

University of Groningen

The human dimension in the assessment of medicines

Beyer, Andrea Rose-Marie

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version

Publisher's PDF, also known as Version of record

Publication date:
2015

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Beyer, A. R.-M. (2015). *The human dimension in the assessment of medicines: Perception, preferences, and decision making in the European regulatory environment*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



SUMMARY AND SAMENVATTING

6

SUMMARY

.

1

For over three decades, the research on risk perception has supported a theory that experts have a one-dimensional view of risk, that is, they focus on the probability and the magnitude of a hazardous occurrence, which when combined is reduced to 'expected loss'. Only when outside their area of expertise does subjectivity impact their judgments. The installation and maintenance of a paternalistic view on the part of scientific experts is further supported by an accepted theory that patients, due to their lack of expert knowledge, only view risk subjectively and therefore must be managed via what has become known as a *knowledge-deficit* model. The hypothesis of the current study is that assessors in the regulatory environment, like patients and indeed like all non-experts, are not one-dimensional but multidimensional in their view of risk, as it is impossible to remove the 'human' aspect of our interaction with risk.

It is the human aspect of risk regulation for experts and non-experts that was examined in this thesis using methods from the fields of behavioural economics and decision theory. The overarching research questions are whether and how decision making for risk regulation, as conducted by experts within the European pharmaceutical regulatory network (a decentralized system for regulation and consensus building for medicinal products and devices across 31 European countries) is influenced by known behavioural and cognitive limitations. In addition, is there an existing method(s) that responds to the challenge of moving public engagement on risk assessment out of the procedural realm (e.g. limited engagement of a small number of representatives working within the current evaluative structure) and into the political realm (i.e. interaction with the public that is reflective of their concerns and their humanity).

In order to test the above hypotheses, two studies were conducted. First, 80 assessors from nine National Competent Authorities (NCAs) in Europe with expertise in the therapeutic areas of cardiology, oncology and central nervous system disorders were invited to participate in a research study. The study was implemented as a web-based questionnaire and was launched between June 2010 and October 2010. Three data collection instruments were used: a questionnaire on the risk perception of 28 types of medicinal products, a rating of several benefit–risk dimensions using data from mock 'clinical dossiers' in the therapeutic areas stated above and a questionnaire on general risk attitudes and risk perceptions (the Domain-Specific Risk Taking or DOSPERT scale). Second, data were collected from a sample of multiple sclerosis (MS) patients and EU regulators who assess medicinal products for MS using a qualitative questionnaire designed to interact with the M-MACBETH software.

In **Chapter 2.1**, assessors were asked to evaluate a list of 28 types of medicinal products using four perception scales: benefit, risk, seriousness of harm to those exposed and the knowledge of potential harm for those exposed. Oncology products scored the highest on

the 'risk perception' scale and on the 'seriousness of harm to patients' scale, while insulin, vaccines and antibiotics had the highest mean scores on the 'benefit' scale. Assessors gave the lowest score for insulin on the 'knowledge of harm' scale, followed by oncology and AIDS medications. Female assessors saw more benefit for almost all the products on the list; the junior assessors (1–3 years of experience) provided statistically different scores on three of the four scales measured but only for a few products. Compared to efficacy assessors, safety assessors reported higher risk scores for almost all the products; there was mixed differences by professional qualification (MD, PhD, pharmacist). The interpretation of the results from all four scales is that, in keeping with their role as gatekeepers of medical products, assessors view these products as predominantly beneficial with mid to low risks when used appropriately but with the potential for serious harm when misused, and they believe that patients are mostly aware of the potential for harm of medicinal products. The results highlight a methodological issue common to risk perception research. The use of broadly defined hazards, such as cholesterol products or biotechnology products, does not in our opinion provide sufficient information for experts to make a real assessment. However, it is evident that there are some differences in risk perception by gender, professional qualification (MD, PhD, other) and years of regulatory experience. While these differences were not seen across all 28 products, it does indicate the need for further exploration of the influence of individual characteristics within this group.

