillende nutsfuncties met elkaar te combineren (meestal via een gewogen som), wordt het multicriteria probleem herleid tot een unicriterium probleem dat met gekende (optimalisatie)technieken opgelost kan worden. In de constructie van de verschillende nutsfuncties dient er echter beroep gedaan te worden op de besluitnemer, die een intensief interactief proces dient te doorlopen. Deze technieken, die dateren van begin jaren '60, alsook hun extensies, worden hoofdzakelijk gebruikt in de Angelsaksische wereld.

Eind jaren '60 is er een alternatieve zienswijze ontstaan op het Europese vaste land onder impuls van B. Roy (Parijs, Frankrijk) en nadien J.P. Brans (Brussel, België) rond het nieuwe paradigma van de overheersingsmethoden ("outranking"-methoden). Kort samengevat, wordt in dit type methoden nagegaan of er voldoende bewijs aanwezig is in de gegevens (evaluaties + besluitnemer-specifieke

informatie) voor uitspraken van de vorm: “alternatief a overheerst alternatief b” of “alternatief a wordt overheerst door alternatief b”. De bewijslast voor elk van de uitspraken wordt nagegaan voor ieder koppel alternatieven, waarna in een volgende stap de verschillende uitspraken gecombineerd worden tot een globaal advies. Ook hier kan het bekomen van de besluitnemer-specifieke informatie een intensief interactieproces zijn tussen onderzoeker en besluitnemer.

Daar de multicriteria methoden expliciet gebruik maken van de preferentiestructuur van de besluitnemer, zijn ze in hoofdzaak enkel maar geschikt in de situatie waarin er slechts 1 besluitnemer aanwezig is. In het geval dat een beslissing dient genomen te worden in groep, moet men overstappen op zogenaamde groepsbesluitvormingsystemen, die al dan niet gebaseerd zijn op een multicriteria methode (Hwang and Lin 1987; De Keyser and Springael 2002). Verschillende structuren zijn hier mogelijk, doch in de analyses uitgevoerd in het kader van het fasenplan werd geopteerd om voor een jury een multicriteria analyse te gebruiken, gebaseerd op gegevens toegeleverd door experts. Voor we in het volgende hoofdstuk nader ingaan op de gebruikte methode willen we eerst schetsen hoe het gebruik van een multicriteria analyse in het concept fasenplan (Koppen et al. 2005) gekaderd was bij aanvang.

2.3 Kadering in het concept fasenplan

In de opzet van het concept fasenplan is de multicriteria analyse als onderdeel opgenomen en deels uitgewerkt. In tabel 2 staan de criteria opgelijst.

De term evaluatiepanel is later vervangen door de expertronde en betreft volgens het concept fasenplan: *personen met ervaring/expertise op vlak van een specifiek deelaspect (bvb. zware metalen, of een bepaald gezondheidseffect of een maatschappelijk aspect) dat een rol speelt in de voorliggende probleemstelling in elk van de verschillende fasen van het actieplan.*

Over het prioriseren van meetresultaten van de biomonitoring voor beleid wordt gezegd: *prioritering is het resultaat van een afwegingsprocedure (op basis van een multicriteria analyse) tussen wetenschappelijke en maatschappelijke criteria en mogelijkheid tot actie. Het houdt rekening met ‘ernst’, kosten-baten, maatschappelijke afwegingen, haalbaarheid van beleid en rechtvaardigheid (‘equity’), mogelijkheid tot actie. Uitkomst van de prioritering is een rangorde van zaken die van belang voor verdere stappen zijn.*

Over de rol van de jury staat beschreven: *de Jury evalueert inschattingen en argumentaties gegeven door experts van het Evaluatiepanel en weegt dit samen met factoren van beleid en maatschappelijk belang. Essentieel ligt hier de uitdaging voor om een afweging te maken tussen ongelijksoortige factoren die van betekenis zijn voor beoordeling van de vraag wat er moet gebeuren. Het gaat enerzijds om kennis van experts die enkel op deelfacetten van de ernstbepalende elementen expert zijn, en niet met betrekking tot het geheel aan ernstbepalende elementen. Anderzijds gaat het bovendien niet enkel om factoren van ernst inzake gezondheidsrisico's. Het gaat ook om factoren van beleid en maatschappelijk belang: wat zijn beleidsprioriteiten, wat is haalbaar in termen van beleid, welke maatschappelijke afwegingen spelen een rol van betekenis? De jury geeft een advies over welk gewicht de verschillende factoren, criteria hebben in de beoordeling van het vraagstuk dat voorligt*

Tabel 2

FASE I: Afwegingscriteria voor het Evaluatiepanel	
Ernst	<i>Mate van afwijking biomerker</i>
	De afwijking van de biomerker in het aandachtsgebied vergeleken met waarden in <i>andere aandachtsgebieden</i> , Berekende referentiewaarden, laagste waarde van alle gebieden, internationale literatuurwaarden...
	Is de afwijking van de biomerker van die mate, dat verdere aanpak nodig is? Suggesties om het

probleem verder aan te pakken?
<i>Aantal afwijkingen in aandachtsgebied</i> Aantal afwijkende blootstellings- en effectmerkers, verband tussen deze merkers?
<i>Betroffen leeftijdsgroepen</i> Merkers afwijkend bij meerdere leeftijdsgroepen?
<i>Verbanden/relaties</i> Zijn er in het aandachtsgebied significante blootstellings-effect relaties gevonden voor de afwijkende biomerker Zijn er meerdere significante blootstelling-effect relaties in het aandachtsgebied? Zijn er op Vlaams niveau significante blootstellings-effect relaties gevonden voor de afwijkende biomerker in het aandachtsgebied Gekende effecten uit dierproeven/ <i>in-vitro</i> onderzoek Dosiswaarbij effect optreedt + onzekerheden
<i>Biomerker in functie van blootstellingstermijn en specifiek gezondheidseffect</i> Blootstellingstermijn waarvoor merker indicator is, duur van gezondheidseffecten Technische kwaliteit biomerker (labo predictieve waarde) Surveillance kwaliteit biomerker (outcome predictieve waarde)
<i>Biomonitoringsgegevens van vroegere campagnes</i> Was toen <u>zelfde/gelijkaardige</u> biomerker afwijkend in het aandachtsgebied? Werd verhoging van biomerker vastgesteld in dezelfde leeftijdsgroepen. Kan er gesproken worden van een trend/evolutie in biomerker (in laatste 5-10 jaar)?
<i>Morbiditeit- en mortaliteitsgegevens uit de regio/aandachtsgebied</i> Morbiditeitsgegevens afwijkend in de regio? Biologisch plausibele relatie tussen sommige van de morbiditeitsgegevens en de gemeten biomerker? Trends in de tijd (laatste 5-10 jaar) voor deze regionale morbiditeitsgegevens? Mortaliteitsgegevens afwijkend in de regio? Biologisch plausibele relatie tussen sommige van de mortaliteitsgegevens en de gemeten biomerker? Trends in de tijd (laatste 5-10 jaar) voor deze regionale mortaliteitsgegevens?
Beleidshaalbaarheid
Beleidsprioriteit? Welk belang hecht de overheid aan deze problematiek?
Interventiemogelijkheden (voor de overheid): - Aanpak bron - Preventie tegen blootstelling - Preventie gezondheidseffecten - Behandelingsmogelijkheden effecten
<i>Kosten baten afweging: wat zijn de kosten en hoe verhoudt zich dit tot het belang dat de overheid hieraan toekent? (vrijwel zeker niet opgenomen, want moeilijk in kaart te brengen)</i>
Maatschappelijke haalbaarheid
Bestaat er maatschappelijke ongerustheid met betrekking tot deze problematiek?
Maatschappelijk belang: welk belang hecht de maatschappij aan deze problematiek?
Maatschappelijk draagvlak: - Is er voldoende draagvlak voor aanpak van deze problematiek? - Hoe hoe verhoudt zich dit tot andere zaken die hier relevant zijn, zoals bijvoorbeeld economische afwegingen?

In het volgende hoofdstuk zullen we ingaan op de door ons gebruikte multicriteria methode en op de wijze waarop we die methode in praktijk gebracht hebben binnen de specifieke context van het fasenplan.

3. Uitwerking in de praktijk

Centraal ligt de uitdaging voor een weg te vinden in een veelheid aan informatie (desk research rond de cases; zie: Morrens et al. 2007) die bovendien divers van aard is (verschillende aandachtspunten vertaald in criteria en subcriteria), complex (milieu & gezondheid betreft erg complexe materie, voor juryleden als ook voor (wetenschappelijke) experts) en beoordeeld vanuit verschillende actorperspectieven (diversiteit aan experts en juryleden; zie: Keune et al. 2007a en 2007b). Hierbij is het niet enkel nodig de informatie te structureren als basis voor de oordeelsvorming van de jury en de overheid, maar ook de cases te wegen op basis van die informatie als ook ten opzichte van elkaar om tot een rangschikking van prioriteit voor beleid te komen. De multicriteria methode biedt de mogelijkheid de ongelijksoortige cases en beoordelingsinformatie vergelijkbaar te maken door middel van het vergelijken van rangschikkingen en scores.

