ARBOKENNISPRODUCTEN

ONDER DE AGREE-LOEP

EEN ONDERZOEK NAAR DE BRUIKBAARHEID VAN EEN BEOORDELINGSINSTRUMENT VOOR ARBOKENNISPRODUCTEN

opdrachtgever	Coronel Instituut Prof.dr. F.J.H.van Dijk
opdrachtnemer	Kwaliteitsbureau NVAB
uitvoering *inhoudelijke begeleiding* *projectleiding*	Drs. J.A.H.J. Hoenen, Bureau Beter Dr. C.T.J.Hulshof, NVAB Dr. C. van Vliet, directeur NVAB
datum	September 2005
status	Eindrapport

VOORWOORD

Als kennis beschikbaar is over de effectiviteit van het handelen van arboprofessionals, is toepassing daarvan in de dagelijkse praktijk geen vanzelfsprekendheid. Voor het verspreiden van kennis worden diverse soorten kennisproducten gebruikt. Los van de inhoudelijke kwaliteit, opbouw, omvang of herkomst van die kennisproducten, is de systematische implementatie in veel arbodiensten en aangrenzende praktijken een probleem. Beschikbare kennis stroomt onvoldoende door, zowel binnen beroepsgroepen als tussen beroepsgroepen. In de arbowereld worden tal van wegen bewandeld om het kwaliteitspeil van de zorg te verbeteren en op peil te houden: scholing, richtlijnen, protocollen, werkvoorschriften, audits, intervisie, onderlinge toetsing, refereerbijeenkomsten en het gebruik van performance indicatoren. Elk van deze methodieken levert weliswaar een bijdrage aan het kwaliteitssysteem, maar welke weg of combinatie van wegen daarbij het beste bewandeld kan worden lijkt meer een zaak van persoonlijke voorkeur of geschiedenis dan van overwegingen van effectiviteit.

Het is voor veel arboprofessionals op dit moment niet duidelijk hoe bijvoorbeeld NVAB-Richtlijnen, Leidraden, STECR-Werkwijzers, interne werkvoorschriften of PAGO-Protocollen zich tot elkaar verhouden en welke waarde aan de verschillende aanbevelingen toegekend mag worden. Om de inhoudelijke waarde van deze documenten te kunnen vergelijken is het van belang dat criteria ontwikkeld en bediscussieerd worden om tot een gefundeerd kwaliteitsoordeel te kunnen komen. De NVAB vindt dit van belang en wil daaraan graag een bijdrage leveren.

In het kader van het 'Programma ter versterking van de kennis- onderzoeksinfrastructuur van arbodiensten' (KIS) van het Coronel Instituut (AMC) is derhalve door het Kwaliteitsbureau NVAB een onderzoek uitgevoerd ter bevordering van de implementatie van hoogwaardige informatie en kennis in de praktijk van arbodiensten. Dit rapport doet verslag van dit onderzoek. Het doel van het onderzoek was arboprofessionals te helpen bij het beoordelen van de waarde van de diverse richtlijnen en kennisproducten (samen 'arbokennisproducten' genoemd). Wat zijn belangrijke en beoordeelbare criteria waar een goed arbokennisproduct aan moet voldoen? Is het mogelijk om tot overeenstemming te komen over die criteria? Op welke manier kunnen wetenschappelijke kwaliteit en praktijkgerichtheid tegen elkaar afgewogen worden?
Voor deze en andere vragen is in dit onderzoek nagegaan of, en hoe arboprofessionals een aantal aspecten van kwaliteit kunnen beoordelen van arbokennisproducten waar zij in de praktijk mee te maken hebben. Er is onderzocht of kwaliteitsverschillen tussen diverse arbokennisproducten zinvol zichtbaar te zijn maken met behulp van het internationaal gevalideerde AGREE-instrument voor richtlijnen (Appraisal of Guidelines for Research and Evaluation).

Een aantal arbokennisproducten zijn beoordeeld door sleutelpersonen uit het arboveld. Hiermee is tegelijkertijd ook een indruk verkregen van een draagvlak voor criteria voor arbokennisproducten en is de bruikbaarheid van deze criteria getest. De sleutelpersonen waren afkomstig uit kringen van NVAB, Kennisinstituten en arbodiensten. Met medewerking van STECR hebben ook vele STECR stafleden en leden van STECR kenniskringen actief in het onderzoek geparticipeerd.

Het AGREE-instrument kan voor de beoordeling van de inhoudelijke waarde van arbokennisproducten een markeringspunt vormen in de toekomstige discussie over dit onderwerp. Over de waarde en interpretatie van de hier getoonde resultaten is uiteraard discussie mogelijk. Ik nodig hierbij dan ook alle betrokkenen en betrokken instanties uit dat te doen. Als dit rapport mede een aanzet is tot die discussie, kan dit project in ieder geval als geslaagd beschouwd worden.

Tot slot wil ik graag alle beoordelaars van harte bedanken voor hun grote inzet en hun waardevolle tijd die zij in dit project gestoken hebben.

Dr C. van Vliet,
directeur Kwaliteitsbureau NVAB

Inhoud

Executive summary **4**

1 Inleiding **8**

1.1 Aanleiding 8
1.2 Probleem- en doelstelling 10

2 Methode **12**

2.1 Selectie van arbokennisproducten 12
2.2 Selectie van beoordelaars 12
2.3 Het AGREE-instrument en vragenformulier 14
2.4 Gegevensverzameling 14
2.5 Gegevensbewerking en –analyse 16

3 Resultaten **20**

3.1 De AGREE-scores van de arbokennisproducten 20
3.2 Overeenstemming tussen beoordelaars 24
3.3 Relevantie van AGREE-items 28
3.4 Toegevoegde vragen 30

4 Beschouwing **37**
4.1 Sterktes en zwaktes van het onderzoek 37
4.2 Betekenis van de gevonden AGREE-scores 40
4.3 Relevantie AGREE-items en mogelijke aanvullingen 46
4.4 Richtlijnigheid, gebruiksgemak en waardering AGREE 47

5 Samenvatting en conclusies **52**

5.1 Doel en vraagstelling van het onderzoek 52
5.2 Methode 52
5.3 Resultaten 52
5.4 Conclusies en aanbevelingen 54

6 Literatuur **57**

Bijlagen
1 NVAB over richtlijnen 62
2 STECR-webteksten 63
3 Vragenformulier 65
4 Interne consistentie van de AGREE-domeinen 66
5 Volledigheid van de ingevulde vragenformulieren 67
6 Gemiddelde AGREE-itemscores per arbokennisproduct: Stecr 68
7 Gemiddelde AGREE-itemscores per arbokennisproduct: NVAB 69
8 Gemiddelde AGREE-itemscores per arbokennisproduct: Overige 70
9 Frequenties maximale ranges per arbokennisproduct over 23 items 71
10 Intraclass Correlatie Coëfficiënten per item en domein 72
11 Scores toegevoegde vragen per arbokennisproduct: Stecr 73
12 Scores toegevoegde vragen per arbokennisproduct: NVAB 74
13 Scores toegevoegde vragen per arbokennisproduct: Overig 75

EXECUTIVE SUMMARY

Arbokennisproducten onder de AGREE-loep

1. Doel en vraagstelling van het onderzoek

Dit onderzoek is uitgevoerd als onderdeel van het “Programma ter versterking van de kennis- en onderzoeksinfrastructuur van de arbeids- en bedrijfsgeneeskunde in arbodiensten” (KIS). Het project is bedoeld om arboprofessionals te helpen bij het beoordelen van de waarde van richtlijnen en andere documenten waarin een aanbevelingen gegeven worden voor de adequate aanpak van arbeidsgerelateerde problemen (hierna aangeduid als ‘arbokennisproducten’). De aanleiding voor het onderzoek is het debat over de verschillen tussen diverse soorten arbokennisproducten zoals de NVAB-Richtlijnen, de STECR-Werkwijzers en de diagnostische protocollen van het NCvB. In de praktijk is niet altijd duidelijk wat de relevantie is van de geconstateerde verschillen, in het bijzonder wanneer de arbokennisproducten eenzelfde onderwerp behandelen. Het centrale vraagstuk is of een bestaand instrument voor de beoordeling van richtlijnen in de gezondheidszorg (AGREE) ook bruikbaar is om de kwaliteit van arbokennisproducten te beoordelen. Dit vraagstuk is in de volgende vraagstellingen uitgewerkt:

1. Wat zijn de AGREE-scores voor de diverse arbokennisproducten en hoe verhouden deze zich tot elkaar?
2. Levert de AGREE ook in een arbocontext valide en betrouwbare scores op? Hoe is de overeenstemming tussen verschillende beoordelaars en wat betekenen eventuele verschillen tussen beoordelaars?
3. Hoe wordt de relevantie van de beoordelingscriteria van de AGREE beoordeeld in deze arbocontext?
4. Kan elk type arbokennisproduct met de AGREE beoordeeld worden of moet het een ‘richtlijn’ zijn?
5. Hoe wordt het gebruiksgemak, de uitvoerbaarheid en de toepassing in de praktijk gewaardeerd?

2. Methode

Het AGREE-instrument (Appraisal of Guidelines for Research and Evaluation) is een beoordelingslijst die is ontwikkeld door experts uit 13 landen, internationaal gevalideerd en aanbevolen door de WHO en de Raad van Europa. De lijst bestaat uit 23 items verdeeld over zes domeinen en een algemene vraag of men de richtlijn aanbeveelt voor toepassing in de praktijk. Op basis hiervan kunnen zes domeinscores berekend worden, die elk verwijzen naar een aparte dimensie van kwaliteit van richtlijnen: onderwerp en doel, betrokkenheid belanghebbenden, methodologie, helderheid en presentatie, toepassing en onafhankelijkheid opstellers.
De AGREE levert domeinscores op tussen 0 en 100 voor deze zes kwaliteitsaspecten van het onderzochte arbokennisproduct. 100 betekent dat aan alle criteria van dat domein goed voldaan is; 0 dat aan geen enkel criterium van dat domein voldaan is. Bij een score van 66,7 wordt globaal genomen geen van de criteria negatief beoordeeld.

In totaal zijn 22 arbokennisproducten afkomstig van de NVAB, STECR en een aantal andere kennisinstituten beoordeeld door 30 sleutelpersonen uit kennisinstituten en de universitaire wereld en door personen die binnen grote arbodiensten betrokken zijn bij deskundigheidsbevordering. Elk van hen ontving drie verschillende arbokennisproducten en een vragenlijst, bestaande uit de 23 AGREE-items, waarbij voor elk item ook gevraagd is hoe relevant men dat criterium in dit specifieke geval vond. Tevens waren een twintigtal vragen toegevoegd over het nut en de toepasbaarheid van de AGREE. Elk arbokennisproduct werd door twee tot zes verschillende mensen beoordeeld, waarbij een product nooit beoordeeld werd door iemand uit de kring van de organisatie die het product uitbracht.

3. Resultaten

De AGREE-scores

De AGREE-domeinscores van de onderzochte arbokennisproducten lopen flink uiteen, maar zijn in het algemeen laag. Voor elk domein zijn er echter voorbeelden voorhanden van arbokennisproducten met scores boven de drempelwaarde van 66,7. Veel scores van de NVAB-richtlijnen zijn boven die drempelwaarde. Lage domeinscores verwijzen naar aspecten die verbeterd kunnen worden. Soms is dat verbeteren alleen een kwestie van meer informatie geven over de manier waarop het arbokennisproduct ontwikkeld is. In andere gevallen duidt de lage score op meer structurele problemen in het ontwikkelproces, waarbij alleen door grondiger aanpassingen hogere AGREE-scores gerealiseerd kunnen worden.

- *Onderwerp en doel*

Deze domeinscores zijn over het algemeen redelijk tot goed. Verbeteringen zijn vaak eenvoudig te realiseren door het aanvullen van de informatie.

- *Betrokkenheid belanghebbenden*

Deze domeinscores beslaan een brede range. Voor verbeteringen op dit vlak is het meestal nodig dat er explicieter gebruik gemaakt wordt van de wensen en voorkeuren van cliënten, patiënten of werknemers (en eventueel werkgevers).

- *Methodologie*

Er zijn tal van arbokennisproducten waarbij de belangrijke domeinscore voor methodologie erg laag is, omdat niet duidelijk gemaakt wordt of bij het ontwikkelproces voldoende systematisch gebruik is gemaakt van wetenschappelijke literatuur en de relatie tussen inhoud van de adviezen is en het onderliggende bewijs voor effectiviteit van die interventies niet duidelijk is.

- *Helderheid presentatie*

Scores zijn over het algemeen redelijk tot goed. Door middel van praktijktesten of pre-testen van de uiteindelijke indeling en vormgeving kunnen verbeteringen gerealiseerd worden.

- *Toepassing*

Veel arbokennisproducten scoren laag op dit domein. Bij het opstellen van arbokennisproducten wordt blijkbaar te weinig rekening gehouden met de organisatorische belemmeringen en kostenimplicaties; hulpmiddelen voor de toepassing ontbreken vaak.

- *Onafhankelijkheid opstellers*

Deze scores zijn vaak laag omdat er weinig informatie over is opgenomen in de arbokennisproducten. We hebben de indruk dat veel opstellers het belang van deze informatie onderschatten. Omdat die informatie meestal wel beschikbaar is, betekent dat echter ook dat verbeteringen eenvoudig te realiseren zijn.

Overeenstemming tussen beoordelaars

De instrumenteigenschappen van de AGREE in deze arbocontext zijn redelijk tot goed. De oordelen van verschillende personen voor hetzelfde arbokennisproduct lopen regelmatig uiteen.

Relevantie van de AGREE-items

Alle AGREE-items worden door de beoordelaars relevant geacht om de kwaliteit van arbokennisproducten in kaart te brengen. Het aantal voorstellen voor aanvullingen is beperkt.

Richtlijnkarakter van arbokennisproducten

Onafhankelijk van de titel, aanduiding of toelichting in de inleiding van de opstellers zelf, is met behulp van vier kenmerken het richtlijnkarakter van arbokennisproducten vastgesteld. Dat de NVAB-Richtlijnen een hoge score voor ‘richtlijnigheid’ hebben is niet verbazingwekkend. Maar ook een aantal STECR-werkwijzers en enkele overige arbokennisproducten worden door de beoordelaars gezien als documenten met een ontegenzeggelijk richtlijnkarakter. Dat betekent dat dergelijke producten met criteria voor richtlijnen goed beoordeeld kunnen worden.

Gebruiksgemak, uitvoerbaarheid en waardering AGREE

Het gebruiksgemak van de AGREE is goed: het instrument kan bij de meeste arbokennisproducten gemakkelijk ingevuld worden. De beoordelaars zijn van mening dat het gebruik van de AGREE op allerlei vlakken meerwaarde kan hebben om de kwaliteit van nieuwe arbokennisproducten te verbeteren of om arboprofessionals inzicht te geven in de kwaliteit van bestaande arbokennisproducten.

4. Beschouwing en conclusies

De AGREE-scores

- Alhoewel het toekennen van AGREE-scores aan de verschillende arbokennisproducten niet het hoofddoel vormt van deze studie, zijn de gevonden scores wel informatief. Ze maken goed zichtbaar wat vanuit het AGREE-perspectief de zwakke plekken van de onderzochte arbokennisproducten zijn en hoe die verbeterd kunnen worden. Acceptabele AGREE-domeinscores zijn niet slechts een fictief ideaal, maar ook feitelijk realiseerbaar: voor elk domein zijn er voldoende voorbeelden van arbokennisproducten met een hoge AGREE-score.
- Domeinscores kunnen verbeterd worden door volledediger informatie op te nemen over het ontwikkelproces of bepaalde stappen in dat proces systematischer uit te voeren. Soms moet het ontwikkelproces uitgebreid worden met nieuwe elementen.
- Ook al is de inhoud zorgvuldig samengesteld, als in het document informatie ontbreekt over manier waarop adviezen onderbouwd zijn, dan leidt dat tot een lage methodologiescore. De geloofwaardigheid van de inhoud van die arbokennisproducten wordt daarmee aangetast. Ook als er weinig wetenschappelijke literatuur beschikbaar is om arboadviezen te schragen zijn goede methodologiescores mogelijk door systematisch te werken en verslag te doen van het zoekproces.

Overeenstemming tussen beoordelaars

- Verschillen tussen beoordelaars – die regelmatig voorkomen – kunnen op diverse manieren geïnterpreteerd worden. Het belang van die verschillen hangt af van de context waarin de AGREE-scores gebruikt worden.
- Gemiddelde scores met behulp van het AGREE-instrument van minimaal 4 beoordelaars van diverse achtergrond leveren een betrouwbare uitslag op.
- Mits voldoende toegelicht, hoeven de AGREE-items niet anders verwoord te worden in een arbocontext. De huidige formulering levert geen bovenmatige spreiding van antwoordscores op.
- Allerlei soorten arbokennisproducten (als zelfstandig document uitgegeven) kunnen op een goede vergelijkbare manier met de AGREE beoordeeld worden.

Relevantie van de AGREE-items

- Volgens de beoordelaars omvatten de AGREE-items alle belangrijke criteria om de kwaliteit van arbokennisproducten te kunnen beoordelen.

Richtlijnkarakter van arbokennisproducten

- Het richtlijnkarakter van arbokennisproducten kan weliswaar verschillen, maar de 6 AGREE-domeinen bestrijken een dermate breed gebied aan kwaliteitscriteria dat die domeinscores een goed totaalbeeld van de diverse kwaliteitsaspecten van het arbokennisproduct geven.
- Vooropgesteld dat het om een arbokennisproduct gaat (een zelfstandig document met aanbevelingen voor een welomschreven arboprobleem, bedoeld voor brede verspreiding onder arboprofessionals als hulp bij het dagelijkse werk, door een onafhankelijke groep experts opgesteld) dat zijn buiten de AGREE vooralsnog geen extra criteria nodig om een goed beeld te krijgen van de kwaliteit van een arbokennisproduct.

Gebruiksgemak, uitvoerbaarheid en waardering AGREE

- Het invullen van de AGREE is eenvoudig en goed uitvoerbaar.

5. Aanbevelingen

- De AGREE-scores uit deze studie kunnen van nut zijn bij het verbeteren van de kwaliteit van de onderzochte arbokennisproducten. De scores uit deze studie moeten vooralsnog niet in absolute zin als waardeoordeel gebruikt worden.
- In arbokennisproducten moet voldoende aandacht besteed worden aan de manier waarop literatuur gezocht en geselecteerd is, en hoe de aanbevelingen onderbouwd zijn. Op dat vlak is tegenwoordig voldoende expertise beschikbaar. Het ontwikkelen van arbokennisproducten met een lage score voor methodologie is onwenselijk, en moet vermeden worden omdat het functioneren van de arbokennisinfrastructuur daardoor verzwakt.

- Arboprofessionals kunnen de AGREE-domeinscores gebruiken om zwakke kanten van arbokennisproducten te herkennen. Betrouwbare AGREE-scores zouden voor veel gebruikte arbokennisproducten openbaar beschikbaar moeten zijn. Arboprofessionals moeten geïnformeerd worden over de betekenis van de AGREE-domeinscores.
- Goed ontwikkelen van arbokennisproducten vraagt om specifieke kennis en voldoende menskracht en tijd. De diverse kennisinstituten en universiteiten moeten samenwerken om zo efficiënt mogelijk gebruik te maken van de beperkte middelen en menskracht.
- Met de informatie uit een praktijktest kunnen diverse domeinscores van arbokennisproducten verbeterd worden, met name die van het domein 'Toepassing'.
- Volledige en transparante informatie over het ontwikkelproces opnemen in een arbokennisproduct kan de andere kwaliteitscriteria in gevaar brengen. Verwijzingen naar internetpagina's voor aanvullende informatie kunnen een goede oplossing zijn.
- De zes AGREE-domeinscores vormen aparte dimensies van kwaliteit van arbokennisproducten, en worden bij voorkeur niet samengebracht in één kwaliteitsscore.Voor betrouwbare scores moeten arbokennisproducten door minimaal vier beoordelaars van diverse achtergrond gescoord worden met behulp van het AGREE-instrument.
- Er is debat nodig over de betekenis van de AGREE-scores omdat het belang van verschillen tussen beoordelaars daarmee samenhangt.
- Beoordelaars moeten meer uitleg over de AGREE krijgen en meer training en instructie in het gebruik van het instrument.
- In de beoordelingsprocedure moet ruimte zijn voor discussie en overleg tussen beoordelaars over niet overeenstemmende scores.
- Er zijn enkele suggesties om items/criteria toe te voegen aan de AGREE als deze in een arbocontext gebruikt wordt. De (internationale) vergelijkbaarheid van AGREE-scores moet echter behouden blijven. Terwijl de items van het domein 'Methodologie' vrij specifiek zijn benoemd, zijn die uit het domein 'Toepassing' vrij breed en algemeen geformuleerd. Hierover zou discussie kunnen plaatsvinden. Desalniettemin zijn er door de beoordelaars geen aanvullingen voorgesteld voor dit domein.
- Er is binnen de groep van beoordelaars in deze studie steun voor een onafhankelijk keurmerk voor arbokennisproducten gebaseerd op de AGREE-scores. Die suggestie verdient nader onderzoek.
- Bedrijfsartsen en arboprofessionals in Nederland moeten op een eenvoudige manier kunnen achterhalen hoe goed een arbokennisproduct op diverse aspecten scoort. Op een onafhankelijke website zouden bijvoorbeeld de AGREE-domeinscores beschikbaar gesteld kunnen worden. Het zou plezierig zijn als op diezelfde site ook de arbokennisproducten van goede kwaliteit te downloaden of te verkrijgen zijn. De mogelijkheid tot downloaden biedt de STECR-website reeds.
- Elke organisatie die arbokennisproducten uitbrengt kan de AGREE-items gebruiken als leidraad om het ontwikkelproces te verbeteren.
- Organisaties die veel arbokennisproducten uitbrengen en een vooraanstaande rol in de arbokennisinfrastructuur willen spelen, moeten met een expliciet kwaliteitsbeleid streven naar het realiseren van hoge AGREE-domeinscores voor hun arbokennisproducten.

De aanleiding voor dit onderzoek werd gevormd door het debat over verschillen tussen diverse soorten arbokennisproducten. We hebben aangetoond dat een breed panel van sleutelpersonen uit de arbokennisinfrastructuur dezelfde criteria hanteert om een arbokennisproduct 'goed' te noemen. De gezamenlijke aandacht kan daarom nu gericht worden op het realiseren van arbokennisproducten van zo goed mogelijke kwaliteit, die met behulp van de AGREE in kaart gebracht kan worden.

Een beter arbokennisproduct, in de zin van een hogere AGREE-score, leidt niet meteen tot arbozorg met meer effectiviteit. De verwachting is wel dat die invloed in de nabije toekomst zal gaan toenemen. De kennisinfrastructuur op het gebied van arbodienstverlening en bedrijfsgeneeskunde kan alleen goed functioneren als de beschikbare arbokennisproducten inhoudelijk onbetwist zijn. De AGREE vormt een belangrijk instrument om een betere kwaliteit van arbokennisproducten te realiseren.

1. Inleiding

1.1 Aanleiding

Arbokennisproducten

NVAB en STECR geven documenten uit waarin aanbevelingen gegeven worden voor de adequate aanpak van arbeidsgerelateerde gezondheidsproblemen. STECR gebruikt daarbij vooral de aanduiding 'Werkwijzers', de NVAB gebruikt in hoofdzaak de aanduiding 'Richtlijn'. In dit verslag gebruiken we voor deze documenten het algemene woord 'arbokennisproduct': een document met een aantal aanbevelingen (specifieke, expliciete uitspraken) voor passend of 'goed' handelen bij een omschreven probleem of doelpopulatie ter ondersteuning van de dagelijkse praktijk van veel arboprofessionals gebaseerd op systematisch verzamelde kennis, steekhoudende argumenten, (senior) ervaringen in de praktijk, kritische beoordeling en consensus, uitgebracht door een belangenvrije organisatie. Het betreft hier dus kennisproducten die deels op ervaring in de praktijk gebaseerd zijn en altijd voor toepassing bedoeld zijn.
Er zijn tal van andere organisaties die vergelijkbare documenten publiceren en verspreiden: kenniscentra, brancheorganisaties, universiteiten en uitgevers. Regelmatig zijn er voor hetzelfde onderwerp diverse producten beschikbaar, bijvoorbeeld voor lage rugklachten en ABBE/rsi/KASN. Arbodiensten maken soms zelf inhoudelijke werkinstructies die intern onder de titel 'richtlijn' of 'protocol' verspreid worden.

De samenstellers van deze arbokennisproducten hebben doorgaans argumenten waarom die tekst een eigen plaats verdient naast andere, maar voor de professionals in het veld is het verschil tussen producten niet altijd duidelijk. STECR positioneert haar werkwijzers met de trefwoorden 'practice-based', 'praktijkgericht' en 'multidisciplinair', de NVAB legt de nadruk op 'evidence-based', 'wetenschappelijkheid' en stelt het handelen van de bedrijfsarts centraal (Hulshof & Weel 1996): zie ook bijlage 1. Hoewel STECR en het Kwaliteitsbureau NVAB vanaf het begin geprobeerd hebben door gezamenlijke communicatie aan het veld duidelijkheid te verschaffen, is dat maar ten dele gelukt.

Inhoudelijke kwaliteit van arbokennisproducten

Niemand trekt in twijfel dat er veel behoefte is aan explicitering van het bedrijfsgeneeskundig handelen en onderbouwing met onderzoeksresultaten naar de effectiviteit van diverse vormen van arbodienstverlening (van Dijk et al. 2003). Maar het uitbrengen en verspreiden van een arbokennisproduct als een formeel document schept ook verplichtingen. Er moet adequaat gebruik gemaakt worden van beschikbare wetenschappelijke literatuur en het ontwikkelproces moet systematisch en transparant zijn. Soms wordt betoogd dat de praktijk niet altijd kan wachten op zo'n grondige en daarom langdurige ontwikkelprocedure. Ook in de periode dat de wetenschappelijke discussie over de meest effectieve aanpak van een arboprobleem nog niet is afgerond, kan er een dringende behoefte zijn aan houvast. De beperkingen van documenten die met zo'n uitgangspunt gepubliceerd worden, moeten goed uitgelegd worden aan professionals.

Het is de ervaring dat de beperkingen van een tekst genoemd in de inleiding in de praktijk echter makkelijk naar de achtergrond verdwijnen. Los van bedoeling en reikwijdte, gaan de aanbevelingen uit een arbokennisproduct dan een eigen leven leiden. Zo wordt er in het toetsingskader voor de wet verbetering Poortwachter bijvoorbeeld verwezen naar STECR-Werkwijzers en NVAB-Richtlijnen: de verzekeringsarts moet beide gebruiken om te beoordelen of bij de begeleiding van een zieke werknemer adequaat gehandeld is door de bedrijfsarts of arbodienst (UWV 2002; Ministerie van SZW 2004). Men nam aan dat de richtlijnen en werkwijzers voor dat doel geschikt waren en dat rapportages voldoende inhoudelijke informatie zouden bevatten om zo'n inhoudelijk oordeel over het handelen te kunnen motiveren. Maar verzekeringsartsen geven aan dat het in de praktijk nog niet zo eenvoudig toe te passen is.
Naarmate er meer arbokennisproducten gepubliceerd worden, neemt de behoefte toe om een beter zicht te krijgen op de kwaliteit en zeggingskracht ervan. Professionals en beleidsmakers weten wat de inhoudelijke waarde is van de verschillende breed beschikbare documenten.

Onderbouwing van arbokennisproducten

Een arboprofessional moet met veel schriftelijke informatie kunnen omgaan. Terecht mag van zo iemand verwacht worden dat hij/zij in staat is de kwaliteit van informatie goed te beoordelen en dienovereenkomstig te hanteren in de praktijk. Voor een goede professional is er een verband tussen de mate van wetenschappelijke onderbouwing van een tekst en de striktheid waarmee een arboprofessional de aanbevelingen hanteert. Kort gesteld: hoe meer bewijs, hoe minder vrijheidsgraden. Daarmee moet Evidence-Based Medicine (EBM) niet als karikatuur worden neergezet of als eenheidsworst of kookboekgeneeskunde afgeschilderd worden (Jansen et al. 2003; van Everdingen et al. 2003). Beargumenteerd afwijken van richtlijnen is altijd mogelijk als een vorm van goed professioneel handelen. EBM betekent in een bepaalde situatie met behoud van de eigen professionele expertise (en dus ook ervaring uit de praktijk!) gebruik maken van het beste beschikbare bewijs, terwijl er ook rekening gehouden wordt met de mogelijkheden van de organisatie en de voorkeuren van de patiënt (Offringa et al. 2000). Maar als een bepaalde aanpak in diverse onderzoeken van hoge kwaliteit het meest effectief is gebleken, dan moet zo'n aanbeveling terecht veel gewicht in de schaal leggen. Een professional die wil beargumenteren dat een cliënt met een alternatieve aanpak beter geholpen is, zal van goede huize moeten komen. Hij kan bijvoorbeeld verwijzen naar literatuur of onderzoeksgegevens die bij het samenstellen van de richtlijn over het hoofd zijn gezien of onvoldoende gewicht hebben gekregen en waar die alternatieve aanpak beter uit de bus komt. De cliënt kan natuurlijk ook zelf - goed voorgelicht over voor- en nadelen - een andere aanpak prefereren.
Dat ligt anders als een aanbeveling vooral op basis van praktijkervaringen en de meningen van experts is gebaseerd. Wil een professional van zo'n advies afwijken, dan is het vooral van belang dat hij overtuigend kan laten zien dat hij via een andere weg minimaal op hetzelfde resultaat uitkomt. Is het arbokennisproduct vooral een weerslag van een grondige discussie tussen professionals en experts en heeft het een plaats verworven omdat er nog weinig over dat onderwerp op schrift gesteld was, dan zijn de vrijheidsgraden voor de professional in de praktijk nog groter: zo'n arbokennisproduct kan als een lijst met aandachtspunten of checklist gebruikt worden. Maar als de professional op basis van goede (eigen) afwegingen tot andere aanpak besluit, kan hem geen slecht handelen verweten worden.

In de literatuur over richtlijnontwikkeling worden hiërarchieën gebruikt om de mate van bewijs te categoriseren. Op basis van de kwaliteit van het onderliggende bewijs kunnen vervolgens de aanbevelingen in een richtlijn geordend worden. Een voorbeeld van zo'n systeem komt van de Scottish Intercollegiate Guidelines Network, het zogenaamde SIGN-systeem om 'levels of evidence' en 'grades of recommendations' te ordenen (SIGN 2004). Nicholson et al (2004) gebruiken dat systeem bijvoorbeeld om aanbevelingen voor identificatie, preventie en behandeling van beroepsastma te ordenen. Er bestaan meer systemen, waarbij de specifieke invulling van de niveaus wat kan verschillen maar die grosso modo op hetzelfde neerkomen: bovenaan staan de aanbevelingen die gebaseerd zijn op meta-analyses en systematic reviews en onderaan staan de aanbevelingen die door de expertgroep zijn opgesteld zonder expliciete grondslag in onderzoeksgegevens, maar gebaseerd op inzicht in werkingsmechanismen, professionele ervaring of algemene principes (van Everdingen 1999; Atkins et al. 2005; Usphur 2003).
Vanaf de richtlijnen KASN en Astma/COPD heeft de NVAB bij elke aanbeveling de mate van bewijs expliciet gespecificeerd door middel van de volgende indeling (Verbeek et al. 2003).

A1	Sterk bewijs uit meta-analyse of systematic review van tenminste enkele onderzoeken van A2 niveau, waarbij de resultaten van afzonderlijke studies eenduidig zijn.
A2	Sterk bewijs uit een gerandomiseerde trial van goede kwaliteit en omvang.
B	Matig bewijs uit een gerandomiseerde trial van matige kwaliteit of onvoldoende omvang of ander vergelijkend onderzoek of een systematische review over dergelijk onderzoek.
C	Beperkt bewijs uit een niet-vergelijkend onderzoek of analogie bij andere aandoeningen.
D	Mening van de deskundigen of consensus in de projectgroep.

Hoewel er altijd professionele ruimte is om af te wijken van aanbevelingen uit een arbokennisproduct, is die vrijheid idealiter gerelateerd aan de kwaliteit van de kennis die in het arbokennisproduct wordt aangeboden. Arbokennisproducten krijgen door een autorisatie van beroepsvereniging of brancheorganisatie snel gewicht en status en vormen dan een belangrijk uitgangspunt voor de praktijk. Arboprofessionals vertrouwen erop dat dergelijke arbokennisproducten zorgvuldig zijn samengesteld en beoordeeld. Het is daarom belangrijk dat de kwaliteit van arbokennisproducten op een heldere en eenduidige manier kan worden vastgesteld.

In het debat over de noodzakelijke kwaliteit van arbokennisproducten komen verschillende uitgangspunten aan het licht. Criteria zoals wetenschappelijkheid en praktijkgerichtheid worden tegenover elkaar gezet, terwijl misschien beide van belang zijn voor een arbokennisproduct van adequate kwaliteit. Met het onderzoek waarover hier wordt gerapporteerd, wordt een nieuwe impuls aan dat debat gegeven en wordt geprobeerd brede overeenstemming te realiseren over kwaliteitscriteria voor arbokennisproducten.

Kennisinfrastructuur

Het ministerie van SZW heeft in 2003 en 2004 het Coronel instituut van het AMC te Amsterdam subsidiemiddelen verstrekt om de infrastructuur te verbeteren voor onderzoek, ontwikkeling, informatievoorziening en kennisoverdracht op het terrein van de arbeids- en bedrijfsgeneeskunde (van Dijk en Hugenholtz 2005). Arbokennisproducten vervullen een belangrijke functie in de kennisinfrastructuur: recente wetenschappelijke inzichten worden daarin samengevoegd tot praktische aanbevelingen voor het handelen.
Het onderzoek waarover hier wordt gerapporteerd was een deelproject in het KIS-kader (verbeteren Kennis InfraStructuur) en heeft als doel om na te gaan of er overeenstemming bereikt kan worden over de criteria voor goede arbokennisproducten. Het project werd uitgevoerd door het Kwaliteitsbureau NVAB. STECR wilde na overleg graag betrokken worden bij het project. Door die verbreding wordt voorkomen dat de criteria voor goede arbokennisproducten te veel als het product van één (deels belanghebbende) organisatie gezien worden. STECR wilde graag dat in het onderzoek voldoende rekening gehouden zou worden met het eigen karakter van de STECR-producten. De kans dat de criteria tot verbetersuggesties leiden wordt door die verbreding van het draagvlak vergroot.

1. 2 Probleem- en doelstelling

Inleiding

Het centrale vraagstuk in dit onderzoek is welke criteria van belang zijn bij het beoordelen van de kwaliteit van arbokennisproducten. Meer in het bijzonder gaat het om de vraag of er criteria geformuleerd kunnen worden met een breed draagvlak in de Nederlandse arbowereld.
Kwaliteit van arbokennisproducten kan niet los gezien worden van bedoeling, toepassingscontext en de belangen van opstellers, uitgevers en gebruikers. Het debat over kwaliteit in de arbodienstverlening, over evidence based occupational medicine en over de meerwaarde van arbokennisproducten is ons inziens echter inmiddels zover gevorderd dat de opinieleiders op dit gebied expliciet, dan wel impliciet grotendeels van eenzelfde set criteria uitgaan. Het doel van dit onderzoek is om die criteria aan het licht te brengen en te laten zien in welke mate er overeenstemming bestaat onder de opinieleiders. De verwachting is dat het aantonen van dat draagvlak een stimulans betekent voor de verdere verspreiding en toepassing van die criteria. Praktijkmensen raken dan beter bekend met wat goede arbokennisproducten zijn en hoe ze die kunnen herkennen, en mensen die betrokken zijn bij het opstellen van arbokennisproducten kunnen die criteria dan een grotere rol laten spelen bij het ontwikkelproces.