Also in **Chapter 2.1**, assessors were further asked to review a mock 'clinical dossier' for three medicinal products, depending on their area of therapeutic and clinical expertise, and to provide responses regarding their perceptions of the risk associated with the product (risk dimension) and seven other dimensions, all measured on a seven-point scale. The results from a combined principal component analysis revealed two latent components explaining 59% of the total variance. The subsequent regression model, explaining 54% of the variability among the assessors, showed that the assessors' ratings on the risk dimension scale is predicted by perceived worry regarding safety, the number of people exposed to the risk and the ethical issues associated with the drug. These dimensions were inversely correlated to benefit and risk acceptability. The precision of the scientific knowledge or unfamiliarity with the potential risk did not predict the assessors' risk dimension responses. In addition, the observed relationship was mediated by gender, the medicinal product under review and number of years in a regulatory role. Traditionally, assessors are believed to focus only on the probability of the risk and the magnitude of the event when reviewing a potential hazard. However, in a regression analysis predicting the risk dimension scores only the seriousness of harm component was a significant predictor of individual risk perception. This is an important finding given that ideally, in their role as regulators, the precision of the scientific knowledge or

the lack thereof and the attending uncertainties would be expected to be very relevant to how a drug is perceived. These results show, as found in a study among nuclear scientists, that risk perception among experts is more complex than previously believed.

In **Chapter 2.2**, the results from the DOSPERT scale showed that assessors do not have a consistent risk attitude (risk-seeking, risk-neutral, risk-averse) across the five life domains measured (social, financial, health/safety, recreational and ethical). Depending on the context, (such as social or financial), assessors changed their appetite for risk-taking and their perceptions of associated risks. However, the results do show a relationship between risk attitude and risk perception in that assessors have a weak but statistically significant perceived risk-averse attitude in four of the five domains, that is, the more risky an activity was perceived to be by the assessors, the less likely they were to engage in it. The lack of a very strong correlation indicates that risk perception of an activity is not the only determinant of whether the assessor would engage in such an activity; however, it does give some insight into what we now believe to be a multidimensional mental map of risk among assessors. It is not only risk attitudes that may impact the perception of the benefit or risk associated with a drug but also personality traits. It was shown that an increase in the conscientiousness score from the Big Five Inventory scale predicted an increase in the perception of a drug's benefit, while extraverted assessors saw fewer risks. Gender is also a predictive factor, as male assessors gave higher scores for drugs' benefit ratings than female assessors, while risk-averse assessors, as defined by the DOSPERT scale (perceived neutral-averse or averse-risk profile) saw greater risks. Individual traits, such as personality, risk attitudes and gender, influence risk perceptions and consequently may impact decision making, even among expert groups within regulatory organisations. However, this subjective component of the decision process is rarely acknowledged.

Work Package 1 of the European Medicines Agency Benefit–Risk (EMA BR) Methodology Project reported that assessors do not have a clear definition of what benefits are and what risks are in relation to medical products. However, it is possible that the methodology used to collect the data from the assessors was not sufficiently concise, as it did not require the assessors to focus on a specific drug or even a specific disease. In **Chapter 2.3**, we report the degree of congruency among the medical assessors in identifying benefits and risks for specific medical products. In the third phase of the study, assessors who had reviewed the mock 'clinical dossier' specific to their area of expertise were asked to identify the risks and benefits for that product. They were also asked to indicate the value they place on various sections of the clinical dossier by distributing 100 points within the sections—pharmacokinetic/pharmacodynamic, toxicology, efficacy, safety and study quality. For both the oncology treatment and the CNS treatment, there appeared to be a low level of agreement between the assessors as to the benefits and the risks. For the cardiology

treatment, there was more consistency in the identified benefits and risks. Estimations of value also varied among assessors, although there appeared to be consistently lower values assigned to pharmacokinetic/pharmacodynamic attributes. Assessors appear to focus on different aspects of a clinical dossier and assign differing levels of value, as can be seen by the number of points they apportioned to the various sections of the clinical dossier. The results of this study appear to confirm what was reported in Work Package 1 of the EMA BR Project—that assessors are not congruent in identifying the benefits and risks of medical products or the degree of value they place on the data provided in clinical dossiers.