3.1 Gebruikte multicriteria methode

Gegeven de verscheidenheid en de complexiteit van het onderwerp en de vraagstelling (priorisering), werd er geopteerd voor operationalisering in de praktijk van een nieuwe multicriteria methode AURORA (Aggregating Unicriterion Rankings into One RAnking) die recent ontwikkeld is aan de Universiteit Antwerpen. Deze methode is gebaseerd op de samenvoeging en vergelijking van rangschikkingen. Experts werd gevraagd om de verschillende cases (alternatieven genoemd binnen multicriteria analyse) te scoren en/of te rangschikken met betrekking tot een bepaald (sub)criterium. De scores van cases op bepaalde criteria kunnen geanalyseerd worden als rangschikking van cases op de scoreschalen van verschillende criteria. De reden om gebruik te maken van rangschikkingen is hoofdzakelijk te zoeken in het feit dat ook voor experts het vaak moeilijk is om over een dergelijke materie in absolute termen te spreken. Tevens laat het werken met rangschikkingen toe om in kwalitatieve termen te antwoorden (vb.: zeer moeilijk, moeilijk, eerder moeilijk, eerder gemakkelijk, enz.). Dit laatste zou echter een probleem opleveren voor de meeste andere multicriteria methoden, daar ze wiskundig niet geschikt zijn om met kwalitatieve gegevens te werken. Het voordeel van het werken met kwalitatieve gegevens is ook dat dit reeds enige robuustheid genereert met betrekking tot de formulering van de inschattingen van experts. Een nadeel is echter dat men geen elementaire rekenkundige bewerkingen meer mag doorvoeren op de gegevens. Dit nadeel weegt echter niet op tegen de kwaliteit en robuustheid van de inschattingen.

Vertrekkende van de gegevens toegeleverd door de experts voor de subcriteria² wordt er door de methode voor ieder hoofdcriterium een rangschikking gebouwd dat een compromis is tussen de verschillende expert-rangschikkingen. Hiervoor maakt men gebruik van een optimalisatieprocedure die gebaseerd is op de maximalisatie van een analytische grootheid, welke de graad van overeenstemming meet tussen de verschillende rangschikkingen enerzijds en de compromis rangschikking anderzijds.

Nadien wordt er voor ieder jurylid apart gebruik gemaakt van de door hem of haar toegeleverde informatie met betrekking tot de belangrijkheden (gewichten) van de hoofdcriteria en opnieuw een compromis rangschikking gebouwd, maar nu met de geconstrueerde compromis rangschikkingen voor de hoofdcriteria. Het resultaat is een rangschikking per jurylid. In de laatste fase kunnen de rangschikkingen bekomen voor de juryleden indien wenselijk opnieuw samengevoegd worden tot een groepscompromis-rangschikking.

Zoals uit het bovenstaande blijkt, is één van de bijkomende voordelen van deze methode dat ze een hiërarchische structuur van het probleem toelaat (subcriteria, hoofdcriteria, groepsrangschikking). In principe is het aantal niveaus in de hiërarchie onbeperkt en hoeft de diepte van iedere tak (het niveau van detail m.b.t. de sub- en subsubcriteria) in de hiërarchie niet dezelfde te zijn. Tevens is het zo dat in iedere fase in de behandeling van de hiërarchie de betrokken actoren kunnen interageren indien wenselijk. Zo kunnen bvb. juryleden hun rangschikking aanpassen indien ze dit wensen vooraleer een groepsrangschikking wordt bepaald.

3.2 Opzet proces

² In geval van het Fasenplan moeten we spreken over hoofd- en subcriteria (zie ook Keune et al. 2007a en Morrens et al. 2007). De hoofdcriteria (Ernst van het gezondheidsrisico, Beleidshaalbaarheid en Maatschappelijke haalbaarheid) zijn elk opgebouwd uit subcriteria die deelaspecten van het hoofdcriterium betreffen.

Oorspronkelijk hadden we voor de analyse van de expertronde een meer beperkte werking van het MCA-model in gedachten dan waar we uiteindelijk op uitgekomen zijn. Schematisch zag dit er voor aanvang van het proces in de praktijk als volgt uit:

Processtap	Beoordelingsbasis
Steunpunt Milieu & Gezondheid	
Selectie significant hoge meetresultaten biomonitoring	- Referentiewaarden - Vergelijking met buitenlandse studies
Groeperen resultaten leeftijdsgroepen biomonitoring	- Resultaten drie campagnes
Preselectie cases voor fase 1 fasenplan	- P90 - Resultaten bij jongeren - Gezondheidsrisico - Ea...
Expertronde	
Individuele scores en/of rangschikkingen van bepaalde cases op bepaalde criteria	- Hun expertise - Deskresearch <i>(per email)</i>
Onderzoeksgroep faseplan	
Synthese van expertinschattingen	- Multicriteria analyse expertinschattingen
Jury	
Zonder desk research informatie uit de criteria, gewichten vragen voor hoofdcriteria en rangschikking cases op individuele basis	- De lijst van cases - Hoofdcriteria <i>(per email)</i>
Onderzoeksgroep faseplan	
Individuele rangschikkingen van de cases op de criteria	- Scores uit expertronde - Input juryleden
Jury	
Per criterium worden de rangschikkingen getoond met zicht op cases, met mogelijkheid tot aanpassen (reflectief moment)	- Resultaten van emailbevraging <i>(individueel gesprek)</i>
Onderzoeksgroep faseplan	
Samenbrengen van de info tot 1 individuele rangschikking	- Rangschikkingen + wegingen
Jury	
- Definitieve individuele wegingen en rangschikkingen op criteria - Definitieve individuele rankings	- Feedback onderzoekers naar jurylid <i>(per email)</i>
Onderzoeksgroep faseplan	
Analyse van rangschikkingen met als doel: - Feedback van proces voor de groep - Mogelijke discussiepunten (goed structureren) naar boven te brengen - Te zoeken naar consensusruimte (clustering)	- Rangschikkingen
Jury	
Groepsadvies voor de overheid	- Synthese input vanuit de expertronde - Synthese individuele input juryleden <i>(groepsdiscussie)</i>

Oorspronkelijk zagen we als belangrijkste ingrediënten de cases, de criteria (desk research), de scores + argumentaties van de experts (vragenlijsten) en de gewichten aan de criteria + argumentaties van de juryleden (vragenlijst en groepsdiscussie).

Aandachtspunten bij aanvang praktische uitwerking

Voor aanvang in de praktijk lag een aantal methodologische vragen voor die we in acht wilden nemen bij de uitwerking. We zullen ze hier overlopen.

Hiërarchische structuur criteria?

Allereerst stelde zich de vraag of we zouden werken met een hiërarchische structuur voor de hoofdcriteria: ernst gezondheidsrisico, beleidshaalbaarheid en maatschappelijke haalbaarheid. Deze drie categorieën kunnen op hun beurt onderverdeeld worden in subcriteria.

Om te voorkomen dat bepaalde cases minder prioriteit (in relatie tot andere cases) toebedeeld krijgen in eerste aanzicht omdat de gemiddelde score op verschillende criteria als het ware een hoge prioriteit op één criterium verbergt (compensatie), kan gewerkt worden volgens het principe van uitsluiting: een hoge score op één criterium boven een bepaalde grenswaarde, ook wel een discordantie-drempel genoemd, maakt dat, ongeacht de scores op andere criteria die bvb. minder prioriteit aangeven, betreffende optie wel tot een eerste selectie van opties gaat behoren van meest prioritaire opties. Zo kan bvb. een hoge urgentiescore op het aspect gezondheidsrisico dermate zwaarwegend worden ingeschat, dat er sowieso meer prioriteit aan wordt gegeven, ongeacht scores op andere criteria.

Het vaststellen van een grenswaarde op bvb. 7, als het bvb. een 10 puntsschaal van Laag – Hoog betreft, vormt zich als het ware automatisch op basis van de expertinschattingen: scoren zij op een criterium 7 of hoger, dan bevelen zij extra prioriteit aan, ondanks scores op andere criteria. De vraag die zij zich hierbij moeten stellen is dan ongeveer als volgt:

Waardeer ik betreffend criterium (of deelaspect daarvan) als dermate belangrijk voor actie door de overheid, dat discussie over de andere criteria eigenlijk van secundair belang is?

Er selecteren zich dus in eerste instantie twee hoofdgroepen van cases uit:

- Cases waar op één van de criteria door experts een score boven een grenswaarde wordt gegeven
- Cases waarbij dat niet het geval is

De eerstgenoemde groep zal prioritaire aandacht krijgen in de MCA, en de kopgroep vormen in de rangorde van cases voor de overheid. Binnen deze kopgroep wordt overigens de rangorde wel bepaald op basis van scores op alle criteria.

Welke eisen stellen de analyse van de expertadviezen en het groepsproces van de jury?

Verschillende strategische opties liggen voor bij het vormgeven van analyse en groepsdiscussie. Globaal voorzien we een werkwijze die vertrekt bij individuele visies en in principe leidt tot een groepsadvies aan de overheid.