We realiseren ons dat consensus onder die beperkte groep opinieleiders geenszins een afspiegeling vormt van wat er leeft onder arboprofessionals in het algemeen. Maar we laten ons inspireren door het cascademodel voor implementatie en gaan er van uit dat deze opinieleiders wel een cruciale sleutelrol spelen bij het verspreiden van gedachtegoed onder praktiserende arboprofessionals (Hoenen & de Buisonjé 2003; Audet et al. 2005; Eccles & Mason 2005; Peters et al. 2003).

AGREE in arbocontext

De criteria voor goede arbokennisproducten hebben we niet zelf ontwikkeld, we hebben hiervoor gebruik gemaakt van de AGREE-lijst, Appraisal of Guidelines for Research and Evaluation (AGREE Collaboration, 2001en 2003; zie ook: Vlayen et al. 2005; Burgers et al. 2003a/2003b; Cluzeau 1999; Graham et al. 2002). Dit instrument is ontwikkeld om de kwaliteit van richtlijnen te beoordelen, en het vindt binnen de curatieve zorg steeds meer opgang. Het wordt in zowel in Nederland als daarbuiten

regelmatig toegepast. Naast tal van studies waarin de AGREE gebruikt wordt om richtlijnen voor artsen te beoordelen zijn er voorbeelden waarin het instrument gebruikt wordt om richtlijnen te beoordelen voor manueel therapeuten, chiropractors, of fysiotherapeuten (Boluyt 2005; Burgers et al. 2002; Burgers & Boluyt 2004; Burgers et al. 2004; Fervers et al 2005; Cates et al 2003; Cranney 2002; MacDermid et al. 2005; Nicholson et al. 2005; Staal et.al 2003; Tuut et al. 2002). In andere beroepsgroepen worden criteria voor richtlijnen gebruikt die veel verwantschap met de AGREE vertonen (AVVV 2004). Het AGREE-instrument is een beoordelingslijst die is ontwikkeld door experts uit 13 landen en de lijst is internationaal gevalideerd (AGREE Collaboration 2003). Het instrument is bekrachtigd door de WHO en wordt formeel aanbevolen door de Raad van Europa (European Health Committee 2001). De AGREE levert scores op voor zes domeinen, elk betrekking hebbend op een aparte dimensie van kwaliteit van richtlijnen: onderwerp en doel, betrokkenheid belanghebbenden, methodologie, helderheid en presentatie, toepassing en onafhankelijkheid opstellers. De criteria hebben betrekking op de informatie in het te beoordelen document, bijvoorbeeld: *de patiëntenpopulatie waarop de richtlijn van toepassing is, is specifiek beschreven*. De beoordelaar geeft aan in welke mate de tekst aan dat criterium voldoet met een oordeel ‘zeer oneens’, ‘oneens’, ‘eens’ of ‘zeer eens’. Met de aanduiding AGREE wordt verwezen naar het oorspronkelijke instrument met de formulering van de items (AGREE Collaboration 2001).

Vraagstellingen

In deze studie willen we nagaan of de AGREE bruikbaar is om de kwaliteit van arbokennisproducten te beoordelen. Meer specifiek vragen we ons af:

1. Wat zijn de AGREE-scores voor diverse arbokennisproducten? Wat valt op als we die scores onderling vergelijken? Hoe scoren de Nederlandse arbokennisproducten vergeleken met andere richtlijnen in de zorg in Nederland en internationaal? Zijn uit de AGREE-scores concrete verbetersuggesties af te leiden? Hoe verhouden de AGREE-scores van de NVAB-producten zich tot die van STECR en die van andere organisaties in het arboveld?
2. Wat zijn de instrumenteigenschappen van de AGREE in de context van de arbokennisproducten die beoordeeld worden? Kan uit die scores een valide en betrouwbaar kwaliteitsoordeel afgeleid worden? Hoe is de overeenstemming tussen verschillende beoordelaars? Welke betekenis moet aan de verschillen tussen beoordelaars toegekend worden?
3. Hoe wordt de relevantie van de AGREE-items beoordeeld in deze arbocontext? Ontbreken er belangrijke criteria in de AGREE?
4. Zijn de AGREE-criteria onverkort toepasbaar voor alle soorten arbokennisproducten of zijn bedoelingen van producten soms zo verschillend dat aangepaste criteria nodig zijn? Kunnen we richtlijnen, werkwijzers en andere arbokennisproducten langs dezelfde meetlat leggen? Betekenen de scores dan ook hetzelfde? Voor welk soort arbokennisproducten is de AGREE niet geschikt, en waarom niet? Wat zijn de kenmerken van arbokennisproducten die niet goed met de AGREE beoordeeld kunnen worden?
5. Hoe wordt het gebruiksgemak en uitvoerbaarheid van de AGREE beoordeeld? Is de benodigde informatie makkelijk te vinden in een arbokennisproduct? Kan de AGREE zonder tekstuele aanpassingen gebruikt worden in een arbocontext en voor de beoordeling van arbokennisproducten? In de AGREE wordt soms gesproken over ‘patiënten’ of ‘richtlijnen’. Door niets aan de formulering te veranderen, is vergelijking van scores uit andere studies beter mogelijk. Maar het nadeel is dat beoordelaars begrippen dan zelf moeten ‘vertalen’. Hoe denken de beoordelaars daar zelf over op basis van hun ervaring met de AGREE?
   Hoe is de waardering voor het AGREE-instrument? Wat verwacht men van de toepassing? Op welke manier zou het instrument in de praktijk ingezet of toegepast kunnen worden?

De nadruk in dit verkennende onderzoek ligt op het draagvlak voor de AGREE-criteria en de uitvoerbaarheid van de AGREE in deze arbocontext.

2. METHODE

2.1 Selectie van arbokennisproducten

Er is een brede range arbokennisproducten in de studie betrokken. Op website van STECR staat een richtlijnenpagina waar, verdeeld over ruim 200 records, honderden documenten te vinden zijn met aanbevelingen voor het handelen van arboprofessionals. We hebben daaruit op een pragmatische manier een selectie gemaakt aan de hand van de volgende *inclusiecriteria*:
- Recent gepubliceerd door NVAB (richtlijnen), STECR (werkwijzers) of andere organisaties.
- Aanwijzingen dat de arbokennisproducten regelmatig gebruikt worden in de praktijk en nascholing.
- Betrekking hebbend op regelmatig voorkomende arboproblemen of aandoeningen.
- Primair bedoeld voor bedrijfsartsen en/of arboprofessionals.
- Openbaar en breed beschikbaar.
- Bij voorkeur elektronisch beschikbaar.

De LESA-Overspanning was weliswaar nog niet gepubliceerd ten tijde van het onderzoek, maar is toch meegenomen omdat de tekst op dat moment wel geautoriseerd was. De daaraan gerelateerde NVAB-Richtlijn Psychische Klachten was ruim vier jaar oud en zou binnenkort in revisie worden genomen. De LESA-Overspanning zorgde voor een moderne variant en vergelijkingsmateriaal.
De NVAB-Leidraad Verwijzen door de bedrijfsarts was ten tijde van het onderzoek al vormgegeven en bijna gedrukt.
Enkele producten zijn omvangrijk, maar bestaan uit diverse (losse) onderdelen. We hebben bijvoorbeeld uit de klapper met SMB-richtlijnen van Arbouw het gedeelte gekozen dat betrekking had op rugklachten, omdat er enkele andere arbokennisproducten in het onderzoek op dat onderwerp betrekking hebben.

Exclusiecriteria zijn :
- Gepubliceerd vóór 1999.
- Documenten die in revisie zijn of inmiddels aanwijsbaar geen juiste inhoud meer vermelden.
- Een omvang die problemen bij verspreiding en beoordeling zou opleveren.

Hoewel enkele van de STECR-producten ook in revisie waren, was dat tijdens het opzetten van het onderzoek nog niet bekend of waren de nieuwe versies nog niet beschikbaar.
De Standaard Geen Duurzaam Benutbare Mogelijkheden van het Lisv/UWV was weliswaar van 1996, maar is meegenomen vanwege de samenhang met enkele werkwijzers van STECR. De standaard is overigens nog steeds van kracht en niet officieel geactualiseerd.

Het uiteindelijke palet aan documenten is breed: brochures, waarvan één met een achterliggende literatuurstudie, korte teksten en uitgebreide boekwerken of een verzameling webpagina's. Sommige zijn zelfstandige publicaties, andere zijn onderdeel van een groter geheel (zie hierna: tabel 1). Als dat nodig was voor een goed begrip en positionering is in die gevallen de algemene inleiding toegevoegd.

2.2 Selectie van beoordelaars

We hebben beoordelaars gezocht die met gezag, kennis van zaken en ervaring konden spreken over het ontwikkelen en toepassen van arbokennisproducten: in de eerste plaats personen die direct betrokken waren bij het opstellen van richtlijnen en werkwijzers. Daarnaast zijn we op zoek gegaan naar mensen met expertise op het gebied van Evidence Based (occupational) Medicine, uitvoeren van literatuuronderzoek, selecteren en beoordelen van onderzoeksresultaten, uitvoeren van deskundigheidsbevorderingactiviteiten voor arboprofessionals, uitwerken en uitvoeren van kwaliteitsbeleid, het evalueren van de effectiviteit van arbo-interventies of het toepassen en invoeren van richtlijnen en werkwijzers. We hebben personen benaderd uit de projectgroepen van NVAB-richtlijn of -stuurgroep (geselecteerd door de staf van het Kwaliteitsbureau NVAB), uit de STECR-kenniskringen (geselecteerd door STECR-staf), samenstellers van de overige arbokennisproducten, stafmedewerkers en managers betrokken bij kwaliteitsbeleid en kennismanagement bij grote arbodiensten, onderzoekers op het gebied van arbodienstverlening en bedrijfsgeneeskunde van de verschillende faculteiten en enkele grote onderzoeksinstituten en personen actief in het Kennisnetwerk Gezondheid faculteiten

Tabel 1. Arbokennisproducten

	Onderwerp, soort document (aantal pagina's)	*Organisatie, Plaats, Jaartal*
1	ABBE-rsi STECR-werkwijzer (33)	STECR, Hoofddorp, 2003
2	Arbeidsconflicten STECR-werkwijzer (62)	STECR, Hoofddorp, 2002
3	Aanpak Problematisch Verzuim STECR-werkwijzer (33)	STECR, Hoofddorp, 2002
4	Lage Rugklachten en Arbeidsomstandigheden STECR-werkwijzer (44)	STECR, Hoofddorp, 2002
5	Reductie van werkstress door de werknemer in de Gezondheidszorg/Onderwijs STECR-werkwijzer (68)	STECR, Hoofddorp, 2003
6	Cultuur of Persoon; Sociaal-medische Begeleiding Allochtonen STECR-Assist (16)	STECR, Hoofddorp, 2003
7	Vrouwen in de WAO + Vrouwen níet in de WAO Voorstudie (33) en STECR–Assist (20)	STECR, Hoofddorp, 2003 en 2004
8	Somatisatie, multidisciplinaire aanpak bij somatoforme verschijnselen door arboprofessionals STECR-werkwijzer (78)	STECR, Hoofddorp, 2004
9	Handelen van de bedrijfsarts bij werknemers met lage rugklachten NVAB-richtlijn (39)	NVAB, Eindhoven, 1999
10	Handelen van de bedrijfsarts bij werknemers met psychische klachten NVAB-richtlijn (91)	NVAB, Eindhoven, 2000
11	Handelen van de bedrijfsarts bij werknemers met Astma en COPD NVAB-richtlijn (128)	NVAB, Amsterdam, 2003
12	Handelen van de bedrijfsarts bij werknemers met klachten aan arm, schouder of nek NVAB-richtlijn (85)	NVAB, Amsterdam, 2003
13	Handelen van de bedrijfsarts bij het verwijzen van werknemers NVAB Leidraad (57)	NVAB, Utrecht, 2004
14	Overspanning Landelijke Eerstelijns Samenwerkingsafspraak (7)	NVAB, NHG, Utrecht, 2004
15	Geen Duurzaam Benutbare Mogelijkheden Standaard (27)	Lisv/UWV, Amsterdam, 1996
16	Zwangerschap en bevalling als oorzaak van ongeschiktheid voor haar arbeid Standaard (23)	Lisv/UWV, Amsterdam, 1999
17	Beroepsgebonden burnout/overspanning voorwoord (8) en registratierichtlijn (5)	NCvB, Amsterdam, 1999
18	Interventies bij acute lage rugklachten Kennisbank http://www.kenniscentrumakb.nl (22)	Nederlands Kenniscentrum Arbeid en Klachten Bewegingsapparaat, Rotterdam, 2002
19	Klachtenreductie en stressmanagement, Arbeidshulpverlening bij psychische klachten: Inleiding (8) en protocol (21)	Nederlands Kenniscentrum Arbeid en Psyche, i.s.m. GGZ Nederland en GGZ Friesland, Utrecht, 2003
20	Diabetes mellitus: Hoofdstuk uit het Handboek Arbeid en Belastbaarheid (44)	Bohn Stafleu Van Loghum, Houten, 2003
21	PAGO voor huisvuilbeladers (artikel uit TBV) (13)	Coronel Instituut, Amsterdam, 2003
22	Sociaal Medische Begeleiding van Werknemers in de Bouw: Introductie (4) en richtlijn rugklachten (13)	Arbouw, Amsterdam, 2004

en enkele grote onderzoeksinstituten en personen actief in het Kennisnetwerk Gezondheid en Arbeid. We hebben gestreefd een groep beoordelaars samen te stellen van brede variëteit: bedrijfsartsen en andere disciplines, onderzoekers en praktijkmensen, zowel uit NVAB- als STECR-kring en daarbuiten. We hebben actief beoordelaars geworven bij de organisaties waarvan arbokennisproducten in de selectie zijn opgenomen.
Namens de directeur van de NVAB zijn 86 personen aangeschreven met het verzoek om medewerking aan het onderzoek. Bijna de helft daarvan zijn personen die STECR ons gesuggereerd heeft. Ruim de helft is bedrijfsarts, de rest heeft een andere arbodiscipline. Ruim 40% is op dat moment

werkzaam bij een arbodienst als praktiserend arboprofessional. Ongeveer 40% is werkzaam bij kennisinstituten, eenzelfde percentage bij universiteiten. In de uitnodiging is de samenwerking met STECR vermeld, het doel van het onderzoek toegelicht en beschreven wat er van hen gevraagd wordt. We zijn ervan uitgegaan dat het beoordelen van drie arbokennisproducten ongeveer twee tot drie uur tijd zou kosten. Via een weblink kon men alvast meer informatie raadplegen over de AGREE, het instrument dat bij de beoordeling gebruikt zou gaan worden.Van de mensen die half september 2004 zijn aangeschreven, zegden 39 personen hun medewerking toe. Hiervan hebben 30 personen ook daadwerkelijk aan het onderzoek deelgenomen.

2.3 Het AGREE-instrument en vragenformulier

Het AGREE-instrument bestaat uit 23 items verdeeld over zes domeinen en een algemene vraag of men de richtlijn aanbeveelt voor toepassing in de praktijk. Op basis van deze items kunnen zes domeinscores berekend worden, die elk verwijzen naar een aparte dimensie van kwaliteit van richtlijnen: onderwerp en doel, betrokkenheid belanghebbenden, methodologie, helderheid en presentatie, toepassing en onafhankelijkheid opstellers. In elk item is een criterium verwoord en de beoordelaar geeft aan in welke mate de voorliggende tekst aan dat criterium beantwoord. Gescoord wordt op een vierpunts Likertschaal van 1: ‘Zeer oneens’ naar 4: ‘Zeer eens’.

Aan het antwoordformulier is in dit onderzoek een kolom toegevoegd, zodat de beoordelaar naast de AGREE-score tevens kon invullen in welke mate hij/zij er mee instemde dat het betreffende criterium voor dat specifieke arbokennisproduct relevant was. Die vraag werd op gelijke wijze als de AGREE gescoord op een vierpunts Likertschaal die liep van 1: ‘Zeer oneens’ tot 4: ‘Zeer eens’ (zie bijlage 2: het vragenformulier).

Tenslotte zijn een aantal vragen toegevoegd over het type arbokennisproduct en de uitvoerbaarheid van de AGREE en de waardering ervoor (zie het vragenformulier). De conceptvragenlijst is door STECR van commentaar voorzien en aangevuld, zodat de uiteindelijke vragenlijst de belangrijke thema’s en vraagstukken rondom de AGREE dekt. Voor deze vragen zijn nagenoeg altijd dezelfde antwoordcategorieën gehanteerd. Enkele vragen zijn in een open vorm gesteld.

2.4 Gegevensverzameling

Elk van de beoordelaars krijgt eind september 2004 een uitgebreide instructie toegestuurd. Een van de onderdelen is de standaard invulinstructie bij de AGREE (AGREE Collaboration 2003). Er is een toelichting toegevoegd hoe bepaalde items in deze arbocontext gelezen en beoordeeld moeten worden (zie tekstkader 1).

Tekstkader 1. Toelichting toegevoegd aan de AGREE-invulinstructie

Omdat de AGREE is opgesteld met klinische richtlijnen in de gezondheidszorg voor ogen moeten sommige woorden in een arbocontext opgevat worden.

- Lees waar ‘richtlijn’ staat, ‘arbokennisproduct’ of ‘dit document’.
- Lees waar ‘patiënt’ staat ‘patiënt en/of werknemer’. Arbokennisproducten hebben soms ook betrekking op mensen die zichzelf geen patiënt noemen of die nog geen klachten hebben.
- Lees i.p.v. ‘klinisch(e)’ ‘bedrijfsgeneeskundig(e)’ of ‘arbokundig’. Een ‘klinische vraag’ kan dan bijvoorbeeld gelezen worden als een ‘bedrijfsgeneeskundig probleem’ of ‘de ‘begeleidingsvraag die in deze arbozorg centraal staat’.
- Lees i.p.v. ‘gezondheid’ ‘functioneren’. ‘Gezondheidswinst’ mag opgevat worden als volwaardiger persoonlijk en sociaal functioneren (ook als dat van toepassing is in een werkcontext) en is dus breder dan alleen klachtenvrij zijn.
- Onder een ‘bijwerking’ kan in bedrijfsgeneeskundige context ‘een ongewenst neveneffect’ van een behandeling of aanpak verstaan worden. Dat ongewenste effect kan betrekking hebben op het lichamelijk, psychisch en het sociaal functioneren.

De AGREE wordt afgesloten met een 24e algemene vraag: Vindt u deze richtlijn aan te bevelen voor gebruik in de praktijk? Men kan vervolgens kiezen uit de antwoorden: ‘Sterk aan te bevelen’, ‘Aan te bevelen (onder voorwaarden of met veranderingen)’, ‘Onzeker’ en ‘Niet aan te bevelen’. Voor de lay-out van het antwoordformulier is het plezierig om dezelfde antwoordcategorieën te gebruiken. Daarom is deze vraag en de antwoordmogelijkheden anders geformuleerd, zodat de antwoordcategorieën ‘Zeer oneens’ – ‘Zeer eens’ toegepast kunnen worden. In het tekstkader 2 ziet u de toelichting.

Tekstkader 2. Toegevoegde toelichting

24: **Deze richtlijn is aan te bevelen voor gebruik in de praktijk.**
Zeer Eens 4 3 2 1 Zeer Oneens; Toelichting

Het algemene oordeel vereist dat de beoordelaar een oordeel geeft over de kwaliteit van de richtlijn waarbij elk beoordelingscriterium wordt meegenomen. Wat is uw algemene indruk van de kwaliteit van het arbokennisproduct en verwacht u positieve effecten van brede toepassing? In welke mate zijn de opstellers erin geslaagd zijn om de inhoud op een transparante en geloofwaardige manier voor het voetlicht te brengen: hebben ze u overtuigd?
Zeer eens = ‘Sterk aan te bevelen’
Eens = ‘Aan te bevelen (onder voorwaarden of met veranderingen)’
Oneens = ‘Onzeker’
Zeer Oneens = ‘Niet aan te bevelen’.

Elke beoordelaar krijgt verschillende arbokennisproducten toegestuurd met drie antwoordformulieren. In overleg met de beoordelaars krijgt men de stukken per post of per e-mail. Men krijgt nooit het product waaraan men zelf direct heeft meegewerkt of nauw bij betrokken is geweest, en men hoeft geen product van de ‘eigen club’ te beoordelen.
De beoordelaars krijgen twee weken de tijd om de antwoordformulieren in te vullen en op te sturen. Er zijn twee reminders verstuurd en de periode om in te sturen is verlengd tot een aantal weken. Het streven was om elk arbokennisproduct door minimaal drie personen te laten beoordelen. Na enkele afmeldingen was er zicht op waar lacunes dreigden. In overleg met enkele beoordelaars die nog niet ingestuurd hadden zijn toen de arbokennisproducten gewijzigd die zij moesten beoordelen. We hebben hen nieuwe arbokennisproducten toegezonden.
De laatste formulieren die verwerkt zijn ontvingen we begin november 2004.Uiteindelijk hebben we 85 ingevulde formulieren verwerkt: een respons van 73% van het aantal uitgezette antwoordformulieren plus arbokennisproducten (zie tabel 2. Tussen de groep die toegezegd heeft en de feitelijke respons zijn de verschillen klein).

De 85 formulieren zijn afkomstig van 30 verschillende beoordelaars: 27 van hen beoordeelde drie producten zoals de bedoeling was, drie personen beoordeelden er minder dan drie. Het merendeel van de arbokennisproducten is door vier verschillende personen beoordeeld (zie tabel 3).

Tabel 2. Respons

aantal %	*A gevraagd*		*B toegezegd*			*C respons (personen)*			
	n	%	n	%N	%A	n	%N	%A	%B
‘STECR’	37	*43*	12	*31*	*32*	8	*26*	*22*	*67*
‘NVAB’	19	*22*	10	*26*	*53*	9	*30*	*47*	*90*
‘Overig’	30	*35*	17	*44*	*57*	16	*53*	*53*	*94*
discipline geen bedrijfsarts	39	*45*	17	*44*	*44*	13	*44*	*33*	*76*
werkzaam bij kennisinstituten	24	*28*	9	*23*	*38*	8	*26*	*33*	*89*
wetenschap/universiteit	25	*29*	11	*29*	*44*	10	*32*	*40*	*90*
N =	86		39			30			

Tabel 3. Aantal beoordelingen per arbokennisproduct

Aantal verschillende beoordelaars	2	3	4	5	6
Aantal arbokennisproducten	1	4	15	1	1

2. 5 Gegevensbewerking en gegevensanalyse

Berekening antwoordscores

De AGREE-scores zijn per item gemiddeld (minimum 1,00-maximum 4,00). In dit verslag worden zowel de numerieke (1, 2, 3 en 4) als de woordelijke aanduiding (Zeer eens, Eens, Oneens en Zeer oneens) voor de itemscores gebruikt. In principe komt 1 overeen met Zeer oneens en 4 met Zeer eens, tenzij de vraag is omgepoold, maar dat staat dan vermeld.
Burgers heeft aangetoond dat de scores per domein een samenhangende schaal vormen en opgeteld mogen worden (Burgers 2002). We hebben die betrouwbaarheidsanalyse voor dit gegevensbestand herhaald. In bijlage 3 staan de Cronbach's alfa's voor interne consistentie van de domeinen vermeld. De interne consistentie van alle domeinschalen is goed tot zeer goed te noemen. In de handleiding voor de AGREE wordt aangegeven hoe de domeinscores omgerekend dienen te worden tot een waarde tussen 0 en 100. Men komt op 0 als alle beoordelaars alle items binnen dat domein minimaal scoren (1) en op 100 als alle items maximaal scoren (4).
In formule:

$$\text{Domeinscore} = \frac{\text{verkregen score - minimaal mogelijke score}}{\text{maximaal mogelijke score - minimaal mogelijk score}} \times 100$$

Als één of twee itemscores ontbreken, is een domeinscore berekend door aan te nemen dat het domein in dat geval uit één of twee items minder is samengesteld. Er kunnen geen specifieke normwaarden voor de AGREE-domeinscores gegeven worden. Maar we moeten ons realiseren dat als alle items van een domein 'Zeer oneens' gescoord worden, dat leidt tot een domeinscore van 0, alle items 'Oneens' leidt tot een domeinscore van 33,3; alle items met een score 'Eens' resulteren in een domeinscore van 66,7 en als alle items de hoogste score 'Zeer eens' krijgen wordt de domeinscore 100. Als we er van uit gaan dat elk van de criteria die in de AGREE-items verwoord worden van belang is en dus minimaal een score 'Eens' verdient, dan moeten we als minimale domeinscore 66,7 hanteren. Een lagere score betekent immers dat er op één van de items 'Oneens' of lager is gescoord. Omgekeerd geredeneerd kunnen we zeggen dat bij een score hoger dan 50 er minimaal bij één item minimaal 'Eens' gescoord is.

Indeling beoordelaars en 'afkomst' arbokennisproducten

Er is nagegaan of er samenhang bestaat tussen de AGREE-scores, de andere antwoorden en de achtergrond van de beoordelaars. Zij zijn daarom door de onderzoekers ingedeeld in diverse categorieën: bedrijfsarts of een andere discipline; werkzaam bij een universiteit in de directe nabijheid van onderzoek en wetenschap of meer in de praktijk; vooral geaffilieerd met STECR of meer met de NVAB. Het betreft niet altijd elkaar uitsluitende categorieën. Bedrijfsartsen kunnen bijvoorbeeld zowel binnen STECR als bij de richtlijnontwikkeling van de NVAB actief zijn. De indeling is tot stand gekomen op basis van de voorkennis van de onderzoekers over het zwaartepunt in werkzaamheden of affiliatie van de beoordelaars. In een enkel geval hebben we de beoordelaars zelf om bevestiging van onze indeling gevraagd. Het lijkt derhalve verstandig om de analyses die met behulp van deze indelingen van beoordelaars gemaakt worden, behoedzaam te interpreteren.

De arbokennisproducten zijn ingedeeld op hun afkomst: 'STECR', 'NVAB' en 'Overige'. De Landelijke Eerstelijns Samenwerkingsafspraak Overspanning, samengesteld door de NVAB en het NHG is onder 'NVAB' geschaard.

Interbeoordelaarsvariatie en consistentie van beoordelingen

De verschillende arbokennisproducten zijn door steeds andere combinaties (en aantallen) van beoordelaars van een score voorzien zijn. Vanwege de brede range aan soorten arbokennisproducten hebben we daarom onderzocht of de AGREE goed vergelijkbare resultaten oplevert als deze aan verschillende beoordelaars wordt voorgelegd. Komen de beoordelaars bij één en hetzelfde arbokennisproduct wel tot consistente of overeenkomstige beoordeling? Het meetniveau van de vierpuntsschaal lopend van 'Zeer oneens' tot 'Zeer eens' in combinatie met de inhoud van de verschillende AGREE-items is ordinaal, ook al worden de itemscores soms op intervalniveau behandeld, bijvoorbeeld als er gemiddelden berekend worden.
Verschil van mening op itemniveau kan op diverse manieren geïnterpreteerd worden.

- *Wel of niet voldaan aan criterium*

Beschouwt men de itemscore op nominaal niveau (het document vermeldt informatie over het betreffende criterium of niet) dan is een zo hoog mogelijke graad van overeenstemming tussen de beoordelaars wenselijk. Gebrek aan overeenstemming tussen beoordelaars moet uitgelegd worden als een mogelijke foutenbron. Het kwaliteitsoordeel gebaseerd op de AGREE wordt verzwakt als beoordelaars het op zo'n feitelijk niveau met elkaar oneens zijn. Om dat in kaart te brengen, worden de scores 'Zeer oneens' en 'Oneens' omgescoord tot ''Voldoet niet' en de scores 'Eens' en 'Zeer eens' tot 'Voldoet wel'. Vervolgens wordt in kaart gebracht hoe vaak er overeenstemming is tussen de beoordelaars.
Er worden percentages overeenstemming per arbokennisproduct (over alle items) berekend. De uitkomsten worden gemiddeld en er wordt een standaarddeviatie berekend. Percentages die meer dan één standaarddeviatie van het gemiddelde verwijderd liggen worden opvallend hoog of laag genoemd. Er zijn percentages overeenstemming berekend per item (over alle arbokennisproducten). Na berekening van een gemiddelde en standaarddeviatie kunnen items geïdentificeerd worden met percentages die meer dan één standaarddeviatie van het gemiddelde verwijderd liggen. Bij die arbokennisproducten is dan sprake van opvallend veel of juist opvallend weinig verschil van mening.

- *Verschil van waardering*

De AGREE-scores bij de diverse items kunnen ook als waardeoordelen worden opgevat (Stemler 2004). Het is dan minder belangrijk of de beoordelaars exact op een zelfde score uitkomen. Bij de genoemde variatie aan achtergronden van de beoordelaars is dat immers onwaarschijnlijk. Maar als de range (het verschil tussen hoogste en laagste score) laag is, verhoogt dat de geloofwaardigheid van de uitkomst en het kwaliteitsoordeel dat op die uitkomsten gebaseerd wordt.
Het verschil tussen hoogste en laagste beoordeling per item bedraagt 0, 1, 2 of 3. We hebben een verschil van 1 categorie acceptabel genoemd en vervolgens opgeteld hoe vaak de range '2' en '3' voorkomt per arbokennisproduct over alle items: hoe vaak komen er bij dat arbokennisproduct sterk uiteenlopende scores voor? Die waardes worden gemiddeld over alle 22 arbokennisproducten en er wordt een standaarddeviatie berekend. Waarden die meer dan één standaarddeviatie van het gemiddelde verwijderd liggen worden opvallend hoog of laag genoemd. Let wel: Het antwoord kan geïnterpreteerd worden als kenmerk van het arbokennisproduct (namelijk: zorgend voor veel verschil van mening tussen de beoordelaars) maar ook als eigenschap van de AGREE in deze context.

Tevens is op dezelfde manier de range bepaald gemiddeld per item over alle arbokennisproducten. Die waarden zijn gemiddeld over alle 23 items en er is een standaarddeviatie berekend. Waarden die meer dan één standaarddeviatie verwijderd liggen van het gemiddelde worden opvallend hoog of laag genoemd. Het antwoord geeft informatie over de toepasbaarheid van de AGREE in deze context. Omdat de kans op ongelijke scores bij het toenemen van het aantal beoordelaars onevenredig snel oploopt, kunnen de frequenties bij een ongelijk aantal beoordelaars niet goed vergeleken worden.

- *Overeenstemming en consistentie in oordelen*

The AGREE Collaboration beveelt aan om elke richtlijn door minimaal twee, maar bij voorkeur door vier beoordelaars te laten scoren. Dat advies is gebaseerd op analyses van de interbeoordelaarsbetrouwbaarheid van de item- en domeinscores. De aanname is dat bij meer beoordelaars de uiteindelijke gemiddelde score de 'ware' score dichter zal benaderen. De mate waarin de gemiddelde score van verschillende beoordelaars een juiste schatting is van de 'ware' score wordt uitgedrukt in de intraclass correlatie coëfficiënt (ICC). In die coëfficiënt wordt de mate waarin variabelen in waarde en variantie gelijk op lopen uitgedrukt in een getal tussen 0 en 1. Een waarde hoger dan 0,75 is zeer goed, 0,6-0,75 is goed, 0,4-0,6 is matig en lager dan 0,4 is slecht (MacDermid 2002). Een lage ICC

betekent dat de gemiddelde score een slechte voorspeller is van de 'ware' score omdat beoordelaars niet op een consistente manier een AGREE-score toekennen aan een arbokennisproduct.

- *Gemiddelde standaarddeviaties per item en domein*

We kunnen de scores van de beoordelaars ook als een waarderingscijfer zien. Het is dan niet van belang of de score een feitelijke kwaliteit of eigenschap van het arbokennisproduct uitdrukt. De gemiddelde of opgetelde waardering van de beoordelaars vormt dan het gemiddelde rapportcijfer waar we naar op zoek zijn. Dat beoordelaars niet allemaal dezelfde waardering voor bepaalde aspecten van een arbokennisproduct hebben is logisch, acceptabel en niet problematisch, maar een vanzelfsprekend gevolg van de verschillende achtergrond van de beoordelaars. Informatie over de spreiding in scores is in beschrijvende zin informatief en wordt weergegeven door middel van de standaarddeviatie per item en per arbokennisproduct en per domein per arbokennisproduct.

- Gemiddelde standaarddeviatie per item

De gemiddelde itemstandaarddeviatie per arbokennisproduct vertelt ons hoe gespreid de scores voor dat arbokennisproduct zijn. Als die waarden gemiddeld worden en er over alle 22 arbokennisproducten een standaarddeviatie berekend wordt, kunnen waarden die meer dan één standaarddeviatie van het gemiddelde verwijderd liggen aangeduid worden als opvallend hoog of laag.
Tevens kunnen de itemstandaarddeviaties gemiddeld worden per item over alle arbokennisproducten: hoe gespreid per arbokennisproduct zijn de scores voor dit item?
Die gemiddelde itemstandaarddeviaties worden (nogmaals) gemiddeld over alle 23 items en er wordt een standaarddeviatie berekend. Waarden die meer dan één standaarddeviatie verwijderd liggen van het gemiddelde worden opvallend hoog of laag genoemd.

- Gemiddelde standaarddeviatie per domein

De gemiddelde domeinstandaarddeviatie per arbokennisproduct vertelt ons hoe groot de spreiding is in de domeinscores per arbokennisproduct.
Om te bepalen welk domein voor veel spreiding veroorzaakt wordt die domeinstandaarddeviatie vergeleken met het gemiddelde van standaarddeviaties per domein over alle arbokennisproducten.
Om te bepalen bij welk arbokennisproduct er veel variatie optreedt in de domeinscores, wordt de domeinstandaarddeviatie vergeleken met het gemiddelde van de standaarddeviaties per arbokennisproduct gemiddeld over alle domeinen van dat kennisproduct.

<u>Aanvullende schaalconstructie</u>

De antwoordscores voor de overige vragen zijn per item gemiddeld (minimum 1,00- maximum 4,00). De vier items die over de kenmerken van het voorliggende arbokennisproduct gingen, bleken onderling sterk samen te hangen en als schaal behandeld te kunnen worden (Cronbach's alfa: 0,83). Het betrof de items:

- *Dit document bevat specifieke, expliciete uitspraken voor goed handelen bij een welomschreven probleem of doelpopulatie (aanbevelingen).*
- *Met dit document kan in principe het handelen van veel arboprofessionals beïnvloed worden (bedoeld voor brede verspreiding).*
- *De organisatie die deze tekst heeft ontwikkeld of uitgegeven, heeft het vertrouwen van het merendeel van de bedrijfsartsen en andere arboprofessionals in de praktijk. ('zij maken inhoudelijk goede producten')*
- *Dit document zal door het merendeel van de praktijk gezien worden als een richtinggevend document op het terrein van de bedrijfsgeneeskunde en/of arbozorg. (los van eventuele titel, inleiding, of omschreven bedoeling)*

We hebben deze schaal met een knipoog '<u>Richtlijnigheid</u>' genoemd. De gemiddelde score (1-4) brengt tot uitdrukking in welke mate het arbokennisproduct door de beoordelaars als een richtlijn gezien wordt, los van de bedoelingen of aanduidingen van de samenstellers. De minimumscore per arbokennisproduct is 2,44 en de maximumscore 3,94.