In its role as the central agency coordinating the activities of 31 NCAs, the EMA provided a unique opportunity via the BR Methodology Project to examine the processes currently in use for judging the benefit–risk balance of medicinal products. The Problem, Objectives, Alternatives, Consequences and Trade-offs (PrOACT-URL) framework is the qualitative framework shown to be most comprehensive and theoretically able to encompass decisions dominated by conflicting objectives. PrOACT provides a generic problem structure, which is adaptable to benefit–risk decision making by regulators, and the ‘-URL’ encompasses the uncertainty, the risk tolerance of the decision makers and linkage to other decisions. **Chapter 3.1** describes the EMA BR Methodology Project at the time of publication, provides an assessment of the use of the PrOACT-URL framework and multi-criteria decision modelling among assessors in five EU agencies. The case is presented that there is now compelling evidence demonstrated in workshops and seminars—for example, Mussen and others who proposed a framework and methodology based on multi-criteria decision analysis, Felli and others who have provided an example with Lilly’s BRAM model, recent work by the BRAT group in developing a framework for selecting, organizing, summarizing and interpreting data—that may help make the assessment process more explicit and less intuitive. The contribution of the EMA BR Project to the above-mentioned initiatives points in a similar direction, that is, that quantitative benefit–risk modelling is possible. The paper concludes that adoption of such methods within the current regulatory environment requires effective technology transfer to the regulatory authorities.

In the previous chapters, the longstanding myth that experts are veridical was challenged and shown not to be supported. However, only accepting that there are human factors that impact the perception of risk for both experts and non-experts does not respond to the challenge of how to move beyond the *knowledge-deficit* model to the desired *dialogue* model between regulators of medicines and patients. Regulatory agencies have in recent years established means for interacting with patients and patient groups, including inviting patients or patient

representatives to be members of regulatory decision-making committees. Despite this evolution, patient groups continue to voice the need for even greater transparency of and involvement in health policy decisions and in some cases challenge the decisions made by regulators to withhold or withdraw market authorizations.

In **Chapter 4.1**, we report the results of a web-based field experiment to apply the Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH) to eliciting preferences for outcomes that are possible with several MS treatments. Sixty-two MS patients were enrolled and preferences were elicited for six treatment outcomes: three efficacy outcomes (number of relapses, time to disease progression, disease progression) and three safety outcomes (risk for progressive multifocal leukoencephalopathy, liver failure and leukemia). Outcome value-function profiles were derived from individual judgments elicited from the patients, and logistic regression models were used to evaluate the relationship between the shapes of the value functions and disease and demographic characteristics. Weighting judgments were also elicited from the patients. The qualitative elicitation procedure used by MACBETH allowed patients to provide consistent judgments, which are clinically and psychologically relevant for decision making, with many patients providing positive comments on the ease of using and responding to the questionnaire. The majority of value functions were non-linear, meaning that patients' preferences are not proportional to outcome levels; additionally patients do not equally weigh the MS outcomes. A decision analytic technique like MACBETH that utilizes a qualitative preference elicitation procedure can be easily integrated into a pre- or post-authorization setting, and the results can easily be included in the regulatory approval process. Regulatory authorities and the industry should consider the application of this non-numerically based method to increase their understanding of patient preferences and to use these data as supportive inputs in their decision-making processes.

On completion of the proof of concept study demonstrating how MACBETH can be used to elicit preferences, the next objective reported in **Chapter 4.2** is the evaluation of whether the resulting decisions using the MACBETH decision models are consistent with what is seemingly logical for several MS treatments; that is, do treatments with similar efficacy and safety profiles obtain similar decision scores and do treatments that are dissimilar in either efficacy or safety profiles obtain dissimilar decision scores? Another objective is to determine if the decisions of the regulators are different from those of the patients in this study population. Value judgments elicited from the sample of MS patients and a sample of medical assessors for the three favourable effects and three unfavourable effects previously mentioned were used to create treatment decision models using data from the pivotal MS trials PRISMS4, EVIDENCE and AFFIRM. Decision models comparing Rebif[®] 44 mg to Rebif[®] 22 mg and a placebo indicated that

patients would choose Rebif[®] 44 mg over other treatments. Similarly, models comparing Rebif[®] 44 mg to Avonex[®] 30 mg and a placebo showed that patients would again choose Rebif[®] 44 mg over the other alternatives. It should be noted that the scores obtained for Rebif[®] 44 mg were very close to those obtained for Avonex[®] 30 mg, indicating similarity in the performance of both treatments. A model comparing Tysabri and Rebif[®] 44 mg and a placebo, with no risk of PML, showed 100% of patients choosing Tysabri versus Rebif[®] 44 mg or the placebo. However, when risk for PML was added, only 80% of patients would now choose Tysabri, while 20% would choose Rebif[®] 44 mg. The MACBETH approach can be easily used to capture the values of patients and regulators and to distinguish the point at which increased levels of risk may separate patients' values from those of regulators or other stakeholders and may give some indication as to the proportion of the groups likely to switch treatment. A comparison of decision models between regulators and patients showed similar treatment choices if no risk data are included; however, when safety risks are added, regulators were more likely than patients to switch to the treatment with lower risk.