Bij de expertinschattingen gebeurt het groeperen door middel van het proces van vergelijking van rangschikkingen welke door het multicriteria model omgevormd wordt tot compromisrangschikkingen. Hierbij worden experts op het niveau van de subcriteria in principe met gelijk gewicht meegewogen en wordt ook geen keuze gemaakt waar het (verschil in) belangrijkheid (gewicht) van (sub)criteria betreft.

Bij de jury komen in principe verschillen naar voren waar het de belangrijkheid van criteria betreft: afhankelijk van hun voorkeur zullen juryleden in principe verschillende gewichten toekennen wat leidt tot verschillende rangschikkingen van de cases. Het groeperen van deze verschillende inschattingen is onder andere afhankelijk van de doelstelling die gekozen wordt voor de groepsdiscussie. Uit de volgende mogelijkheden kan gekozen worden:

- De doelstelling is het bereiken van een consensus: het formuleren (of berekenen) van één optimaal voorstel (een combinatie van opties, rangorde) aan de overheid
- De doelstelling is het in kaart brengen van diversiteit: een overzicht van verschillende opinies inzake de cases
- De doelstelling is onderhandeling tussen verschillende gezichtspunten, op basis waarvan een herschikking van voorkeuren, of verzamelingen van voorkeuren kan plaatsvinden

Onzekerheid en argumentaties

Naast een kwantitatieve vertaling van verscheidenheid aan inschattingen, aandachtspunten en voorkeuren, willen we ook aandacht en ruimte reserveren voor kwalitatieve informatie. We denken hierbij aan argumenten die naar voren gebracht worden. We denken hierbij ook aan onzekerheid van inschattingen zoals die bvb. bij de experts aan de orde kan zijn.

Wat onzekerheid betreft hadden we optioneel de volgende kadering binnen de vragenlijsten voor experts in gedachten: experts geven een optimistische en een pessimistische inschatting van bvb. gezondheidsrisico of beleidshaalbaarheid. Naarmate het verschil tussen beide inschattingen kleiner is, is er minder sprake van inschattingsonzekerheid. Een andere optie was om de mate van onzekerheid in kaart te brengen door per subcriterium een extra criterium in te voeren, waarop een expert de mate van onzekerheid van de inschatting kon aangeven. Omwille van de complexiteit van de vragenlijsten en de hoeveelheid inspanning die hiermee gevraagd wordt van experts, hebben we beide opties laten vallen in de concrete uitwerking.

Analyse van het model

Om de gevoeligheid van het model te analyseren is het gebruikelijk een *gevoeligheidsanalyse* uit te voeren in de ontwikkelingsfase. Hierbij wordt bekeken in hoeverre belangrijke variabelen van het model invloed hebben op de uitkomsten van het model en hoe zich dat verhoudt tot de invloed van andere variabelen. Dit kan bvb. bekeken worden op het niveau van de (hoofd- en sub-) criteria: welk effect hebben verschillende gewichtsverhoudingen tussen de criteria op het model? Dit kan ook bekeken worden op het niveau van de experts: in hoeverre is het model gevoelig voor standpunten die duidelijk afwijken van de meerderheid en geformuleerd worden door een minderheid van experts? Op het niveau van de scores kan bekeken worden in hoeverre het model gevoelig is voor wijzigingen in scores op een criterium.

Naast een gevoeligheidsanalyse is het ook mogelijk een *diversiteitanalyse* te doen. Hierbij kan bekeken worden in hoeverre het model de cases niet enkel individueel ten opzichte van elkaar bekijkt, maar ook in subgroepen kan denken, subgroepen die dan bvb. niet noodzakelijk bepaald worden door clustering op basis van hiërarchische posities van cases. Hiermee kan mogelijk een (extra) basis voor onderhandeling tussen actoren betrokken bij advies- of besluitvorming geboden worden.

Learning by doing

Het gaat hier om de operationalisering en toepassing van deze multicriteria methode binnen de probleemspecifieke context van het fasenplan. We zullen dit iteratieve proces van 'learning by doing' (zie bvb.: Belton and Stewart 2002) in de volgende paragrafen nader beschrijven. De noodzakelijke interdisciplinaire samenwerking daarbij tussen een multicriteria methode expert en een sociaal wetenschappelijk expert zal in het laatste hoofdstuk bij de evaluatie behandeld worden.

3.3 Expert ronde

De vragenlijsten

Tijdens de opzet van de expertronde hebben we een aantal keuzen gemaakt die van belang waren voor de multicriteria analyse. Aandachtspunten hierbij waren categorieën die structuur geven aan de inhoud van het informatiemateriaal dat door de experts beoordeeld moest worden en categorieën die vorm moesten geven aan de expertinschattingen. De inhoudelijke categorieën betroffen de specifieke vragen die we de experts (per hoofdcriterium) hebben voorgelegd. De antwoordcategorieën betroffen het soort antwoorden dat experts werd gevraagd.

Om de vergelijkbaarheid van expertinschattingen in termen van rangschikkingen mogelijk te maken, hebben we gekozen voor een combinatie van scores op voor alle cases gelijke antwoordschalen en rangschikkingsvragen. Dit verschilde per vraag. In het geval van scores konden experts op een kwalitatieve schaal met 6 opties een inschatting aanduiden. Hierbij was steeds sprake van een ordinale schaal. Dit is noodzakelijk om in de analyse een rangorde tussen de cases te kunnen bepalen. De vraagstelling betrof dan de inschatting van een individuele case op de ordinale antwoordschaal. Bij het vragen van een *rangschikking* tussen de cases vroegen we naar een vergelijking van cases ten opzichte van de vraagstelling.

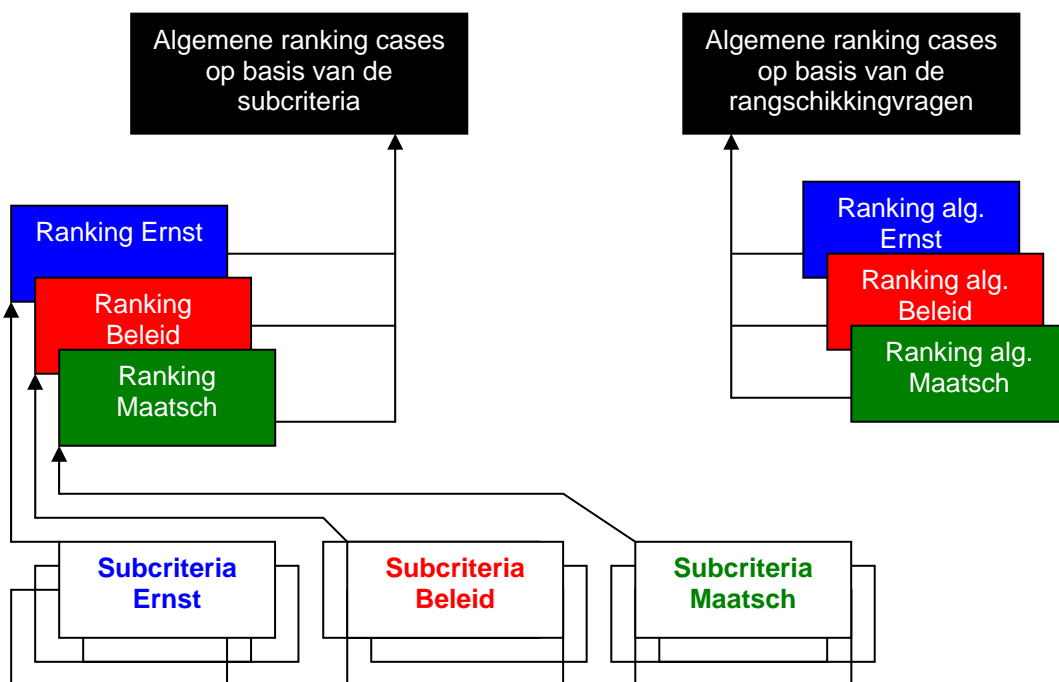
Bij zowel de criteria 'Ernst gezondheidsrisico' als 'Beleidshaalbaarheid' hebben we de experten bij de inhoudelijke inschattingsvragen antwoordschalen voorgelegd. Bij het criterium 'Maatschappelijke haalbaarheid' hebben we enkel rangschikkingvragen gesteld. De reden hiervoor was dat het bepalen van een waarde op een ordinale schaal voor de onderwerpen die bij de maatschappelijke haalbaarheid aan bod kwamen te moeilijk was. Voor ons gevoel konden we hier beter de cases in plaats van antwoordmogelijkheden vergelijken. Bij de twee andere hoofdcriteria leek ons dat beter mogelijk omdat er meer concreet materiaal voor handen was, ofwel middels de desk research, ofwel middels de concrete praktijkervaringen van experten zelf. Bij de maatschappelijke haalbaarheid moesten we meer vertrouwen op een gevoelsmatige inschatting van de experten op basis van 'circumstantial evidence'. De elementen die we binnen onze mogelijkheden konden inventariseren in de desk research zijn omwille van de complexe eigenheid van maatschappelijke fenomenen minder makkelijk concreet te maken dan bij de andere twee hoofdcriteria het geval was. Hoewel we bvb. bij het criterium 'Beleidshaalbaarheid' beslist hebben dat desk research voor ons omwille van de complexiteit onbegonnen werk was, leek het naar onze inschatting voor ingewijden, experten werkzaam binnen die beleidspraktijk, concreter in te schatten. Maatschappelijke haalbaarheid leek ons kortom een relatief moeilijker in te schatten fenomeen, en is bovendien een thema dat binnen beleid inzake milieu & gezondheid relatief veel minder aandacht krijgt en dus ook minder makkelijk te duiden is.