De drie items over het gemak waarmee de AGREE ingevuld kon worden, zijn tot de schaal '<u>Gebruiksgemak AGREE</u>' samengevoegd (Cronbach's alfa: 0,53).

- *Mijn deskundigheid op het inhoudelijke terrein van deze tekst was voldoende om de AGREE-vragen te kunnen beantwoorden.*
- *Het beantwoorden van de AGREE-vragenlijst voor deze tekst was eenvoudig.*

- *De informatie die ik nodig had voor het beantwoorden van de AGREE-vragen was goed te vinden in dit document (of de bijgevoegde inleiding.)*

De minimumscore per arbokennisproduct is 2,67 en de maximumscore 3,67.

Acht items waarin naar de waarde van de AGREE gevraagd werd hangen onderling samen en kunnen als een aanduiding van de gebruikswaarde van de AGREE gezien worden. Wat zou het toepassen van de AGREE bij dit arbokennisproduct op kunnen leveren? Het gaat om de items:

- *Voor mij is de AGREE vragenlijst een zinvol instrument om de kwaliteit dit arbokennisproduct te beoordelen.*
- *Het is hinderlijk dat de AGREE-vragen onvoldoende zijn toegespitst op de arbocontext van dit specifieke kennisproduct. (omgepoold)*
- *De AGREE is een te gedetailleerd instrument in verhouding tot de bedoeling en reikwijdte van dit arbokennisproduct .(omgepoold)*
- *Omdat dit document niet als 'richtlijn' bedoeld is, zijn de vragen van de AGREE niet relevant. (omgepoold)*
- *Een kwaliteitsoordeel over dit arbokennisproduct kan zonder de AGREE eenvoudiger en net zo goed tot stand komen; (omgepoold)*
- *Als dit arbokennisproduct verspreid wordt onder veel arboprofessionals is het wenselijk dat het voorzien wordt van een onafhankelijk keurmerk (bijvoorbeeld gebaseerd op de AGREE-beoordeling).*
- *Als een bedrijfsarts of arboprofessional weet in welke mate een aanbeveling in dit arbokennisproduct gebaseerd is op wetenschappelijk bewijs dan zal dat invloed hebben op de ruimte die hij/zij professioneel neemt om in de praktijk van deze aanbeveling af te wijken.*
- *De vragen van de AGREE zouden arboprofessionals en bedrijfsartsen in de praktijk helpen bij het zelf beoordelen van de inhoud van dit arbokennisproduct.*

Deze schaal hebben we 'Waarde AGREE' genoemd. (Cronbach's alfa 0,74). De minimumscore per arbokennisproduct is 2,33 en de maximumscore 3,69.

Statistische analyse

We behandelen de basale itemscores steeds op ordinaal niveau, de gemiddelde itemscores, de domein- en schaalscores op intervalniveau. Systematisch verschillen in scores en antwoorden op domeinniveau en schaalniveau worden getoetst met one way anova ('compare means'-procedure van SPSS 10.0.7). Verschillen tussen subgroepen op itemniveau worden nonparametrisch getoetst met behulp van de Kruskal-Wallis H toets. Bij het bepalen van correlaties tussen variabelen is rekening gehouden met het ordinale of intervalniveau van de maten die in beschouwing worden genomen. Dat betekent dat er soms sprake is van de Pearsons correlatie coëfficiënt en soms van Kendals Tau β. Intraclass Correlatie Coëfficiënten worden per item en domein berekend op basis van de mate van overeenstemming over verschillende arbokennisproducten. De ICC voor twee beoordelaars is berekend op basis van de eerste twee ingevoerde beoordelingen van alle 22 arbokennisproducten. Informatie van beoordelingen die als derde of meer zijn ingevoerd, zijn niet gebruikt en dus is 48% van de beoordelingen buiten beschouwing gelaten. De ICC voor drie beoordelaars is berekend op basis van de eerste drie ingevoerde beoordelingen van 21 arbokennisproducten (dus uitgezonderd de NVAB-richtlijn Psychische Klachten). Informatie van beoordelingen die als vierde of meer zijn ingevoerd, zijn niet gebruikt. In totaal is dus 26% van de beoordelingen buiten beschouwing gelaten. De ICC voor vier beoordelaars is berekend op basis van de eerste vier ingevoerde beoordelingen van 17 arbokennisproducten. (dus zonder de gegevens van de STECR-werkwijzer Lage Rugklachten en Arbeidsomstandigheden; de NVAB-richtlijn Psychische Klachten; de UWV/Lisv standaard Geen Duurzaam Benutbare Mogelijkheden, het hoofdstuk Diabetes uit het Handboek Arbeid en Belastbaarheid en de SMB-richtlijn Rugklachten van Arbouw). Tevens is informatie beoordelingen die als vijfde en zesde zijn ingevoerd niet gebruikt. In totaal is 15% van de beoordelingen niet gebruikt.
ICC voor vijf en zes beoordelaars waren niet mogelijk omdat er maar twee arbokennisproducten waren die door meer dan vier beoordelaars zijn gescoord.
In de regel wordt getoetst op $p<0,05$, tenzij dat anders wordt vermeld.

3. Resultaten

3.1 De AGREE-scores van de arbokennisproducten

68% van de arbokennisproducten werd door vier verschillende personen van een oordeel voorzien, 9% werd door meer dan vier personen beoordeeld, 23% door minder personen (minimaal twee beoordelaars). Ruim 90% van de formulieren was volledig ingevuld (zie bijlage 4).

In tabel 4 staan de item- en domeinscores gemiddeld over 22 verschillende arbokennisproducten. De gemiddelde itemscores lopen flink uiteen. Of voorkeuren en het perspectief van patiënten zijn nagegaan, wordt kennelijk zelden vermeld. Dat item (nr.5) scoort laag: 2,15. Ook de criteria voor het selecteren van wetenschappelijk bewijsmateriaal zijn volgens de beoordelaars vaak niet duidelijk beschreven: 2,06. Regelmatig ontbreekt informatie over de herzieningsprocedure zodat niet duidelijk is wat de 'houdbaarheid' van de gepresenteerde kennis is: 1,90. Kostenimplicaties worden niet vaak vermeld (2,07). Conflicterende belangen van de werkgroepleden blijven vaak onvermeld (1,88).
Het doel van het arbokennisproduct wordt meestal wel goed beschreven (3,18) en de beoogde gebruikers heeft men goed voor ogen (3,29). Het domein 'Onderwerp en Doel' scoort van de zes domeinen het hoogste (71,49); 'Toepassing' het laagste (35,31).

In bijlage 5 staan de gemiddelde itemscores per arbokennisproduct vermeld. De interne consistentie van de antwoordscores is hoog. Cronbach's alfa variëert van 0,85 tot 0,90 per domein. In bijlage 3 staan de waarden voor Cronbach's alfa per domein vermeld. De resultaten zijn redelijk vergelijkbaar met de waarden die Burgers vond (Burgers 2002).

De AGREE-domeinscores per arbokennisproduct staan in tabel 5.
De domeinscores van de STECR-producten variëren tussen 11,1 en 70,4 per domein. Dat minimum en maximum komen overigens bij dezelfde werkwijzer voor: Lage Rugklachten en Arbeidsomstandigheden! Alle STECR-producten scoren laag op 'Methodologie' en redelijk op 'Helderheid en Presentatie'. Het domein 'Toepassing' scoort over de hele linie vrij laag. De onafhankelijk van de opstellers krijgt alleen in het geval van de werkwijzer 'Werkstress zorg en onderwijs' een redelijke score.
Van alle 48 STECR-domeinscores is 31,3% hoger dan 50.

De domeinscores van de NVAB-producten liggen tussen de 36,1 en 100. Het domein 'Onderwerp en Doel' krijgt bij alle NVAB-producten een hoge score. De domeinscore 'Methodologie' varieert van 40,5 tot 78,6. Duidelijk is dat de laatste uitgebrachte NVAB-Richtlijnen (Astma en COPD en KASN) hierop het hoogste scoren. De 'Helderheid en Presentatie' krijgt hoge scores. De score voor de richtlijn Psychische Klachten en de LESA-Overspanning is het laagst. De scores voor 'Toepassing' variëren sterk: van laag (KASN en de LESA Overspanning) tot hoog (Leidraad Verwijzen door de bedrijfsarts).
Van alle 36 NVAB-domeinscores is 83,3% hoger dan 50.

De domeinscores van de overige arbokennisproducten varieert van 0 tot 96,3. 'Onderwerp en Doel' van beide standaarden van het UWV zijn hoog, evenals die domeinscore voor de Registratierichtlijnen van het NCvB. De domeinscores 'Methodologie' zijn over de hele linie vrij laag, die voor het hoofdstuk Diabetes uit het Handboek Arbeid en Belastbaarheid redelijk.

Opvallend laag zijn de domeinscores 'Toepassing' voor de Interventies Acute Lage rugklachten van het NKAB en de SMB Rugklachten van Arbouw. De domeinscore 'Onafhankelijk Opstellers' voor alle overige arbokennisproducten is vrij laag, de registratierichtlijnen van het NCvB vormen een uitzondering met een redelijke score van 55,6.
Van alle 48 'Overige'-domeinscores is 37,5% hoger dan 50.

Tabel 4. AGREE scores van de 22 arbokennisproducten

	Domein	*gemiddel de score*	*min*	*max*	*SD*
	Onderwerp en Doel	**71,49**	**27,78**	**100,00**	**20,16**
1	*Het doel van de richtlijn is specifiek beschreven.*	3,18	1,75	4,00	0,64
2	*De klinische vraag/vragen die in de richtlijn aan de orde komt/komen, is/zijn specifiek.*	2,93	1,33	4,00	0,71
3	*De patiëntenpopulatie waarop de richtlijn van toepassing is, is specifiek beschreven.*	3,22	1,50	4,00	0,77
	Betrokkenheid Belanghebbenden	**50,46**	**2,08**	**81,25**	**17,27**
4	*De leden van de werkgroep die de richtlijn heeft ontwikkeld, komen uit alle relevante beroepsgroepen.*	2,80	1,00	4,00	0,72
5	*Het perspectief en de voorkeuren van patiënten zijn nagegaan.*	2,15	1,00	3,75	0,64
6	*De beoogde gebruikers van de richtlijn zijn duidelijk benoemd.*	3,29	1,00	4,00	0,71
7	*De richtlijn is getest onder de beoogde gebruikers.*	2,29	1,00	4,00	0,88
	Methodologie	**40,60**	**11,11**	**78,57**	**17,43**
8	*Er zijn systematische methoden gebruikt voor het zoeken naar wetenschappelijk bewijsmateriaal.*	2,33	1,50	3,75	0,62
9	*De criteria voor het selecteren van het wetenschappelijk bewijsmateriaal zijn duidelijk beschreven.*	2,06	1,00	3,75	0,68
10	*De gebruikte methoden om de aanbevelingen op te stellen, zijn duidelijk beschreven.*	2,46	1,25	3,75	0,62
11	*Gezondheidswinst, bijwerkingen en risico's zijn overwogen bij het opstellen van de aanbevelingen.*	2,66	1,00	3,50	0,61
12	*Er bestaat een expliciet verband tussen de aanbevelingen en het onderliggende wetenschappelijke bewijsmateriaal.*	2,51	1,00	3,75	0,65
13	*De richtlijn is vóór publicatie door externe experts beoordeeld.*	2,77	1,00	4,00	0,75
14	*Een procedure voor herziening van de richtlijn is vermeld.*	1,90	1,00	3,25	0,50
	Helderheid en Presentatie	**61,66**	**41,67**	**93,75**	**13,78**
15	*De aanbevelingen zijn specifiek en ondubbelzinnig.*	2,88	1,67	4,00	0,66
16	*De verschillende beleidsopties zijn duidelijk vermeld.*	2,85	1,75	4,00	0,55
17	*De kernaanbevelingen zijn gemakkelijk te herkennen.*	3,11	1,67	4,00	0,58
18	*De toepassing van de richtlijn wordt ondersteund met hulpmiddelen.*	2,48	1,25	4,00	0,71
	Toepassing	**35,31**	**0,00**	**86,11**	**19,90**
19	*De mogelijke organisatorische belemmeringen bij het toepassen van de aanbevelingen zijn besproken.*	2,53	1,00	4,00	0,68
20	*De mogelijke kostenimplicaties van het toepassen van de aanbevelingen zijn overwogen.*	2,07	1,00	3,50	0,60
21	*De richtlijn geeft de belangrijkste criteria om na te gaan en te toetsen of de richtlijn wordt gevolgd.*	2,31	1,00	4,00	0,76
	Onafhankelijkheid Opstellers	**37,01**	**11,11**	**55,56**	**15,16**
22	*De richtlijn is niet beïnvloed door de opvattingen of belangen van de financierende instantie.*	2,86	1,67	4,00	0,71
23	*Conflicterende belangen van leden van de werkgroep zijn vastgelegd.*	1,88	1,00	3,00	0,51
24	*Deze richtlijn is aan te bevelen voor gebruik in de praktijk.*	3,06	2,00	4,00	0,64

Tabel 5. AGREE domeinscores per arbokennisproduct

	N = aantal beoordelaars	Onderwerp en Doel	Betrokkenheid Belanghebbenden	Methodologie	Helderheid Presentatie	Toepassing	Onafhankelijkheid Opstellers	N
	STECR							
1	ABBE-rsi: STECR-werkwijzer	38,9	33,3	31,0	41,7	25,0	29,2	2
2	Arbeidsconflicten; STECR-werkwijzer	67,8	56,7	42,9	60,0	44,4	43,3	5
3	Aanpak Problematisch Verzuim; STECR-werkwijzer	37,5	45,8	15,5	54,2	27,8	25,0	4
4	Lage Rugklachten en Arbeidsomstandigheden; STECR-werkwijzer	70,4	52,8	44,4	63,9	33,3	11,1	3
5	Werkstress werknemer Gezondheidszorg /Onderwijs; STECR-werkwijzer	63,9	60,4	40,5	64,6	41,7	50,0	4
6	Cultuur of Persoon; Sociaal-medische Begeleiding Allochtonen; STECR-Assist	66,7	48,6	22,2	54,2	36,1	29,2	3
7	Vrouwen in de WAO + Vrouwen níet in de WAO; Voorstudie en STECR–Assist	63,9	41,7	21,4	47,9	13,9	41,7	4
8	Somatisatie, multidisciplinaire aanpak bij somatoforme verschijnselen door arboprofessionals; STECR-werkwijzer	55,6	47,9	36,9	52,1	47,2	29,2	3
	NVAB							
9	Handelen van de bedrijfsarts bij werknemers met lage rugklachten; NVAB-richtlijn	94,4	58,3	40,5	89,6	44,4	50,0	4
10	Handelen van de bedrijfsarts bij werknemers met psychische klachten; NVAB-richtlijn	100,0	70,8	59,5	62,5	66,7	50,0	2
11	Handelen van de bedrijfsarts bij werknemers met Astma en COPD; NVAB-richtlijn	91,7	72,9	78,6	77,1	61,1	50,0	4
12	Handelen van de bedrijfsarts bij werknemers met klachten aan arm, schouder of nek; NVAB-richtlijn	86,1	60,4	69,1	77,1	36,1	41,7	4
13	Handelen van de bedrijfsarts bij het verwijzen van werknemers; NVAB Leidraad	86,1	81,3	58,3	93,8	86,1	62,5	4
14	Overspanning; Landelijke Eerstelijns Samenwerkingsafspraak NHG/NVAB	74,1	58,3	43,7	56,9	38,9	66,7	6
	Overige							
15	Geen Duurzaam Benutbare Mogelijkheden; Lisv/UWV Standaard	96,3	58,3	11,1	61,1	40,7	22,2	3
16	Zwangerschap en bevalling als oorzaak van ongeschiktheid voor haar arbeid; Lisv/UWV Standaard	91,7	60,4	39,3	41,7	25,0	20,8	4
17	Beroepsgebonden burnout/overspanning; voorwoord en registratierichtlijn; NCvB	91,7	33,3	38,8	58,3	27,8	55,6	4
18	Interventies bij acute lage rugklachten; van de kennisbank http://www.kenniscentrumakb.nl; NKAB	27,8	02,1	25,0	47,9	00,0	20,8	4
19	Klachtenreductie en stressmanagement, Arbeidshulpverlening bij psychische klachten; Inleiding en protocol; NKAP	69,4	60,4	48,8	77,1	22,2	37,5	4
20	Diabetes mellitus; Hoofdstuk uit het Handboek Arbeid en Belastbaarheid; BSL	59,3	41,7	60,0	58,3	40,7	22,2	3
21	PAGO voor huisvuilbeladers; Artikel uit TBV)	80,6	31,3	45,2	58,3	13,9	33,3	4
22	Sociaal Medische Begeleiding van Werknemers in de Bouw; Introductie en richtlijn rugklachten	59,3	33,3	20,6	58,3	03,7	22,2	3
	Gemiddeld over 22 arbokennisproducten:	71,49	50,46	40,60	61,66	35,31	37,01	3,7

Samenvattend

22 arbokennisproducten beoordeeld met de AGREE op zes verschillende domeinen leveren 132 domeinscores op tussen 0 en 100. Hoger dan 66,7 is 21,2% van dat aantal domeinscores en tussen 50 en 66,7 ligt 28,0% van de scores.
Het domein 'Onderwerp en Doel' scoort het hoogste (14 van de 22 van de arbokennisproducten hoger dan 66,7). Het domein 'Toepassing' scoort het laagste (19 van de 22 arbokennisproducten lager dan 50). Voor 'Methodologie' scoren slechts drie arbokennisproducten hoger dan 66,7; 17 scoren lager dan 50!

13 van de 22 arbokennisproducten scoren gemiddeld hoger dan 3,00 als het gaat om de algemene vraag of deze 'richtlijn' is aan te bevelen voor de praktijk (zie ook bijlage 6). De STECR-Werkwijzer 'Lage Rugklachten en Arbeidsomstandigheden' en de werkwijzer 'ABBE–rsi' scoren gemiddeld 2,00. Dat betekent dat nagenoeg geen van de beoordelaars de algemene vraag voor deze arbokennis-producten positief heeft beantwoord.

Alle AGREE domeinscores correleren hoog met deze algemene vraag of de richtlijn aanbevolen moet worden voor de praktijk (tabel 6).

Tabel 6. Correlaties AGREE-domeinscores met algemene vraag

Deze richtlijn is aan te bevelen voor gebruik in de praktijk	Kendall's tau b
Onderwerp en Doel	0,532**
Betrokkenheid Belanghebbenden	0,499**
Methodologie	0,432**
Helderheid en Presentatie	0,493**
Toepassing	0,386*
Onafhankelijkheid Opstellers	0,530**

*: $p < 0,05$; **: $p < 0,01$

In tabel 7 ziet u de gemiddelde AGREE-scores van de producten van NVAB, STECR en Overige vergeleken. De gemiddelde domeinscores van de NVAB-producten vallen hoger uit dan de gemiddel-de domeinscores van de STECR-producten.

Tabel 7. Verschillen domeinscores per organisatie arbokennisproduct

Organisatie AKP	NVAB	STECR	Overige	Totaal
Doel en Onderwerp*	88,8	58,1	72,0	71,5
Betrokkenheid Belanghebbenden**	67,0	48,4	40,1	50,5
Methodologie**	58,3	31,8	36,1	40,6
Helderheid en Presentatie**	76,2	54,8	57,6	61,7
Toepassing**	55,6	33,7	21,8	35,3
Onafhankelijkheid Opstellers**	53,5	32,3	29,3	37,0

Verschillen tussen groepen significant; ** $p<0,01$, * p: $< 0,05$

Tussen de AGREE scores voor de arbokennisproducten en de achtergrondkenmerken van de beoor-delaars kan in het algemeen geen verband worden aangetoond. Werkzaam in wetenschap of in de praktijk, bedrijfsarts of een andere discipline, uit NVAB-kring of niet: geen enkel verschil in gemiddelde AGREE-scores blijkt significant. Met één uitzondering: de scores toegekend door niet-NVAB-mensen op het domein 'Onafhankelijkheid Opstellers' zijn significant hoger dan de anderen 41,96 versus 30,13; $p<0,05$.

Ook als we de beoordelaars indelen op NVAB, STECR of overig, blijkt het verschil in gemiddelde AGREE-score voor twee domeinen significant. De gemiddelde scores voor 'Onderwerp en Doel' zijn bij de STECR-beoordelaars wat hoger, 'Onafhankelijkheid Opstellers' scoort gemiddeld bij de NVAB lager (zie tabel 8).

Tabel 8. Verschillen domeinscores per organisatie beoordelaar

Organisatie beoordelaar	NVAB	STECR	Overige	Totaal
Doel en Onderwerp*	60,0	84,3	67,4	70,5
Betrokkenheid Belanghebbenden	37,2	52,3	52,3	49,9
Methodologie	29,0	46,3	41,5	41,0
Helderheid en Presentatie	48,8	66,3	63,4	61,7
Toepassing	25,2	30,7	39,4	34,7
Onafhankelijkheid Opstellers*	22,2	39,4	41,7	38,2

*: p<0,05

3.2 Overeenstemming tussen beoordelaars

In totaal zijn er 506 itemscores verzameld. 80 Maal (15,8%) waren alle beoordelaars het exact met elkaar eens. In 53,8% van de gevallen bedroeg het verschil maximaal 1. Bij 16,2% van de itemscores was het verschil maximaal: in dat geval was er zowel een beoordelaar die 'Zeer eens' had gescoord, als iemand die 'Zeer oneens' had ingevuld. Zie bijlage 7.

Overeenstemming en spreiding per arbokennisproduct

De beoordelaars zijn het regelmatig oneens op een basaal niveau: voldoet het arbokennisproduct naar hun mening nu wel ('Zeer eens' en 'Eens') of niet ('Oneens' en 'Zeer oneens') aan het gestelde AGREE-criterium. In tabel 9, kolom 1 ziet u hoe vaak er per arbokennisproduct op dat niveau geen overeenstemming was. Bij de STECR-Werkwijzer 'Somatisatie' zijn de beoordelaars het bij 22 van 24 items oneens. Bij de LESA-Overspanning bij 18 items, bij de STECR werkwijzer 'ABBE-rsi' bij 20 items.
Gemiddeld is er bij 13,6 item per arbokennisproduct geen overeenstemming op dit niveau.

In bijlage 7 vindt u hoe vaak een maximale range van 1, 2 of 3 voor de items per arbokennisproduct voorkwam. Een maximale range van 2 of 3 komt relatief vaak voor bij de itemscores van de STECR-Werkwijzer ABBE-rsi, de STECR-werkwijzer Problematisch Verzuim en de STECR-Werkwijzer Somatisatie. Ook bij de LESA-Overspanning en het protocol arbeidshulpverlening van het NKAP komt de range van 2 of 3 relatief vaak voor.

De gemiddelde standaarddeviatie geeft een indruk van de spreiding in oordelen bij een bepaald arbokennisproduct. In tabel 9 in kolom 2 (itemniveau) en kolom 3 (domeinniveau) ziet u welke arbokennisproducten er uitspringen. Met een '+' zijn arbokennisproducten gemarkeerd waar de spreiding erg laag was, met een '-' de arbokennisproducten met een relatief hoge spreiding. De NVAB-Richtlijn 'Psychische klachten' is slechts door twee beoordelaars bekeken en zij waren het vaak eens, maar als zij het oneens waren, verschilden de scores maximaal!

Overeenstemming en spreiding per item en domein

In tabel 10 ziet welke items voor veel spreiding verantwoordelijk waren of juist heel weinig. In kolom 1 wordt getoond of de beoordelaars het op een basaal niveau niet eens waren ('Zeer eens' samengenomen met 'Eens' en 'Zeer oneens' met 'Oneens'). Dat varieert van 4 tot 17 keer, gemiddeld: 12,46. Item 2, 10, 11 en 18 liggen flink boven het gemiddelde. 3, 9, 14 en 23 daaronder.

Kijken we naar de gemiddelde standaarddeviatie (van 0,47 tot 0,96, gemiddeld 0,75), dan vallen dezelfde items 2, 11, 18 op met een relatief hoge gemiddelde standaarddeviatie en item 3, 14 en 23 met een lage gemiddelde standaarddeviatie.

Tabel 9. Overeenstemming en spreiding AGREE scores Arbokennisproducten

		voldoet wel/ voldoet niet	*gemiddelde SD Item niveau*	*gemiddelde SD Domein niveau*
1	ABBE-rsi: STECR-werkwijzer	-	-	-
2	Arbeidsconflicten; STECR-werkwijzer			
3	Aanpak Problematisch Verzuim; STECR-werkwijzer			
4	Lage Rugklachten en Arbeidsomstandigheden; STECR-werkwijzer		-	
5	Reductie van werkstress door de werknemer in de Gezondheidszorg /Onderwijs; STECR-werkwijzer			
6	Cultuur of Persoon; Sociaal-medische Begeleiding Allochtonen; STECR-Assist			-
7	Vrouwen in de WAO + Vrouwen níet in de WAO; Voorstudie en STECR–Assist			
8	Somatisatie, multidisciplinaire aanpak bij somatoforme verschijnselen door arboprofessionals; STECR-werkwijzer	-	-	-
9	Handelen van de bedrijfsarts bij werknemers met lage rugklachten; NVAB-richtlijn	+	+	+
10	Handelen van de bedrijfsarts bij werknemers met psychische klachten; NVAB-richtlijn		-	-
11	Handelen van de bedrijfsarts bij werknemers met Astma en COPD; NVAB-richtlijn			
12	Handelen van de bedrijfsarts bij werknemers met klachten aan arm, schouder of nek; NVAB-richtlijn			
13	Handelen van de bedrijfsarts bij het verwijzen van werknemers; NVAB Leidraad	+	+	
14	Overspanning; Landelijke Eerstelijns Samenwerkingsafspraak NHG/NVAB	-		
15	Geen Duurzaam Benutbare Mogelijkheden; Lisv/UWV Standaard	+	+	+
16	Zwangerschap en bevalling als oorzaak van ongeschiktheid voor haar arbeid; Lisv/UWV Standaard			
17	Beroepsgebonden burnout/overspanning; voorwoord en registratierichtlijn; NCvB	+		
18	Interventies bij acute lage rugklachten; van de kennisbank http://www.kenniscentrumakb.nl; NKAB		+	
19	Klachtenreductie en stressmanagement, Arbeidshulpverlening bij psychische klachten; Inleiding en protocol; NKAP			
20	Diabetes Mellitus; Hoofdstuk uit het Handboek Arbeid en Belastbaarheid; BSL			-
21	PAGO voor huisvuilbeladers; (artikel uit TBV)			
22	Sociaal Medische Begeleiding van Werknemers in de Bouw; Introductie en richtlijn rugklachten			
	Gemiddeld:	13,59	0,75	18,42
	Ondergrens - SD	9,40	0,58	14,72
	Bovengrens + SD	17,78	0,92	22,12

Tabel 10. Spreiding AGREE scores per item

	Kolom 1 en 2: + = lage spreiding: waarde lager dan gemiddelde – standaarddeviatie - = hoge spreiding: waarde hoger dan gemiddelde + standaarddeviatie kolom 3: +: ICC4: > 0,6 -: ICC4 < 0,6	voldoet wel/ voldoet niet	SD	ICC
	Onderwerp en Doel			
1	*Het doel van de richtlijn is specifiek beschreven.*			-
2	*De klinische vraag/vragen die in de richtlijn aan de orde komt/komen, is/zijn specifiek.*	-	-	-
3	*De patiëntenpopulatie waarop de richtlijn van toepassing is, is specifiek beschreven.*	+	+	+
	Betrokkenheid Belanghebbenden			
4	*De leden van de werkgroep die de richtlijn heeft ontwikkeld, komen uit alle relevante beroepsgroepen.*			+
5	*Het perspectief en de voorkeuren van patiënten zijn nagegaan.*			+
6	*De beoogde gebruikers van de richtlijn zijn duidelijk benoemd.*			+
7	*De richtlijn is getest onder de beoogde gebruikers.*			+
	Methodologie			
8	*Er zijn systematische methoden gebruikt voor het zoeken naar wetenschappelijk bewijsmateriaal.*			-
9	*De criteria voor het selecteren van het wetenschappelijk bewijsmateriaal zijn duidelijk beschreven.*	+		+
10	*De gebruikte methoden om de aanbevelingen op te stellen, zijn duidelijk beschreven.*	-		+
11	*Gezondheidswinst, bijwerkingen en risico's zijn overwogen bij het opstellen van de aanbevelingen.*	-	-	
12	*Er bestaat een expliciet verband tussen de aanbevelingen en het onderliggende wetenschappelijke bewijsmateriaal.*			+
13	*De richtlijn is vóór publicatie door externe experts beoordeeld.*			+
14	*Een procedure voor herziening van de richtlijn is vermeld.*	+	+	-
	Helderheid en Presentatie			
15	*De aanbevelingen zijn specifiek en ondubbelzinnig.*			-
16	*De verschillende beleidsopties zijn duidelijk vermeld.*			-
17	*De kernaanbevelingen zijn gemakkelijk te herkennen.*			-
18	*De toepassing van de richtlijn wordt ondersteund met hulpmiddelen.*	-	-	+
	Toepassing			
19	*De mogelijke organisatorische belemmeringen bij het toepassen van de aanbevelingen zijn besproken.*			+
20	*De mogelijke kostenimplicaties van het toepassen van de aanbevelingen zijn overwogen.*			-
21	*De richtlijn geeft de belangrijkste criteria om na te gaan en te toetsen of de richtlijn wordt gevolgd.*			+
	Onafhankelijkheid Opstellers			
22	*De richtlijn is niet beïnvloed door de opvattingen of belangen van de financierende instantie.*		-	-
23	*Conflicterende belangen van leden van de werkgroep zijn vastgelegd.*	+	+	-
	Gemiddeld over alle arbokennisproducten:	12,46	0,75	
	Ondergrens - SD	9,20	0,63	
	Bovengrens + SD	15,72	0,87	
24	Deze richtlijn is aan te bevelen voor gebruik in de praktijk.	12	0,65	

In tabel 11 staat in de eerste kolom de gemiddelde standaarddeviatie per domein. 'Onderwerp en Doel' heeft de hoogste standaarddeviatie, 'Betrokkenheid Belanghebbenden' en 'Methodologie' komen laag uit wat spreiding betreft. In de kolom daarnaast staat de ICC voor 4 beoordelaars

(berekend over 17 arbokennisproducten). De waarden voor 'Helderheid Presentatie' en 'Onafhankelijkheid Opstellers' liggen beneden de grens van 0,6.

Tabel 11. Spreiding AGREE domeinscores arbokennisproducten

Gemiddelde domeinstandaarddeviatie over alle arbokennisproducten	SD	ICC4
Onderwerp en Doel	21,11	0,71
Betrokkenheid Belanghebbenden	16,81	0,84
Methodologie	16,82	0,80
Helderheid en Presentatie	20,34	0,59
Toepassing	16,92	0,72
Onafhankelijkheid Opstellers	18,42	0,47
Gemiddeld over alle domeinen	18,13	

In tabel 12 hebben de resultaten ten aanzien van de intraclass correlatie coëfficiënten per item en domein samengevat (zie bijlage 7 voor meer detail). Er is een duidelijke trend dat de ICC's verbeteren bij een toename van het aantal beoordelaars van 2 naar 4. De item-ICC's liggen bij 4 beoordelaars vaker in het acceptabele gebied dan bij 2 en ook zijn er meer ICC-waarden significant. Dezelfde trend is waarneembaar op domeinniveau. De ICC waarden voor 4 beoordelaars zijn het meest indicatief omdat bij die berekening de kleinste portie gegevens (15%) buiten beschouwing is gelaten.
De item-ICC waarden lopen van slecht tot zeer goed. De item-ICC waarden voor item 3, 4, 5, 6, 7, 9, 10, 12, 13, 18, 19 en 21 zijn bevredigend, die voor item 1, 2, 8, 14, 15, 16, 17, 20 en 22, en 23 problematisch.
Op domeinniveau zijn de ICC waarden bij vier beoordelaars voor Onderwerp en Doel, Betrokkenheid Belanghebbenden, Methodologie en Toepassing bevredigend, die voor Helderheid Presentatie en Onafhankelijkheid Opstellers problematisch.

Tabel 12. Intraclass correlatie coëfficiënten per item en per domein

ICC bij 2, 3 en 4 beoordelaars	ICC2 N=22		ICC3 N=21		ICC4 N=17	
23 items	1	K	1	K	1	K
'zeer goed' ICC hoger dan 0,75	0	0	0	1	0	5
'goed' ICC tussen 0,60 en 0,75	0	3	0	4	0	7
'matig' ICC tussen 0,40 en 0,60	6	8	4	12	6	8
'slecht' item ICC onder 0,40	17	12	19	6	17	3
aantal significant	6	6	17	17	16	16
6 domeinen	1	K	1	K	1	K
'zeer goed' ICC hoger dan 0,75	0	0	0	1	0	2
'goed' ICC tussen 0,60 en 0,75	0	2	0	2	0	2
'matig' ICC tussen 0,40 en 0,60	2	0	3	2	4	2
'slecht' item ICC onder 0,40	4	4	3	1	2	0
aantal significant	2	2	5	5	5	5

3.3 Relevantie van AGREE items

Relatie tussen relevantiescores en achtergrondkenmerken

In tabel 13 worden de relevantiescores getoond. De kolom ≤ 2 geeft aan welk percentage beoordelaars hier 'Zeer oneens' of 'Oneens' scoorde en dus aangaf dit item niet of minder relevant te vinden. De gemiddelde relevantiescores op itemniveau voor alle arbokennisproducten variëren van 3,08 tot 3,88. De beoordelaars vinden dus alle items (gemiddeld over alle arbokennisproducten) relevant.