The aspiration of reflective and consequently democratic institutions of science as outlined by Beck, Wynne and others has not been realized. After two decades of advocating the need for such a model to govern interactions between scientists and laypeople, social scientists are beginning to wonder if they have overpromised on what could be delivered. This thesis is based on the premise that such a model is still achievable if misconceptions of expert infallibility and non-expert incomprehensibility are debunked and new methodologies for supporting such interactions are allowed to flourish within our scientific institutions.

6.2

SAMENVATTING

Al meer dan drie decennia heeft het onderzoek met betrekking tot risicoperceptie een theorie ondersteund dat experts een eendimensionale visie hebben op risico's. Zij richten zich op de waarschijnlijkheid en de omvang van een gevaarlijke gebeurtenis. Gecombineerd is dat dan het 'verwachte verlies'. Alleen buiten hun kennisgebied heeft subjectiviteit invloed op hun beoordelingen. De gevestigdheid van een paternalistische visie op het deel van wetenschappelijke experts wordt verder ondersteund door een geaccepteerde theorie dat patiënten, vanwege hun gebrek aan kennis, risico's alleen subjectief bekijken en daarom beheerd moeten worden via iets dat nu bekend staat als een *kennisgebrek*-model. De hypothese van dit onderzoek is dat beoordelaars in een regulerende omgeving, zoals patiënten en alle niet-experts, niet eendimensionaal, maar multidimensionaal zijn in hun visie op risico. Het is immers onmogelijk om het 'menselijke' aspect te verwijderen uit onze interactie met risico.

Het menselijke aspect van risicoregulatie voor experts en niet-experts is in deze scriptie onderzocht is, gebruikmakend van methoden uit gedragseconomie en beslissingstheorie. De overkoepelende onderzoeksvragen zijn of en hoe besluitvorming voor risicoregulatie, zoals uitgevoerd door experts binnen het Europese farmaceutische regelgevende netwerk (een gedecentraliseerd systeem voor regulatie en opbouw van consensus voor medicinale producten en apparaten in 31 Europese landen) wordt beïnvloed door bekende gedrags- en cognitieve beperkingen. Daarnaast is er een bestaande methode(n) die reageert op de uitdaging van de publieke betrokkenheid bij risicobeoordeling verplaatsen uit de procedurele sfeer (b.v. beperkte betrokkenheid van een klein aantal vertegenwoordigers dat actief is binnen de huidige evaluerende structuur) naar de politieke sfeer (b.v. interactie met het publiek dat hun zorgen en hun menselijkheid weergeeft).

Om de bovenstaande hypothesen te testen werden twee onderzoeken uitgevoerd. Eerst werden 80 beoordelaars uit negen Nationale Competente Autoriteiten (NCA's) binnen Europa, met expertise in de therapeutische gebieden van cardiologie, oncologie en stoornissen in het centrale zenuwstelsel uitgenodigd om deel te nemen aan het onderzoek. Het onderzoek werd geïmplementeerd als een online vragenlijst en werd tussen juni en oktober 2010 gelanceerd. Er werden drie instrumenten voor dataverzameling gebruikt: een vragenlijst over de risicoperceptie van 28 soorten medicinale producten, een waardering van verschillende voordeel vs. risico-dimensies, waarbij data gebruikt werd uit nep 'klinische dossiers' in de therapeutische gebieden die hierboven genoemd werden en een vragenlijst inzake algemene risicohoudingen en risicopercepties (de DOSPERT-schaal, Domain-Specific Risk Taking). Ten tweede werd data verzameld uit een selectie multiple sclerosis (MS) patiënten en EU-regulators die medicinale producten beoordelen voor MS, waarbij een kwalitatieve vragenlijst gebruikt werd die ontworpen is om te interacteren met de M-MACBETH-software.