Het voorgaande betreft overigens de *inhoudelijke vragen* op subcriterium niveau. Deze vragen werden steeds aan het begin van een vragenlijst gesteld (voor een overzicht van de vragenlijsten zie: bijlage bij Morrens et al. 2007) Hiernaast hebben we ook (bij elk hoofd criterium) een *algemene rangschikkingvraag* gesteld. Deze vraag werd ter afsluiting gesteld. We vroegen de experten daarbij op basis van de voorgaande (inhoudelijk) vraagstelling een rangschikking van de cases op het hoofd criterium te geven.

Bij beantwoording van de vragen werd experten ook steeds expliciet gevraagd argumenten voor hun scores en/of rangschikkingen te geven. Bovendien werd hen expliciet de mogelijkheid geboden in geval van gebrek aan expertise met betrekking tot een onderwerp dit aan te geven.

Hiërarchische structuur criteria

We hebben voor de volgende hiërarchische structuur gekozen bij de analyse van de expert inschattingen van de subcriteria.



De optie om met grenswaarden (discordantiedrempels) te werken hebben we omwille van de complexiteit van de vragenlijsten in dit project laten vallen.

Compromisrangschikkingen criteria

De computer zoekt naar de beste compromisrangschikking (of rangschikkingen) op basis van de expertrangschikkingen die voorhanden zijn in het kader van een bepaald hiërarchisch niveau van criteria. Het model doet dit door de verschillende rangschikkingen op een paarsgewijze manier te vergelijken met een voorstel tot compromisrangschikking. Dit levert voor iedere rangschikking een waarde op van een vergelijkingsmaatstaf (i.e. een rangcorrelatiecoëfficiënt) die weergeeft hoe sterk een rangschikking lijkt op het voorstel tot compromis. Van de zo bekomen lijst van rangcorrelatiecoëfficiënten wordt de mediane waarde³ bepaald, welke weergeeft dat 50% van de rangschikkingen een graad van overeenstemming heeft die hoger is dan deze mediane waarde. Hoe groter deze mediane waarde, hoe groter de graad van consensus binnen de groep van rangschikkingen m.b.t. de voorgestelde compromisrangschikking. Deze stappen worden in een algoritme herhaald tot wanneer 1 (of meerdere) compromisrangschikking(en) gevonden word(t)(en) met de maximale waarde voor deze mediane waarde van de lijst van rangcorrelatiecoëfficiënten. Deze compromisrangschikking(en) kan (kunnen) al dan niet gelijkgerangschikte cases bevatten. Dit is een mogelijke keuze in het programma, waarbij we ervoor geopteerd hebben om gelijkgerangschikte cases toe te laten om zo het keuzeproces dat leidt tot een compromisrangschikking te vergemakkelijken.

Om de scores van experts op de vragen bij de criteria 'Ernst gezondheidsrisico' als 'Beleidshaalbaarheid' om te zetten in rangschikkingen, hebben we de scores op een subcriterium van de cases onderling vergeleken en op basis van die vergelijking een rangorde bepaald. Als voorbeeld: als op de vraag naar beleidshaalbaarheid van bestrijding aan de bron voor 5 cases werd geantwoord dat de haalbaarheid klein is en voor 1 case dat de haalbaarheid zeer klein is, krijgen de 5 cases positie 1 in de rangorde, en de overgebleven case positie 2 in de rangorde. We hebben dus in de analyse ten behoeve van het berekenen van compromisrangschikkingen enkel de volgordes van de cases op de antwoordschalen gebruikt, en niet de reële posities op die schalen. Hierdoor vallen bij het berekenen van de compromisrangschikkingen de kwalificaties per case buiten de analyse. Het kan bvb. zo zijn dat een case in rangorde bij een expert 'hoog scoort' op een criterium, terwijl ze als relatief moeilijk haalbaar ingeschat wordt, maar dan nog altijd minder moeilijk dan de overige cases.

De reden voor deze analytische keuze is tweërlei:

- Kwalificaties kun je niet eenduidig vergelijkbaar achten. De preferentiestructuur van de ene expert hoeft niet dezelfde dan die van een andere te zijn. Dus wat bvb. moeilijk haalbaar is bij de één zal misschien redelijk goed haalbaar bij een ander zijn
- De kwalificaties per case voegen *an sich* geen onderscheidende informatie toe aan oplossing van het vraagstuk welke rangschikking cases hebben in prioriteit; de rangschikking blijft identiek, ongeacht waar de groep zich op een schaal bevindt

Deze laatstgenoemde reden impliceert ook dat bij het berekenen van de compromisrangschikkingen enkel rekening is gehouden met experts die een onderscheid in kwalificatie maken tussen de cases, een onderscheid op basis waarvan een volgorde bepaald kan worden. De (verschillen in) kwalificaties zijn wel interessant op case niveau als we spreiding tussen inschattingen bekijken: hier hebben we in de analyse wel gewerkt met de exacte kwalificaties per case, aangezien we daar analyse deden op de verschillen op case niveau, en niet op het niveau van een groep van cases (zie onder bij beschrijving 'Spreiding expertinschattingen').

Om tot een hoger niveau van rangschikking te komen wordt een compromis gezocht tussen de compromisrangschikkingen op het niveau eronder. In geval van compromisrangschikking van verschillende (sub)criteria op een hoger niveau wordt geen verschil in gewicht toegepast tussen de verschillende (sub)criteria. Verschil in gewicht is een mogelijkheid die de jury heeft. Ook wordt er hier geen verschil in gewicht tussen de experts gemaakt. In geval experts geen rangschikking hebben gegeven of in hun scores geen onderscheid maken tussen de cases, worden ze niet meegenomen in het zoeken naar een compromisrangschikking.

³ In plaats van de mediaan had men ook kunnen opteren voor een ander type compromis-maatstaf (33ste of 66ste percentiel) hetgeen aan outliers een groter of kleiner gewicht zou hebben toegekend, hetgeen in het huidige proces ter ondersteuning van de groepsbesluitvorming te vermijden valt.

In geval van de algemene rangschikkingsvraag die experten aan het einde van de vragenlijsten gesteld is, is mogelijks wel sprake van gewichtsverschillen met betrekking tot de eraan voorafgaande vragen op subcriterium niveau. Experten kunnen hierbij bvb. ongelijk gewicht aan de subcriteria toegekend hebben. Dit betekent dat de rangschikking op het niveau van de hoofdcriteria en op het allerhoogste niveau van de algemene rangschikking (beide op basis van de subcriteria) niet noodzakelijk hetzelfde resultaat geeft als de algemene rangschikking op basis van de rangschikkingsvraag. Verschil op deze twee niveaus kan overigens ook verklaard worden door in ieder geval drie andere oorzaken:

1. Het kan zijn dat bij beantwoording van de algemene rangschikkingvraag andere informatie is meegewogen dan in de vragen over de subcriteria.
2. Het kan zijn dat experten een rangschikkingsvraag anders ervaren dan een inschattingvraag op het niveau van één case.
3. Verschillende experten gaven op de subcriteria ofwel geen duidelijk onderscheid aan tussen cases en gaven dus geen onderscheidend oordeel als basis voor een rangorde. Ofwel gaven experten op onderdelen aan geen expertise te hebben. In beide gevallen telden deze inschattingen van experten niet mee bij het opmaken van de rangschikkingen op het niveau van de hoofdcriteria. Dit betekent dat het aantal experten dat op een subcriterium gewicht had in sommige gevallen beduidend lager was dan het aantal experten dat een rangschikking op meer globaal niveau aangaf.

Clustering expertinschattingen

De clustering werkt als volgt: de computer zoekt steeds naar de kleinste verschillen en maakt daar een compromis voor. Dit proces gaat net zolang door tot er geen verschillen meer gevonden worden. Dat de computer in elke stap het kleinste van de nog aanwezige verschillen zoekt zegt niks over de grootte van het verschil. Het aantal stappen dat door de computer gezet moet worden zegt in principe ook niks over de uiteenlopendheid van de inschattingen van de experten binnen de groep. Dit proces is afhankelijk van de diversiteit aan verschillen (des te meer verscheidenheid, des te groter het aantal stappen) maar niet van de mate van verschil. Het kan dus zijn dat in geval van een groot aantal stappen op zich binnen een beperkte reikwijdte een beperkte mate van verschil wordt overbrugd, terwijl het ook kan zijn dat in een beperkt aantal stappen juist een grote mate van verschil wordt overbrugd. De diversiteit is mede afhankelijk van de potentiële verscheidenheid op basis van het aantal experten in de groep.