Tabel 13. Relevantiescores van de AGREE-items

		gemiddelde score	SD	score ≤ 2
	Onderwerp en Doel			
1	*Het doel van de richtlijn is specifiek beschreven.*	3,88	0,39	2,4
2	*De klinische vraag/vragen die in de richtlijn aan de orde komt/komen, is/zijn specifiek.*	3,66	0,68	4,7
3	*De patiëntenpopulatie waarop de richtlijn van toepassing is, is specifiek beschreven.*	3,84	0,43	2,4
	Betrokkenheid Belanghebbenden			
4	*De leden van de werkgroep die de richtlijn heeft ontwikkeld, komen uit alle relevante beroepsgroepen.*	3,47	0,68	8,2
5	*Het perspectief en de voorkeuren van patiënten zijn nagegaan.*	3,24	0,80	15,3
6	*De beoogde gebruikers van de richtlijn zijn duidelijk benoemd.*	3,73	0,52	1,2
7	*De richtlijn is getest onder de beoogde gebruikers.*	3,62	0,69	7,1
	Methodologie			
8	*Er zijn systematische methoden gebruikt voor het zoeken naar wetenschappelijk bewijsmateriaal.*	3,55	0,75	10,6
9	*De criteria voor het selecteren van het wetenschappelijk bewijsmateriaal zijn duidelijk beschreven.*	3,41	0,79	11,6
10	*De gebruikte methoden om de aanbevelingen op te stellen, zijn duidelijk beschreven.*	3,52	0,70	9,4
11	*Gezondheidswinst, bijwerkingen en risico's zijn overwogen bij het opstellen van de aanbevelingen.*	3,38	0,77	10,7
12	*Er bestaat een expliciet verband tussen de aanbevelingen en het onderliggende wetenschappelijke bewijsmateriaal.*	3,32	0,82	15,5
13	*De richtlijn is vóór publicatie door externe experts beoordeeld.*	3,55	0,65	5,9
14	*Een procedure voor herziening van de richtlijn is vermeld.*	3,17	0,80	20,2
	Helderheid en Presentatie			
15	*De aanbevelingen zijn specifiek en ondubbelzinnig.*	3,67	0,63	6,0
16	*De verschillende beleidsopties zijn duidelijk vermeld.*	3,49	0,74	9,6
17	*De kernaanbevelingen zijn gemakkelijk te herkennen.*	3,81	0,48	16,9
18	*De toepassing van de richtlijn wordt ondersteund met hulpmiddelen.*	3,11	0,89	22,4
	Toepassing			
19	*De mogelijke organisatorische belemmeringen bij het toepassen van de aanbevelingen zijn besproken.*	3,21	0,80	16,5
20	*De mogelijke kostenimplicaties van het toepassen van de aanbevelingen zijn overwogen.*	3,08	0,83	18,8
21	*De richtlijn geeft de belangrijkste criteria om na te gaan en te toetsen of de richtlijn wordt gevolgd.*	3,31	0,74	11,8
	Onafhankelijkheid Opstellers			
22	*De richtlijn is niet beïnvloed door de opvattingen of belangen van de financierende instantie.*	3,50	0,65	6,0
23	*Conflicterende belangen van leden van de werkgroep zijn vastgelegd.*	3,19	0,78	13,1
24	Deze richtlijn is aan te bevelen voor gebruik in de praktijk.	3,74	0,54	2,4

Slechts in 10% van alle scores is gebruik gemaakt van de antwoordmogelijkheid ‘Zeer oneens’ of ‘Oneens’. Voor 8 van de 23 items is het percentage beoordelaars dat het betreffende item niet relevant vindt (score ‘Zeer oneens’ en ‘Eens’) lager dan 10%. Voor 7 van de 23 items is dat percentage ‘(Zeer) oneens’ hoger dan 15%. Dat het doel specifiek is beschreven, de populatie of het probleem waarop het arbokennisproduct van toepassing is en een goede herkenbaarheid van de kernaanbevelingen krijgt de hoogste gemiddelde relevantiescores. Dat er rekening is gehouden met eventuele kostenoverwegingen en dat de toepassing wordt ondersteund met hulpmiddelen vindt men (gemiddeld over alle arbokennisproducten) het minst belangrijk.
Bij 10 arbokennisproducten komt een gemiddelde itemrelevantiescore lager dan 3 (of ‘Eens’) niet voor. Het hoofdstuk Diabetes Mellitus uit het Handboek Arbeid en belastbaarheid scoort wat dat betreft het slechtste: 13 items scoren daar lager dan 3,00 en daarmee is het het enige arbokennis-product waarvan de gemiddelde itemrelevantiescore over alle AGREE-items lager uitkomt dan 3,00.
Over de relevantie van het nagaan van de patiëntenvoorkeuren (item 5) en de ondersteuning van de toepassing met hulpmiddelen (item 18) bestaat het meeste verschil van mening tussen de beoordelaars als we afgaan op de gemiddelde standaarddeviatie van de itemrelevantiescores per arbokennisproduct.

Als we de gemiddelde relevantiescores vergelijken voor de verschillende arbokennisproducten gegroepeerd per uitgever (NVAB, STECR en Overige) zijn die verschillen voor geen enkel item significant.

Op domeinniveau blijken er geen significante verschillen aantoonbaar tussen gemiddelde relevantie-score en de achtergrondkenmerken van de beoordelaars. Voor één gebied is het verschil wel signifi-cant. ‘Toepassing’ scoort hoger relevant bij beoordelaars die werkzaam zijn in de praktijk dan bij beoordelaars uit de groep ‘universiteit/wetenschap’ (3,37 en 3,02 gemiddeld: 3,19; p <0,05). Analyse op itemniveau maakt duidelijk dat het verschil vooral blijkt te komen doordat meer belang gehecht wordt aan de bespreking van de mogelijke organisatorische belemmeringen bij toepassing van de aanbevelingen en het vermelden van criteria om na te kunnen toetsen of men de richtlijn heeft gevolgd. Beide items worden belangrijker gevonden door de mensen werkzaam in de praktijk (zie tabel 14).

De relevantie van de items van het domein ‘Toepassing’ worden door de beoordelaars uit NVAB-kring wat minder relevant geacht dan door de overige beoordelaars. Voor elk van de drie items waaruit dit domein is samengesteld zijn de verschillen in gemiddelde relevantiescore significant.

Tabel 14. Verschillen per organisatie beoordelaars

	NVAB	*STECR*	*Overig*	*Totaal*
Domeinscore ‘Toepassing’*	2,78	3,36	3,26	3,20
De mogelijke organisatorische belemmeringen bij het toepassen van de aanbevelingen zijn besproken.	2,80	3,60	3,27	3,21
*De mogelijke kostenimplicaties van het toepassen van de aanbevelingen zijn overwogen.**	2,60	3,14	3,21	3,08
*De richtlijn geeft de belangrijkste criteria om na te gaan en te toetsen of de richtlijn wordt gevolgd.**	2,93	3,59	3,29	3,31

(toetsing itemniveau: Kruskal-Wallis H; getoond wordt middelde score) *:p<0,05

Voor vijf items (uit diverse domeinen) vinden we een significant verschil tussen de beoordelaars werkzaam in de praktijk en werkzaam in ‘wetenschap/universiteit’. Praktijkmensen lijken bij het beoordelen van arbokennisproducten een aantal criteria wat meer belangrijk te vinden dan personen werkzaam aan de universiteit/wetenschap. Beoordelaars uit NVAB-kring scoren de herkenbaarheid van kernaanbevelingen significant minder relevant dan anderen. Voor de overige items wordt geen significant verschil gevonden. Bedrijfsartsbeoordelaars scoren de relevantie van het systematisch zoeken naar wetenschappelijk bewijsmateriaal significant hoger dan anderen.

Voor de overige items zijn geen significante verschillen aantoonbaar tussen de relevantiescores van beoordelaars die bedrijfsarts zijn of niet.

Tabel 15. Verschillen beoordelaars werkzaam in praktijk of wetenschap

	Praktijk	*Wetenschap*	*Totaal*
*Het perspectief en de voorkeuren van patiënten zijn nagegaan.**	3,43	3,05	3,24
*Een procedure voor herziening van de richtlijn is vermeld.***	3,43	2,90	3,17
*De toepassing van de richtlijn wordt ondersteund met hulpmiddelen.***	3,36	2,86	3,11
*De mogelijke organisatorische belemmeringen bij het toepassen van de aanbevelingen zijn besproken.**	3,38	3,02	3,20
*De richtlijn geeft de belangrijkste criteria om na te gaan en te toetsen of de richtlijn wordt gevolgd.***	3,50	3,10	3,30

(toetsing op itemniveau Kruskal-Wallis H; getoond wordt gemiddelde itemscore) *: p<0,05; **p<0,01

Tabel 16. Verschil itemscore uit NVAB-kring of niet

	NVAB	*Anderen*	*Totaal*
*De kernaanbevelingen zijn gemakkelijk te herkennen.**	3,64	3,89	3,81

(toetsing op itemniveau Kruskal-Wallis H; getoond wordt gemiddelde itemscore) *: p<0,05

Tabel 17. Verschil itemscore bedrijfsarts of andere achtergrond

	Bedrijfsarts	*Andere achtergrond*	*Totaal*
*Er zijn systematische methoden gebruikt voor het zoeken naar wetenschappelijk bewijsmateriaal.**	3,67	3,44	3,55

(toetsing op itemniveau Kruskal-Wallis H; getoond wordt gemiddelde itemscore) *: p<0,05

3.4 Toegevoegde vragen

Inleiding

De spreiding in de antwoorden op de toegevoegde vragen is groot (zie tabel 18). Voor bijna alle items wordt de hele beschikbare range van antwoordcategorieën benut. Dat blijkt ook uit het hoge aandeel antwoorden in de categorieën ‘Zeer oneens’ en ‘Oneens’(kolom 3).

Richtlijnigheid

In tabel 19 kunt u vinden hoe de diverse arbokennisproducten scoren op richtlijnigheid. STECR-Werkwijzers en assists scoren van 2,44 tot 3,30 op deze schaal; NVAB-Richtlijnen, -Leidraad en -LESA van 3,25 tot 3,94; de overige arbokennisproducten van 2,50 tot 3,69. De STECR-Werkwijzer ‘Arbeidsconflicten’ scoort op deze schaal dus hoger dan de NVAB-Richtlijn ‘COPD en Astma’.

Voorbeelden van arbokennisproducten met een hoge richtlijnigheidsscore zijn: de NVAB-Richtlijnen Lage Rugklachten (3,94) en Psychische Klachten (3,88) en de NVAB-Leidraad Verwijzen door de bedrijfsarts (3,88). Bij de niet-NVAB producten valt de hoge score op van het artikel in TBV waarin de PAGO voor huisvuilbeladers wordt toegelicht (3,50). Ook de registratierichtlijnen van het NCvB voor burnout scoren hoog op richtlijnigheid (3,69).

Aanpak problematisch verzuim, Somatisatie, Reductie werkstress en Arbeidsconflicten zijn STECR-werkwijzers met een redelijk hoog richtlijnkarakter: respectievelijk: 3,00; 3,00; 3,13 en 3,30. Relatief laag scoren de Voorstudie met STECR–Assist Vrouwen in de WAO; het hoofdstuk Diabetes mellitus uit het Handboek Arbeid en Belastbaarheid (2,67) en de STECR-Asstist Cultuur of Persoon;

Sociaal-medische Begeleiding Allochtonen (2,88). Uit de hoge standaarddeviatie blijkt dat de beoordelaars het bij de volgende arbokennisproducten flink oneens waren over het richtlijnkarakter: de STECR werkwijzer Aanpak Problematisch Verzuim, de beide Lisv/UWV Standaarden Geen duurzaam benutbare mogelijkheden en Zwangerschap en arbeidsongeschiktheid, en wederom het hoofdstuk Diabetes Mellitus uit het handboek arbeid en belastbaarheid.

Gemiddeld worden de NVAB-documenten in deze studie meer als richtlijnen gezien dan de andere kennisproducten (zie tabel 20). De minimumscore voor de NVAB-producten is 2,75 voor de STECR-producten 1,25 voor de overige producten 1,50.

Tabel 18. Toegevoegde vragen

		Gemiddelde score	SD	score ≤ 2
26	*Dit document bevat specifieke, expliciete uitspraken voor goed handelen bij een welomschreven probleem of doelpopulatie (aanbevelingen).*	3,18	0,86	22,4
27	*Met dit document kan in principe het handelen van veel arboprofessionals beïnvloed worden (bedoeld voor brede verspreiding).*	3,21	0,87	17,6
28	*De organisatie die deze tekst heeft ontwikkeld of uitgegeven, heeft het vertrouwen van het merendeel van de bedrijfsartsen en andere arboprofessionals in de praktijk. ('zij maken inhoudelijk goede producten')*	3,18	0,86	19,3
29	*Dit document zal door het merendeel van de praktijk gezien worden als een richtinggevend document op het terrein van de bedrijfsgeneeskunde en/of arbozorg. (los van evt. titel, inleiding, of omschreven bedoeling)*	2,94	0,99	34,5
30	*Mijn deskundigheid op het inhoudelijke terrein van deze tekst was voldoende om de AGREE-vragen te kunnen beantwoorden.*	3,36	0,59	3,5
31	*Het beantwoorden van de AGREE-vragenlijst voor deze tekst was eenvoudig.*	2,99	0,73	20,0
32	*De informatie die ik nodig had voor het beantwoorden van de AGREE-vragen was goed te vinden in dit document (of de bijgevoegde inleiding).*	3,22	0,79	15,3
33	*Naast de informatie die in het document stond, heb ik aanvullende informatie gebruikt (bijvoorbeeld van internet of persoonlijke kennis) om sommige AGREE-vragen te beantwoorden.*	2,32	0,92	23,8*
34	*Mijn algemene oordeel over deze tekst is nu, na het invullen van de AGREE-vragenlijst, anders dan mijn eerste indruk.*	2,16	0,81	21,2*
36	*Voor mij is de AGREE vragenlijst een zinvol instrument om de kwaliteit dit arbokennisproduct te beoordelen.*	3,11	0,89	20,2
37	*Het is hinderlijk dat de AGREE-vragen onvoldoende zijn toegespitst op de arbocontext van dit specifieke kennisproduct.*	1,80	0,89	34,5*
38	*De AGREE is een te gedetailleerd instrument in verhouding tot de bedoeling en reikwijdte van dit arbokennisproduct.*	1,82	0,80	32,9*
39	*Omdat dit document niet als 'richtlijn' bedoeld is, zijn de vragen van de AGREE niet relevant.*	2,13	0,80	16,5*
40	*Een kwaliteitsoordeel over dit arbokennisproduct kan zonder de AGREE eenvoudiger en net zo goed tot stand komen.*	1,81	0,75	22,6*
41	*In de AGREE ontbreken vragen die noodzakelijk zijn om dit arbokennisproduct goed in zijn context te kunnen beoordelen.*	1,95	0,73	24,1*
43	*Als dit arbokennisproduct verspreid wordt onder veel arbo-professionals is het wenselijk dat het voorzien wordt van een onafhankelijk keurmerk (bijvoorbeeld gebaseerd op de AGREE-beoordeling).*	3,09	0,84	23,5
44	*Het merendeel van de bedrijfsartsen en arboprofessionals in de praktijk heeft onvoldoende kennis om de inhoudelijke kwaliteit van dit arbokennisproduct goed te kunnen beoordelen.*	2,26	0,89	36,5*
45	*Als een bedrijfsarts of arboprofessional weet in welke mate een aanbeveling in dit arbokennisproduct gebaseerd is op wetenschappelijk bewijs dan zal dat invloed hebben op de ruimte die hij/zij professioneel neemt om in de praktijk van deze aanbeveling af te wijken.*	3,05	0,77	22,4
46	*De vragen van de AGREE zouden arboprofessionals en bedrijfsartsen in de praktijk helpen bij het zelf beoordelen van de inhoud van dit arbokennisproduct.*	2,88	0,81	29,4

* Deze items zijn negatief gesteld: antwoordcategorieën zijn daarom omgepoold.

Tabel 19. Richtlijnigheid arbokennisproducten

		gemiddelde score	*min*	*range*	*SD*
1	ABBE-rsi: STECR-werkwijzer	2,44	1,25	1,75	0,80
2	Arbeidsconflicten; STECR-werkwijzer	3,30	2,25	1,75	0,76
3	Aanpak Problematisch Verzuim; STECR-werkwijzer	3,00	2,00	2,00	0,91
4	Lage Rugklachten en Arbeidsomstandigheden; STECR-werkwijzer	2,75	2,00	1,25	0,66
5	Reductie van werkstress door de werknemer in de Gezondheidszorg /Onderwijs; STECR-werkwijzer	3,13	2,75	1,00	0,43
6	Cultuur of Persoon; Sociaal-medische Begeleiding Allochtonen; STECR-Assist	2,88	2,50	0,75	0,32
7	Vrouwen in de WAO + Vrouwen níet in de WAO; Voorstudie en STECR–Assist	2,44	1,50	2,00	0,83
8	Somatisatie, multidisciplinaire aanpak bij somatoforme verschijnselen door arboprofessionals; STECR-werkwijzer	3,00	2,75	0,50	0,20
9	Handelen van de bedrijfsarts bij werknemers met lage rugklachten; NVAB-richtlijn	3,94	3,75	0,25	0,13
10	Handelen van de bedrijfsarts bij werknemers met psychische klachten; NVAB-richtlijn	3,88	3,75	0,25	0,18
11	Handelen van de bedrijfsarts bij werknemers met Astma en COPD; NVAB-richtlijn	3,25	2,50	1,50	0,75
12	Handelen van de bedrijfsarts bij werknemers met klachten aan arm, schouder of nek; NVAB-richtlijn	3,56	3,00	1,00	0,52
13	Handelen van de bedrijfsarts bij het verwijzen van werknemers; NVAB Leidraad	3,88	3,50	0,50	0,25
14	Overspanning; Landelijke Eerstelijns Samenwerkingsafspraak NHG/NVAB	3,45	2,75	1,00	0,45
15	Geen Duurzaam Benutbare Mogelijkheden; Lisv/UWV Standaard	2,75	1,75	1,75	0,90
16	Zwangerschap en bevalling als oorzaak van ongeschiktheid voor haar arbeid; Lisv/UWV Standaard	2,81	1,75	2,25	1,01
17	Beroepsgebonden burnout/overspanning; voorwoord en registratierichtlijn; NCvB	3,69	3,00	1,00	0,47
18	Interventies bij acute lage rugklachten; van de kennisbank http://www.kenniscentrumakb.nl; NKAB	2,50	2,00	0,75	0,43
19	Klachtenreductie en stressmanagement, Arbeidshulpverlening bij psychische klachten; Inleiding en protocol; NKAP	3,25	2,25	1,50	0,68
20	Diabetes mellitus; Hoofdstuk uit het Handboek Arbeid en Belastbaarheid; BSL	2,67	1,50	2,00	1,04
21	PAGO voor huisvuilbeladers; (artikel uit TBV)	3,50	2,25	1,75	0,84
22	Sociaal Medische Begeleiding van Werknemers in de Bouw; Introductie en richtlijn rugklachten	2,83	2,50	0,75	0,38

Tabel 20. Richtlijnigheid arbokennisproducten per organisatie

Organisatie arbokennisproduct	**NVAB**	**STECR**	**overig**	**Totaal**
Gemiddelde Richtlijnigheid**	3,71	2,85	3,02	3,14
Minimum	2,75	1,25	1,50	1,25
Range	1,25	2,75	2,50	2,75
N=	23	33	26	82

**:p<0,001

Achtergrondkenmerken van de beoordelaars hebben weinig samenhang met de richtlijnigheidsscore: bedrijfsartsen en niet bedrijfsartsen, werkzaam in wetenschap of praktijk: de verschillen in richtlijnigheidsscore zijn niet significant. Niet-NVAB-beoordelaars scoren hun producten een beetje hoger op richtlijnigheid (3,26 versus 2,96; p<0,05). Waarschijnlijk wordt dat verklaard door het feit dat niet NVAB-beoordelaars meer kans hadden om NVAB-producten te beoordelen.

Gebruiksgemak AGREE

De uitvoerbaarheid van het scoren met de AGREE komt naar voren uit de antwoorden op vraag 30 tot en met 33. De items 30 tot en met 32 zijn samengevoegd tot de schaal Gebruiksgemak AGREE (zie ook tabel 21). De gemiddelde score per arbokennisproduct ligt tussen 2,67 en 3,67. Slechts vier arbokennisproducten scores lager dan 3,00: De webteksten interventies acute lage rugklachten van het NVAB (benodigde informatie slecht vindbaar), de voorstudie met Assist Vrouwen in de WAO (benodigde informatie slecht vindbaar), De NVAB-Richtlijn Psychische Klachten (AGREE-vragen beantwoorden niet eenvoudig) en de werkwijzer arbeidsconflicten (beantwoorden AGREE-vragen niet eenvoudig). Voor het overige blijkt de lijst gemakkelijk ingevuld te kunnen worden. Er is een significant verband tussen de richtlijnigheid van een tekst en het gemak waarmee de AGREE gehanteerd kan worden: (Pearsons correlatiecoëfficiënt 0,23 p<0,05.). Ook blijkt het gemak waarmee benodigde informatie gevonden kon worden gecorreleerd aan de domeinscores 'Onderwerp en Doel', 'Betrokkenheid Belanghebbenden' en 'Helderheid Presentatie' (respectievelijk: 0,305**, 0,216* en 0,291**; * p<0,05 en ** p < 0,01).

Tabel 21. Gebruiksgemak AGREE arbokennisproducten

		gemiddelde score	*min*	*range*	*SD*
1	ABBE-rsi: STECR-werkwijzer	3,17	2,67	1,00	0,43
2	Arbeidsconflicten; STECR-werkwijzer	2,80	2,33	1,00	0,38
3	Aanpak Problematisch Verzuim; STECR-werkwijzer	3,17	3,00	0,33	0,19
4	Lage Rugklachten en Arbeidsomstandigheden; STECR-werkwijzer	3,33	2,67	1,33	0,67
5	Reductie van werkstress door de werknemer in de Gezondheidszorg /Onderwijs; STECR-werkwijzer	3,08	2,33	1,33	0,57
6	Cultuur of Persoon; Sociaal-medische Begeleiding Allochtonen; STECR-Assist	3,25	3,00	0,67	0,32
7	Vrouwen in de WAO + Vrouwen níet in de WAO; Voorstudie en STECR–Assist	2,67	2,33	0,67	0,38
8	Somatisatie, multidisciplinaire aanpak bij somatoforme verschijnselen door arboprofessionals; STECR-werkwijzer	3,08	2,33	1,33	0,57
9	Handelen van de bedrijfsarts bij werknemers met lage rugklachten; NVAB-richtlijn	3,67	3,33	0,67	0,38
10	Handelen van de bedrijfsarts bij werknemers met psychische klachten; NVAB-richtlijn	2,83	2,67	0,33	0,24
11	Handelen van de bedrijfsarts bij werknemers met Astma en COPD; NVAB-richtlijn	3,00	2,33	1,00	0,47
12	Handelen van de bedrijfsarts bij werknemers met klachten aan arm, schouder of nek; NVAB-richtlijn	3,50	3,33	0,33	0,19
13	Handelen van de bedrijfsarts bij het verwijzen van werknemers; NVAB Leidraad	3,50	2,67	1,33	0,58
14	Overspanning; Landelijke Eerstelijns Samenwerkingsafspraak NHG/NVAB	3,28	2,67	1,00	0,44
15	Geen Duurzaam Benutbare Mogelijkheden; Lisv/UWV Standaard	3,33	3,00	0,67	0,33
16	Zwangerschap en bevalling als oorzaak van ongeschiktheid voor haar arbeid; Lisv/UWV Standaard	3,42	3,00	1,00	0,50
17	Beroepsgebonden burnout/overspanning; voorwoord en registratierichtlijn; NCvB	3,33	2,67	1,33	0,54
18	Interventies bij acute lage rugklachten; van de kennisbank http://www.kenniscentrumakb.nl; NKAB	2,92	2,00	2,00	0,92
19	Klachtenreductie en stressmanagement, Arbeidshulpverlening bij psychische klachten; Inleiding en protocol; NKAP	3,17	2,33	1,33	0,64
20	Diabetes mellitus; Hoofdstuk uit het Handboek Arbeid en Belastbaarheid; BSL	3,11	2,33	1,33	0,69
21	PAGO voor huisvuilbeladers; (artikel uit TBV)	3,25	2,33	1,67	0,69
22	Sociaal Medische Begeleiding van Werknemers in de Bouw; Introductie en richtlijn rugklachten	3,33	2,67	1,33	0,67

Er is gevraagd of beoordelaars aanvullende informatie hebben gebruikt bij het invullen van de AGREE (naast de beschikbare tekst van het arbokennisproduct, soms aangevuld met een inleiding). Dat bleek nauwelijks het geval: de gemiddelde scores variëren per arbokennisproduct van 1,00 tot 2,75, gemid-

deld 1,68 (zie bijlage 8). Slechts 5 arbokennisproducten scoren gemiddeld hoger dan 2,00: Bij het invullen van de AGREE voor de webteksten interventies acute lage rugklachten van de kennisbank NKAB, de STECR-Werkwijzer Reductie werkstress, de NVAB-Richtlijn Astma en COPD, de Lisv/UWV standaard Zwangerschap en arbeidsongeschiktheid en de NVAB-Leidraad Verwijzen door de bedrijfsarts hebben beoordelaars soms aanvullende informatie gezocht en gebruikt. Voor de overige arbokennisproducten kon de AGREE op basis van de informatie in de tekst zelf ingevuld worden.

Waardering AGREE

Er zijn een aantal vragen gesteld over het nut en de waardering van de AGREE en de manier waarop de AGREE-score toegepast zou kunnen worden. Acht items (36, 37, 38, 39, 40, 43, 44 en 46) zijn samengevoegd tot een schaal: Waardering AGREE.

Tabel 22. Waardering AGREE arbokennisproducten

		gemiddelde score	*min*	*range*	*SD*
1	ABBE-rsi: STECR-werkwijzer	3,06	2,63	0,88	0,36
2	Arbeidsconflicten; STECR-werkwijzer	2,58	1,75	1,25	0,48
3	Aanpak Problematisch Verzuim; STECR-werkwijzer	3,22	2,88	1,00	0,45
4	Lage Rugklachten en Arbeidsomstandigheden; STECR-werkwijzer	3,25	2,50	1,13	0,65
5	Reductie van werkstress door de werknemer in de Gezondheidszorg /Onderwijs; STECR-werkwijzer	2,94	2,88	0,25	0,13
6	Cultuur of Persoon; Sociaal-medische Begeleiding Allochtonen; STECR-Assist	3,03	2,38	1,13	0,48
7	Vrouwen in de WAO + Vrouwen níet in de WAO; Voorstudie en STECR–Assist	2,94	2,38	0,75	0,38
8	Somatisatie, multidisciplinaire aanpak bij somatoforme verschijnselen door arboprofessionals; STECR-werkwijzer	3,17	3,00	0,50	0,29
9	Handelen van de bedrijfsarts bij werknemers met lage rugklachten; NVAB-richtlijn	3,69	3,13	0,88	0,39
10	Handelen van de bedrijfsarts bij werknemers met psychische klachten; NVAB-richtlijn	3,19	3,13	0,13	0,09
11	Handelen van de bedrijfsarts bij werknemers met Astma en COPD; NVAB-richtlijn	3,29	3,13	0,38	0,19
12	Handelen van de bedrijfsarts bij werknemers met klachten aan arm, schouder of nek; NVAB-richtlijn	3,06	2,63	1,00	0,43
13	Handelen van de bedrijfsarts bij het verwijzen van werknemers; NVAB Leidraad	2,91	2,63	0,63	0,33
14	Overspanning; Landelijke Eerstelijns Samenwerkingsafspraak NHG/NVAB	2,79	2,00	1,63	0,57
15	Geen Duurzaam Benutbare Mogelijkheden; Lisv/UWV Standaard	2,83	2,50	0,63	0,31
16	Zwangerschap en bevalling als oorzaak van ongeschiktheid voor haar arbeid; Lisv/UWV Standaard	2,84	2,13	1,75	0,78
17	Beroepsgebonden burnout/overspanning; voorwoord en registratierichtlijn; NCvB	2,91	2,63	0,50	0,26
18	Interventies bij acute lage rugklachten; van de kennisbank http://www.kenniscentrumakb.nl; NKAB	3,19	2,88	0,75	0,33
19	Klachtenreductie en stressmanagement, Arbeidshulpverlening bij psychische klachten; Inleiding en protocol; NKAP	2,50	1,63	1,38	0,60
20	Diabetes mellitus; Hoofdstuk uit het Handboek Arbeid en Belastbaarheid; BSL	2,33	1,88	1,13	0,59
21	PAGO voor huisvuilbeladers; (artikel uit TBV)	2,81	2,38	0,88	0,44
22	Sociaal Medische Begeleiding van Werknemers in de Bouw; Introductie en richtlijn rugklachten	3,00	2,38	1,25	0,63

De gemiddelde score ‘Waardering AGREE’ per arbokennisproduct bedraagt 2,96 voor de scores per arbokennisproduct (zie tabel 22). Relatief laag scoort de waardering AGREE voor het protocol arbeidshulpverlening van het NKAP (2,50), het hoofdstuk Diabetes Mellitus uit het handboek arbeid en belastbaarheid (2,33) en de STECR-werkwijzer arbeidsconflicten (2,58). In deze gevallen vragen de

beoordelaars zich af of de AGREE wel het geschikte instrument is om de kwaliteit van het product te beoordelen. Men mist de toespitsing op de arbocontext, met vindt het geen richtlijn, dus de AGREE-items niet relevant of men twijfelt over de beoogde reikwijdte van het document.

De waardering voor de AGREE is gecorreleerd aan het gemak waarmee de beoordelaars de AGREE invullen (Pearsons Correlatie Coëfficiënt 0,27, p<0,01). De waardering voor de AGREE is tevens gecorreleerd aan de gemiddelde relevantiescores per domein voor 'Betrokkenheid Belanghebbenden' (0,325, p<0,01), Methodologie (0,471, p<0,01), 'Helderheid en Presentatie' (0,486, p<0,01) en 'Toepassing' (0,422, p<0,01). Dus naarmate men die AGREE-items meer relevant vindt, stijgt ook de waarde die men van toepassing van de AGREE verwacht.
Bij de LESA-Overspanning en de Lisv/UWV standaard zwangerschap en arbeidsongeschiktheid is de range relatief hoog: bij die arbokennisproducten loopt de waardering voor de AGREE bij de beoordelaars onderling flink uiteen.

De waardering voor de AGREE hangt niet samen met de achtergrondkenmerken van de beoordelaars zoals bedrijfsarts of niet of afkomstig uit NVAB-richtlijnkringen. Mensen uit de praktijk lijken de AGREE iets meer te waarderen (3,07 tegen 2,86, p=0,051) dan mensen werkzaam in de wetenschap. Dat verschil komt vooral omdat de beoordelaars uit de praktijk meer waarde toekennen aan het onafhankelijk keurmerk (2,91 en 3,29, Kruskal wallis H: p<0,05) en verwachten dat betere informatie over de mate van onderbouwing de vrijheidsgraden van de professionals zal beïnvloeden (2,74 en 3,36, Kruskal Wallis H: p< 0,01).

Beoordelaars uit STECR-kringen komen tot een hogere waardering voor de AGREE dan anderen.

Tabel 23. Waardering AGREE per organisatie beoordelaars

	NVAB	STER	Overige	Totaal
Waardering AGREE**	2,70	3,22	2,93	2,96
N	15	22	46	83

**:p <0,01

Analyseren we dat op itemniveau dan blijken drie items vooral verantwoordelijk te zijn voor dat verschil.
- Het is hinderlijk dat de AGREE-vragen onvoldoende zijn toegespitst op de arbocontext van dit specifieke kennisproduct. (omgepoold)
- Omdat dit document niet als 'richtlijn' bedoeld is, zijn de vragen van de AGREE niet relevant. (omgepoold)
- Als een bedrijfsarts of arboprofessional weet in welke mate een aanbeveling in dit arbokennisproduct gebaseerd is op wetenschappelijk bewijs dan zal dat invloed hebben op de ruimte die hij/zij professioneel neemt om in de praktijk van deze aanbeveling af te wijken.

De beoordelaars uit STECR-kring hebben op basis van deze items beduidend positievere verwachtingen van de AGREE dan de nadere beoordelaars.

De waardering voor de AGREE is weliswaar steeds gekoppeld aan het arbokennisproduct dat de beoordelaar voor zich had liggen, maar dat levert weinig verschil op als we de producten van diverse aanbieders naast elkaar leggen: het verschil is klein en niet significant.

Tabel 24. Waardering AGREE organisatie arbokennisproducten

	NVAB	STECR	Overige	Totaal
Waardering AGREE	3,11	3,00	2,80	2,96
N	23	32	28	83

p = 0,067

We hebben de beoordelaars gevraagd of het invullen van de AGREE hun (eerste) oordeel over het arbokennisproduct veranderde en ten aanzien van welk aspect men dan positiever of negatiever was gaan denken. In 18 gevallen hebben de beoordelaars zo'n verandering gemeld; zij konden maximaal 3 punten noemen. 22 Maal is het oordeel negatiever geworden en 2 maal positiever. Door het invullen van de AGREE is men onder andere kritischer gaan denken over onderbouwing en methodologie, de praktische toepasbaarheid, de invloed van patiënten of kosteneffecten.

We hebben de beoordelaars gevraagd of zij dachten dat bedrijfsartsen en arboprofessionals in de praktijk zelf onvoldoende(!) kennis zouden hebben om de inhoudelijke kwaliteit van het arbokennisproduct goed te kunnen beoordelen. Dat zou immers een argument voor het gebruiken van de AGREE zijn. In 36,5 % van de gevallen meent men dat de praktijk problemen kon hebben met het zelf beoordelen van de inhoudelijke kwaliteit van het arbokennisproduct (score 3: 'Eens' en 4: 'Zeer eens'). De gemiddelde scores per arbokennisproduct variëren van 1,5 tot 3,25 (zie bijlage 8).
Laag scoort bijvoorbeeld de webtekst over de effectiviteit van interventies bij acute lage rugklachten van de kennisbank NKAB (1,50), de NVAB-Richtlijn lage rugklachten (1,50) en STECR-Werkwijzer arbeidsconflicten (1,60): in die gevallen verwacht men dus dat de meeste arboprofessionals weinig problemen zullen hebben om zelf tot een goed oordeel over de inhoudelijke kwaliteit van het arbokennisproduct te komen.

Hoog scoort bijvoorbeeld de STECR-Werkwijzer Somatisatie (3,25), de STECR-Werkwijzer ABBE-rsi (3,00), de NVAB-Richtlijn Psychische Klachten (3,00) en de STECR-Werkwijzer Lage Rugklachten en Arbeidsomstandigheden (3,00). Bij die arbokennisproducten bestaat er twijfel bij de beoordelaars of arboprofessionals in de praktijk voldoende kennis hebben om zelfstandig de inhoud van die producten goed te kunnen beoordelen. In die gevallen zou een onafhankelijk oordeel met behulp van de AGREE meerwaarde hebben.

Een van de overige vragen gaat over de wenselijkheid van een onafhankelijk keurmerk op basis van de AGREE voor arbokennisproducten die breed verspreid worden. Op 76,5% van de formulieren oordeelt men (zeer) positief over die mogelijkheid. (Gemiddelde score over 22 arbokennisproducten: 3,08.)