In **Hoofdstuk 2.1** wordt beoordelaars gevraagd om een lijst met 28 soorten medicinale producten te evalueren, waarbij ze gebruik moeten maken van vier perceptieschalen: voordeel, risico, ernst van de schade aan diegenen die er bloot aan worden gesteld en de kennis van het potentiële gevaar voor hen die eraan blootgesteld worden. Oncologie-producten scoorden het hoogst op 'risicoperceptie' en 'ernst van schade aan patiënten', terwijl insuline, vaccins en antibiotica de hoogste gemiddelde scores hadden op 'voordeel'. Beoordelaars gaven de laagste score voor insuline op de 'kennis van schade'-schaal, gevolgd door medicatie voor oncologie en AIDS. Vrouwelijke beoordelaars zagen meer voordelen voor praktisch alle producten op de lijst: de junior beoordelaars (1-3 jaar ervaring) geven statistisch gezien andere scores op drie van de vier schalen die gemeten werden, maar alleen voor een aantal producten. In vergelijking met de efficiëntie-beoordelaars gaven de veiligheidsbeoordelaars hogere risicoscores voor bijna alle producten. Er waren gemengde verschillen per professionele kwalificatie (MD, PhD, apotheker). De interpretatie van de resultaten uit alle vier de schalen is dat, wat volgt uit hun rol als poortwachter van medicinale producten, beoordelaars deze producten als voornamelijk gunstig zien met gemiddelde tot lage risico's als ze op de juiste manier gebruikt worden, maar met de potentie voor ernstige schade bij misbruik. Ook geloven ze dat patiënten grotendeels op de hoogte zijn van het potentiële gevaar van medicinale producten. De resultaten markeren een methodologisch probleem dat hoort bij onderzoek naar risicoperceptie. Het gebruik van breed gedefinieerde gevaren, zoals cholesterolproducten of biotechnologische producten, biedt naar onze mening niet voldoende informatie voor experts om een echte beoordeling te maken. Het is echter duidelijk dat er bepaalde verschillen zijn in risicoperceptie, op geslacht, professionele kwalificatie (MD, PhD, overig) en jaren regelgevende ervaring. Hoewel deze verschillen niet gezien konden worden in alle 28 producten, laat het zien dat er behoefte is aan meer onderzoek naar de invloed van individuele kenmerken binnen deze groep.

In **Hoofdstuk 2.1** worden beoordelaars ook gevraagd om een nep 'klinisch dossier' te bekijken voor drie medicinale producten, afhankelijk van hun therapeutische gebied en klinische expertise, en om reacties te bieden met betrekking tot hun risicopercepties horend bij het product (risicodimensie) en zeven andere dimensies, allen gemeten op een zevenpuntsschaal. De resultaten van een gecombineerde analyse van het hoofdonderdeel liet twee latente onderdelen zien, die samen goed waren voor 59% van de totale variantie. Het daaropvolgende regressiemodel, dat 54% van de variabelen tussen de beoordelaars verklaarde, toonde dat de waarderingen van de beoordelaars op de risicodimensieschaal is voorspeld door vermeende zorgen met betrekking tot veiligheid, het aantal mensen dat wordt blootgesteld aan het risico en de ethische problemen die samengaan met het medicijn. Deze dimensie werden verbonden aan de aanvaardbaarheid van voordeel en risico. De precisie van wetenschappelijke kennis of onbekendheid met het potentiële

risico voorspelde de risicodimensiereacties van de beoordelaars niet. Daarnaast werd de waargenomen relatie gemiddeld op geslacht, het medicinale product dat geëvalueerd werd en het aantal jaren in een regelgevende rol. Traditioneel gezien wordt geloofd dat beoordelaars zich alleen richten op de waarschijnlijkheid van het risico en de omvang van de gebeurtenis tijdens het evalueren van een potentieel gevaar. In een regressieanalyse voor het voorspellen van de risicodimensiescores is echter alleen de ernst van de schade een significante voorspeller van individuele risicoperceptie. Dit een belangrijk gegeven, aangezien ideaal gezien, in hun rol als regulator, van de precisie van de wetenschappelijke kennis of het gebrek daarvan en de aanwezige onzekerheden wordt verwacht dat ze zeer relevant zijn in hoe een medicijn waargenomen wordt. Deze resultaten laten zien, zoals ook te zien was in een onderzoek onder kernwetenschappers, dat de risicoperceptie onder experts complexer is dan eerder werd gedacht.