De clustering in grotere groepen zegt op zich nog niks over de inhoud van de clustering: tussen welke verschillen wordt een compromis gezocht; betreft het een groot of klein verschil (aantal verschillen dat overbrugd wordt in één stap, grootte van de verschillen), welk soort verschil (bvb. tussen switchen van twee cases tussen de plaatsen 5 en 6 in een rangschikking of switchen van twee cases tussen de plaatsen 1 en 6). Het is dus van belang ook te kijken tussen welke rangschikkingen welk compromis gegenereerd wordt. Bovendien is het goed een maat te formuleren voor de mate van verschil die overbrugd wordt in een stap. Tevens moeten we de potentiële verscheidenheid op basis van het aantal experten in rekening brengen.

We maken gebruik van het berekenen van correlatiecoëfficiënten (Kendall 1938, Kendall & Gibbons 1990) met betrekking tot verschillende inschattingen. Op basis van de mediane correlatiecoëfficiënt worden clusters gevormd daar waar relatief de sterkste correlatie berekend wordt. Door ook de minimum correlatie binnen een nieuw gevormde cluster te berekenen, kunnen we de moeilijkheidsgraad van de clustering typeren. Hieronder volgt een duiding van de verschillende correlaties (in geval een correlatiecoëfficiënt negatief is, gaat het om een negatieve correlatie):

Correlatiecoëfficiënt	Kracht lineair verband
< 0,3	zeer zwak
0,3 - 0,5	zwak
0,5 - 0,7	matig

0,7 - 0,85	sterk
0,85 - 0,95	zeer sterk
> 0,95	uitzonderlijk sterk

Analytische elementen

Op basis van de analyse van de informatie van de expertronde hebben we ook een aantal analytische elementen geïnventariseerd die interessant zijn naast de analyse in rangschikkingen. We behandelen de verwerking van die elementen in het multicriteria model in de volgende paragraaf bij de jurydiscussie. Hier behandelen we de wijze waarop we sommige elementen in kaart gebracht hebben op een wijze die verwerking binnen het multicriteria model mogelijk maakt, en waarbij duiding nodig is. Het gaat dan specifiek over:

- Spreiding tussen expertinschattingen: de mate waarin experts van mening verschillen
- Kennis- en inschattingonzekerheid: welke soort onzekerheid geven experts aan?

Spreiding expertinschattingen

De *inhoudelijke spreiding* van antwoorden van experts op een vraag hebben we berekend door het verschil in posities te nemen tussen de minimum en de maximum score (maximum – minimum + 1). Als bijvoorbeeld deze schaal wordt gebruikt:

zeer moeilijk	moeilijk	eerder moeilijk	eerder goed	goed	zeer goed
---------------	----------	-----------------	-------------	------	-----------

Als de meest negatieve beoordeling ‘moeilijk’ is (positie 2) en de meest positieve beoordeling ‘zeer goed’ (positie 6) dan is de inhoudelijke spreiding: 6 – 2 + 1 = 5.

De *verspreiding* van experts binnen de range aan voorkomende antwoorden hebben we berekend aan de hand van een spreidingscoëfficiënt. Wiskundig is dat berekend volgens deze formule (Leik, 1966; Thijssen, 2006):

$$ODI = 1 - \frac{\sum_{i=1}^{k-1} (2F_i - 1)^2}{k - 1}$$

Deze ordinale dispersie-index geeft het percentage weer van de maximale verspreiding, dus hoe hoger hoe meer verspreiding.

Kennis- en inschattingonzekerheid

Voor de verschillende typen onzekerheid die we geïnventariseerd hebben verwijzen we naar het rapport van de expertronde (Keune et al. 2007a). Hier beschrijven we enkel hoe we deze gekwantificeerd hebben om te kunnen gebruiken binnen het multicriteria model. We hebben in de analyse zowel vormen van onzekerheid geanalyseerd op het niveau van cases als op het niveau van experts. We kunnen dus iets zeggen over onzekerheid van experts en over onzekerheid inzake bepaalde cases (voor bepaalde (sub)criteria).

Als vuistregel hebben we steeds onzekerheid geteld in het aantal experts dat een vorm van onzekerheid aangeeft op het analyseniveau (case, criterium, ...) dat we willen beschrijven. Hierbij hebben we experts steeds maar één keer meegeteld. Als een expert bvb. op het niveau van een subcriterium bij meerdere cases een bepaalde vorm van onzekerheid aangeeft, tellen we deze expert op het niveau van dat subcriterium één keer. Als bvb. een expert bij een case meerdere vormen van onzekerheid aangeeft tellen we deze expert maar één keer als we enkel onzekerheid in algemene zin willen duiden. In geval van duiding van een bepaald type van onzekerheid kan dan wel meer specifiek gekeken worden naar het aantal experts dat een bepaald type onzekerheid benoemt binnen een criterium, zij het dat experts ook dan maximaal één keer geteld worden.

In geval een expert op een hoger analyseniveau dan één of een deel van de cases aangeeft dat er sprake is van onzekerheid, door bijvoorbeeld aan te geven geen expertise te hebben inzake een subcriterium of een rangschikking, dan hebben we dit geïnterpreteerd als geldend voor alle cases op dat subcriterium.

3.4 Jury discussie

We hebben in de aanloop naar de jurydiscussie verschillende stappen doorlopen waarbij een aantal contactmomenten voorzien was in de opzet bij aanvang. Gaandeweg is hier echter zowel inhoudelijk als wat tussenstappen betreft één en ander toegevoegd op basis van zowel de analyse van de expertronde als de contacten met de jury zelf. In essentie komt het erop neer dat er een extra bevraging van de juryleden is ingelast ten opzichte van de oorspronkelijke opzet waarin onder andere elementen bevraagd worden die we op basis van analyse van de expertinformatie hebben geïnventariseerd en welke, zo bleek tijdens de interviews die we met de juryleden afzonderlijk hadden, de belangstelling hadden van de juryleden om mee hun inbreng te kleuren voor de groepsdiscussie. Voor een overzicht van de stappen voorafgaand aan de groepsdiscussie zie onderstaand schema. Voor een nadere duiding van die analytische elementen verwijzen we naar het rapport van de expertronde en het juryrapport (Keune et al. 2007a, b). Hier duiden we de wijze waarop rekenschap is gegeven van die analytische element binnen het multicriteria model. Het gaat dan om de volgende analytische elementen:

- Diversiteit aan achtergronden van experts
- Spreiding tussen expertinschattingen: de mate waarin experts van mening verschillen
- Kennis- en inschattingonzekerheid: welke soort onzekerheid geven experts aan?
- Kennisbasis: welk aantal experts geeft een inschatting die onderscheid maakt tussen de cases?

Deze elementen bleken voor de meeste juryleden relevante informatie bij de beoordeling van de vraag naar priorisering van cases voor overheidsbeleid, en zijn bevraagd in een tweede vragenlijst voor de juryleden, voorafgaand aan de groepsdiscussie.

Overzicht verschillende stappen in het juryproces

Processtap	Beoordelingsbasis
Jury: vragenlijst 1 (per email)	
Zonder desk research informatie uit de criteria, gewichten vragen voor hoofdcriteria en rangschikking cases op individuele basis	<ul style="list-style-type: none"> - De lijst van cases - Hoofdcriteria
Onderzoeksgroep faseplan	
Individuele rangschikkingen van de cases op de criteria	<ul style="list-style-type: none"> - Scores uit expertronde - Input juryleden
Jury: individueel gesprek	
Per criterium worden de rangschikkingen getoond met zicht op cases, met mogelijkheid tot aanpassen (reflectief moment)	<ul style="list-style-type: none"> - Resultaten van emailbevraging
Onderzoeksgroep faseplan	
<ul style="list-style-type: none"> - Synthese van individueel gesprek - Opmaak tweede vragenlijst: oa. subcriteria, analytische elementen 	<ul style="list-style-type: none"> - Resultaten gesprekken juryleden - Informatie desk research, subcriteria - Overzicht eigen inbreng tot dan
Jury: vragenlijst 2 (per email)	
<ul style="list-style-type: none"> - Rangschikking cases - Gewichten subcriteria + hoofdcriteria - Gewichten analytische elementen 	<ul style="list-style-type: none"> - Desk research informatie - Informatie over argumentaties experts en juryleden
Onderzoeksgroep faseplan	
Samenbrengen van de info tot overzicht individuele rangschikkingen	Rangschikkingen + wegingen obv voorgaande stappen
Jury: feedbackdocument (per email)	
Bepalen definitieve individuele input voor de jurydiscussie	<ul style="list-style-type: none"> - Feedback onderzoekers naar jurylid

Onderzoeksgroep faseplan	
Analyse van rangschikkingen met als doel: <ul style="list-style-type: none"> - Feedback van proces voor de groep - Mogelijke discussiepunten (goed structureren) naar boven te brengen - Te zoeken naar consensusruimte (clustering) Analyse van voorkeuren analytische elementen	<ul style="list-style-type: none"> - Informatie voorgaande stappen - Individuele input die juryleden voor de groepsdiscussie willen inbrengen
Jury: groepsdiscussie	
Groepsadvies voor de overheid	<ul style="list-style-type: none"> - Synthese input vanuit de expertronde - Synthese individuele input juryleden
Onderzoeksgroep faseplan	
Synthese groepsadvies	Groepsdiscussie
Jury: evaluatievragenlijst (per email)	
<ul style="list-style-type: none"> - Keuze formulering groepsadvies - Keuze rapporteerwijze eigen voorkeuren 	<ul style="list-style-type: none"> - Synthese groepsadvies - Evaluatievragen
Onderzoeksgroep faseplan	
Opmaak definitief juryadvies	Alle voorgaande stappen

We beschrijven hieronder hoe we de voorkeuren van de juryleden zoals ze die in vragenlijsten en groepsdiscussie naar voren brachten binnen het multicriteria model methodologisch behandeld hebben. Vanzelfsprekend bespreken we hier enkel die zaken die van belang zijn voor het multicriteria model.