<u>Ervaren omissies in het AGREE-instrument</u>

De respondenten hebben aangegeven of in AGREE criteria ontbreken, die naar hun oordeel wel essentieel zijn om tot een goed kwaliteitsoordeel over een arbokennisproduct te komen. In 76% van de gevallen wordt hierop ontkennend gereageerd. De gemiddelde score op deze vraag per arbokennisproduct komt nooit boven de 3,00 uit. Ruim 20 maal werd een aanvulling voorgesteld. Soms werden deze suggesties bij de andere open vragen opgeschreven.
Voorbeelden van voorgestelde aanvullingen waren:
- Prijs en gemak van beschikbaarheid van het arbokennisproduct.
- Verspreiding onder cliënten.
- Prioritering van verschillende aanbevolen maatregelen: de AGREE wekt soms de indruk uit te gaan van één (klinische) interventie, terwijl in de arbopraktijk vaak meerdere vormen van begeleiding tegelijk uitgevoerd worden.
- Multidisciplinaire aspecten: men vroeg zich af of een arbokennisproduct niet per discipline gescoord moest worden met behulp van de AGREE.
- Arbocuratieve samenwerkingsaspecten, relatie met diagnostiek en behandeling door andere disciplines; de algemene gerichtheid op samenwerking.
- Mate van commitment of participatie van werkgevers en werknemers.
- Implicaties bij toetsing door UWV als professional richtlijn niet volgt.
- Kwaliteit van de verwijzingen en referenties.
- Volledigheid informatie.
- Praktijkrelevantie.

4. Beschouwing

4.1 Sterktes en zwaktes van het onderzoek

In deze studie zijn we nagegaan welke waarde het AGREE-instrument kan hebben om de kwaliteit van arbokennisproducten te bepalen. NVAB-Richtlijnen, STECR-Werkwijzers en vergelijkbare producten van andere organisaties zijn met behulp van het AGREE-instrument beoordeeld door een breed panel van 30 sleutelpersonen betrokken bij NVAB, STECR, richtlijnontwikkeling, kwaliteitsbeleid in arbodiensten en wetenschap. De AGREE resulteert in domeinscores voor verschillende aspecten: Onderwerp en Doel, Betrokkenheid Belanghebbenden, Methodologie, Helderheid en Presentatie, Toepassing en Onafhankelijkheid Opstellers.

De beoordelaars hebben ook aangegeven hoe relevant ze elk van de AGREE-items vonden om het betreffende arbokennisproduct op kwaliteit te beoordelen. Daarnaast beantwoordden zij een aantal toegevoegde vragen over het richtlijnkarakter van het arbokennisproduct, het gemak waarmee ze de AGREE konden invullen en welke toepassingsmogelijkheden men voor de AGREE zag.

In deze paragraaf gaan we in op de positieve en negatieve gevolgen van de gekozen onderzoeksopzet met name voor de gevonden AGREE-scores. Mogelijke foutenbronnen en de validiteit van de AGREE-scores in een arbocontext worden besproken.

Onderzoeksopzet

Het onderzoek is opgezet om na te gaan of de AGREE een geschikt instrument is om de kwaliteit van arbokennisproducten op een uniforme en relevante manier te beoordelen. Hoewel het feitelijk beoordelen van de arbokennisproducten geen primair doel van het onderzoek was, zijn de AGREE-scores voor de diverse producten in deze studie wel informatief. Hoe serieus moeten die scores genomen worden?

Er is geen sprake van een aselecte steekproef, noch bij de beoordelaars, noch bij de arbokennisproducten. Met het oog op het testen van bruikbaarheid en draagvlak voor de AGREE is gekozen voor een brede variatie aan beoordelaars en kennisproducten. We hebben beoordelaars uitgenodigd waarvan we verwachtten dat zij op basis van hun kennis en ervaring konden beoordelen wat de waarde was van de diverse arbokennisproducten. Mogelijk zijn zij vanuit hun betrokkenheid strenger en kritischer dan arboprofessionals in de praktijk. Zij hebben geen producten beoordeeld waar men zelf aan mee gewerkt had of die uitgegeven werden door de eigen organisatie. We gaan er van uit dat de beoordelaars geen belang hadden bij het afzwakken of overdrijven van hun oordeel. Zowel uit NVAB- als STECR-sfeer hebben er een flink aantal beoordelaars deelgenomen aan de studie. Als we de respondenten vergelijken met de groep beoordelaars die aangeschreven is en de groep die hun medewerking toezegde dan is de proportie mensen uit STECR-kring relatief kleiner. Maar omdat de scores slechts weinig samenhangen met de achtergrond van de beoordelaar verwachten we niet dat hierdoor een systematische vertekening van de resultaten is ontstaan.

Er is gewerkt met een beoordelingsinstrument dat uitgebreid gevalideerd is. Om richtlijnen in de gezondheidszorg te beoordelen is de AGREE een internationaal aanvaard instrument. Ook in deze arbocontext toont het instrument goede schaaleigenschappen met goede waarden voor interne consistentie.

De selectie van 22 arbokennisproducten omvatte de meeste recente NVAB-Richtlijnen en STECR-producten. De overige acht beoordeelde producten vormden noodzakelijkerwijs een kleine selectie van wat er op de markt was en diende vooral als vergelijkingsmateriaal. In het geheel overheerst het richtlijnachtige karakter van de beoordeelde documenten, maar er waren er ook bij waar dat minder op de voorgrond stond. De brede variatie aan arbokennisproducten was bedoeld om te testen hoe uitvoerbaar de AGREE-beoordeling was en of er grenzen aan de toepasbaarheid waren van het instrument in de arbocontext. Het ligt niet in de aard van deze studie om de gevonden resultaten te generaliseren naar alle arbokennisproducten die op dit moment in Nederland beschikbaar zijn. Maar het scala aan arbokennisproducten in de studie is wel zodanig breed en de aanpak ons inziens

dermate solide dat de scores niet alleen duidelijk maken hoe de afzonderlijke arbokennisproducten verbeterd kunnen worden, maar de resultaten ook in bredere zin betekenisvol voor organisaties die regelmatig arbokennisproducten uitgeven zoals STECR en de NVAB.

Het antwoordformulier was zo opgezet dat de relevantie van elk item gescoord werd direct nadat men de AGREE-score had toegekend. De kans dat de relevantie in abstracte zin is beoordeeld, is daardoor kleiner. Daar staat tegenover dat beide scores elkaar beïnvloed kunnen hebben: 4 van de 6 domein-relevantiescores zijn significant gecorreleerd aan de domeinscores: van 0,267 (p:<0,05) voor 'Toepassing' tot 0,517 (p:<0,01) voor 'Betrokkenheid Belanghebbenden'. Op itemniveau vinden we voor 8 van 23 items een significante correlatie tussen beide scores. Opvallend is dat voor geen van de methodologie-items er een significante correlatie gevonden wordt tussen AGREE-score en relevantie-score. Omdat de relevantiescores over de hele linie hoog liggen, is voorstelbaar dat de beoordelaars naarmate de vragenlijst vorderde minder kritisch zijn geworden en alleen nog maar hoge scores hebben uitgedeeld. Omdat bij het scoren van de relevantie van de AGREE-items aan het einde van de lijst nog regelmatig gebruik wordt gemaakt van de antwoordmogelijkheden 'Zeer oneens' of 'Oneens', nemen we niet aan dat een dergelijk effect is opgetreden.
In principe is relevantie een open begrip dat door de beoordelaars met verschillende achtergronden op vele manieren ingevuld kan worden. De itemrelevantiescores vertonen per domein echter een grote samenhang: Cronbach's alfa's van 0,60, 0,67, 0,83, 0,79, 0,73 en 0,66. Dat wijst op een goede consistente hantering van het relevantiebegrip.

De toegevoegde vragen zijn ten behoeve van dit onderzoek geformuleerd. Een belangrijke bron voor die vragen vormden de discussies van de laatste jaren over de verschillen en overeenkomsten tussen NVAB-Richtlijnen en STECR-Werkwijzers. We hebben geprobeerd de diverse afwegingen en criteria die impliciet of expliciet in die discussie een rol speelden te verwoorden in de vragen. De vragen zijn vooraf met vertegenwoordigers van STECR besproken en aangevuld. De schalen voor Richtlijnigheid, Gebruiksgemak AGREE en Waardering AGREE zijn later pragmatisch geconstrueerd om de verslaglegging door middel van datareductie eenvoudiger te maken. Het gaat om niet-gevalideerde schalen en de schaalscores moeten met terughoudendheid worden beoordeeld. De beschrijvende scores per vraag zijn minstens zo informatief.
We moeten ervan uit gaan dat de beoordelaars in de toegevoegde vragen direct of indirect thema's konden herkennen uit de discussies tussen STECR en NVAB over werkwijzers en richtlijnen. Sociaal wenselijke antwoorden zijn niet uit te sluiten. Door de beantwoording van de vragen echter te koppelen aan de beoordeling van een specifiek arbokennisproduct hebben we gepoogd de kans daarop te verminderen. De antwoorden op de toegevoegde vragen verschillen soms aanzienlijk per arbokennis-product. We nemen daarom aan dat er nagenoeg geen systematische vertekening van antwoorden heeft plaatsgevonden.

Foutenbronnen in de AGREE-scores

De AGREE-score drukt idealiter uit in welke mate het criterium een rol heeft gespeeld bij de ontwikkeling en uitwerking van het arbokennisproduct. Als dat criterium geen rol heeft gespeeld dan is een lage score juist. Een lage AGREE-domeinscore betekent niet dat de inhoud van de richtlijn slecht of onbetrouwbaar is. Zo'n lage score maakt wel duidelijk dat beoordelaars van mening zijn dat er over dat criterium onvoldoende informatie in de tekst opgenomen was.

Spreiding in scores (voor hetzelfde item bij hetzelfde arbokennisproduct) kan verwijzen naar verschillen van inzicht tussen beoordelaars, maar ook naar mogelijke foutenbronnen. De gevonden AGREE-score wijkt dan af van wat die feitelijk zou moeten zijn. Als een beoordelaar bij toeval of systematisch informatie 'mist' in de tekst heeft dat veel gevolgen voor de uitkomst. Wat is de kans dat beoordelaars daardoor arbokennisproducten te laag of te hoog gescoord hebben?
Een AGREE-score ongelijk aan de 'juiste' score kan in principe op drie manieren ontstaan:

A. De onderzoekers hebben onvolledige of gewijzigde teksten aan de beoordelaar verstrekt en in de tekstdelen die weggelaten of veranderd zijn, staat de benodigde informatie vermeld.
B. De beoordelaar heeft slecht gelezen of geredeneerd. De benodigde informatie staat bij nauwgezette lezing wel degelijk in de tekst of de beoordelaar heeft op basis van de beschikbare informatie assumpties gemaakt die bij nadere beschouwing aantoonbaar onjuist zijn.

C. De gepubliceerde en beschikbare tekst van het arbokennisproduct geeft geen volledig of juist beeld van de ontwikkeling van het kennisproduct. In werkelijkheid is er bijvoorbeeld wel rekening gehouden met een aspect, maar daarover is in de tekst geen informatie opgenomen.
Ook het omgekeerde is mogelijk: in de tekst staat dat er bijvoorbeeld systematisch naar literatuur gezocht is, terwijl dat feitelijk niet op een juiste manier is uitgevoerd.

We gaan er van uit dat foutenbron A geen rol van betekenis gespeeld heeft. Alle beoordelaars kregen de NVAB-Richtlijnen en STECR-producten in de volledige versie zoals ze op het internet beschikbaar zijn. Bij de overige producten zijn soms keuzes gemaakt als de tekst onderdeel was van een groter geheel. Dan werd bijvoorbeeld de inleiding en het betreffende gedeelte opgestuurd en de rest niet. We hebben echter geen aanwijzingen dat op die manier essentiële informatie niet beschikbaar is geweest voor de beoordelaars.

Foutenbron B vormt in deze studie een reële mogelijkheid. Het beoordelen van drie verschillende, soms uitgebreide, arbokennisproducten met behulp van een vragenlijst van ruim 40 items was geen sinecure. Voor de meeste van de beoordelaars was het AGREE instrument nieuw. Zij moesten aan de hand van de bijgevoegde schriftelijke instructie en het AGREE-handboek leren werken met het instrument. Het is mogelijk dat er fouten zijn gemaakt bij het interpreteren van de items of het beoordelen van de teksten, hetgeen zowel tot hogere als lagere beoordelingen geleid kan hebben. Het is aannemelijk dat men voor de veiligheid eerder aan de lage kant is gebleven. We hebben geen feitelijk zicht op de zorgvuldigheid waarmee de beoordelaars de taak hebben uitgevoerd. Alle beoordelaars waren mensen met volle agenda's. Enkele beoordelaars hebben afgehaakt met een expliciete verklaring dat ze de boordeling niet konden combineren met hun andere werkzaamheden. Als beoordelaars hun taak te snel hebben uitgevoerd is het waarschijnlijk dat ze over informatie in de teksten hebben heen gelezen met als resultaat lagere AGREE-scores dan bedoeld. Daar staat echter tegenover dat de AGREE bedoeld is om de beschikbaarheid van informatie die nodig is om een goed kwaliteitsoordeel over een arbokennisproduct te vormen in een maat uit te drukken. Want ook arboprofessionals zullen vaak in een drukke praktijk snel moeten beoordelen wat de kwaliteit van een arbokennisproduct is. Als dergelijke informatie niet snel en makkelijk te vinden is in een arbokennisproduct, dan is het terecht dat de AGREE-score daardoor lager uitvallen. Het is mogelijk dat beoordelaars op basis van vooroordelen over de uitgevende organisatie of bekendheid met de auteurs een hogere of lagere score heeft toegekend. We hebben de beoordelaars gevraagd of na het invullen van de AGREE-vragenlijst het algemene oordeel over het arbokennisproduct veranderd was ten opzichte van hun eerste indruk. Dat was in ruim 20% van de gevallen het geval. Voor het overgrote merendeel dacht men na scoren van AGREE negatiever over een arbokennisproduct, omdat men het belang van diverse AGREE-criteria had gezien.
We hebben geen aanwijzingen om aan te nemen dat foutenbron B tot systematisch hogere of lagere AGREE-scores heeft geleid.

Ten aanzien van foutenbron C kan men betogen dat een beoordelaar geen rekening hoeft te houden met informatie die niet of niet juist is weergegeven. De AGREE probeert de transparantie van de tekst in een score uit te drukken. Het is terecht dat niet opgenomen informatie tot een lagere score leidt.

Sommige AGREE-items zijn dermate concreet geformuleerd dat we moeten aannemen dat er één juiste score bestaat. Andere bevatten in meerdere of mindere mate waardeoordelen, zodat het waarschijnlijk is dat de ware scores een zekere range beslaan. Er bestaat geen absolute, objectieve manier om de 'juiste' AGREE-score vast te stellen. We moeten het dus doen met relatieve gegevens: waar ontstaat er relatief gezien veel verschil van mening over de scores? De resultaten ten aanzien van de overeenstemming in scores (en het gebrek daaraan) heeft betekenis voor de toepasbaarheid van het AGREE-instrument in deze arbocontext.
Ten eerste vertelt de spreiding ons voor welke arbokennisproducten de AGREE als beoordelingsinstrument van de kwaliteit geschikt is. De variatie in arbokennisproducten in dit onderzoek was breed om na te gaan of er documenten met aanbevelingen zijn waar de AGREE-beoordeling niet tot consistente scores leidt. Ten tweede geeft de spreiding informatie over welke AGREE- items en –domeinen dicht bij elkaar liggende scores opleveren. Die items of domeinen zijn dus voor de beoordelaars eenduidig geformuleerd en gemakkelijk te scoren. In de AGREE-items wordt bijvoorbeeld gesproken over een klinische vraag, patiënten en gezondheidswinst: dergelijke begrippen passen niet zo goed in een arbocontext. We hebben die begrippen daarom toegelicht in de invulinstructie. De spreidingsmaten geven aanwijzingen over de instrumenteigenschappen van de AGREE in combinatie

met de effectiviteit van de invulinstructie: hoe goed konden de beoordelaars met de formulering van AGREE-items uit de voeten?
Ten derde geeft de spreiding informatie over de mate waarin de beoordelaars consistent en uniform beoordeeld hebben. Alle beoordelaars hebben de drie arbokennisproducten kort na elkaar doorgelezen en de formulieren ingevuld. In het begin kan de interne maatstaf per beoordelaar nog verschuiven, terwijl dat na een flink aantal beoordelingen minder het geval zal zijn. We hebben geen gegevens die duidelijk maken of arbokennisproducten als eerste, tweede of derde beoordeeld zijn. Vanwege het kleine aantal beoordelingen en het feit dat de meeste beoordelaars geen directe ervaring hadden met de AGREE nemen we aan dat deze foutenbron geen rol heeft gespeeld.
Ten vierde kan er onbedoeld spreiding zijn ontstaan door invoerfouten en vergissingen met nacoderen. We hebben bij de eerste analyse van de gegevens geen onverklaarbare resultaten gevonden die mogelijk een aanwijzing zouden kunnen vormen voor dergelijke fouten. We gaan er van uit dat deze foutenbron geen rol van betekenis heeft gespeeld.

De invloed van foutenbronnen zou voor een deel teruggedrongen kunnen worden door nader onderzoek. De beoordelaars zou gevraagd kunnen worden hun oordeel te onderbouwen met voorbeelden uit de tekst. Bij ongelijke scores zou overleg tussen beoordelaars georganiseerd kunnen worden. In andere studies blijkt dat beoordelaars het in het begin soms oneens zijn, maar na discussie nagenoeg altijd tot gelijke scores komen (Burgers & Boluyt 2002; Tuut et al. 2002). Daarnaast zou de beoordeelde tekstversie samen met de AGREE-scores opgestuurd kunnen worden aan de organisatie die het arbokennisproduct heeft uitgegeven. Zij zouden aperte onjuistheden in de AGREE-scores kunnen opsporen en dat kunnen aantonen met (aanvullende) documentatie. Voor beide vormen van controle is in deze onderzoeksopzet geen ruimte geweest

Validiteit AGREE-scores

We hebben in deze studie de validiteit van het AGREE-instrument niet uitgebreid opnieuw onderzocht. Wel hebben we gevonden dat alle domeinscores zeer sterk correleren met de algemene vraag of de ‘richtlijn’ is aan te bevelen voor gebruik in de praktijk. De domeinscores onderling correleren sterk met elkaar. De relevantieoordelen, de antwoorden op de overige vragen over de waardering voor de AGREE en de opmerkingen die beoordelaars op de formulieren maken, geven geen aanleiding om te twijfelen of de AGREE op een goede wijze meet, waarvoor het instrument bedoeld is. We gaan er vanuit dat de AGREE ook in de arbocontext een valide instrument is.

De indeling van de items in domeinen is niet door middel van een factoranalyse ter discussie gesteld. Gezien de nieuwe context was dat wel een mogelijkheid geweest (MacDermid et al. 2005), maar een nieuwe indeling van domeinen had een vergelijking van scores met de literatuur onmogelijk gemaakt. De Cronbach’s alfa van de domeinschalen zijn goed; dat wijst op een goede interne samenhang.

4.2 Betekenis van de gevonden AGREE-scores

Achtereenvolgens zullen we per onderwerp de gevonden resultaten samenvatten, vergelijken met andere onderzoeksgegevens (als die beschikbaar zijn) en de beleidsimplicaties van de resultaten van dit onderzoek toelichten.

AGREE-scores

- *Samenvatting van de resultaten en vergelijking*

De AGREE-domeinscores voor deze 22 arbokennisproducten beslaan een brede range tussen het minimum van 0 en het maximum van 100. Iets meer dan de helft van alle domeinscores ligt beneden de 50. Het domein ‘Onderwerp en Doel’ scoort gemiddeld over alle arbokennisproducten het hoogste; ‘Toepassing’ het laagste. De gemiddelde domeinscores van de zes NVAB-Richtlijnen, de acht STECR-werkwijzers en de acht Overige arbokennisproducten verschillen structureel. Voor alle zes de domeinen is de gemiddelde score voor de NVAB-producten significant hoger dan voor de arbokennisproducten uitgegeven door STECR en de overige organisaties.

Een absolute grenswaarde voor minimale kwaliteit van arbokennisproducten is niet te bepalen, maar wij vinden de grens van 66,7 betekenisvol (AGREE Collaboration 2003). Een lagere domeinscore betekent immers dat het arbokennisproduct in de ogen van enkele beoordelaars niet aan alle genoemde criteria van dat domein voldoet. Zelfs als we stellen dat een arbokennisproduct aan minimaal de helft van de criteria moet voldoen (minimaal score 'Eens') en daarnaast een lagere score voor één van de andere criteria kan compenseren met een hogere score voor één van de andere criteria) dan blijft de grenswaarde 66,7. Veel domeinen in ons onderzoek halen die drempelscore niet. Geen enkele van de nu onderzochte kennisproducten scoort voldoende op alle domeinen. De helft van de 36 domeinscores van de NVAB-producten ligt boven die grenswaarde, terwijl dat voor 2 van de 48 domeinscores voor de STECR-producten geldt en voor 7 van de 48 domeinscores 'Overig'. Hanteren we een soepeler norm van minimaal een gemiddelde domeinscore van 50, dan scoren 30 van de 36 NVAB-domeinen voldoende; 16 van de 48 STECR-domeinen en 18 van de 'Overige'-domeinen. Van de scores voor het belangrijke domein Methodologie liggen er 17 van de 22 beneden de 50.
In reactie op de stelling: *Deze 'richtlijn' is aan te bevelen voor gebruik in de praktijk* antwoorden de beoordelaars zo, dat slechts 13 van de 22 onderzochte arbokennisproducten gemiddeld 3,00 of hoger scoren.

De AGREE-scores voor de arbokennisproducten uit deze studie kunnen vergeleken worden met AGREE-scores uit andere studies: in dat geval gaat het steeds om documenten die expliciet als richtlijnen bedoeld waren en ook zo werden genoemd. Er zijn diverse cijfers beschikbaar:

1. Een studie naar 100 richtlijnen uit diverse landen die de voorzet vormde voor de ontwikkeling van de AGREE. Bij publicatie over deze gegevens zijn de richtlijnen gesplitst in oncologische en overige richtlijnen (Burgers et al. 2004; Fervers et al. 2005). (tabel 25, kolom 1 en 2).
2. De AGREE-validatiestudie naar 33 richtlijnen uit 11 verschillende landen (AGREE Collaboration 2003). (tabel 25, kolom 3).
3. De AGREE-beoordeling van 5 NHG-Standaarden bedoeld voor huisartsen en 5 recente multidisciplinaire CBO-richtlijnen (Burgers & Boluyt 2004). (tabel 25, kolom 4 en 5)

De resultaten van deze studie vindt u in tabel 25, kolom 6, 7 en 8.

Als we 'onze' AGREE-scores naast de gegevens uit deze studies zetten, valt op hoe de scores voor de NVAB-Richtlijnen op alle domeinen boven de andere uitkomen. Alleen de vijf CBO-producten scoren op 'Methodologie' gemiddeld iets beter dan de NVAB-Richtlijnen. De STECR-producten scoren relatief hoog op 'Doel en Onderwerp'. De domeinscores van de NHG-Standaarden en CBO-Richtlijnen zijn lager. De gemiddelde Methodologie-score van de acht STECR-producten is vergelijkbaar met de score voor de NHG-Standaarden. Het NHG wordt vaak als voorbeeld gebruikt als het om evidence based richtlijnontwikkeling gaat! Voor alle duidelijkheid: deze scores zeggen niet dat de methodologie van de richtlijnontwikkeling gebrekkig is verlopen en de aanbevelingen twijfelachtig van kwaliteit zijn, maar dat er in de richtlijn onvoldoende terug te vinden is over de methodologie.
De overige scores bevinden zich in het middengebied.

Tabel 25. Vergelijking van AGREE-scores uit andere studies

	1	2	3	4	5	6	7	8
	Oncologie	*Other*	*Validatie*	*NHG*	*CBO*	*STECR*	*NVAB*	*Overig*
Doel en Onderwerp	63,2	65,5	69,3	57,8	53,3	56,6	86,6	72,2
Betrokkenheid Belanghebbenden	34,3	29,5	36,1	53,3	61,7	48,4	66,0	37,3
Methodologie	42,2	29,4	40,7	33,3	60,0	32,8	56,9	36,4
Helderheid en Presentatie	57,2	52,3	65,8	80,0	68,3	53,8	75,7	59,0
Toepassing	25,8	29,6	36,9	13,3	44,4	35,0	52,7	19,4
Onafhankelijkheid Opstellers	47,5	40,9	30,3	0,00	43,3	33,3	54,9	28,9
N: aantal 'richtlijnen'	32	68	33	5	5	8	6	8

Zo'n vergelijking van AGREE-scores heeft beperkingen, maar zet zaken wel wat in perspectief. De AGREE-scores voor de diverse arbokennisproducten maken duidelijk dat er op de beoordeelde domeinen ruimte is voor verbetering. Maar deze scores voor arbokennisproducten lopen niet schokkend veel uit de pas met de scores voor richtlijnen in de curatieve zorg. Zelfs een ervaren richtlijnorganisatie als het NHG, die wetenschappelijkheid hoog in het vaandel voert, scoort met 5 recente Standaarden slechts matig op het gebied van 'Methodologie'.

De gemiddelde scores voor ‘Toepassing’ of ‘Helderheid en Presentatie’ van de STECR-producten zijn niet opvallend hoog. In het debat wordt de praktijkgerichtheid van STECR-Werkwijzers tegenover wetenschappelijkheid van richtlijnen gezet, maar deze vergelijking laat geen opvallend hoge STECR-scores zien voor domeinen die met praktijkgerichtheid geassocieerd kunnen worden.
Omdat de AGREE-criteria nauwelijks enige samenhang vertonen met de achtergrondkenmerken of affiliatie van de beoordelaars, mag worden aangenomen dat de beoordelaars naar de inhoud van de documenten gekeken hebben en voorkennis of vooroordelen over bepaalde organisaties geen rol van betekenis hebben gespeeld. We gaan er van uit dat het oordeel op een onafhankelijke manier tot stand is gekomen.

Als beoordelaars worden ingedeeld op hun achtergrond (NVAB, STECR of overig) is dat niet geheel neutraal ten aanzien van de hoogte de AGREE-scores. De AGREE-scores voor NVAB-producten zijn immers systematisch hoger dan voor de andere arbokennisproducten. Omdat NVAB-mensen geen NVAB-Richtlijnen hebben beoordeeld is het mogelijk dat scores van deze beoordelaars gemiddeld wat lager uitkomen dan de overige beoordelaars. Een systematische vertekening van de resultaten door dit effect, verwachten we echter niet, omdat we geen (negatieve) correlatie hebben kunnen aantonen tussen affiliatie van de beoordelaar en afkomst van de arbokennisproducten.

- *Beleidsimplicaties van de gevonden AGREE-scores*

De domein- en itemscores voor elk van de arbokennisproducten maken steeds goed duidelijk op welke punten (soms eenvoudig) verbeteringen door te voeren zijn. In een aantal gevallen is onze stellige indruk dat het een kwestie is van beter opschrijven hoe er gewerkt is bij het ontwikkelen van het arbokennisproduct. We hebben dan aanwijzingen dat er wel degelijk met een criterium rekening is gehouden, alleen staat daarover niets vermeld in de tekst.

De scores voor ‘Onderwerp en Doel’ zijn voor de meeste arbokennisproducten goed. De inleiding van arbokennisproducten met een lage score kan aan de hand van de betreffende AGREE-items gemakkelijk aangevuld worden. ‘Betrokkenheid Belanghebbenden’ levert een gevarieerder beeld op aan scores. We hebben de indruk dat het probleem hier meer is dat bepaalde zaken onvoldoende een rol hebben gespeeld bij het ontwikkelproces (en niet dat men alleen vergeten is het op te schrijven). Om tot betere scores te komen zal het ontwikkelproces structureel uitgebreid moeten worden, bijvoorbeeld met het nagaan van de wensen en voorkeuren van patiënten (en dat misschien opvatten als de wensen en voorkeuren van zowel werkgevers als werknemers). Of het standaard uitvoeren van een goede praktijktest. Zo’n test levert niet alleen meer punten op voor het domein ‘Betrokkenheid Belanghebbenden’ maar levert tevens zinvolle informatie op die voor andere domeinen van belang is, zoals ‘Helderheid Presentatie’ en ‘Toepassing’. Het is op voorhand niet meteen gezegd dat daarmee de inhoud van de aanbevelingen drastisch zal wijzigen. Maar zelfs een gelijkblijvende inhoud kan dan wel met betere argumenten en toelichting onderbouwd worden.

Alle AGREE-domeinscores zijn van belang, maar het domein ‘Methodologie’ verwijst het meeste naar onderbouwing en wetenschappelijkheid van de aanbevelingen die gepresenteerd worden in het arbokennisproduct. Omdat er de laatste jaren een steeds luidere roep klinkt om meer effectonderzoek naar diverse vormen van arbodienstverlening en beter gebruik van onderzoeksresultaten in de praktijk verdient het domein ‘Methodologie’ speciale aandacht. Eigenlijk zouden er geen arbokennisproducten meer uitgebracht moeten worden met een Methodologiescore onder de maat. Er is inmiddels voldoende know-how op dit gebied beschikbaar. De NSPOH verzorgt cursussen Evidence Based Medicine. Voor deelnemers aan kenniskringen of richtlijnwerkgroepen is zo’n cursus of iets vergelijkbaars verplichte kost. In de opleidingen wordt steeds meer aandacht besteed aan het zoeken van literatuur met Pubmed en het kritisch beoordelen van onderzoeksartikelen. Samenwerking met universiteiten (of HBO-opleidingen die ook steeds meer taken op het gebied van onderzoek gaan uitvoeren) is onontbeerlijk. Afstemming tussen kennisinstituten is noodzakelijk om te voorkomen dat op diverse plaatsen hetzelfde werk gedaan wordt. Want om tot goede Methodologiescores te komen moet er grondig en systematisch gewerkt worden. Dat kost tijd, inspanning en deskundigheid. Bij de planning en begroting van een ontwikkeltraject moet daar rekening mee gehouden worden.

In een aantal gevallen zal het ook betekenen dat een arbokennisproduct (volgens de regelen der kunst) niet haalbaar is. Als een werkgroep dan toch een publicatie wil maken van het inhoudelijk werk dat ze gedaan hebben, dan moet in de inleiding heel expliciet vermeld moeten worden waar de aanbevelingen op gebaseerd zijn en wat dat betekent voor de wijze waarop een professional daarmee in

de praktijk kan of mag omgaan. In veel gevallen kan ons inziens beter afgezien worden van het willen maken van een arbokennisproduct. De energie van zo'n werkgroep kan dan veel beter besteed worden aan het maken van goede nascholingsmaterialen, een concreet verbetertraject of een audit op basis van een bestaand arbokennisproduct.
Lage Methodologiescores bieden de arboprofessional navenant veel vrijheidsgraden in het hanteren van zo'n arbokennisproduct in de praktijk. In de huidige vorm kan aan de aanbevelingen uit de STECR-producten, de NVAB-Richtlijn Lage Rugklachten, de LESA-Overspanning en alle arbokennisproducten uit de groep 'Overig' onvoldoende gewicht worden toegekend. Een arboprofessional die in de praktijk een andere aanpak kiest, kan op die grond alleen geen slecht handelen worden verweten.

De scores voor 'Helderheid en Presentatie' zijn over de hele linie redelijk tot goed, met enkele uitschieters naar beneden. Door het uitvoeren van een praktijktest met een conceptversie of een pretest van de nagenoeg definitieve tekst en goed luisteren naar de verbetersuggesties van potentiële gebruikers moet deze domeinscore voor vele arbokennisproducten makkelijk te verbeteren zijn. Werkgroepen verliezen zich soms in de inhoud van het arbokennisproduct waar ze aan werken. Terugbrengen van de inhoud tot de essentie en alle toelichting verplaatsen naar een achtergronddocument (dat eventueel via internet beschikbaar gemaakt wordt) kan een goede oplossing zijn. Ook met samenvattingskaarten, stroomschema's en andere hulpmiddelen kan de score voor dit domein verbeterd worden.

Het is teleurstellend dat de scores voor het domein 'Toepassing' voor de meeste arbokennisproducten zo laag uitvallen. Maar drie arbokennisproducten scoren boven de 50! Het is onvermijdelijk dat arbokennisproducten een ideale aanpak beschrijven en er een zekere kloof is tussen die inhoud en de dagelijkse praktijk. Maar het lijkt erop dat werkgroepen zich nu wel gemakzuchtig van dat probleem afmaken. Als de werkgroep onvoldoende heeft onderzocht hoe de kloof tussen ideale en huidige praktijk overbrugd kan worden, blijft de waarde van zo'n arbokennisproduct beperkt. Het is overigens waarschijnlijk dat als er wél zorgvuldig onderzocht wat de organisatorische belemmeringen zijn bij toepassing van de aanbevelingen, hoe kostenimplicaties een rol spelen en er expliciete toetscriteria toegevoegd worden om na te gaan of de richtlijn wordt gevolgd, dat dan de inhoud van het arbokennisproduct zal veranderen. Aanbevelingen zullen aangepast worden en misschien minder ideaaltypisch, maar wel veel beter uitvoerbaar en inpasbaar moeten worden opgeschreven (Michie & Johnston 2005).

De lage scores voor 'Onafhankelijkheid Opstellers' tenslotte vragen ook om de nodige aandacht. Hier is de verwachting dat flinke verbeteringen gerealiseerd kunnen worden door eenvoudige aanvullingen in de tekst. Heel vaak is er gewoon niets over opgeschreven, terwijl die informatie wel voor handen is, omdat men zich niet gerealiseerd heeft dat die informatie voor de kwaliteit van het arbokennisproduct van belang is. Binnen de curatieve zorg is men al sinds jaar en dag bewust van de invloed die bijvoorbeeld farmaceutische industrieën kunnen hebben op de inhoud van wetenschappelijke publicaties en soms de inhoud van richtlijnen proberen te sturen. Daarom wordt in de AGREE (ontwikkeld in die curatieve zorg) daar zoveel aandacht voor gevraagd. In de arbowereld spelen minstens zoveel financiële belangen, maar het risico op beïnvloeding van onderzoek, kennis en de inhoud van richtlijnen wordt nog weinig gezien, maar is natuurlijk wel degelijk aanwezig. Het is voor de geloofwaardigheid van arbokennisproducten essentieel dat arboprofessionals erop kunnen vertrouwen dat de inhoud strikt onafhankelijk tot stand gekomen is en niet direct of indirect door financiers of bepaalde belangenpartijen is beïnvloed.

<u>Spreiding en interbeoordelaarsbetrouwbaarheid</u>

Elk arbokennisproduct is door meerdere beoordelaars gescoord (meestal vier). Die waren het onderling overigens lang niet altijd eens. Uit de maten voor het gebrek aan overeenstemming en de spreiding in scores (voldoet wel/niet criterium, maximale range, ICC en standaarddeviatie) komt naar voren dat er een aanzienlijke variatie is.

- *Beschouwing overeenstemming tussen beoordelaars*

De officiële AGREE-instructie beveelt minimaal twee verschillende beoordelaars aan per arbokennisproduct. In dit geval is het merendeel van de arbokennisproducten door vier mensen beoordeeld.

De beoordelaars komen zelden op exact dezelfde itemscore uit bij een en het zelfde arbokennisproduct: in 14% van alle 506 itemscores voor 22 arbokennisproducten. Als een verschilscore van 1 categorie hoger of lager geaccepteerd wordt, dan vallen 52,4% van de 506 scores in het acceptabele gebied. In 31,0% van de gevallen is de maximale range 2. In 16,6% van de gevallen gebruiken de beoordelaars beide uitersten: zowel 'Zeer eens' als 'Zeer oneens' (een range van 3) bij hetzelfde item bij hetzelfde arbokennisproduct. Burgers en Boluyt (2004) vinden in hun studie voordat de twee (!) beoordelaars overleg gehad hebben in 10,4% van de gevallen een maximaal verschil van score.
De interbeoordelaarsbetrouwbaarheid wordt vaak met Cohen's Kappa in getal uitgedrukt als het gaat om 2 beoordelaars. In dit geval zijn het er meer dan twee, varieert het aantal per arbokennisproduct en is de verdeling van waarden regelmatig scheef. Daarom is in dit geval de intraclass correlatie coëfficiënt het meest aangewezen (Yaffee 1998; Hoehler 2000; McGraw & Wong 1996; Nicols 1998; Stemler 2004).