In **Hoofdstuk 2.2** lieten de resultaten van de DOSPERT-schaal zien dat beoordelaars geen consistente risicohouding hebben (risico-zoekend, risiconeutraal, risico-afkerig) over de vijf levensdomeinen die gemeten werden (sociaal, financieel, gezondheid/veiligheid, recreatief en ethisch). Afhankelijk van de context (zoals sociaal of financieel) veranderde bij beoordelaars hun lust om risico te nemen en hun perceptie van bijbehorende risico's. De resultaten tonen echter een relatie tussen risicohouding en risicoperceptie in dat beoordelaars een zwak maar statistisch significant gepercipieerde risico-afkerige houding hebben in vier van de vijf domeinen. Dat wil zeggen dat des te riskanter een activiteit wordt gezien door de beoordelaar, des te minder groot de kans dat ze erin mee zouden doen. Het gebrek aan een zeer sterke correlatie doet denken dat risicoperceptie van een activiteit is niet de enige bepalende factor in of een beoordelaar met zo'n activiteit mee zou doen. Het geeft echter wel bepaald inzicht in wat we nu zien als een multidimensionale mentale kaart van risico onder beoordelaars. Het zijn niet alleen risicohoudingen die invloed hebben op de perceptie van voordeel of risico horend bij een medicijn, maar ook persoonlijkheidskenmerken. Duidelijk was dat een toename in plichtsgetrouwheidsscore van de Big Five Inventory-schaal een voorspellende factor was voor een toename in de perceptie van het voordeel van een drug. Extraverte beoordelaars zagen minder risico's. Geslacht is ook een voorspellende factor, aangezien mannelijke beoordelaars hogere scores gaven voor de voordelen van medicijnen dan vrouwelijke beoordelaars, terwijl risico-afkerige beoordelaars, zoals bepaald door de DOSPERT-schaal (waargenomen neutraal-afkerig of afkerig-risico profiel), grotere risico's zagen. Individuele kenmerken, zoals persoonlijkheid, risicohoudingen en geslacht, hebben invloed op risicopercepties en kunnen als gevolg daarvan impact hebben op besluitvorming, zelfs onder expertgroepen binnen regulerende organisaties. Dit subjectieve onderdeel van het besluitvormingsproces wordt echter zelden erkend.

Werkpakket 1 van het European Medicines Agency Benefit-Risk (EMA BR)-methodologieproject gaf aan dat beoordelaars geen helder begrip hebben van wat voordelen zijn en wat risico's zijn in relatie tot medische producten. Het is echter mogelijk dat de methodologie die gebruikt is om de data te verzamelen van beoordelaars niet voldoende beknopt was, aangezien de beoordelaars zich niet hoefden te richten op een specifiek medicijn of zelfs een specifieke ziekte. In **Hoofdstuk 2.3** melden we de mate van congruentie onder de medische beoordelaars in het identificeren van voordelen en risico's voor specifieke medische producten. In de derde fase van het onderzoek werd aan beoordelaars die het nep 'klinische dossier' voor hun vakgebied hadden bekeken gevraagd om de risico's en voordelen voor dat product te identificeren. Hen werd ook gevraagd om aan te geven hoeveel waarde ze hechtten aan de verschillende delen van het klinische dossier, door 100 punten te verdelen binnen de onderdelen - farmacokinetisch/farmacodynamisch, toxicologie, efficiëntie, veiligheid en onderzoekskwaliteit. Voor zowel de oncologiebehandeling als de CNS-behandeling leek er maar weinig overeenstemming te zijn tussen de beoordelaars voor wat betreft de voordelen en de risico's. Voor de cardiologiebehandeling was er meer consistentie in de geïdentificeerde voordelen en risico's. Schattingen van waarde verschilden ook onder de beoordelaars, hoewel er steeds lagere waarden werden toegewezen aan farmacokinetische/farmacodynamische kenmerken. Beoordelaars lijken zich te richten op verschillende aspecten van een klinisch dossier en wijzen verschillende niveaus van waarde toe, zoals gezien kan worden aan het aantal punten dat ze toewijzen aan verschillende delen van het klinische dossier. De resultaten van dit onderzoek lijken te bevestigen wat al gemeld werd in Werkpakket 1 van het EMA BR-project - dat beoordelaars niet congruent zijn in het identificeren van de voordelen en risico's van medische producten of de waarde die ze toekennen aan de data die in de klinische dossiers gegeven wordt.