Verwerking voorkeuren juryleden analytische elementen

In de tweede vragenlijst hebben we juryleden gevraagd of ze bepaalde analytische elementen al dan niet in rekening willen brengen binnen het multicriteria model. We hebben dit voorgesteld als *informatieknoppen*, die ze al dan niet aan konden zetten, aangezien het om toegevoegde analytische informatie ging ten opzichte van de informatie waarover ze in voorgaande stappen geïnformeerd waren (voor de vragenlijst, zie: Keune et al. 2007b). Behalve aanzetten (of uit laten) werd hen ook de mogelijkheid gegeven *richting* en *sterkte* aan te geven van doorwerking van die informatie binnen het multicriteria model. Onder *richting* verstaan we dempen of versterken: als men bijvoorbeeld een onzekere experten minder zwaar wil laten meewegen in de beoordeling van de cases, dan kan men de invloed daarvan dempen. De *sterkte* van de demping kan bovendien gevarieerd worden van een kleine mate van effect tot een grote mate van effect.

In algemene zin kunnen we een aantal typen verwerkingen beschrijven. Voor alle duidelijkheid, dit wordt steeds per subcriterium apart bekeken aangezien analytische kwalificaties op dat niveau geïnventariseerd zijn en er per expert verschil in kwalificatie kan bestaan tussen de subcriteria. Als een expert bijvoorbeeld op een bepaald subcriterium aangeeft onzeker te zijn over de inschatting van een bepaalde case, hoeft dat bij een ander subcriterium niet ook het geval te zijn. Naast een onderscheid tussen experten binnen subcriteria kan dus bovendien een onderscheid tussen cases binnen subcriteria relevant zijn.

Allereerst gaan we nu in op de verwerking op het *niveau van de experten*. Hierbij wordt relevante analytische informatie verwerkt door experten verschillend te behandelen op basis van de voorkeur die bvb. een jurylid aangeeft met betrekking tot een kenmerk zoals wetenschappelijke onzekerheid. Wil een jurylid de inbreng van een expert met een bepaalde kwalificatie versterken, dan wordt de rangschikking van een expert die voldoet aan die kwalificatie extra meegewogen in de analyse. Voor een klein versterkend effect wordt de rangschikking van betreffende expert één keer extra meegewogen in de analyse, voor een middelmatig versterkend effect twee keer extra, voor een groot versterkend effect vier keer extra. In geval van een dempend effect worden dezelfde grootordes gebruikt maar dan voor alle experten uitgezonderd de expert die gedempt moet worden. Als voorbeeld: als een expert in kleine mate gedempt moet worden, worden alle andere experten (die niet aan die kwalificatie voldoen) één keer extra meegewogen in de analyse.

Relevante analytische informatie wordt verwerkt op het *niveau van de cases* indien het een kwalificatie betreft die enkel van toepassing is op een beperkt aantal cases (dus niet op alle cases en dus op de hele rangschikking van betreffende expert op betreffend subcriterium), dan moet de informatie verwerkt worden door cases binnen betreffende rangschikking van de expert anders te behandelen. Wil een jurylid de inbreng van een expert met een bepaalde kwalificatie versterken, en doet die kwalificatie zich enkel voor op case niveau, dan wordt een case die binnen de rangschikking van een expert, die voldoet aan die kwalificatie, binnen de analyse meer prioriteit toegekend door de case een hogere positie in de rangschikking te geven. Voor een klein versterkend effect stijgt een case één positie, voor een middelmatig versterkend effect twee, en voor een groot versterkend effect vier. Als een case, die moet stijgen, al een relatief hoge positie binnen de rangschikking heeft (en dus maar in beperkte mate kan stijgen) dan kan, in geval de stijging niet in de rangschikking verwerkt kan worden door positiestijging, het versterkend effect bekomen worden door alle andere cases in gelijke mate in de rangschikking te laten dalen. Zo wordt hetzelfde verschil in positie, en dus in prioriteit, bereikt. Dit kan wel betekenen dat het aantal posities binnen de rangschikking vergroot wordt ten opzichte van de oorspronkelijke rangschikking. In geval van een dempend effect worden dezelfde grootordes gebruikt maar dan in omgekeerde richting.

We bespreken hieronder per type analytisch element specifieke kenmerken van de berekening in het multicriteria model.

Diversiteit aan achtergronden van experts

Vraag Infoknop achtergrond experts						
Expertise	Knop aanzetten ? (ja/nee)	Versterkend (+) of dempend (-) effect? (kruis aan)		Mate effect (kruis aan)		
		+	-	Klein	Middelmatig	Groot
<i>Medisch</i>				(1)	(2)	(4)
<i>Milieukundig</i>				(1)	(2)	(4)
<i>Medisch - milieukundig</i>				(1)	(2)	(4)

We hebben een typologie van experts in kaart gebracht op basis van de driedeling medisch, milieukundig en medisch milieukundig. In het geval van beleidshaalbaarheid zijn enkel de eerste twee typen gebruikt aangezien de beleidsdomeinen op die manier opgedeeld zijn. De verwerking van deze analytische informatie gebeurde op basis van de voorkeuren van de juryleden op het *niveau van de experts*.

Spreiding tussen expertinschattingen

Vraag Infoknop spreiding						
Spreiding inschattingen experts	Knop aanzetten ? (ja/nee)	Versterkend (+) of dempend (-) effect? (kruis aan)		Mate effect (kruis aan)		
		+	-	Klein	Middelmatig	Groot
<i>Spreiding groot maar ongelijk verdeeld</i>				(1)	(2)	(4)
<i>Spreiding groot maar gelijk verdeeld</i>				(1)	(2)	(4)
<i>Spreiding klein</i>				(1)	(2)	(4)

Hierbij hebben we eerst de *inhoudelijke spreiding* (hoe groot is het verschil tussen de meest uiteenlopende inschattingen) en de *verspreiding van de experts* binnen dat bereik (in hoeverre zijn de expertopinions gelijk verdeeld over de verschillende inschattingen binnen het inhoudelijk bereik, of tekent zich een meerderheid van experts af op een bepaalde inschatting?) geïnventariseerd en gekwalificeerd. We hebben de volgende typologie gebruikt voor kwalificatie van de *inhoudelijke spreiding*:

Verschil minimum – maximum inschatting (max – min + 1)	Kwalificatie spreiding
1 - 2	Klein
3 - 4	Middelmatig
5 - 6	Groot

Voor de *verspreiding van de experts* (verdeling) hebben we de volgende typologie gebruikt:

Mate van verspreiding experts (ordinale dispersie-index)	Kwalificatie verspreiding
0 < 0,35	Relatief gelijke verdeling
0,35 < 0,65	Gemiddelde verdeling
0,65 - 1	Relatief ongelijke verdeling

Deze analytische informatie is op basis van de voorkeuren van juryleden verwerkt op het *niveau van de cases* indien de kwalificatie enkel een beperkt aantal cases betrof. Indien de kwalificatie gelijk was voor alle cases vond de verwerking plaats op het *niveau van de experten*.

Kennis- en inschattingonzekerheid

Vraag Infoknop kennisonzekerheid						
Type kennis- onzekerheid expert	Knop aanzetten ? (ja/nee)	Versterkend (+) of dempend (-) effect? (kruis aan)		Mate effect (kruis aan)		
		+	-	Klein	Middelmatig	Groot
<i>Type 1</i>				(1)	(2)	(4)
<i>Type 2</i>				(1)	(2)	(4)
...				(1)	(2)	(4)

Voor de kwalificatie van de typen onzekerheid verwijzen we naar paragraaf 3.4.1 *Kennis- en inschattingonzekerheid* in het rapport over de expertronde (Keune et al. 2007a). Voor de inventarisatie van de typen onzekerheid verwijzen we naar het deel *Kennis- en inschattingonzekerheid* onder paragraaf 3.3 van dit rapport.

Aangezien het hier zowel kwalificaties op het niveau van gehele rangschikkingen van experten kon betreffen als op het niveau van individuele cases, hebben we hier ook een mix van verwerkingen op het *niveau van de experten* en op het *niveau van de cases* toegepast.