Een ICC kan alleen berekend worden voor de arbokennisproducten als AGREE-itemscores door alle beoordelaars zijn ingevuld. Omdat niet alle arbokennisproducten door vier beoordelaars zijn gescoord en er soms itemscores niet zijn ingevuld, kan een deel van de gegevens niet worden betrokken in de berekening van de ICC. De portie gegevens die buiten de boot valt wordt bepaald door de invoervolgorde. De ICC waarden voor vier beoordelaars zijn het meest indicatief omdat bij die berekening de kleinste portie gegevens (15%) buiten de berekening valt.
De ICC's voor één beoordelaar zijn over de hele linie laag. We constateren een duidelijke trend dat de ICC's voor het gemiddelde van een aantal beoordelaars verbeteren bij een toename van het aantal beoordelaars van twee naar vier personen. De item-ICC's liggen bij vier beoordelaars vaker in het acceptabele gebied (> 0,6) dan bij twee en ook zijn er meer ICC-waarden significant. Dezelfde trend is waarneembaar op domeinniveau.
Voor de berekening van ICC's voor vijf of zes beoordelaars zijn onvoldoende gegevens aanwezig. MacDermid ea. constateerde dat toename van twee naar drie beoordelaars de ICC flink verbeterde maar verdere toename incoherente resultaten opleverde: soms een verbetering, soms niet (MacDermid 2005). Zij had gegevens van zes in het gebruik van de AGREE getrainde fysiotherapeuten die elk zes fysiotherapierichtlijnen beoordeelden. Burgers (2002) vermeldt in zijn proefschrift ICC waarden uit de AGREE-validatiestudie. De ICC waarden voor 'Helderheid Presentatie' en 'Onafhankelijkheid Opstellers' zijn substantieel lager dan in de andere twee studies. De overige ICC waarden zijn grofweg vergelijkbaar.

Tabel 26. Vergelijking IntraClass Correlatie Coëfficiënten uit diverse studies

ICC gemiddelde voor 4 beoordelaars	*Deze studie 2005*	*Burgers ea 2002*	*MacDermid e.a. 2005*
Doel en Onderwerp	0,71	0,76	0,73
Betrokkenheid Belanghebbenden	0,84	0,72	0.88
Methodologie	0,80	0,88	0,93
Helderheid Presentatie	0,59	0,69	0,80
Toepassing	0,72	0,79	0,82
Onafhankelijkheid opstellers	0,47	0,64	0,69
N	17	33	55

- *Beschouwing spreiding van de AGREE-scores per item en domein*

Als we het 'voldoet wel/niet criterium' gebruiken, de maximale range, de ICC en de standaarddeviatie dan zijn er een aantal items die regelmatig tot uiteenlopende scores per arbokennisproduct leiden:

1. Het doel van de richtlijn is specifiek beschreven.
2. De klinische vraag/vragen die in de richtlijn aan de orde komt/komen, is/zijn specifiek.
10. De gebruikte methoden om de aanbevelingen op te stellen, zijn duidelijk beschreven.
11. Gezondheidswinst, bijwerkingen en risico's zijn overwogen bij het opstellen van de aanbevelingen.
17. De kernaanbevelingen zijn gemakkelijk te herkennen.
18. De toepassing van de richtlijn wordt ondersteund met hulpmiddelen.

De volgende items zijn relatief eenduidig, afgaande op de spreidingsmaten:

3. De patiëntenpopulatie waarop de richtlijn van toepassing is, is specifiek beschreven.
6. De beoogde gebruikers van de richtlijn zijn duidelijk benoemd.

9. De criteria voor het selecteren van het wetenschappelijk bewijsmateriaal zijn duidelijk beschreven.
12. Er bestaat een expliciet verband tussen de aanbevelingen en het onderliggende wetenschap pelijke bewijsmateriaal.
14. Een procedure voor herziening van de richtlijn is vermeld.
23. Conflicterende belangen van leden van de werkgroep zijn vastgelegd.

Het karakter van de items verschilt: sommige zijn heel concreet geformuleerd (bijvoorbeeld 14 en 23), in andere is meer ruimte voor eigen interpretatie en waardeoordelen (bijvoorbeeld 12 en 18). In enkele van de items met veel spreiding komen woorden voor die in een arbocontext enige 'vertaling' behoeven (2 en 11). Uit andere studies komen soms andere items naar voren met een lagere overeenstemming in scores voordat er overleg geweest is (MacDermid et al. 2005; Burgers & Boluyt 2004). Daar is geen duidelijk patroon uit af te leiden. De verwachtingen dat de formulering van sommige AGREE-items mogelijk problemen op kunnen leveren, wordt maar deels bevestigd. Het is plezierig dat de items van een belangrijk domein als 'Methodologie' relatief weinig verschil van mening tussen de beoordelaars laten zien.

- *Beschouwing spreiding van de AGREE-scores per arbokennisproduct*

Bij de volgende arbokennisproducten is er sprake van relatief veel verschil van mening tussen de beoordelaars: De STECR-Werkwijzer ABBE-rsi, vooral in de domeinen 'Methodologie' en 'Helderheid Presentatie'. De STECR-Werkwijzer Somatisatie vooral in diezelfde twee domeinen en ook bij het domein 'Toepassing'. De LESA-Overspanning vooral in de domeinen 'Onderwerp en Doel' en 'Betrokkenheid Belanghebbenden'. De scores Gebruiksgemak AGREE voor deze arbokennisproducten zijn gemiddeld. Het is dus niet zo dat deze arbokennisproducten wanordelijker in elkaar zitten of benodigde AGREE-informatie lastiger is te vinden. De beschikbare informatie in deze producten wordt blijkbaar verschillend beoordeeld in het licht van de AGREE-criteria. De NVAB-Richtlijn Lage Rugklachten en de Lisv/UWV standaard Geen Duurzaam Benutbare Mogelijkheden leveren de meest uniforme scores op. Afgaande op de spreidingsmaten komen er geen algemene kenmerken bovendrijven van arbokennisproducten die beter niet met de AGREE beoordeeld kunnen worden. Er zijn twee STECR-Werkwijzers waar de scores tussen de beoordelaars weinig overeenstemmen, maar er zijn ook werkwijzers - en andere STECR-producten - waar de scores weinig verschillen. Het is eerder omgekeerd: zelfs een Lisv/UWV standaard – bepaald een buitenbeentje in dit scala producten - levert weinig meningsverschil op tussen de beoordelaars.

- *Beleidsimplicaties naar aanleiding van de gevonden spreiding*

De spreiding in scores is aan de hoge kant en verdient nader onderzoek. In onze studie is er geen training van de beoordelaars geweest: men kreeg een AGREE-handleiding en invulinstructie. Het is waarschijnlijk dat iets meer instructie en training in het gebruik van de AGREE tot meer overeenstemmende resultaten zal leiden. Bovendien verdient het aanbeveling om overleg te stimuleren of daadwerkelijk te organiseren tussen beoordelaars als uiteenlopende itemscores geconstateerd zijn. De overeenstemming zal dan toenemen en de ICC waarden zullen dan verbeteren. Het is echter de vraag of we moeten streven naar geheel gelijkluidende scores. Door beoordelaars te rekruteren met diverse achtergronden wordt recht gedaan aan de multidisciplinaire en veelzijdige context waarin de arbokennisproducten toegepast moeten worden. De wijze waarop de AGREE-items geformuleerd zijn, biedt uitdrukkelijk ruimte voor waardeoordelen Het is normaal dat beoordelaars met andere achtergronden tot andere afwegingen komen. Daar wordt de AGREE-score niet door verzwakt, maar versterkt.

Op basis van deze gegevens doen we het advies om bij de AGREE-beoordeling van arbokennisproducten minimaal vier beoordelaars in te zetten met een gevarieerde achtergrond. Later onderzoek moet aantonen of het optimale aantal beoordelaars per arbokennisproduct nog boven die vier ligt.
Als we spreidingsmaten vergelijken voor de verschillende items en domeinen dan zijn er voorlopig geen aanwijzingen die herformuleren van items voor de arbocontext noodzakelijk maken. Dat is een belangrijke bevinding!
Op basis van diezelfde spreidingmaten kunnen niet op voorhand soorten arbokennisproducten geïdentificeerd worden waarvoor de AGREE niet geschikt is. Het lijkt erop dat een heel breed spectrum aan arbokennisproducten op een gelijkluidende manier met de AGREE beoordeeld kan worden.

4.3 Relevantie AGREE-items en mogelijke aanvullingen

De 23 AGREE-items scoren worden alle relevant geacht door de beoordelaars. In 90% van de gevallen is gekozen voor ‘Zeer eens’ of ‘Eens’. Er zijn enkele items waar relatief wat vaker voor ‘Zeer oneens’ of ‘Oneens’ gekozen is, maar dat is nooit meer dan een kwart (gemiddeld over alle items en beoordeelde arbokennisproducten). Er zijn enkele verschillen in de mate waarin de items voor de diverse arbokennisproducten relevant geacht worden, maar die zijn klein. Er is geen systematische samenhang gevonden van de relevantiescore met het arbokennisproduct dat beoordeeld is. Alleen voor het Hoofdstuk Diabetes Mellitus uit het handboek Arbeid en Belastbaarheid wordt de relevantie van de AGREE-items zo vaak betwijfeld dat de gemiddelde relevantiescore over alle items bij dat arbokennisproduct onder de 3,00 uitkomt.

Er wordt nagenoeg geen samenhang gevonden tussen de relevantiescores en de achtergrondkenmerken van de beoordelaars. Iedereen vindt dezelfde criteria van belang. Voor zover die samenhang wel in geringe mate aanwezig is, wijst die erop dat mensen werkzaam in de praktijk de items die betrekking hebben op toepassing iets belangrijker vinden dan anderen.

Ons is geen onderzoek bekend waarin de relevantie van de AGREE-items in een specifieke context beoordeeld werd. Er zijn geen onderzoeksgegevens waarmee de resultaten uit dit onderzoek vergeleken kunnen worden. Het merendeel van de respondenten heeft niets toe te voegen aan de criteria van de AGREE. Er worden enkele aanvullingen op de AGREE-items voorgesteld.

- *Beschouwing ervaren omissies in het AGREE-instrument*

Van de aanvullingen die worden voorgesteld op de AGREE-items heeft een aantal ons inziens betrekking op aspecten die al door de AGREE-items gedekt worden. Bijvoorbeeld de kwaliteit van de verwijzingen en referenties, de volledigheid van de informatie, de praktijkrelevantie, reacties vragen aan een steekproef van de beoogde gebruikers, of de mate waar ineen richtlijn aansluit bij een van tevoren gepeilde behoefte.

De mate van commitment of participatie van werkgevers en werknemers is een interessante aanvulling specifiek voor de arbowereld. Feitelijk valt dat criterium onder ‘Betrokkenheid Belanghebbenden’. Men zou kunnen overwegen om item 5 (nagaan voorkeuren van patiënten) met wat toelichting in de invulinstructie die betekenis te geven. Een andere mogelijkheid zou zijn om dat aspect door middel van extra toelichting onder item 19 en/of 20 te vatten (het bespreken van organisatorische belemmeringen bij toepassing van aanbevelingen en het overwegen kostenimplicaties). Maar let wel: de onafhankelijkheid van de opstellers, een ander AGREE-criterium kan in gevaar komen als dergelijke partijen daadwerkelijk invloed gaan uitoefenen op het samenstellen van een arbokennisproduct.
Een beoordelaar merkt op dat de AGREE lijkt uit te gaan van één (klinische) interventie, terwijl in de arbopraktijk vaak meerdere vormen van begeleiding tegelijk uitgevoerd worden. Hij vraagt zich af of in de AGREE niet gevraagd moet worden naar prioritering van maatregelen of aanbevelingen. Wij zijn van mening dat wat opgemerkt wordt geen exclusief kenmerk van de arbopraktijk is. Bij richtlijnen voor de gezondheidszorg is eveneens vaak sprake van meerdere interventies die tegelijk uitgevoerd worden of kunnen worden. Een goede richtlijn geeft dan aanwijzingen voor de prioritering van die interventies of interactie daartussen. Die richtlijninhoud kan gewoon met de bestaande AGREE-criteria beoordeeld worden. Daar zijn geen aanvullende items voor nodig.

Enkele beoordelaars stellen aanvullingen voor op het gebied van multidisciplinaire aspecten, eventueel scoren van de AGREE per discipline of aandacht voor arbocuratieve samenwerkingsaspecten en de mate waarin aandacht besteed wordt aan diagnostiek en behandeling door andere aanpalende disciplines en de algemene gerichtheid op samenwerking. Voor deze punten volgen we dezelfde redenering als hiervoor: multidisciplinaire samenwerking is geen exclusief arbothema, maar is ook in de reguliere zorg belangrijk en wordt daar steeds belangrijker. De AGREE-criteria zijn volgens ons voldoende om de kwaliteit van aanbevelingen op dat gebied te beoordelen. Overigens zijn alle STECR-Werkwijzers multidisciplinair van aard en geen van de beoordelaars heeft opmerkingen gemaakt dat zij die aspecten met behulp van de AGREE niet goed konden beoordelen.
Één van de beoordeelde arbokennisproducten was zelfs expliciet opgesteld met het doel om de samenwerking en afstemming van beleid tussen huisarts en bedrijfsarts te verbeteren. De AGREE (zonder aanvullingen) bleek goed geschikt te zijn om de kwaliteit van zo’n landelijke eerstelijns samenwerkingsafspraak (LESA-Overspanning) te beoordelen.

Eén beoordelaar (die vanuit zijn functie zicht had op veel verschillende arbokennisproducten) stelde voor om de AGREE-criteria uit te breiden met prijs en het gemak van beschikbaarheid. Dat is een sympathiek idee, maar eigenlijk vinden we dat goede arbokennisproducten minimaal in elektronische vorm gratis voor de individuele arboprofessional beschikbaar moeten zijn (en eventueel tegen geringe betaling in gedrukte vorm). Dat kan wel betekenen dat een organisatie betaalt voor die beschikbaarheid voor haar personeel. Op zo'n manier wordt bijvoorbeeld tegenwoordig bij universiteiten veel wetenschappelijke artikelen uit tijdschriften online beschikbaar gemaakt voor de medewerkers. Er zijn ook al arbodiensten die op die manier de toegang tot informatie voor haar medewerkers geregeld heeft (van Dijk et al. 2003; Maassen & Noort 2004).
Goede arbokennisproducten zijn gebaseerd op openbaar beschikbare (wetenschappelijke) kennis en worden vaak met overheidsmiddelen ontwikkeld. Dat strookt niet met hoge prijzen voor arbokennisproducten. Het zou de onafhankelijkheid van de opstellers in gevaar kunnen brengen. Dat criterium is al onderdeel van de AGREE. We hebben de openbare brede beschikbaarheid onderdeel gemaakt van de definitie van een arbokennisproduct. Als er te veel drempels opgeworpen worden voor de beschikbaarheid, zoals bijvoorbeeld het geval kan zijn bij interne protocollen van een arbodienst, dan vinden wij dat de aanduiding 'richtlijn' of 'arbokennisproduct' op een onjuiste manier gebruikt wordt.

Iemand stelt voor een criterium toe te voegen aan de AGREE met de implicaties bij toetsing door UWV als de professional een richtlijn niet volgt. Volgens ons wordt dan het paard achter de wagen gespannen. Het UWV zou bij het bepalen van de gevolgen juist rekening moeten houden met het de kwaliteit van het arbokennisproduct. We komen daar later op terug.

- *Beleidsimplicaties relevantie AGREE-items*

De groep beoordelaars is breed samengesteld. Het gaat om personen die meer of minder direct verbonden zijn aan tal van belangrijke kennisorganisaties op arbogebied. De groep is gevarieerd wat achtergrondkenmerken betreft. De arbokennisproducten die zijn beoordeeld, zijn eveneens gevarieerd. Omdat we nagenoeg geen samenhang gevonden hebben van de relevantiescores met de achtergrondkenmerken, kunnen de gevonden resultaten ten aanzien van de relevantie van de AGREE-items goed gegeneraliseerd worden.
We kunnen concluderen dat de AGREE-items op een breed draagvlak mogen rekenen in de Nederlandse arbowereld. In de AGREE-items worden belangrijke criteria verwoord om de kwaliteit van allerlei soorten arbokennisproducten te kunnen beoordelen. De achtergrond van de beoordelaar kleurt het relevantieoordeel een beetje, maar niet noemenswaardig of verbazingwekkend: mensen uit de praktijk hechten meer belang aan toepassingsitems.
Eventueel kunnen apart enkele extra items aan de AGREE toegevoegd worden als deze gebruikt wordt in een arbocontext, maar veel aanleiding zien wij daar niet voor. Voor een deel kan die inhoud ook opgevangen worden door een goede toelichting te geven bij het AGREE-instrument. Het is onwenselijk om de vergelijkbaarheid van AGREE-scores in gevaar te brengen door het instrument te veranderen.

4.4 Richtlijnigheid, gebruiksgemak en waardering AGREE

Het richtlijnkarakter van de onderzochte arbokennisproducten loopt uiteen: uitgedrukt in score voor richtlijnigheid van 2,44 tot 3,94. De gemiddelde score voor Richtlijnigheid van de STECR-producten (doorgaans minder opgezet als richtlijn) is 2,85, voor de NVAB-producten, 3,71 en voor de overige producten 3,02.

Het Gebruiksgemak van de AGREE scoort tussen de 2,67 en 3,67 per arbokennisproduct, gemiddeld 3,19. Men heeft de vragen goed kunnen beantwoorden en de benodigde informatie makkelijk kunnen vinden. Zelden heeft men aanvullende informatie opgezocht en gebruikt. Bij 4 arbokennisproducten is die score lager dan 3,00. Dat wijst erop dat de benodigde informatie bij die arbokennisproducten moeilijker te vinden is in het document.

De gemiddelde score Waardering AGREE per arbokennisproduct bedraagt 2,96 (minimum 2,00 en maximum 4,00). Ten aanzien van enkele arbokennisproducten bestaat twijfel of de AGREE wel een geschikt instrument is om de kwaliteit te beoordelen. Beoordelaars uit STECR-kring komen tot gemiddeld tot een hogere waardering voor het AGREE-instrument dan beoordelaars van NVAB of andere organisaties.

Er zijn geen vergelijkingsgegevens voor deze toegevoegde vragen beschikbaar.
Door de brede variatie in respondenten en arbokennisproducten kunnen de resultaten op dit terrein goed gegeneraliseerd worden.

- *Beschouwing richtlijnigheid van arbokennisproducten*

In het debat over de kwaliteit van arbokennisproducten wordt naar onze mening soms onterecht betoogd dat een document geen richtlijn is en dus niet met die richtlijncriteria beoordeeld moet worden. Ook het omgekeerde komt voor: iets wordt een richtlijn genoemd, terwijl dat bij nadere beschouwing ons inziens niet terecht is .Daarom zijn in dit onderzoek bewust arbokennisproducten opgenomen waarvan het richtlijnkarakter ter discussie kan worden gesteld, zoals bijvoorbeeld de beide wettelijk georiënteerde standaarden van UWV, het hoofdstuk Diabetes Mellitus uit het Handboek Arbeid en Belastbaarheid en de webteksten van de kennisbank met interventies voor acute lage rugklachten van het NKAB. We wilden nagaan of het richtlijnkarakter op een meer objectieve manier kon worden vastgesteld. Als de geraadpleegde experts het document als een richtlijn beschouwen, zijn de criteria van de AGREE van toepassing en kan geen aanspraak gemaakt worden op een eigen set criteria. We hebben daarvoor de score voor Richtlijnigheid geïntroduceerd, waarin vier essentiële kenmerken van richtlijnen worden samengevat:

1. *Het gaat om een document met specifieke, expliciete uitspraken voor goed handelen bij een welomschreven probleem of doelpopulatie (aanbevelingen).*
   Dit in tegenstelling tot een discussienota, een verslag met onderzoeksgegevens, een pamflet of algemeen leerboek.
2. *Het document beoogt het handelen van veel arboprofessionals te beïnvloeden. Het is bedoeld voor brede verspreiding.*
   Dit in tegenstelling tot een heel beperkt of besloten gebruiksgebied.
3. *De organisatie die de tekst heeft ontwikkeld of uitgegeven het vertrouwen van het merendeel van de bedrijfsartsen en arboprofessionals in de praktijk.*
   Dit in tegenstelling tot reclame en herkenbaar commerciële informatie, waarbij professionals zelf meewegen dat de inhoud gekleurd zal zijn door de belangen van de uitgever.
4. *Het document zal door het merendeel van de praktijk gezien worden als een richtinggevend document op het terrein van de bedrijfsgeneeskunde en/of arbozorg, los van eventuele aanduidingen in de titel, inleiding of omschreven bedoeling*
   Hiermee leggen we de nadruk op de functie en het feitelijk gebruik in de ogen van de arboprofessionals en leggen we allerlei verschillende aanduidingen die de opstellers gebruiken terzijde.

Omdat het merendeel van de STECR-Werkwijzers door de respondenten volgens de bovengenoemde kenmerken als richtlijn gezien wordt (scores hoger dan 3,00) kunnen we ervan uitgaan dat de AGREE-criteria ook voor die arbokennisproducten opgaan. Discussie is mogelijk over de match tussen AGREE-criteria en doel en functie van de Werkwijzer ABBE-rsi en, omdat het richtlijnkarakter daarvan door de respondenten wordt betwijfeld (Richtlijnigheid 2,44). De beoordelaars vinden dat het document onvoldoende expliciete aanbevelingen bevat en door de praktijk niet als richtinggevend document gezien wordt.

Ook de Werkwijzer Lage Rugklachten en Arbeidsomstandigheden scoort laag op Richtlijnigheid: 2,75). STECR gaf zelf vooraf al aan dat de Assist over allochtonen vooral bedoeld was om discussie aan te wakkeren en het onderwerp op de agenda te zetten. Dat de Assist wat lager scoort op Richtlijnigheid (2,88) is dan in overeenstemming met doel en functie. (Hetzelfde geldt voor de voorstudie met veel literatuuronderzoek en de bijbehorende Assist Vrouwen niet in de WAO). Zo'n combinatie van eigenlijk twee aparte producten is moeilijk in één oordeel te vatten. De lage richtlijnigheidsscore van 2,44 is daarom niet verbazingwekkend.

De richtlijnigheid van de NVAB-Richtlijnen scoort hoog, maar dat is gezien die aanduiding weinig verrassend. Bij de overige arbokennisproducten valt op dat de SMB-richtlijn Rugklachten van Arbouw laag op richtlijnigheid scoort, ondanks de aanduiding 'richtlijn' in de titel.

- *Gebruiksgemak AGREE*

In het algemeen levert het invullen van de AGREE weinig problemen op. Men heeft de vragen goed kunnen beantwoorden en de benodigde informatie makkelijk kunnen vinden. Bij enkele arbokennisproducten is het wat lastiger. Zelden is er aanvullende informatie opgezocht en gebruikt.

- *Waardering AGREE*

De score van bijna 3 voor Waardering AGREE betekent dat men veel positiefs verwacht van het toepassen van de AGREE als instrument om de kwaliteit van arbokennisproducten te beoordelen. Dat oordeel verschilt wel wat per arbokennisproduct, maar dat verschil is niet structureel als we de gemiddelde scores van de arbokennisproducten van STECR, NVAB of de overige organisaties vergelijken. Ten aanzien van enkele arbokennisproducten bestaat twijfel of het gebruiken van de AGREE veel meerwaarde heeft.
Mogelijk dat het AGREE-instrument voor de beoordelaars uit STECR-kring wat nieuwer was en daarom gemiddeld wat meer gewaardeerd werd dan door beoordelaars van NVAB of andere organisaties. Enkele beoordelaars uit STECR-kring schreven als commentaar op de beoordelingsformulieren dat ze nu ze kennis hadden genomen van de AGREE-criteria een volgend ontwikkeltraject van een werkwijzer anders zouden aanpakken.

De schaal bestaat uit acht vragen, waarvan enkele scores ook apart betekenis hebben. In twee daarvan wordt bijvoorbeeld verwezen naar potentiële problemen die beoordelaars kunnen ervaren bij het invullen van de AGREE:

- *Het is hinderlijk dat de AGREE-vragen onvoldoende zijn toegespitst op de arbocontext van dit specifieke kennisproduct.*
- *De AGREE is een te gedetailleerd instrument in verhouding tot de bedoeling en reikwijdte van dit arbokennisproduct.*

In zijn algemeenheid blijken deze problemen geen grote rol te spelen. Maar in ongeveer een derde van de gevallen wordt toch gekozen voor '(Zeer) eens'. Bij sommige arbokennisproducten zoals de STECR-Werkwijzer Arbeidsconflicten en het Protocol arbeidshulpverlening klachten reductie en stressmanagement van het NKAP heeft men hinder van de geringe toespitsing op de arbocontext. Voor het beoordelen van de registratierichtlijnen beroepsgebonden burnout/overspanning van het NCvB vindt men de AGREE te uitgebreid.

- *Omdat dit document niet als 'richtlijn' bedoeld is, zijn de vragen van de AGREE niet relevant.*

Slechts in 16% van de formulieren wordt hier bevestigend gereageerd. De gemiddelde scores voor alle arbokennisproducten zijn laag, behalve voor het Hoofdstuk Diabetes Mellitus uit het Handboek Arbeid en Belastbaarheid (3,33). We hadden eerder al gevonden dat de richtlijnigheidsscore van dit arbokennisproduct laag was. Dus ook als een document niet expliciet als richtlijn bedoeld is, worden de AGREE-criteria relevant gevonden.

- *Als dit arbokennisproduct verspreid wordt onder veel arboprofessionals is het wenselijk dat het voorzien wordt van een onafhankelijk keurmerk (bijvoorbeeld gebaseerd op de AGREE-beoordeling).*

Een ruime meerderheid van de beoordelaars steunt deze suggestie, al wordt een onafhankelijk keurmerk bij bepaalde arbokennisproducten van meer belang geacht dan bij andere. De overwegingen om een kleurmerk niet zo nodig te vinden, lijken afgaande op de toelichting die respondenten soms er bij schreven van verschillende aard te zijn. Soms vindt men dat de uitgevende organisatie (in dit geval de NVAB) zelf al voldoende kwaliteitswaarborg is; een keurmerk zou dan dubbelop zijn. In andere gevallen vindt men een keurmerk een te zwaar middel vanwege het type arbokennisproduct, zoals bij het Hoofdstuk Diabetes Mellitus uit het Handboek Arbeid en Belastbaarheid.

- *Als een bedrijfsarts of arboprofessional weet in welke mate een aanbeveling in dit arbokennisproduct gebaseerd is op wetenschappelijk bewijs dan zal dat invloed hebben op de ruimte die hij/zij professioneel neemt om in de praktijk van deze aanbeveling af te wijken.*

Deze vraag raakt in het hart van dit onderzoek, namelijk de relatie tussen onderbouwing van een arbokennisproduct en de professionele vrijheidsgraden in de praktijk. De beoordelaars zijn het voor de meeste arbokennisproducten eens met deze stelling. Met name bij de arbokennisproducten van STECR en de NVAB zijn nauwelijks mensen die het oneens zijn. Dat betekent dat de beoordelaars een transparante onderbouwing van aanbevelingen in deze arbokennisproducten van groot belang vinden.

- *De vragen van de AGREE zouden arboprofessionals en bedrijfsartsen in de praktijk helpen bij het zelf beoordelen van de inhoud van dit arbokennisproduct.*

Daarnaast nog is er nog een vraag in dit verband gesteld, die niet is meegenomen in de schaal:
- *Het merendeel van de bedrijfsartsen en arboprofessionals in de praktijk heeft onvoldoende kennis om de inhoudelijke kwaliteit van dit arbokennisproduct goed te kunnen beoordelen.*

We vinden hier een gedifferentieerd beeld. De score voor de eerste vraag varieert per arbokennisproduct: bij 13 arbokennisproducten is de gemiddelde score lager dan 3,0. Van de beoordelaars betwijfelt 37% of bedrijfsartsen en arboprofessionals voldoende kennis hebben om de inhoudelijke kwaliteit goed te kunnen beoordelen. Niet bij de overige arbokennisproducten: daar scoort men nagenoeg altijd 'oneens'. Maar bij de STECR-Werkwijzers ABBE-rsi, Lage Rugklachten en arbeidsomstandigheden, het rapport met Assist Vrouwen in de WAO, de NVAB-Richtlijn Astma en COPD en de Leidraad Verwijzen door de bedrijfsarts vindt men impliciet dat bij een beoordeling beter experts ingeschakeld kunnen worden.

- *Beleidsimplicaties Richtlijnigheid, gebruiksgemak en waardering AGREE*

Een lagere Richtlijnigheidsscore opent in principe de mogelijkheid om andere, nog niet genoemde criteria in de beoordeling van arbokennisproducten te betrekken. Maar de zes AGREE-domeinen bestrijken een dermate breed gebied aan kwaliteitscriteria voor arbokennisproducten dat niet meteen duidelijk is welk nieuw relevant criterium toegevoegd zou moeten worden om bijvoorbeeld een specifieke functie wel goed voor het voetlicht te brengen. Er zijn ons in dat verband geen voorstellen voor nieuwe criteria bekend. Bij de onderzochte brede range aan arbokennisproducten hebben de beoordelaars maar zelden last van het feit dat het document niet als 'richtlijn' bedoeld was en kunnen nog steeds goed uit de voeten met de AGREE-criteria. We gaan ervan uit dat de AGREE-criteria onverkort van toepassing zijn, ook bij een lagere score voor richtlijnigheid.

Naarmate een document een hogere Richtlijnigheidsscore heeft, is een lage Methodologiescore meer problematisch. In de Methodologiescore wordt immers uitgedrukt hoe de aanbevelingen op een solide en transparante manier met onderzoeksgegevens zijn onderbouwd. Een lage score ondergraaft de geloofwaardigheid van de inhoud van een arbokennisproduct. Juist bij documenten die door de praktijk als richtlijn gezien en gehanteerd worden is dat onaanvaardbaar. Een voorbeeld van zo'n problematische combinatie van scores is de STECR-Werkwijzer Somatisatie met een Methodologiescore van 36,9 en een Richtlijnigheidsscore van 3,00. Dat arbokennisproduct wordt door de beoordelaars wel degelijk als een richtlijn gezien, maar die is in hun ogen slecht onderbouwd.

In het debat over de benodigde kwaliteit van arbokennisproducten zijn soms te gemakkelijke argumenten gebruikt en geaccepteerd. Richtlijncriteria werden in algemene zin soms niet van toepassing verklaard door uitgaven in een bijzondere, eigen categorie te plaatsen. Die redenering is ons inziens door de richtlijnigheidsgegevens ontkracht. De richtlijnigheid van bijvoorbeeld STECR-Werkwijzers mag dan wat minder zijn dan arbokennisproducten die expliciet het woord 'richtlijn' in de titel dragen, het richtlijnkarakter blijft desalniettemin ruimschoots overeind. En adeldom verplicht: een document met een aanbevelingen voor passend of 'goed' handelen bij een omschreven probleem of doelpopulatie ter ondersteuning van de dagelijkse praktijk van veel arboprofessionals gebaseerd op systematisch verzamelde kennis, steekhoudende argumenten, seniorervaring, kritische beoordeling en consensus, uitgebracht door een belangenvrije organisatie - kortom: een arbokennisproduct - moet nu eenmaal aan minimale kwaliteitscriteria voldoen.

Op basis van de gevonden resultaten ten aanzien van het gebruiksgemak van de AGREE lijkt er geen belemmering te zijn om de AGREE te gebruiken voor de kwaliteitsbeoordeling van arbokennisproducten.
De AGREE wordt door de NVAB al actief gebruikt bij het ontwikkelen van richtlijnen. Het is goed dat nu ook veel mensen uit STECR-kring op de hoogte zijn van het instrument en de waarde ervan beseffen. Ook andere organisaties die arbokennisproducten uitbrengen doen er goed aan de lijst met AGREE-criteria nog een goed door te nemen en te bekijken hoe zij hun uitgaven kunnen opwaarderen met die criteria als leidraad.

De beoordeling van arbokennisproducten door middel van de AGREE kan op allerlei manieren ingezet en gebruikt worden. Omdat er veel steun is voor een onafhankelijk keurmerk gebaseerd op de

AGREE-scores, verdient dat voorstel nader onderzoek. Ook al omdat men van mening is dat het beoordelen van arbokennisproducten met behulp van de AGREE in veel gevallen expertise vraagt die niet bij elke arboprofessional in de praktijk beschikbaar is. Het realiseren van zo'n keurmerk vraagt echter wel de nodige voorbereiding en een goede infrastructuur. Het UWV zou met zo'n keurmerk een goed instrument in handen hebben om te kunnen bepalen welke richtlijnen daadwerkelijk richtinggevend voor het handelen van arboprofessionals dienen te zijn en welke (nog) niet. Een arboprofessional die afwijkt van aanbevelingen uit een arbokennisproduct met een lage AGREE-score (met name voor 'Methodologie') kan ons in ziens op die grond alleen geen slecht handelen verweten worden. Op een zelfde manier zou een AGREE-keurmerk een rol kunnen gaan spelen bij de toenemende juridisering van richtlijnen (Hoge Raad der Nederlanden 2005; Crul & Legemaate 2005).
De kennisinfrastructuur voor arbodiensten in Nederland zou een goede impuls krijgen als allerlei mensen, van bedrijfsartsen en arboprofessionals tot verzekeringsartsen, P&O-functionarissen, beleidsmedewerkers van werknemers- en werkgeversorganisaties, op een eenvoudige manier kunnen achterhalen hoe goed een arbokennisproduct nu eigenlijk is.

5 Samenvatting en conclusies

5.1 Doel en vraagstelling van het onderzoek

Als onderdeel van het Programma ter versterking van de kennis- en onderzoeksinfrastructuur van de arbeids- en bedrijfsgeneeskunde in arbodiensten (KIS) is dit project uitgevoerd om arboprofessionals te helpen bij het beoordelen van de waarde van richtlijnen en andere documenten waarin aanbevelingen gegeven worden voor de adequate aanpak van arbeidsgerelateerde problemen (hierna aangeduid als 'arbokennisproducten'). Het project vindt zijn aanleiding in het debat over de verschillen tussen diverse soorten arbokennisproducten zoals de NVAB-Richtlijnen, de STECR-Werkwijzers en de diagnostische protocollen van het NCvB. In de praktijk is niet altijd duidelijk wat de relevantie is van de geconstateerde verschillen in het bijzonder wanneer ze eenzelfde onderwerp behandelen. Het centrale vraagstuk daarbij was of een bestaand instrument voor de beoordeling van richtlijnen in de gezondheidszorg (AGREE) ook bruikbaar is om de kwaliteit van arbokennisproducten te beoordelen. Dit vraagstuk is in de volgende vraagstellingen uitgewerkt:

1. Hoe zijn de AGREE-scores voor de diverse arbokennisproducten en hoe verhouden deze zich tot elkaar?
2. Levert de AGREE ook in een arbocontext valide en betrouwbare scores op? Hoe is de overeenstemming tussen verschillende beoordelaars en wat betekenen eventuele verschillen tussen beoordelaars?
3. Hoe wordt de relevantie van de beoordelingscriteria van de AGREE beoordeeld in deze arbocontext?
4. Kan elk type arbokennisproduct met de AGREE beoordeeld worden of moet het een 'richtlijn' zijn?
5. Hoe wordt het gebruiksgemak, de uitvoerbaarheid en de toepassing in de praktijk gewaardeerd?