In de rol als centraal bureau dat de activiteiten van 31 NCA's coördineert biedt de EMA een unieke kans via het BR Methodologieproject om de processen te beoordelen die op dit moment gebruikt worden voor het beoordelen van de voordeel-risicobalans van medicinale producten. Het Probleem, Doelen, Alternatieven, Gevolgen en Compromissen (PrOACT-URL)-kader is het kwalitatieve kader dat getoond wordt om zo uitgebreid en theoretisch mogelijk beslissingen te omvatten die gedomineerd worden door conflicterende doelen. PrOACT biedt een generieke probleemstructuur, die aan te passen is aan de besluitvorming omtrent voordeel-risico door regulators, en de '-URL' omvat de onzekerheid, de risicotolerantie van de besluitvormers en link met andere beslissingen. **Hoofdstuk 3.1** beschrijft het EMA BR-Methodologieproject op het moment van publicatie, biedt een beoordeling van het gebruik van het PrOACT-URL-kader en multi-criteria beslissingmodellen onder beoordelaars in vijf Europese bureaus. De zaak laat zien dat er nu overtuigend bewijs wordt getoond in workshops en seminars - bijvoorbeeld Mussen

en anderen, die een kader en methodologie voorstelden op basis van een multi-criteria besluitanalyse, Felli en anderen, die een voorbeeld hebben gegeven met Lilly's BRAM-model, recent werk door de BRAT-groep in het ontwikkelen van een kader voor het selecteren, organiseren, samenvatten en interpreteren van data - dat kan helpen om het beoordelingsproces explicieter en minder intuïtief te maken. De bijdrage van het EMA BR-project aan de hierboven genoemde initiatieven wijst in een gelijksoortige richting, namelijk dat kwantitatieve voordeel-risico-modellering mogelijk is. Het verslag concludeert dat aanname van dergelijke methodes binnen de huidige regelgevende sfeer vraagt om effectieve technologieoverdracht naar de regelgevende autoriteiten.

In eerdere hoofdstukken werd de aanhoudende mythe dat experts waarheidlievend waren uitgedaagd en bleek dat deze niet ondersteund werd. Alleen accepteren dat er menselijke factoren zijn die invloed hebben op de perceptie van risico voor zowel experts als niet-experts reageert echter niet op de uitdaging van hoe we verder kunnen gaan van het *kennisgebrek*-model naar het gewenste *dialog*-model tussen regulators van medicijnen en patiënten. Regelgevende autoriteiten hebben in de afgelopen jaren methoden vastgesteld om te interacteren met patiënten en patiëntengroepen, inclusief het uitnodigen van patiënten of patiëntvertegenwoordigers om deel uit te maken van commissies die regelgevende beslissingen nemen. Ondanks deze evolutie blijven patiëntengroepen aangeven dat ze behoefte hebben aan meer transparantie van en betrokkenheid bij beslissingen in gezondheidsbeleid en in sommige gevallen de beslissingen van regulatoren om marktautorisaties achter te houden of terug te trekken aanvechten.

In **Hoofdstuk 4.1** vermelden we de resultaten van een online veldexperiment om toe de Measuring Attractiveness by a Categorical Based Evaluation Technique (MACBETH, Techniek van meten van Aantrekkelijkheid op basis van een Categorie-evaluatie) om voorkeuren te ontlocken voor resultaten die mogelijk zijn met verschillende MS-behandelingen. Tweeënzestig MS-patiënten deden mee en voorkeuren werden gebruikt om zes behandelresultaten: drie effectiviteitsresultaten (aantal terugvallen, tijd tot progressie van de ziekte, progressie van de ziekte) en drie veiligheidsresultaten (risico op progressieve multifocale leukencefalopathie, leverfalen en leukemie) te ontlocken. Resultaatwaarde-functieprofielen werden ontleend aan individuele beoordelingen van patiënten en logistische regressiemodellen werden gebruikt om de relatie tussen de vormen van waardefuncties en ziekte- en demografische kenmerken te evalueren. Patiënten werd ook gevraagd naar wegingsbeoordelingen. De kwalitatieve ontlockingsprocedure gebruikt door MACBETH gaf patiënten de mogelijkheid om consistente beoordelingen te geven, die klinisch en psychologisch relevant zijn voor besluitvorming, waarbij veel patiënten positieve opmerkingen gaven met betrekking tot het gemak van het gebruik van en reageren op de vragenlijst. De meeste waardefuncties waren niet-lineair,

wat wil zeggen dat de voorkeuren van de patiënten niet in verhouding waren met de resultaatniveaus; daarnaast wogen patiënten de MS-resultaten niet op dezelfde manier. Een besluitanalysetechniek zoals MACBETH die gebruikmaakt van een kwalitatieve procedure voor het uitlokken van voorkeuren, kan eenvoudig worden geïntegreerd in een pre- of post-geautoriseerde setting en de resultaten kunnen eenvoudig worden opgenomen in het regelgevende goedkeuringsproces. Regelgevende autoriteiten en de branche dienen rekening te houden met de toepassing van deze niet-numerieke methode om meer begrip te krijgen van patiëntvoorkeuren en het gebruiken van deze data als ondersteunende input in de besluitvormingsprocessen.