Kennisbasis

Vraag Infoknop kennisonzekerheid						
Type kennis- onzekerheid subcriteria	Knop aanzetten ? (ja/nee)	Versterkend (+) of dempend (-) effect? (kruis aan)		Mate effect (kruis aan)		
		+	-	Klein	Middelmatig	Groot
<i>Beperkt aantal experten maakt onderscheid tussen cases (*)</i>				(1)	(2)	(4)

(*) In geval een expert alle cases op een subcriterium gelijk scoort, weegt deze niet mee omdat er geen rangorde van gemaakt kan worden.

We hebben de kennisbasis in termen van het aantal experten dat een uitspraak doet met betrekking tot een subcriterium en daarbij bovendien onderscheid maakt tussen de cases als volgt geïnventariseerd en gekwalificeerd:

Aantal experten dat op niveau van een case WEL een inschatting geeft			
	Ernst	Beleid	Maatsch
geen uitval	15	10	11
kleine uitval: <20%	12 < 15	8 < 10	8,8 < 11
matige uitval: -20%<40%	9 < 12	6 < 8	6,6 < 8,8
grote uitval: -40<60%	6 < 9	4 < 6	4,4 < 6,6
zeer grote uitval: -60<80%	3 < 6	2 < 4	2,2 < 4,4
extreem grote uitval: -80 tot 100%	0 < 3	0 < 2	0 < 2,2

In geval er sprake was van een grote tot zeer grote uitval hebben we dat geïnterpreteerd als een '*Beperkt aantal experten maakt onderscheid tussen cases*'. Omdat dit enkel op het *niveau van de cases* voorkwam, hebben we dit op dat niveau verwerkt.

Clusteranalyse voorkeursrangschikkingen juryleden

Analoog aan de clusteranalyse die we gebruikt hebben voor clustering van expertinschattingen (par. 3.3) hebben we ook de input die verschillende juryleden gaven als basis voor de groepsdiscussie aan een clusteranalyse onderworpen. Het betrof hier dan wel enkel de input in termen van rangschikkingen. Niet alle juryleden echter gaven aan een rangschikking als input voor de groepsdiscussie te willen geven (zie: Keune et al. 2007b).

De clustering werkt als volgt: de computer zoekt steeds naar de kleinste verschillen tussen rangschikkingen en maakt daar een compromisrangschikking voor. Dit proces gaat net zolang door tot er geen verschillen meer gevonden worden. Dat de computer in elke stap het kleinste van de nog aanwezige verschillen zoekt, zegt niks over de grootte van het verschil. Het aantal stappen dat door de computer gezet moet worden zegt in principe ook niks over de uiteenlopendheid van de inschattingen van de juryleden binnen de groep. Dit proces is afhankelijk van de diversiteit aan verschillen (des te meer verscheidenheid, des te groter het aantal stappen) maar niet van de mate van verschil. Het kan dus zijn dat in geval van een groot aantal stappen op zich binnen een beperkte reikwijdte een beperkte mate van verschil wordt overbrugd, terwijl het ook kan zijn dat in een beperkt aantal stappen juist een grote mate van verschil wordt overbrugd. De diversiteit is mede afhankelijk van de potentiële verscheidenheid op basis van het aantal juryleden in de groep. Om de moeilijkheidsgraad van een clusterstap te definiëren maken we gebruik van een rangcorrelatiecoëfficiënt. Deze beschrijft de mate waarin de rangschikkingen die geclusterd worden op elkaar lijken. De rangcorrelatiecoëfficiënt die in deze procedure wordt gebruikt is de zgn. Kendall rangcorrelatiecoëfficiënt (Kendall, 1938). Naarmate de coëfficiënt dichterbij 1 ligt is het verschil klein, naarmate de coëfficiënt dichterbij nul ligt of zelfs negatief wordt kunnen we zeggen dat het verschil groot is.

Bij iedere clusterstap groepeerde de computer op basis van de oorspronkelijke individuele rangschikkingen. Dus ook wanneer één of meer van de geclusterde rangschikkingen zelf al een compromisrangschikking betreft voor een cluster van expert rangschikkingen, bekijkt de computer het geheel van individuele rangschikkingen dat binnen de nieuw te vormen cluster zal worden opgenomen.

De clustering in grotere groepen zegt op zich nog niks over de inhoud van de clustering: tussen welke verschillen wordt een compromis gezocht; om welk soort verschil gaat het (bvb. switchen twee cases tussen de plaatsen 5 en 6 in een rangschikking of switchen twee cases tussen de plaatsen 1 en 6). Het is dus altijd van belang te kijken tussen welke rangschikkingen welk compromis gegenereerd wordt.

4. Evaluatie

Algemene evaluatie

Van het beter structureren van afwegingen en keuzen van juryleden en/of besluitnemers, heeft de aanpak in de praktijk stap voor stap een ontwikkeling doorgemaakt naar het beter structureren van inherente complexiteit en complexe informatie, en de inpassing ervan in het complexe groepsproces van de jury. Dit had tot gevolg dat verschillende, bij de opzet deels onvoorziene of onvoldoende uitgekristalliseerde, aandachtspunten die uit de analyse van de expertronde en uit de contacten met de juryleden naar voor kwamen binnen een krappe tijds marge ingepast moesten worden. Dit zette niet enkel de ontwikkeling en praktische toepassing van de methodologie onder druk in termen van complexiteit en werk/tijdsdruk. Dit zette vooral de communicatie van de methode en haar resultaten naar de gebruikers van die informatie onder druk. Hoewel deze analyse en informatie op zich goed aansloten bij de interesse van de juryleden, bleek de overdracht in een aantal opzichten te ingewikkeld van vorm om bevattelijk te kunnen zijn voor vragenlijsten en de groepsdiscussie. Hier zou meer toelichting in face to face contact tussen onderzoekers - juryleden en meer ontwikkeltijd voor een gesynthetiseerde presentatie van informatie een oplossing kunnen zijn. Waar de methode duidelijk wel in geslaagd is, is in het omarmen van de complexiteit van het vraagstuk (prioriseren van ongelijksoortige complexe cases op basis van complexe informatie, ongelijksoortige criteria en een diversiteit aan beoordelaars). Ook bood het een structuur voor het synthetiseren van verscheidenheid aan opinies door middel van clustering. De methode heeft duidelijk winst geboekt door de toepassing in de praktijk die een meer routineuze en snellere werkwijze mogelijk maakt, maar heeft nog winst te boeken in het vinden van vertaling naar de praktijk van de gebruikers van die informatie.

Specifieke reflecties

We bespreken tenslotte een paar meer specifieke reflecties.

Complexiteit gebruik multicriteria analyse

We hebben een aantal hypothesen ontwikkeld die mogelijk mede verklaren waarom het gebruik maken van de multicriteria analyse (MCA) voor een deel complex overkomt bij gebruikers:

- Ook al begrijpt men tot op zeker hoogte het concept van MCA en ook al staat men erachter, dan nog kan men er moeite mee hebben in het daadwerkelijke gebruik. Mogelijk kan men er zich onvoldoende een voorstelling van maken of kan men onvoldoende aanvoelen hoe de rekenwijze en de behandeling van hun input leidt tot bepaalde uitkomsten. Men blijft wellicht verveeld zitten met een blinde vlek alsook met de idee zich over te geven aan een machine; duidelijker proberen te maken hoe de MCA werkt lijkt hier op zijn plaats, als ook nadere duiding dat het model eigenlijk niks anders doet dan wat men zelf wil, zij het in een meer complexe behandeling van de complexe informatie dan waartoe het menselijke brein zelf in staat is. Dit zal ook gewenning vragen van de kant van de gebruikers. In eerste instantie bestaat misschien de neiging zich beter vertrouwd te voelen bij een (reken-)methode die 'navoelbaar' is, die men zelf als het ware uit het hoofd kan narekenen (het werken met rekenkundige gemiddelden van kwalitatieve data), ook al heeft een dergelijke simplificatie het nadeel dat aan kwaliteit van de analyse danig ingeboet wordt door wiskundige beperkingen in het werken met kwalitatieve gegevens.
- Het maken van een intuïtieve inschatting is binnen de huidige praktijk misschien gebruikelijker en daarom op dit moment wellicht makkelijker te aanvaarden dan het werken met een omvattende, expliciterende en structurende methode. Het expliciteren (argumenten, afwegingen, keuzen) kan als moeilijk ervaren worden omdat men het niet gewoon is en omdat men zich misschien gedwongen voelt duidelijke keuzen (op onderdelen) te maken waar een meer intuïtieve benadering de argumentaties en achterliggende keuzen van een standpunt misschien minder ter discussie stelt.
- In het kader van een groepsproces kan men zich als het ware verbergen achter de groep, in een gestructureerde bevraging waarin bovendien om duidelijke uitspraken gevraagd wordt, niet. Deze gestructureerde en expliciete werkwijze die onderdeel is van de MCA-methode, kan daarom misschien als bedreigend ervaren worden.