5.2 Methode

Het AGREE-instrument (Appraisal of Guidelines for Research and Evaluation) is een beoordelingslijst die is ontwikkeld door experts uit 13 landen, internationaal gevalideerd en aanbevolen door de WHO en de Raad van Europa. De lijst bestaat uit 23 items verdeeld over 6 domeinen en een algemene vraag of men de richtlijn aanbeveelt voor toepassing in de praktijk. Op basis hiervan kunnen 6 domeinscores berekend worden, die elk verwijzen naar een aparte dimensie van kwaliteit van richtlijnen: onderwerp en doel, betrokkenheid belanghebbenden, methodologie, helderheid en presentatie, toepassing en onafhankelijkheid opstellers.

In totaal zijn 22 arbokennisproducten afkomstig van de NVAB, STECR en een aantal andere kennisinstituten beoordeeld door 30 sleutelpersonen uit kennisinstituten en de universitaire wereld en door personen die binnen grote arbodiensten betrokken zijn bij deskundigheidsbevordering. Elk van hen ontving 3 verschillende arbokennisproducten en een vragenlijst, bestaande uit de 23 AGREE-items, waarbij voor elk item ook gevraagd is hoe relevant men dat criterium in dit specifieke geval vond. Tevens waren een twintigtal vragen toegevoegd over het nut en de toepasbaarheid van de AGREE. Elk arbokennisproduct werd door 2 tot 6 verschillende mensen beoordeeld, waarbij een product nooit beoordeeld werd door iemand uit de kring van de organisatie die het product uitbracht.

5.3 Resultaten

De AGREE-scores

De AGREE-domeinscores van de onderzochte arbokennisproducten lopen flink uiteen. NVAB-producten scoren over het algemeen goed. De domeinscores van de andere arbokennisproducten zijn soms laag. De lage scores verwijzen aspecten die verbeterd kunnen en moeten worden. Soms is dat verbeteren alleen een kwestie van beter informatie geven over de manier waarop het arbokennisproduct ontwikkeld is. In andere gevallen duidt de lage score op meer structurele problemen in het ontwikkelproces, die alleen door grondiger aanpassingen verbeterd kunnen worden.

- *Onderwerp en Doel*

Deze domeinscores zijn over het algemeen redelijk tot goed. Verbeteringen zijn vaak eenvoudig te realiseren.

- *Betrokkenheid Belanghebbenden*

Deze domeinscores beslaan een brede range. Verbeteringen op dit vlak zijn meestal een kwestie van het explicieter gebruik maken van de wensen en voorkeuren van cliënten, patiënten of werknemers (en eventueel werkgevers).

- *Methodologie*

Er zijn tal van arbokennisproducten waarbij de belangrijke domeinscore voor Methodologie onder de maat is, omdat niet duidelijk gemaakt wordt of bij het ontwikkelproces voldoende systematisch gebruik gemaakt is van wetenschappelijke literatuur en de relatie tussen inhoud van de adviezen is en het onderliggende bewijs voor effectiviteit van die interventies niet duidelijk is.

- *Helderheid Presentatie*

Scores zijn over het algemeen redelijk tot goed. Door middel van praktijktesten of pretesten van de uiteindelijke indeling en vormgeving kunnen verbeteringen gerealiseerd worden.

- *Toepassing*

Veel arbokennisproducten scores laag op dit domein. Bij het opstellen van arbokennisproducten wordt blijkbaar te weinig rekening gehouden met de organisatorische belemmeringen en kostenimplicaties; hulpmiddelen voor de toepassing ontbreken vaak.

- *Onafhankelijkheid Opstellers*

Deze scores zijn vaak laag omdat er weinig informatie over is opgenomen in de arbokennisproducten. We hebben de indruk dat veel opstellers het belang van deze informatie onderschatten. Omdat die informatie meestal wel beschikbaar is, betekent dat echter ook dat verbeteringen eenvoudig zijn te realiseren.

<u>Overeenstemming tussen beoordelaars</u>

De instrumenteigenschappen van de AGREE zijn redelijk tot goed. Er zijn weliswaar nogal wat verschillen tussen beoordelaars maar de verwachting is dat die afnemen als de beoordelaars wat meer uitleg, training en instructie krijgen. Daarnaast verwachten we dat discussie en overleg tussen beoordelaars over niet overeenstemmende scores een zinvolle aanvulling is op de beoordelingsprocedure.

<u>Relevantie van de AGREE-items</u>

Alle AGREE-items worden door de beoordelaars relevant geacht om de kwaliteit van arbokennisproducten in kaart te brengen. Het aantal voorstellen voor aanvullingen is beperkt.

<u>Richtlijnkarakter van arbokennisproducten</u>

Met behulp van de maat ‘richtlijnigheid’ is het richtlijnkarakter van arbokennisproducten vastgesteld los van eventuele titel, beoogde functie of toelichting in een inleiding. NVAB-Richtlijnen scoren hoog op richtlijnigheid, STECR-Werkwijzers en de overige arbokennisproducten gemiddeld wat lager, maar er zijn ook documenten met een hoge richtlijnigheidsscore.

Gebruiksgemak, toepassing en waardering AGREE

Het gebruiksgemak van de AGREE is goed: het instrument kan bij de meeste arbokennisproducten gemakkelijk ingevuld worden. De beoordelaars zijn van mening dat het gebruik van de AGREE op allerlei vlakken meerwaarde kan hebben om de kwaliteit van nieuwe arbokennisproducten te verbeteren of om arboprofessionals inzicht te geven in de kwaliteit van bestaande arbokennisproducten.

5.4 Conclusies en aanbevelingen

De AGREE-scores

We moeten benadrukken dat de feitelijke AGREE-scores voor de verschillende arbokennisproducten niet de hoofdzaak vormen van deze studie, het ging immers vooral om een verkenning van de gebruikswaarde van de AGREE in een arbocontext: kan de AGREE gebruikt worden om de kwaliteit van diverse arbokennisproducten op een goede manier met elkaar te vergelijken?
De AGREE-scores in dit onderzoek maken echter goed zichtbaar wat de kwaliteit van de onderzochte arbokennisproducten is en waar verbeteringen mogelijk zijn. Het opstellen en publiceren van arbokennisproducten is geen vrijblijvende bezigheid. Adeldom verplicht: een goed arbokennisproduct stelt eisen aan het ontwikkelproces en de expertise van de werkgroepleden. Omdat er voor elk domein arbokennisproducten zijn waarbij een goede scores is gerealiseerd, hoeft er geen twijfel te zijn dat het mogelijk is om op elk vlak een goede kwaliteit te realiseren.
Het verbeteren van arbokennisproducten is soms alleen een kwestie van vollediger informatie geven over het ontwikkelproces. In andere gevallen moet dat proces zelf structureel verbeterd worden door bijvoorbeeld systematisch commentaar te vragen aan potentiële gebruikers of cliënten. Een goede praktijktest levert informatie op die voor diverse AGREE-domeinen van belang is en die mag daarom eigenlijk nooit ontbreken.

In een goed functionerende arbokennisinfrastructuur behoren arbokennisproducten altijd methodologisch van acceptabel niveau te zijn. Omdat er inmiddels voldoende expertise op dit vlak beschikbaar is, zijn lage domeinscores eigenlijk niet meer nodig. Elk kennisinstituut dat een arbokennisproduct wil uitbrengen kan door samenwerking met universiteiten ervoor zorgen dat de adviezen op een heldere manier onderbouwd worden. Professionals moeten eenvoudig kunnen zien waarop aanbevelingen gebaseerd zijn, omdat het consequenties heeft voor de manier waarop de adviezen in de praktijk gehanteerd moeten worden. In een aantal gevallen is naar onze mening sprake van documenten met pretenties die niet geschraagd worden door de inhoud. Het is van belang dat arboprofessionals leren hoe zij arbokennisproducten van onvoldoende kwaliteit kunnen herkennen.

Een goede Methodologiescore voor een arbokennisproduct kan alleen gerealiseerd worden door middel van een grondige ontwikkelprocedure. Efficiënt om gaan met de beperkte middelen (financiën en menskracht) vereist samenwerking en afstemming tussen de diverse kennisinstituten. Afstemming om te voorkomen er diverse arbokennisproducten over hetzelfde onderwerp gemaakt worden. En samenwerking om goed gebruik te maken van de beschikbare expertise.
Door een goede praktijktest uit te voeren kunnen diverse domeinscores verbeterd worden, met name die van het domein Toepassing. Het is aannemelijk dat de inhoud van de adviezen verandert als de toepassingscontext serieuzer genomen gaat worden. Hulpmiddelen kunnen makkelijk via het internet beschikbaar gemaakt worden (zie de voorbeelden van de NVAB op www.NVAB-online.nl).

De 6 AGREE-domeinscores worden als aparte dimensies van kwaliteit van arbokennisproducten behandeld. The AGREE Collaboration wijst het optellen of sommeren van de afzonderlijke domeinen af. Er zit een inherent spanningsveld in de AGREE-criteria: volledige rapportage over bijvoorbeeld alle Methodologie-items kan de Helderheid Presentatie in gevaar brengen. Omdat richtlijnen soms in vakbladen gepubliceerd worden is de publicatieruimte vaak begrensd. Nu al wordt bij het uitbrengen van veel richtlijnen gekozen voor een korte inhoudelijke tekst met aanbevelingen en een uitgebreider achtergronddocument met toelichting. Door middel van verwijzing naar internetpagina's kan zowel voldaan worden aan de eis van hanteerbaarheid en beperkte omvang als aan de eis van volledige transparantie ten aanzien van het ontwikkelproces, overwegingen en achtergronden.

Overeenstemming tussen beoordelaars

We adviseren om arbokennisproducten door minimaal 4 beoordelaars van diverse achtergrond te laten scoren voor een betrouwbaar resultaat. Nader onderzoek moet aantonen of meer beoordelaars de betrouwbaar doet toenemen.

Maar er is ook debat nodig over de betekenis die we willen toekennen aan de AGREE-scores omdat daarmee ook het gewicht van beoordelingsverschillen samenhangt. Een flink aantal AGREE-items biedt door de formulering uitdrukkelijk ruimte voor waardeoordelen. Door een gevarieerde samenstelling van de groep beoordelaars wordt er recht gedaan aan de multidisciplinaire en veelzijdige context waarvoor de arbokennisproducten bedoeld zijn. Maar dan is het ook onvermijdelijk dat scores soms verschillend zijn.

De verwachting dat bepaalde AGREE-items door de formulering in een arbocontext lastig te scores zouden zijn is niet uitgekomen. Afgaande op de spreidingsmaten voor de verschillende items en domeinen is er geen aanleiding om items te herformuleren.

Er zijn weinig aanwijzingen op om op voorhand soorten arbokennisproducten te identificeren waar de AGREE bovenmatig gespreide scores oplevert. Allerlei soorten arbokennisproducten kunnen op een goede vergelijkbare manier met de AGREE beoordeeld worden. Wel is het zo dat de AGREE vooral geschikt lijkt voor zelfstandig uitgegeven kennissamenvattende documenten. Hoofdstukken in handboeken of een verzameling webteksten zijn mogelijk minder geschikt. De domeinscores voor 'Methodologie' zijn vooral bedoeld om de wetenschappelijke betrouwbaarheid van de inhoud te kwantificeren. Als richtlijnen zijn afgeleid van wettelijke regelingen, zoals bijvoorbeeld de standaarden van het Lisv/UWV, is nader onderzoek nodig hoe de AGREE-scores dan geïnterpreteerd moeten worden.

Relevantie van de AGREE-items

De AGREE-items kunnen rekenen op een breed draagvlak AGREE in de Nederlandse arbowereld. Het is een instrument dat nagenoeg alle belangrijke criteria omvat om de kwaliteit van arbokennisproducten te kunnen beoordelen. Er zouden eventueel enkele items toegevoegd kunnen worden, maar dan liefst zo, dat de (internationale) vergelijkbaarheid van AGREE-scores niet in gevaar gebracht wordt.

Het richtlijnkarakter van arbokennisproducten

Arbokennisproducten kunnen meer of minder 'richtlijnig' zijn, maar de 6 AGREE-domeinen bestrijken een dermate breed gebied aan kwaliteitscriteria voor arbokennisproducten dat ze altijd van toepassing zijn. De AGREE-criteria brengen niet alleen in beeld of de aanbevelingen op een deugdelijke, wetenschappelijke manier tot stand zijn gekomen, maar geven ook inzicht in de gebruiksvriendelijkheid en in welke mate er rekening is gehouden met mogelijkheden voor toepassing in de praktijk en de eventuele belemmeringen. Volgens de beoordelaars zijn de AGREE-criteria daarom ook bruikbaar voor de STECR-werkwijzers en een aantal overige arbokennisproducten: het richtlijnkarakter is nog steeds van zo'n niveau dat de gebruikte AGREE-criteria een goed beeld van de kwaliteit in de breedte geven. Er lijken vooralsnog geen aanvullende criteria nodig.

Gebruiksgemak, uitvoerbaarheid en waardering AGREE

Het invullen van de AGREE levert dermate weinig problemen dat daar geen belemmering ligt voor een brede toepassing van de AGREE.

De beoordeling van arbokennisproducten door middel van de AGREE kan op allerlei manieren ingezet en gebruikt worden. Omdat er veel steun is voor een onafhankelijk keurmerk gebaseerd op de AGREE-scores, verdient dat plan nader onderzoek. Ook al omdat men van mening is dat het beoordelen van arbokennisproducten met behulp van de AGREE in veel gevallen expertise vraagt die niet bij elke arboprofessional in de praktijk beschikbaar is. Het realiseren van zo'n keurmerk vraagt echter wel de nodige voorbereiding en een goede infrastructuur. Mogelijk kan het voorbeeld uit Canada ons

verder helpen, waar een samenwerkingsverband van regering en artsenorganisatie nieuw uitkomende richtlijnen op basis van de AGREE-scores voorziet van een samenvattende aanbeveling.[1] Een pool van beoordelaars is opgeleid in de richtlijnen met behulp van de AGREE te scoren. Een sterke aanbeveling wordt uitgedrukt met 4 appels, 3 appels staan voor 'aanbevolen', 2 appels betekent dat de professional de richtlijn met het nodige voorbehoud in speciale omstandigheden kan toepassen en richtlijnen met 1 appel worden niet aanbevolen. Artsen (en beleidsmedewerkers en patiënten!) kunnen op een website dat oordeel (en de richtlijn) achterhalen. Ook voor Nederland zou de kennisinfrastructuur voor arbodiensten een goede impuls krijgen als bedrijfsartsen en arboprofessionals op een eenvoudige manier kunnen achterhalen hoe goed een arbokennisproduct nu eigenlijk is. Zo'n soort voorziening voor arbokennisproducten zou goed aansluiten bij de plannen voor het verbeteren van de arbokennisinfrastructuur (van Dijk & Hügenholtz 2005). Omdat STECR op dit moment al een webpagina beheert waarop nagenoeg alle beschikbare arbokennisproducten in Nederland te vinden zijn, zou daarbij aangesloten kunnen worden.[2]

De AGREE kan waarschijnlijk de meeste waarde hebben bij het ontwikkelen en herzien van arbokennisproducten. Nu al heeft de NVAB actief gebruik gemaakt van de AGREE-criteria bij het opstellen van richtlijnen KASN en ASTMA/COPD in de commentaarronde en praktijktest. De mogelijkheden om een AGREE-beoordeling in te voeren als onderdeel van de autorisatiefase worden onderzocht. STECR is aan het bekijken of bij de herziening van werkwijzers rekening gehouden kan worden met de AGREE-criteria. Nieuwe versies van Werkwijzers die eind 2004 uitkwamen, wekken de stellige indruk dat er in de toegevoegde inleiding meer rekening gehouden is met de AGREE-criteria dan bij de vorige versie. Dat is daarom niet denkbeeldig omdat veel mensen die betrokken zijn bij het opstellen en herzien van de STECR-Werkwijzers tevens beoordelaar waren in dit onderzoek. Beoordelaars uit STECR-kring schreven ons dat de kennismaking met de AGREE hun kijk op de werkwijze van de STECR-kenniskringen grondig heeft veranderd.

Elke organisatie die arbokennisproducten uitbrengt kan de AGREE-items gebruiken als leidraad om het ontwikkelproces te verbeteren. De AGREE zou op het bureau van elke schrijver of maker van arbokennisproducten moeten liggen, zodat de criteria tijdens het ontwikkelproces richtinggevend kunnen zijn. De AGREE is een richtlijn voor arbokennisproductmakers!

Alhoewel de AGREE op allerlei manieren ingezet kan worden om de kwaliteit van arbokennisproducten te verbeteren, moeten we het belang daarvan ook wat relativeren. De weg van aanbevelingen in arbokennisproducten naar routinematige uitvoering in de praktijk is lang en weerbarstig. Tal van factoren zoals de persoon van de arboprofessional, de organisatie waar hij werkt en kenmerken van werkgever en werknemer beïnvloeden welke kwaliteit van arbozorg uiteindelijk wordt gerealiseerd en welk effect dat heeft op het herstel van zieke werknemers, het terugdringen van onnodig verzuim en beperken van WAO-instroom. De kwaliteit van het arbokennisproduct staat aan het begin van een keten van gebeurtenissen en de invloed op het eindresultaat is waarschijnlijk gering. Andere factoren zijn veel belangrijker (Eccles & Mason 2005; Franx et al. 2004; Graham et al. 2002; Grimshaw et al. 2004; Grol & Grimshaw 2003; Hoenen & de Buisonjé 2003; Peters et al. 2003; Schouten et al. 2004; SIGN 2004). Dat hangt echter samen met de geringe implementatie van de diverse arbokennisproducten: er gaapt een flinke kloof tussen ideaal en praktijk.

Beschikbare arbokennisproducten vormen een steeds belangrijker deel van de opleiding van arboprofessionals. Er zullen de komende tijd steeds meer en betere arbokennisproducten gepubliceerd worden. De beschikbaarheid op de werkplek zal door ICT-oplossingen flink kunnen verbeteren. Kwaliteitsbeleid in arbodiensten zal meer en meer gebruik gaan maken van performance-indicatoren die afgeleid worden van de inhoud van arbokennisproducten, niet in de laatste plaats omdat werkgevers en werknemers op de hoogte raken van welke optimale aanpak van arboproblemen daarin wordt aanbevolen. In zo'n situatie zal de invloed van arbokennisproducten op het eindresultaat in de praktijk kunnen toenemen en wordt de kwaliteit van arbokennisproducten navenant belangrijker. De AGREE vormt een belangrijk instrument om een betere kwaliteit van arbokennisproducten te realiseren.

[1] http://gacguidelines.ca/

[2] http://www.stecr.nl/bis/page-1-1-3.html

6. Literatuur

AGREE Collaboration, AGREE Instrument voor beoordeling van richtlijnen. September 2001 www.agreecollaboration.org

AGREE Collaboration, AGREE Instrument Training Manual. 2003

AGREE Collaboration, Writing Group: Cluzeau FA, Burgers JS, Brouwers M, Grol R, Mäkelä M, Littlejohns P, Grimshaw J, Hunt C, Development and validation of an international appraisal instrument for assessing the quality of clinical practice guidelines: the AGREE project. Qual Saf Health Care 2003;12:18–23.

Atkins D, Briss PA, Eccles M, Flottorp S, Guyatt GH, Harbour RT, Hill S, Jaeschke R, Liberati A, Magrini N, Mason J, O'Connell D, Oxman AD, Phillips B, Schünemann H, Tantorres Edejer T, Vist GE, Williams JW jr, The Grade Working Group, Systems for grading the quality of evidence and strength of recommendations II: Pilot study of a new system, BMC Health Services Research 2005, 5:25 doi:10.1186/1472-6963-5-25.

Audet AMJ, Doty MM, Shamasdin J, Schoenbaum SC, Measure, Learn, And Improve: Physicians' Involvement In Quality Improvement. Health Affairs 2005;24(3):843-53.

AVVV (Algemene Vereniging Verpleegkundigen en Verzorgenden), Criteria voor richtlijnen. Januari 2004, Utrecht.

Boluyt N, Lincke CR, Offringa M, Quality of Evidence-Based Pediatric Guidelines. Pediatrics 2005;115(5):1378-91.

Branche Begeleidings Commissie voor het Arboconvenant Kappers, Reïntegratiehandreiking Kappers. Hulpmiddel voor de bedrijfsarts om bij kappers verzuim of arbeidsongeschiktheid te voorkomen en arbeidsparticipatie zoveel mogelijk in stand te houden. VHP ergonomie, Coronel Instituut voor Arbeid, Milieu en Gezondheid, AMC Amsterdam, 2004.

Burgers JS. Quality of Clinical Practical Guidelines. Proefschrift Katholieke Universiteit, Nijmegen, 2002.

Burgers JS, Bailey JV, Klazinga NS, van der Bij AK, Grol R, Feder G, for the AGREE Collaboration, Inside Guidelines. Comparative analysis of recommendations and evidence in diabetes guidelines from 13 countries. Diabetes Care 2002;25(11):1933-9.

Burgers JS, Grol RP, Zaat JO, Spies TH, van der Bij AK, Mokkink HG., Characteristics of effective clinical guidelines for general practice. Br J Gen Prac 2003a;486:15-9.

Burgers JS, Grol RP, Klazinga NS, Mäkelä M, Zaat J, for the AGREE Collaboration. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. Int J Quality Health Care 2003b;15(1):31-45.

Burgers JS, Boluyt N. De kwaliteit van CBO-richtlijnen en NHG-Standaarden beoordeeld met het AGREE-instrument. Huisarts Wet 2004;47(9):394-9.

Burgers JS, van Barneveld TA, van Everdingen JJE. Behoud het goede. Medisch Contact 2004;59(42):1634-7.

Burgers JS, Fervers B, Haugh M, Brouwers M, Browman G, Philip T, Cluzeau FA. International Assessment of the Quality of Clinical Practice Guidelines in Oncology Using the Appraisal of Guidelines and Research and Evaluation Instrument. J Clin Oncol 2004;22:2000-7.

Cates JR, Young DN, Guerriero DJ, Jahn WT, Armine JP, Korbett AB, Bowerman DS, Porter RC, Sandman T, King RA. An Independant Assessment of Chiropractic Practice Guidelines. J Manipulative Physiol Ther 2003;26:282-86.

Cluzeau FA, Littlejohns P, Grimshaw J, Feder G, Moran, SE. Development and application of a generic methodology to assess the quality of clinical guidelines. Int J Quality Health Care 1999;11(1):21-28.

Coomarasamy A, Ola B, Gee H, Khan, KS. Quality of brief guidelines produced by professional bodies: a study of the 'green top' guidelines by the UK Royal College of Obstetricians and Gynaecologists, J Obst Gynaec 2003;23:479-483.

Cranney A, Waldegger L, Graham ID, Man-Son-Hing M, Byszewski A, Ooi DS. Systematic assessment of the quality of osetoproris guidelines. BMC Musculoskeletal Disorders 2002;3:20, www.biomedcentral.com/1471-1-2474/3/20.

Crul BVM. Kwaliteit van de zorg. Medisch Contact 2004;59(42):1627.

Crul BVM, Legemaate J. Juridisering van richtlijnen en protocollen. Medisch Contact 2005;60(19):812.

Dijk F van, Reinders G, Schaafsma F, Spreeuwers D, Hulshof C, Abspoel M, Vlek F, Verbeek J. Een virtuele bibliotheek voor arbo en sociale zekerheid. TBV 2003;11(12):368-372.

Dijk F van, Hügenholtz N. Programma ter versterking van de kennis- en onderzoeksinfrastructuur van de arbeids- en bedrijfsgeneeskunde in arbodiensten. Eindrapportage, Coronel Instituut, Amsterdam, 2005.

Eccles M, Mason J. How to develop cost-conscious guidelines. Health Technol Assess 2001;5(16):1-69.

Eccles M, Grimshaw J, Walker A, Johnston M, Pitts N., Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. J Clin Epidemiol 2005;58:107–112.

European Health Committee (CSDP) Recommendation Rec (2001)13 on developing a methodology for drawing up guidelines on best medical practices and explanatory memorandum. Strasbourg: Council of Europe Publishing, 2001.

Everdingen JJE. Van consensus naar CBO-richtlijn. Ned Tijdschr Geneesk 1999;143(42):2086-9.

Everdingen JJE, Timmermans AE, Swinkels JA, Barneveld TA, Assendelft WJJ. Geleide vrijheid. Medisch Contact 2003;58(10): 386.

Fervers B, Burgers JS, Haugh MC, Brouwers M, Browman G, Cluzeau F, Philip T. Predictors of high quality clinical practice guidelines: examples in oncology. Int J Qual Health Care 2005;17(2):123-132.

Franx G, van Karsbergen K, Eland, A, Verbrug H, Grol R. Implementatie van richtlijnen in de geestelijke gezondheidszorg: welke (on)mogelijkheden zien hulpverleners in de praktijk? TSG 2004;82(5):317-323.

Franx G, Eland A, Verburg H. Transparante zorg in de GGZ: van de bomen en het bos. Onderscheid en samenhang tussen richtlijnen, protocollen en zorgprogramma's. MGV 2000;57:1036-1050.

Graham ID, Calder LA, Hébert PC, Carter AO, Tetroe JM. A comparison of clinical practice guideline appraisal instruments. Int J Techn Ass Health Care 200.;16(4):1024-38.

Graham ID, Beardall S, Carter AO, Tetroe J, Davies B. The state of science and art of guideline development, disseminiation and evaluation in Canada. J evaluation in Clinical Practice 2002;9(2):195-202.

Grimshaw JM, Thomas RE Maclennan G, Fraser C, Ramsay CR, Vale L, Whitty P, Eccles MP, Matowe L, Shirran L, Wensing M, Dijkstra R, Donaldsons C. Effectiveness and efficiency of guideline dissemination and implementation strategies. Executive summary. Health Technology Assessment 2004;8 nr 6.

Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. Lancet 2003;362(11):1225-30.

Hanney SR, Gonzalez-Block MA, Buxton MJ, Kogan M. The utilisation of health research in policy-making: concepts, examples and methods of assessment. Health Research Policy and Systems 2003, 1:2 doi:10.1186/1478-4505-1-2.

Hoehler FK. Bias and prevalence effects on kappa reviewed in terms of sensitivity and specificity. J Clinic Epidem 2000;53:499-503.

Hulshof CTJ, Weel ANH. Richtlijnontwikkeling in de bedrijfsgezondheidszorg: schets voor een infrastructuur, Amsterdam, SKB/NVAB, 1996.

Hoenen J, de Buisonjé C. Van richtlijn naar routine. Het strategisch implementeren van NVAB-richtlijnen, Bureau Richtlijnen NVAB, Amsterdam 2003.

Jansen TLThA, Went K, Tanja PW. Richtlijnen onder vuur. Toename van protocollen leidt tot kookboekgeneeskunde. Medisch Contact 2003;58(7):253-5.

Katrak P, Bialocerkowski AE, Massy-Westropp N, Kumar VSS, Grimmer, KA. A systematic review of the content of critical appraisal tools, BMC Medical Research Methodology 2004, 4-22; doi:10.1186/147-2288-2-22.

Kuiper J, Burdorf L, Frings-Dresen M, Kuijer P, Lötters F, Spreeuwers D, Miedema H. Rugklachten door het werk. Instrument voor het bepalen van de arbeidsgerelateerdheid van aspecifieke lage rugklachten. Nederlands Kenniscentrum Arbeid en Klachten Bewegingsapparaat, november 2004.

Maasen JHW, Noort A. Een startpunt voor vakinhoudelijke informatie. TBV 2004;12(5):143-145.

MacDermid JC, StudySelf Test on the AGREE Instrument.2003, http://www.biomedcentral.com/content/supplementary/1472-6963-5-18-S1.doc

MacDermid JC, Brooks D, Solway S, Switzer-McIntyre S, Brosseau L, Graham ID. Reliability and validity of the AGREE instrument used by physical therapists in assessment of clinical practice guidelines. BMC Health Services Research 2005, 5:18 doi:10.1186/1472-6963-5-18.

McGraw KO, Wong SP. Forming Inferences About Some Intraclass Correlations Coefficients. Psychological Methods 1996;1(1):30-46.

Michie S, Johnston M, Abraham C, Lawton R, Parker D, Walker A and on behalf of the "Psychological Theory" Group. Making psychological theory useful for implementing evidence based practice: a consensus approach. Qual. Saf. Health Care 2005;14;26-33.

Michie S, Johnston M, Changing clinical behaviour by making guidelines specific. BMJ 2004;328(7):343-345.

Minister van Sociale Zaken en Werkgelegenheid, AJ de Geus, Wijziging Regeling procesgang eerste ziektejaar. Staatscourant 4 november 2004;213:19.

Nichols DP. Choosing an intraclass correlation coefficient. SPSS Keywords, number 67, 1998.

Nicholson PJ, Cullinan P, Newman Taylor AJ, Burge PS, Boyle C. Evidence based guidelines for the prevention, identification, and management of occupational Asthma. Occup Environ Med 2005;62:290-9.

Nolen WA, Laveren tussen aanname en bewijs. inaugurele rede, 30 november 2004, Groningen.

Offringa M, Assendelft WJJ, Scholten RJPM (red.). Inleiding in evidence-based medicine: klinisch handelen gebaseerd op bewijsmateriaal. Houten: Bohn Stafleu Van Loghum, 2000.

Peters MAJ, Harmsen M, Laurant MGH, Wensing M. Ruimte voor verandering? Knelpunten en mogelijkheden voor verbeteringen in de patiëntenzorg. Uitgave: Afdeling Kwaliteit van zorg (WOK), UMC St Radboud, Nijmegen, 2003.

Schaafsma F, Hulshof C, van Dijk F, Verbeek J. Information demands of occupational health physicians and their attitude towards evidence-based medicine. Scand J Work Environ Health 2004;30(4):327-30.

Schouten LMT, Seesing F, Tijink H, Everdingen JJE van. De achilleshiel van de Doorbraakprojecten: verspreiding van resultaten en opgedane kennis stagneert. Medisch Contact 2004;59(42):1642-4.

Scottish Intercollegiate Guidelines Network, 6. Forming guideline recommendations, March 2004.

Scottish Intercollegiate Guidelines Network, 9. Implementation, March 2004.

Staal JB, Hlobil H, van Tulder MW, Waddell G, Burton AK, Koes BW, van Mechelen W. Occupational health guidelines for the management of low back pain: an international comparison. Occup Environ Med 2003;60:618–26.

Stemler SE. Comparison of Consensus, Consistency, and Measurement Approaches to Estimating Interrater Reliability. Practical Assessment, Research & Evaluation 9 (4) Yale University PACE Center, 2004. http://pareonline.net/getvn.asp?v=9&n=4.

Tuut MK, Tulder MW van, Everdingen JJE van, Sanders R, Assendelft WJJ. Aspecifieke lage-rugklachten: bestaande Nederlandse en buitenlandse richtlijnen vergeleken. Medisch Contact 2002;57(26/27):1016-9.

Uitspraak Hoge Raad der Nederlanden d.d. 1 april 2005, Medisch Contact 2005;60(19):812-4.

Uitvoeringsinstituut werknemersverzekeringen (UWV). Beleidsregels beoordelingskader poortwachter: besluit, toelichting en bijlage met kader voor de beoordeling van reïntegratie-inspanningen, Amsterdam, 2002.

Upshur REG. Are all evidence-based practices alike? Problems in the ranking of evidence. JAMA 2003;169(7):672-3.

Velden van J, Zoon C. Waardering activiteiten STECR en platformverbreding. Managementsamenvatting. Ecorys Nei, Rotterdam, 2004.

Verbeek JH, van Dijk FJ, Malmivaara A, Hulshof CT, Räsänen K, Kankaanpää EE, Mukala K, Evidence-based medicine for occupational health. Scand J Work Environ Health 2002;28(3):197-204.

Verbeek J, Husman K, van Dijk F, Jauhiainen M, Pasternack I, Vainio H. Building an evidence base for occupational health interventions. Scand J Work Environ Health 2004;30(2):164-8.

Vlayen J, Aertgeerts B, Hannes K, Sermeus W, Eamaekers D. A systematic review of appraisal tools for clinical practice guidelines: multiple similarities and one common deficit. Int J Qual Health Care 2005;17(3):235-42.

Waddell G, Burton AK, Occupational Health guidelines for the management of low back pain at work: evidence review. Occup Med 2001;51(2):124-35.

Weggeman M., In de kenniseconomie lopen de hazen anders. NRC 11-05-2004.

Yaffee RA, Enhancement of Reliability Analysis: Application of Intraclass Correlations with SPSS/Windows v.8. New York University, 1998.

Bijlage 1 NVAB over richtlijnen

Uit het **Professioneel Statuut** (2003):

3.2. *De bedrijfsarts is werkzaam conform de richtlijnen en standaarden zoals ze door de beroepsgroep algemeen zijn aanvaard.*

6.5. *De arbodienst verschaft de bedrijfsarts zowel de benodigde tijd als de benodigde personele, instrumentele, ruimtelijke voorzieningen en opleidingen om zijn of haar werk te verrichten (conform professionele standaarden en richtlijnen van de bedrijfsarts, stand van de wetenschap en techniek, algemene normen en waarden en wat gebruikelijk is in de gezondheidszorg).*

9.1. *De Bedrijfsarts is onafhankelijk in zijn of haar professionele oordeelsvorming, medisch handelen, verwijsbeleid en de inhoud van de advisering en is persoonlijk verantwoordelijk voor zijn of haar professionele beroepsuitoefening en de kwaliteit van handelen.*
De bedrijfsarts baseert zich hierbij mede op richtlijnen en standaarden van de beroepsgroep, die kunnen fungeren als inhoudelijke toetssteen van zijn of haar professionele autonomie.

Uit het **Beroepsprofiel van de bedrijfsarts** (2004):

Taken en competenties van de bedrijfsarts
De analyse, vaststelling van de diagnose, de advisering en de begeleiding vinden zoveel mogelijk plaats op basis van evidence based methoden, of indien nog niet voorhanden op basis van best practices.
Daarnaast bieden de reeds ontwikkelde richtlijnen een belangrijk instrument voor gestandaardiseerd en geprotocoliseerd handelen (de richtlijn voor psychische klachten en voor lage rugklachten). Hierbij is er aandacht voor terugvalpreventie.