Na voltooiing van het bewijs van conceptonderzoek dat laat zien hoe MACBETH gebruikt kan worden om voorkeuren te ontlocken, wordt het volgende doel vermeld in **Hoofdstuk 4.2**. Dat is de evaluatie van of de resulterende beslissingen met het gebruik van de MACBETH-beslissingsmodellen consistent zijn met wat logisch lijkt voor verschillende MS-behandelingen. Dat wil zeggen, krijgen behandelingen met gelijksoortige efficiëntie- en veiligheidsprofielen gelijksoortige beslissingsscores en krijgen behandelingen die ongelijk zijn in efficiëntie- veiligheidsprofielen even ongelijke beslissingsscores? Een ander doel is het bepalen of de beslissingen van de regulatoren verschillen van die van de patiënten in dit onderzoek. Waardebeoordelingen die ontlokt zijn uit de steekproef van MS-patiënten en een steekproef van medische beoordelaars voor de drie gunstige effecten en de drie ongunstige effecten die eerder genoemd werden, werden gebruikt om de modellen voor behandelbeslissingen te maken. Hierbij werd gebruik gemaakt van data uit de doorslaggevende MS-proeven PRISMS4, EVIDENCE en AFFIRM. Beslissingsmodellen vergelijken Rebif® 44 mg met Rebif® 22 mg en een placebo gaf aan dat patiënten Rebif® 44 mg zouden verkieszen boven andere behandelingen. Ook worden modellen gebruikt om Rebif® 44 mg te vergelijken met Avonex® 30 mg en een placebo liet zien dat patiënten weer Rebif® 44 mg zouden verkieszen boven andere alternatieven. Er moet worden opgemerkt dat de scores die verkregen zijn voor Rebif® 44 mg dichtbij die van Avonex® 30 mg lagen, wat de indruk werkt dat beide behandelingen op gelijke wijze presteren. Een model waarin Tysabri en Rebif® 44 mg en een placebo zonder risico op PML met elkaar vergeleken werden, liet zien dat 100% van de patiënten kozen voor Tysabri versus Rebif® 44 mg of het placebo. Toen het risico voor PML werd toegevoegd, koos nog maar 80% van de patiënten voor Tysabri, terwijl 20% zou kiezen voor Rebif® 44 mg. De MACBETH-benadering kan eenvoudig gebruikt worden om de waarden van patiënten en regelgevers vast te leggen en onderscheid te maken tussen het punt waarop toegenomen risico's onderscheid kunnen maken tussen de waarden van patiënten en die van regelgevers of andere belanghebbenden en enige indicatie kunnen geven met betrekking tot het aandeel van de groepen die waarschijnlijk de behandeling zouden wisselen. Een vergelijking van beslissingsmodellen tussen regelgevers en patiënten toonden gelijke

behandelkeuzes als er geen risicodata was inbegrepen. Als echter veiligheidsrisico's toegevoegd worden, dan is de kans groter dat regelgevers kiezen voor een behandeling met een lager risico dan dat patiënten dat doen.

De aspiratie van reflectieve en vervolgens democratische wetenschappelijke instituten zoals aangegeven door Beck, Wynne en anderen, is niet gerealiseerd. Na twee decennia van het promoten van de behoefte aan een model om interacties tussen wetenschappers en leken te leiden, beginnen sociale wetenschappers zich af te vragen of ze meer hebben beloofd dan ze kunnen leveren. Deze scriptie is gebaseerd op de stelling dat een dergelijk model nog steeds bereikbaar is, als misvattingen omtrent onfeilbaarheid van experts en onbegrijpelijkheid van niet-experts worden ontkracht en nieuwe methodes voor het ondersteunen van dergelijke interacties mogen opbloeien binnen onze wetenschappelijke instituten.