Vraagstellingen

We stellen vast dat de vraagstelling in de expertronde tot bepaalde analytische problemen aanleiding gaf:

- Omdat het werken in de expertconsultatie met een combinatie van specifieke inschattingsvragen bij de subcriteria en een algemene rangschikkingvraag over het geheel aan het einde van de vragenlijsten, de analyse en interpretatie voor problemen stelde (welke op basis van die verschillende vraagstellingen gegenereerde compromisrangschikkingen zijn het meest zinvol?) moeten we naar de toekomst een keuze maken. Wellicht dat de inschattingen op het niveau van de subcriteria hiervoor het meest relevant zijn, omdat zo het dichtst bij de voorliggende achtergrondinformatie vanuit de desk research gebleven wordt.
- Wat de vraagstelling van de vragen op subcriteria niveau betreft, is meer eenduidigheid wenselijk. Nu hebben we zowel vragen naar rangschikkingen als naar inschattingen op antwoordschalen gebruikt. Beide hebben hun eigen voor- en nadelen. In geval van rangschikkingvragen weet men bvb. niet welke kwalificatie op caseniveau gegeven wordt. Hierdoor kan spreiding van inschattingen moeilijker geduid worden. Hiermee verdwijnt mogelijk ook het zicht op onzekerheid van expertinschattingen: een vraagteken kan bvb. onbenoemd blijven en bewust een hoge of lage positie toegekend krijgen. In geval van inschattingvragen weet men echter niet of bvb. dezelfde volgorde gegeven zou worden als bij een rangschikkingvraag. We moeten wellicht een meer eenduidige keuze maken of naar een combinatie zoeken.
- Niet alle analytische problemen konden voorzien worden. We waren bij de start op zich redelijk tevreden met de vraagvormen, maar hadden nog geen vermoeden van implicaties voor de analyse, zoals hierboven beschreven. Bovendien zal zelfs indien men meer ervaring of inzicht heeft, er deels in imperfectie gekozen moeten worden: voor de meeste opties die voorliggen, valt wat te zeggen, zoals bvb. voor vraagvorm in rangschikking of in kwalificatie (inschattingsvragen). Een keuze voor de één levert nadelen op die door de andere keuze gecompenseerd kunnen worden, maar beide samen kiezen gaat niet altijd omdat dat teveel werk vraagt van respondenten. Bovendien stelt zich indien men beide samen neemt het probleem dat bij de interpretatie toch nog voor één van beide gekozen moet worden.

Gevoeligheidsanalyse

De gevoeligheid van het MCA-model voor de inhoudelijke elementen hebben we slechts in zeer beperkte mate kunnen bekijken in dit stadium en binnen de tijdsraming van het project. Een paar reflecties:

- Gevoeligheid op niveau van drie hoofdcriteria blijkt erg beperkt; als één van de drie de andere twee samen overheerst in gewicht, tellen de andere twee niet meer mee, en maakt ook de maat van overgewicht geen verschil. De vraag is of dit te verhelpen is en wat het nut daarvan kan zijn? Het werken met een beperkte set van hoofdcriteria kan ook als voordeel hebben dat het makkelijker te volgen is voor betrokkenen. Op het niveau van de subcriteria hebben we de gevoeligheid (nog) niet bekeken. Hetzelfde geldt voor de invloed van individuele experten: welke invloed gaat er bvb. uit van extreme expertinschattingen?
- Informatieknoppen inzake de analytische elementen (kennisonzekerheid, spreiding inschattingen, ...) blijken effect te hebben; de exacte mate van gevoeligheid (groot of klein) hebben we echter nog niet getest wegens tijdgebrek.

Interdisciplinaire samenwerking

Interdisciplinaire methodologische samenwerking tussen onderzoekers van sociaalwetenschappelijke en modelmatige (wiskundige) achtergrond bij dit besluitvormingsondersteunende beoordelingsproces bleek een duidelijke ontwikkeling door te maken gedurende de aanloop naar het project en tijdens het project zelf. In de kern kan gesteld worden dat de verschillende invalshoeken in de complexe praktijk naar elkaar toegegroeid zijn en er sprake is van een co-constructie van de toepassing in de praktijk van de methodiek, die de samenwerking daadwerkelijk een interdisciplinair karakter heeft gegeven. Om het interdisciplinaire karakter uitgebreid te duiden voert hier te ver. We geven hier enkel een aanzet. Belangrijkste verschil van invalshoek was bij aanvang wellicht dat de modelmatige onderzoeker meer accent legde op de functionaliteit van modelmatige ondersteuning van het leerproces van de gebruikers/deelnemers. Hierbij kunnen we spreken van een model dat in essentie belerend en objectiverend van karakter is: het denkvermogen van de betrokkenen wordt als het ware overstegen door de analytische kracht van het model die de 'betere' uitkomst genereert. De sociaal wetenschappelijke onderzoeker legde meer accent op de groepsdiscussie ondersteunende functionaliteit. Hierbij sturen de (opinions, inschatting van de) betrokkenen het MCA-model aan. In samenspraak heeft de (belangrijkste) functionaliteit zich in de praktijk meer en meer ontwikkeld naar het structureren, toegankelijk en interactief maken van complexe problematiek, complexe informatie en verschillende actor perspectieven.

Kwalitatief onderzoek

In algemene zin kan ervan uitgegaan worden dat er makkelijk kritiek te geven is op dit type kwalitatief onderzoek omdat het nooit perfect gedaan kan worden. Tegelijk moeten er wel keuzen gemaakt worden, deels in imperfectie en onwetendheid. Hierbij spelen ook factoren als tijdsdruk een rol. Deels vaart men hierbij op een (goed geïnformeerde) intuïtie, op basis waarvan zo verstandig mogelijk keuzen gemaakt worden, met goede argumenten. Het openstaan voor en vragen van feedback als ook het zo transparant mogelijk te zijn over keuzen, argumenten en werkwijze zijn elementen die de kwaliteit bevorderen. Bovendien is een zelfkritische houding gedurende het hele proces en nadien van belang.

5. Referenties

- Aerts D. (2005), Actieplan Vlaams Humaan Biomonitoringprogramma wordt opgestart, <http://www.milieu-en-gezondheid.be/nieuwsbrief/archief/biomonitor%207/faseplan.pdf>, in: *Biomonitor* 7
- Aminal Cel Milieu en Gezondheid (2005), Actieplan Vlaams Humaan Biomonitoringprogramma: stappenplan voor actie bij het vrijkomen van de meetresultaten, <http://www.milieu-en-gezondheid.be/nieuwsbrief/archief/biomonitor%205/fasenplan.pdf>, in: *Biomonitor* 5
- Belton V. and Stewart T. (2002), *Multiple Criteria Decision Analysis: An Integrated Approach*, Kluwer Academic, Dordrecht
- De Keyser W. and Springael J. (2002): *Another way of looking at group decision making opens new perspectives*, Universiteit Antwerpen, D/2002/1169/015.
- Dodgson J., Spackman M., Pearman A. and Phillips L. (2000), *Multi-criteria Analysis Manual*, Department for Environment, Transport and the Regions, United Kingdom
- Hwang C. and Lin M. (1987): "Group Decision-Making under multiple criteria", *Lecture notes in Economics and Mathematical Systems* 281, Springer-Verlag
- Kagan D. M. (1988): "Measurements of divergent and complex thinking.", *Educational and Psychological Measurement*, 48, 873-884
- Kendall, M. 1938. A new measure of rank correlation. *Biometrika* 30, 91–93.
- Kendall M and Gibbons J.D. (1990), *Rank Correlation Coefficients*, 5th ed., Oxford University Press
- Keune H. en Goorden L. (2004), *Praktijkcyclus Beleidsscenario's, voor het Steunpunt Milieu & Gezondheid*, Steunpunt Milieu & Gezondheid.
- Keune H. en Goorden L. (2005), *Multi criteria analyse beleidsscenario's biomonitoring*, Steunpunt Milieu & Gezondheid.
- Keune H, Morrens B, Springael J (2007a), *Verslag Expert Ronde fase 1, Priorisering cases voor beleid, Fasenplan Biomonitoring*, Steunpunt Milieu & Gezondheid
- Keune H, Morrens B, Springael J (2007b), *Verslag Jury fase 1, Priorisering cases voor beleid, Fasenplan Biomonitoring*, Steunpunt Milieu & Gezondheid
- Koppen G., Keune H., Casteleyn L. (2005), *Faseplan voor actie Biomonitoringsresultaten*, Steunpunt Milieu & Gezondheid, België
- Leik R. K. (1966), A Measure of Ordinal Consensus, in: *The Pacific Sociological Review*, Vol. 9, No. 2 (Autumn, 1966), pp. 85-90
- Miller, GA (1956): 'The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information', *The Psychological Review*, Vol.101, No. 2, 343-352
- Morrens B, Colles A, Keune H, Koppen G, Loots I (2007), *Verslag Desk Research fase1, Priorisering cases voor beleid, Fasenplan Biomonitoring*, Steunpunt Milieu & Gezondheid
- Roy B. (1996), *Multicriteria methodology for decision aiding*, Kluwer, Dordrecht.
- Thijssen P. (2006), *Van beschrijving naar inzicht - Inleiding in de statistiek*, Acco, Leuven.