Uit de **Statuten** (2001), Taak van de NVAB:

Artikel 4 lid 1b
Het bevorderen van de professionele kwaliteit van de individuele beroepsbeoefenaar en van de randvoorwaarden voor de beroepsuitoefening, onder meer door het bevorderen van de kwaliteit van de beroepsopleiding en de bij- respectievelijk nascholing, alsmede door het ontwikkelen van richtlijnen voor de beroepsuitoefening;

Op de **NVAB-website**:

Het belang van richtlijnen:
Richtlijnontwikkeling draagt bij aan verbetering van de kwaliteit van het arbeidsgezondheidskundig handelen. Het doel hiervan is beroepsziekten, werkgebonden aandoeningen en onnodig ziekteverzuim te voorkómen en reïntegratie van zieke werknemers, wellicht op een aangepaste manier, te bevorderen. Dat is niet alleen van belang voor de werknemer zelf. Ook de werkgever is erbij gebaat: risico's door het werk worden verminderd en zijn uitgevallen werknemer is, op een verantwoorde manier, eerder beschikbaar. Voor de Arbodienst betekent kwalitatief goed werk van de bedrijfsarts: tevreden klanten. En ook maatschappelijk gezien is die kwaliteit belangrijk: werknemers kunnen langer gezond blijven werken en als werknemers met een verminderde belastbaarheid langer aan het werk kunnen blijven, nemen ziekteverzuim en WAO-instroom af.

Bijlage 2 STECR-webteksten

Toelichting op doel en werkwijze

Kenniskringen

Kennis en ervaring uitwisselen tussen professionals. Met als resultaat een Werkwijzer, Assist of onderzoekrapport. Dát is waar het in de STECR-Kenniskringen om draait. De Kenniskringen zijn multidisciplinair van opzet en ontwikkelen succesvolle interventiemethodieken voor reïntegratie. Professionals van arbodiensten aangesloten bij de Brancheorganisatie Arbodiensten (BOA) maken deel uit van de Kenniskringen. Daarnaast nodigen de Kenniskringen gastleden uit, afkomstig uit de wereld van de reïntegratie. Inmiddels telt STECR 10 Kenniskringen (zie links in het blauwe navigatiescherm).

1. “Good practices” eventueel uitgesplitst naar een bepaalde branche;
2. Een overzicht van relevante literatuur/jurisprudentie/richtlijnen van het onderwerp waarop de Kenniskring zich richt;
3. Richtlijnen of identificatie van succesvolle methodieken t.b.v reïntegratie, zowel inhoudelijk als ten aanzien van de organisatie van het reïntegratie traject (Kenniskringen sluiten aan of bouwen voort op de systematiek van bestaande richtlijnen);
4. Een plan voor de implementatie van de resultaten van de Kenniskring: dit kan bijvoorbeeld zijn: input voor de website, onderzoeksvoorstellen, programmavoorstellen voor workshops op basis van een geconstateerde behoefte etc.;
5. Grotere bekendheid van succesvolle interventiemethodieken in het eigen professionele netwerk, zoals de eigen arbodienst en/of beroepsgroep. Hiertoe organiseert ieder kenniskringlid voor de eigen arbodienst bijvoorbeeld een bijeenkomst over de resultaten van de Kenniskring of zorgt voor publicaties over de Kenniskring binnen de eigen arbodienst (bijvoorbeeld via intranet of het personeelsblad).

Het netwerk van de Kenniskring kan door tweedelijns instituten worden benaderd met het verzoek te participeren in onderzoek of te helpen bij de ontwikkeling van methodieken rond het onderwerp van de Kenniskring. De Kenniskring zal hieraan in principe haar medewerking verlenen.

STECR Kenniskringen richten zich op onderwerpen of problematieken die:

- arbeidsrelevant zijn;
- verantwoordelijk zijn voor een groot deel van de WAO instroom;
- multidisciplinaire samenwerking noodzakelijk maken;
- beïnvloed kunnen worden door arboprofessionals;
- problemen opleveren in de praktijk van arboprofessionals.

Informatie over nieuwe kringen is te lezen op deze website in het STECR Webmagazine. Deelnemers voor de nieuwe Kenniskringen kunnen zich aanmelden bij de STECR-correspondent van de eigen arbodienst. De naam van de correspondent kunt u vinden onder het Colofon van deze website (het zwarte navigatiescherm aan de linkerkant van de pagina). Iedere Kenniskring bestaat 1 jaar. Eén jaar na verschijning van de Werkwijzer actualiseert de Kenniskring de inhoud.

Het management van een arbodienst wordt uitgenodigd 1 deelnemer aan te wijzen voor een Kenniskring. Vervolgens stelt het bestuur van STECR vast wie de trekker is van de Kenniskring. De trekker van de Kenniskring is verantwoordelijk voor het tot stand komen van het programma van de kring en stelt in samenwerking met de Kenniskringdeelnemers een plan van aanpak. Onderdelen van dit plan zijn onder andere het organiseren van workshops/bijeenkomsten (vijf keer per jaar) en het optimaal benutten van de website. Het plan van aanpak wordt ter goedkeuring voorgelegd aan de STECR-organisatie. Voor de bijeenkomsten van de Kenniskring worden ook experts uit de onderzoekswereld uitgenodigd om een goede interactie tussen theorie en praktijk te bevorderen.

Iedere STECR Kenniskring wordt ondersteund door de STECR-organisatie. Dit gebeurt op de volgende manieren:

- De kenniskringtrekker wordt geholpen bij het formuleren van een plan van aanpak en het samenstellen van een programma voor de te organiseren workshops/bijeenkomsten.

- Eén keer per kwartaal vindt een intervisiebijeenkomst plaats van de kenniskringtrekkers.
- Tijdens deze bijeenkomst wordt de output van de Kenniskring door de kenniskringtrekkers gepresenteerd en evalueren de deelnemers gezamenlijk het resultaat tot op dat moment. Er worden inhoudelijke en procesmatige tips uitgewisseld.
- Er kan een beroep worden gedaan op het secretariaat van STECR. Het secretariaat assisteert met name bij de organisatie van bijeenkomsten.
- Zalen en apparatuur worden desgewenst door de STECR organisatie geregeld.
- Iedere kenniskring krijgt op de website een eigen pagina.

Werkwijzers

STECR werkt sinds 2001 met een groot aantal professionals aan de Werkwijzers. Inmiddels vinden de Werkwijzers steeds meer hun ingang in de dagelijkse praktijk van verschillende (arbo)professionals. Daarnaast is van de zijde van bedrijven, het UWV, Verzekeringsmaatschappijen en Reïntegratiebedrijven ook belangstelling getoond voor gebruik en implementatie van STECR Werkwijzers. De Werkwijzers zijn alleen via de website te verkrijgen.

De volgende Werkwijzers zijn verschenen:
- ABBE-rsi
- Arbeidsconflicten, herzien 2004
- Aanpak Problematisch Verzuim (APV), herzien 2004
- Lage Rugklachten en Arbeidsomstandigheden, herzien 2004
- Reductie van Werkstress in Onderwijs en Zorg
- Somatisatie

Assists

STECR kent 3 Assists:
- Sociaal Medische Begeleiding van Allochtonen
- Handvatten voor samenwerken tussen professionals
- Vrouwen niet in de WAO

De STECR-Assist geeft een impuls aan een versnelde ontwikkeling van arbodienstverlening door kennisoverdracht in de vorm van tips en aanbevelingen. Het betreft relatief nieuwe en nog weinig gestructureerde vraagstukken in een kennisdomein waar weinig overeenstemming en/of zekerheid bestaat. Een Assist geeft tips voor professionals en/of aanbevelingen voor directies rondom organisatie en geeft een korte samenvatting van literatuur. Bij het ontwikkelen van een Assist moet rekening worden gehouden met de wensen van het klantsysteem, met professionele standaarden en aansluiten bij de praktijk van de arbodienst-organisatie.
De input voor een Assist is gebaseerd op recente literatuur/onderzoek en eigen praktijkervaring.

Rapporten

STECR heeft inmiddels 2 rapporten uitgegeven; ‘Vrouwen in de WAO' en 'De oudere werknemer'. Een studierapport is een rapport dat ‘the state of the art' rondom een bepaald vraagstuk in kaart brengt.

Referentiekaarten

STECR heeft van de Werkwijzers Arbeidsconflicten, ABBE-Rsi, Lage Rugklachten, Aanpak Problematisch Verzuim, Werkstress en Somatisatie Referentiekaarten uitgebracht. Deze kaarten, gebundeld in een map, geven een beknopt overzicht welke stappen kunnen leiden tot een optimale aanpak van ziekteverzuim.

Invulformulier

	Naam beoordelaar:														
	Onderwerp arbokennisproduct:														
	Nummer kennisproduct:														
	Organisatie die het document ontwikkeld of uitgegeven heeft:														
			(AGREE-vragen)								**(Toegevoegd)**				
	ONDERWERP EN DOEL		**Aan criterium wordt voldaan**								**Criterium is relevant**				
1	Het doel van de richtlijn is specifiek beschreven.	1a	Zeer Eens	4	3	2	1	Zeer Oneens		1b	Zeer Eens	4	3	2	1	Zeer Oneens
2	De klinische vraag/vragen die in de richtlijn aan de orde komt/komen, is/zijn specifiek.	2a	Zeer Eens	4	3	2	1	Zeer Oneens		2b	Zeer Eens	4	3	2	1	Zeer Oneens
3	De patiëntenpopulatie waarop de richtlijn van toepassing is, is specifiek beschreven.	3a	Zeer Eens	4	3	2	1	Zeer Oneens		3b	Zeer Eens	4	3	2	1	Zeer Oneens
	BETROKKENHEID VAN BELANGHEBBENDEN														
4	De leden van de werkgroep die de richtlijn heeft ontwikkeld, komen uit alle relevante beroepsgroepen.	4a	Zeer Eens	4	3	2	1	Zeer Oneens		4b	Zeer Eens	4	3	2	1	Zeer Oneens
5	Het perspectief en de voorkeuren van patiënten zijn nagegaan.	5a	Zeer Eens	4	3	2	1	Zeer Oneens		5b	Zeer Eens	4	3	2	1	Zeer Oneens
6	De beoogde gebruikers van de richtlijn zijn duidelijk benoemd.	6a	Zeer Eens	4	3	2	1	Zeer Oneens		6b	Zeer Eens	4	3	2	1	Zeer Oneens
7	De richtlijn is getest onder de beoogde gebruikers.	7a	Zeer Eens	4	3	2	1	Zeer Oneens		7b	Zeer Eens	4	3	2	1	Zeer Oneens
	METHODOLOGIE														
8	Er zijn systematische methoden gebruikt voor het zoeken naar wetenschappelijk bewijsmateriaal.	8a	Zeer Eens	4	3	2	1	Zeer Oneens		8b	Zeer Eens	4	3	2	1	Zeer Oneens
9	De criteria voor het selecteren van het wetenschappelijk bewijsmateriaal zijn duidelijk beschreven.	9a	Zeer Eens	4	3	2	1	Zeer Oneens		9b	Zeer Eens	4	3	2	1	Zeer Oneens
10	De gebruikte methoden om de aanbevelingen op te stellen, zijn duidelijk beschreven.	10a	Zeer Eens	4	3	2	1	Zeer Oneens		10b	Zeer Eens	4	3	2	1	Zeer Oneens
11	Gezondheidswinst, bijwerkingen en risico's zijn overwogen bij het opstellen van de aanbevelingen.	11a	Zeer Eens	4	3	2	1	Zeer Oneens		11b	Zeer Eens	4	3	2	1	Zeer Oneens
12	Er bestaat een expliciet verband tussen de aanbevelingen en het onderliggende wetenschappelijke bewijsmateriaal.	12a	Zeer Eens	4	3	2	1	Zeer Oneens		12b	Zeer Eens	4	3	2	1	Zeer Oneens
13	De richtlijn is voor publicatie door externe experts beoordeeld.	13a	Zeer Eens	4	3	2	1	Zeer Oneens		13b	Zeer Eens	4	3	2	1	Zeer Oneens
14	Een procedure voor herziening van de richtlijn is vermeld.	14a	Zeer Eens	4	3	2	1	Zeer Oneens		14b	Zeer Eens	4	3	2	1	Zeer Oneens
	HELDERHEID EN PRESENTATIE														
15	De aanbevelingen zijn specifiek en ondubbelzinnig.	15a	Zeer Eens	4	3	2	1	Zeer Oneens		15b	Zeer Eens	4	3	2	1	Zeer Oneens
16	De verschillende beleidsopties zijn duidelijk vermeld.	16a	Zeer Eens	4	3	2	1	Zeer Oneens		16b	Zeer Eens	4	3	2	1	Zeer Oneens
17	De kernaanbevelingen zijn gemakkelijk te herkennen.	17a	Zeer Eens	4	3	2	1	Zeer Oneens		17b	Zeer Eens	4	3	2	1	Zeer Oneens

Bijlage 4 Interne consistentie van de AGREE-domeinen

Cronbach's alfa:	AGREE Collaboration 2003	Dit onderzoek 2005
Onderwerp en Doel (3 items)	0,88	0,90
Betrokkenheid Belanghebbenden (4 items)	0,72	0,80
Methodologie (7 items)	0,88	0,92
Helderheid en Presentatie (4 items)	0,69	0,88
Toepassing (3 items)	0,79	0,85
Onafhankelijkheid Opstellers (2 items)	0,64	0,85
N=	33	22

Bijlage 5 Volledigheid van de ingevulde vragenformulieren

Ontbrekende antwoorden	Geen (Volledig ingevuld)	1 item niet ingevuld	2 of meer items niet ingevuld (max)
AGREE-oordelen (23+1)	77	5	3 (6)
AGREE-relevantiescores (24)	77	6	2 (6)
Overige vragen (18)	79	4	2 (2)
Alle items	66	11	8 (12)

Bijlage 6 Gemiddelde AGREE-itemscores per arbokennisproduct: STECR

		1	2	3	4	5	6	7	8	
	Voor feitelijke formulering AGREE-items zie vragenformulier	Werkwijzer ABBE-rsi	Werkwijzer Arbeidsconflicten	Werkwijzer Aanpak Problematisch verzuim	Werkwijzer Lage rugklachten en Arbeidsomstandigheden	Werkwijzer Reductie Werkstress gezondheidszorg/onderwijs	Assist Cultuur of Persoon. SMB Allochtonen	Voorstudie en Assist Vrouwen niet in de WAO	Werkwijzer Somatisatie	Alle arbokennisproducten gemiddeld
	Onderwerp en Doel	38,89	67,78	37,50	70,37	63,89	66,67	63,89	55,56	71,49
1	*doel specifiek*	2,25	3,20	2,75	3,3	3,2	3,2	2,75	2,50	3,18
2	*klinische vraag specifiek*	2,00	2,25	1,33	2,6	2,7	2,2	2,50	2,75	2,93
3	*patiëntenpopulatie specifiek*	2,25	3,20	1,50	3,3	2,7	3,7	3,50	2,75	3,22
	Betrokkenheid Belanghebbenden	33,33	56,67	45,83	52,78	60,42	48,61	41,67	47,92	50,46
4	*leden werkgroep uit alle relevante beroepsgroepen*	2,50	3,60	2,00	3,0	3,2	3,0	2,00	2,50	2,80
5	*perspectief patiënten nagegaan*	1,50	2,20	1,25	2,3	2,2	2,2	3,75	2,25	2,15
6	*beoogde gebruikers benoemd*	3,00	3,40	2,75	3,0	3,7	3,7	2,75	3,50	3,29
7	*De richtlijn getest*	1,75	1,80	1,25	2,3	2,5	2,0	1,50	2,25	2,29
	Methodologie	30,95	42,86	15,48	44,44	40,48	22,22	21,43	36,90	40,60
8	*systematische gezocht wetenschappelijk bewijs*	1,50	2,00	1,50	2,3	2,2	2,0	2,25	2,25	2,33
9	*criteria selecteren bewijs beschreven.*	1,50	1,40	1,00	2,3	2,0	2,0	1,75	2,25	2,06
10	*duidelijke methoden opstellen aanbevelingen*	2,25	2,80	2,00	2,3	2,5	2,0	2,00	2,25	2,46
11	*gezondheidswinst, bijwerkingen, risico's overwogen*	1,75	2,60	1,50	3,0	2,7	2,6	2,00	2,50	2,66
12	*expliciet verband aanbevelingen en bewijs*	1,75	2,20	1,00	2,6	2,7	2,0	2,00	2,50	2,51
13	*richtlijn door externe experts beoordeeld*	2,50	3,20	1,25	2,3	2,5	3,3	2,50	3,00	2,77
14	*procedure herziening vermeld*	1,75	1,80	1,25	1,6	1,7	2,0	1,75	2,25	1,90
	Helderheid en Presentatie	41,67	60,00	54,17	63,89	64,58	54,17	47,92	52,08	61,66
15	*aanbevelingen specifiek en ondubbelzinnig*	2,00	2,60	2,00	2,3	2,7	2,5	2,75	2,25	2,88
16	*verschillende beleidsopties vermeld*	2,25	3,00	2,50	3,0	3,0	2,7	2,75	2,75	2,85
17	*kernaanbevelingen gemakkelijk te herkennen*	2,75	2,80	2,50	3,3	3,2	3,5	2,75	2,50	3,11
18	*toepassing ondersteund met hulpmiddelen*	2,50	2,40	2,00	2,3	3,2	2,5	2,00	2,25	2,48
	Toepassing	25,00	44,44	27,78	33,33	41,67	36,11	13,89	47,22	35,31
19	*organisatorische belemmeringen besproken*	2,50	2,80	2,25	2,0	2,5	2,5	2,50	2,50	2,53
20	*mogelijke kostenimplicaties overwogen*	1,50	2,20	1,50	1,6	2,2	2,7	1,50	2,50	2,07
21	*belangrijkste criteria om te toetsen*	1,50	2,00	1,50	3,0	2,2	2,7	1,50	3,00	2,31
	Onafhankelijkheid Opstellers	29,17	43,33	25,00	11,11	50,00	29,17	41,67	29,17	37,01
22	*niet beïnvloed door financierende instantie*	2,25	2,80	2,00	1,6	3,0	3,0	3,00	2,25	2,86
23	*conflicterende belangen vastgelegd*	1,75	1,80	1,00	1,3	2,0	2,0	1,75	1,75	1,88
24	*richtlijn aan te bevelen voor gebruik in de praktijk*	2,00	3,20	2,50	2,0	3,0	3,5	2,75	2,50	3,06
	Aantal beoordelaars:	2	5	4	3	4	3	4	3	3,7

Bijlage 7 Gemiddelde AGREE-itemscores per arbokennisproduct: NVAB

		9	10	11	12	13	14	
	Voor feitelijke formulering AGREE-items zie vragenformulier	Richtlijn Lage Rugklachten	Richtlijn Psychische Klachten	Richtlijn Astma en COPD	Richtlijn Klachten van Arm Nek en Schouder	Leidraad Verwijzen	LESA Overspanning	Alle arbokennisproducten gemiddeld
	Onderwerp en Doel	**94,44**	**100,00**	**91,67**	**86,11**	**86,11**	**74,07**	**71,4**
1	*doel specifiek*	4,00	4,00	3,50	3,75	4,00	3,33	3,18
2	*klinische vraag specifiek*	3,50	4,00	4,00	3,25	3,25	2,83	2,93
3	*patiëntenpopulatie specifiek*	4,00	4,00	3,75	3,75	3,75	3,33	3,22
	Betrokkenheid Belanghebbenden	**58,33**	**70,83**	**72,92**	**60,42**	**81,25**	**58,33**	**50,4**
4	*leden werkgroep uit alle relevante*	2,00	4,00	2,75	2,75	4,00	3,00	2,80
5	*perspectief patiënten nagegaan*	2,25	2,50	3,25	2,25	2,75	2,00	2,15
6	*beoogde gebruikers benoemd*	4,00	4,00	3,50	4,00	4,00	3,50	3,29
7	*De richtlijn getest*	4,00	2,50	3,00	3,50	3,75	2,33	2,29
	Methodologie	**40,48**	**59,52**	**78,57**	**69,05**	**58,33**	**43,65**	**40,6**
8	*systematische gezocht wetenschappelijk bewijs*	2,75	2,50	3,75	3,50	2,75	3,00	2,33
9	*criteria selecteren bewijs beschreven.*	1,75	2,50	3,75	3,50	2,75	1,83	2,06
10	*duidelijke methoden opstellen aanbevelingen*	2,75	2,50	3,75	3,50	3,00	2,67	2,46
11	*gezondheidswinst, bijwerkingen en risico's*	3,25	3,50	3,50	3,25	4,00	2,67	2,66
12	*expliciet verband aanbevelingen en bewijs*	3,75	2,50	3,50	3,75	3,00	2,50	2,51
13	*richtlijn door externe experts beoordeeld*	2,50	3,50	3,25	3,50	4,00	2,83	2,77
14	*procedure herziening vermeld*	1,75	2,50	2,00	2,25	2,75	1,50	1,90
	Helderheid en Presentatie	**89,58**	**62,50**	**77,08**	**77,08**	**93,75**	**56,94**	**61,6**
15	*aanbevelingen specifiek en ondubbelzinnig*	3,50	4,00	3,25	3,75	4,00	3,33	2,88
16	*verschillende beleidsopties vermeld*	3,75	2,50	3,25	3,75	4,00	2,50	2,85
17	*kernaanbevelingen gemakkelijk te herkennen*	4,00	2,50	3,00	3,75	4,00	3,33	3,11
18	*toepassing ondersteund met hulpmiddelen*	4,00	2,50	3,00	3,25	3,75	1,83	2,48
	Toepassing	**44,44**	**66,67**	**61,11**	**36,11**	**86,11**	**38,89**	**35,3**
19	*organisatorische belemmeringen besproken*	3,50	4,00	3,50	2,50	3,75	2,50	2,53
20	*mogelijke kostenimplicaties overwogen*	2,50	2,50	2,00	2,75	3,50	1,67	2,07
21	*belangrijkste criteria om te toetsen*	2,00	4,00	2,50	2,25	3,75	2,50	2,31
	Onafhankelijkheid Opstellers	**50,00**	**50,00**	**50,00**	**41,67**	**62,50**	**66,67**	**37,0**
22	*niet beïnvloed door financierende instantie*	3,75	4,00	3,50	3,50	3,50	3,67	2,86
23	*conflicterende belangen vastgelegd*	2,50	2,50	1,75	1,75	3,00	2,33	1,88
24	*richtlijn aan te bevelen voor gebruik in de praktijk*	4,00	4,00	3,50	3,75	4,00	3,33	3,06
	Aantal beoordelaars	4	2	4	4	4	6	3,7

Bijlage 8 Gemiddelde AGREE-itemscores per arbokennisproduct: Overige

		15	16	17	18	19	20	21	22	
	Voor feitelijke formulering AGREE-items zie vragenformulier	Geen benutbare mogelijkheden UWV	Zwangerschap en arbeids-ongeschiktheid UWV	Registratierichtlijn Burnout NCvB	Interventies Acute Lage Rugklachten NKAB	Arbeidshulpverlening Psychische Klachten NKAP	Diabetes Handboek Arbeid en Belastbaarheid BSL	PAGO Huisvuilbeladers Coronel	SMB Richtlijn Rugklachten Arbouw	Alle arbokennisproducten gemiddeld
	Onderwerp en Doel	**96,30**	**91,67**	**91,67**	**27,78**	**69,44**	**59,26**	**80,56**	**59,26**	**71,49**
1	*doel specifiek*	3,67	3,75	3,75	1,75	3,25	2,33	3,25	2,33	3,18
2	*klinische vraag specifiek*	4,00	3,75	3,75	2,25	2,75	2,67	3,00	3,00	2,93
3	*patiëntenpopulatie specifiek*	4,00	3,75	3,75	1,50	3,50	2,33	3,75	2,67	3,22
	Betrokkenheid Belanghebbenden	**58,33**	**60,42**	**33,33**	**2,08**	**60,42**	**41,67**	**31,25**	**33,33**	**50,46**
4	*leden werkgroep uit alle relevante beroepsgroepen*	3,33	3,75	2,50	1,00	2,75	2,67	3,00	2,33	2,80
5	*perspectief patiënten nagegaan*	2,00	2,00	1,75	1,00	3,00	1,67	1,50	1,67	2,15
6	*beoogde gebruikers benoemd*	4,00	4,00	3,00	1,00	3,00	3,00	2,50	3,00	3,29
7	*De richtlijn getest*	3,33	2,50	1,75	1,00	3,25	1,00	1,50	1,67	2,29
	Methodologie	**11,11**	**39,29**	**38,81**	**25,00**	**48,81**	**60,00**	**45,24**	**20,63**	**40,60**
8	*systematische gezocht wetenschappelijk bewijs*	2,00	2,25	2,00	1,50	2,75	1,50	3,00	2,00	2,33
9	*criteria selecteren bewijs beschreven.*	2,00	2,25	2,00	1,00	2,50	1,67	2,00	1,67	2,06
10	*duidelijke methoden opstellen aanbevelingen*	3,00	2,25	2,00	1,25	3,25	1,67	2,75	1,67	2,46
11	*gezondheidswinst, bijwerkingen, risico's overwogen*	2,33	2,25	3,00	2,50	3,00	2,33	2,00	2,33	2,66
12	*expliciet verband aanbevelingen en bewijs*	2,00	2,25	2,33	2,75	3,00	2,33	2,75	2,00	2,51
13	*richtlijn door externe experts beoordeeld*	2,33	4,00	2,75	1,00	3,25	2,50	3,00	2,00	2,77
14	*procedure herziening vermeld*	2,00	2,00	2,00	1,00	3,25	1,33	1,50	1,67	1,90
	Helderheid en Presentatie	**61,11**	**41,67**	**58,33**	**47,92**	**77,08**	**58,33**	**58,33**	**58,33**	**61,66**
15	*aanbevelingen specifiek en ondubbelzinnig*	3,33	2,50	3,25	2,50	3,50	1,67	2,50	3,00	2,88
16	*verschillende beleidsopties vermeld*	3,33	2,50	2,33	1,75	3,25	2,33	2,50	3,00	2,85
17	*kernaanbevelingen gemakkelijk te herkennen*	3,67	3,00	3,67	2,50	3,50	1,67	3,50	3,00	3,11
18	*toepassing ondersteund met hulpmiddelen*	2,00	1,75	2,25	1,25	3,50	1,67	2,00	2,67	2,48
	Toepassing	**40,74**	**25,00**	**27,78**	**,00**	**22,22**	**40,74**	**13,89**	**3,70**	**35,31**

Bijlage 9 Frequenties maximale ranges per arbokennisproduct over 23 items

	STECR								NVAB						Overig								
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	
Range	Werkwijzer ABBE-rsi	Werkwijzer Arbeidsconflicten	Werkwijzer Aanpak Problematisch verzuim	Werkwijzer Lage rugklachten en Arbeidsomstandigheden	Werkwijzer Reductie Werkstress gezondheidszorg/onderwijs	Assist Cultuur of Persoon. SMB Allochtonen	Voorstudie en Assist Vrouwen niet in de WAO	Werkwijzer Somatisatie	Richtlijn Lage Rugklachten	Richtlijn Psychische Klachten	Richtlijn Astma en COPD	Richtlijn Klachten van Arm Nek en Schouder	Leidraad Verwijzen	LESA Overspanning	Geen benutbare mogelijkheden UWV	Zwangerschap en arbeidsongeschiktheid UWV	Registratierichtlijn Burnout NCvB	Interventies Acute Lage Rugklachten NKAB	Arbeidshulpverlening Psychische Klachten NKAP	Diabetes Handboek Arbeid en Belastbaarheid BSL	PAGO Huisvuilbeladers Coronel	SMB Richtlijn Rugklachten Arbouw	Gemiddeld
0	0	1	2	1	2	1	3	0	9	10	1	2	4	0	12	3	0	12	0	3	7	7	3,48
1	5	10	6	9	10	12	11	9	9	6	17	12	14	9	8	12	11	4	8	10	4	8	9,33
2	16	10	11	5	9	6	9	6	5	1	4	9	3	12	4	2	11	6	14	6	9	6	7,52
3	3	3	5	9	3	5	1	9	1	9	2	1	3	3	1	7	2	4	2	5	4	3	3,90
N	4	5	4	3	4	4	4	4	4	2	4	4	4	6	3	4	4	4	4	3	4	3	

Bijlage 10 Intraclass Coëfficiënten per item en domein bij 2, 3 en 4 beoordelaars

	B = aantal beoordelaars; N = aantal arbokennisproducten	B=2 N=22		B=3 N=21		B=4 N=17		
	1 = ICC 1 beoordelaar, K = ICC gemiddelde K beoordelaars	1	K	1	K	1	K	
	Onderwerp en Doel	0,445	0,715	0,523	0,767	0,380	0,710	*
1	Het doel van de richtlijn is specifiek beschreven.	0,362	0,532	0,384	0,651	0,214	0,527	
2	De klinische vraag/vragen die in de richtlijn aan de orde komt/komen, is/zijn specifiek.	0,024	0,048	0,093	0,234	0,110	0,340	*
3	De patiëntenpopulatie waarop de richtlijn van toepassing is, is specifiek beschreven.	0,477	0,646	0,529	0,771	0,570	0,845	*
	Betrokkenheid Belanghebbenden	0,214	0,352	0,406	0,672	0,570	0,841	*
4	De leden van de werkgroep die de richtlijn heeft ontwikkeld, komen uit alle relevante beroepsgroepen.	0,176	0,300	0,237	0,482	0,427	0,756	*
5	Het perspectief en de voorkeuren van patiënten zijn nagegaan.	0,053	0,101	0,298	0,560	0,348	0,681	*
6	De beoogde gebruikers van de richtlijn zijn duidelijk benoemd.	0,262	0,415	0,254	0,505	0,430	0,751	*
7	De richtlijn is getest onder de beoogde gebruikers.	0,404	0,576	0,466	0,724	0,402	0,729	*
	Methodologie	0,177	0,300	0,359	0,627	0,496	0,797	*
8	Er zijn systematische methoden gebruikt voor het zoeken naar wetenschappelijk bewijsmateriaal.	0,211	0,349	0,288	0,548	0,385	0,714	*
9	De criteria voor het selecteren van het wetenschappelijk bewijsmateriaal zijn duidelijk beschreven.	0,170	0,291	0,302	0,565	0,490	0,794	*
10	De gebruikte methoden om de aanbevelingen op te stellen, zijn duidelijk beschreven.	0,449	0,620	0,407	0,673	0,297	0,628	*
11	Gezondheidswinst, bijwerkingen en risico's zijn overwogen bij het opstellen van de aanbevelingen.	-0,270	-0,738	0,080	0,287	0,172	0,453	*
12	Er bestaat een expliciet verband tussen de aanbevelingen en het onderliggende wetenschappelijke bewijsmateriaal.	0,184	0,310	0,425	0,689	0,508	0,805	*
13	De richtlijn is vóór publicatie door externe experts beoordeeld.	0,503	0,669	0,235	0,470	0,390	0,719	*
14	Een procedure voor herziening van de richtlijn is vermeld.	0,146	0,255	0,291	0,552	0,233	0,549	
	Helderheid en Presentatie	-0,172	-0,416	0,156	0,356	0,261	0,586	*
15	De aanbevelingen zijn specifiek en ondubbelzinnig.	0,326	0,491	0,111	0,272	0,170	0,453	
16	De verschillende beleidsopties zijn duidelijk vermeld.	-0,069	-0,148	0,137	0,322	0,144	0,401	*
17	De kernaanbevelingen zijn gemakkelijk te herkennen.	-0,293	-0,828	0,077	0,200	0,040	0,142	
18	De toepassing van de richtlijn wordt ondersteund met hulpmiddelen.	0,171	0,292	0,247	0,496	0,375	0,706	*
	Toepassing	0,492	0,659	0,429	0,692	0,385	0,715	*
19	De mogelijke organisatorische belemmeringen bij het toepassen van de aanbevelingen zijn besproken.	0,447	0,618	0,361	0,629	0,342	0,675	*
20	De mogelijke kostenimplicaties van het toepassen van de aanbevelingen zijn overwogen.	0,316	0,480	0,251	0,502	0,243	0,563	*
21	De richtlijn geeft de belangrijkste criteria om na te gaan en te toetsen of de richtlijn wordt gevolgd.	0,308	0,470	0,339	0,606	0,474	0,704	*
	Onafhankelijkheid Opstellers	0,192	0,323	0,289	0,549	0,184	0,474	
22	De richtlijn is niet beïnvloed door de opvattingen of belangen van de financierende instantie.	-0,037	-0,078	0,087	0,223	0,111	0,334	*
23	Conflicterende belangen van leden van de werkgroep zijn vastgelegd.	0,402	0,573	0,306	0,570	0,167	0,444	
24	**Deze richtlijn is aan te bevelen voor gebruik in de praktijk.**	0,238	0,384	0,226	0,466	0,322	0,655	*

*: ICC gemiddelde beoordelaars stijgt bij toename aantal beoordelaars
xxx: onderstreept: niet significant p>0,05

	item ICC2 N=22		item ICC3 N=21		item ICC4 N=17		domein ICC2 N=22		domein ICC3 N=21		domein ICC4 N=17	
	1	K	1	K	1	K	1	K	1	K	1	K
'zeer goed' ICC hoger dan 0,75	0	0	0	1	0	5	0	0	0	1	0	2
'goed' ICC tussen 0,60 en 0,75	0	3	0	4	0	7	0	2	0	2	0	2
'matig' ICC tussen 0,40 en 0,60	6	8	4	12	6	8	2	0	3	2	4	2
'slecht' item ICC onder 0,40	17	12	19	6	17	3	4	4	3	1	2	0
aantal significant	6	6	17	17	16	16	2	2	5	5	5	5

Bij 3 beoordelaars laagste domein ICC: Helderheid Presentatie 0,156 - 0,356 (1-K)
Bij 4 beoordelaars laagste domein ICC: Onafhankelijkheid Opstellers 0,184 - 0,474 (1-K)
Bij 3 beoordelaars hoogste domein ICC: Onderwerp en Doel 0,523 - 0,767 (1-K)
Bij 4 beoordelaars hoogste domein ICC: Betrokkenheid Belanghebbenden 0,570 - 0,841 (1-K)

Bijlage 11 Scores toegevoegde vragen per arbokennisproduct: STECR

		1	2	3	4	5	6	7	8	
Voor feitelijke formulering vragen zie vragenformulier		Werkwijzer ABBE-rsi	Werkwijzer Arbeidsconflicten	Werkwijzer Aanpak Problematisch verzuim	Werkwijzer Lage rugklachten en Arbeidsomstandigheden	Werkwijzer Reductie Werkstress gezondheidszorg/onderwijs	Assist Cultuur of Persoon. SMB Allochtonen	Voorstudie en Assist Vrouwen niet in de WAO	Werkwijzer Somatisatie	Alle arbokennisproducten gemiddeld
	Richtlijnigheid (26,27,28,29)	**2,44**	**3,30**	**3,00**	**2,75**	**3,13**	**2,88**	**2,44**	**3,00**	**3,14**
26	*bevat aanbevelingen goed handelen bij probleem*	2,00	3,60	3,00	2,67	3,25	2,50	2,50	2,50	3,18
27	*handelen veel arboprofessionals wordt beïnvloed*	2,75	3,40	3,25	3,33	3,50	3,00	2,75	3,50	3,21
28	*uitgevende organisatie heeft vertrouwen praktijk.*	3,00	3,00	2,75	2,67	3,25	3,00	2,50	3,00	3,18
29	*wordt gezien als richtinggevend*	2,00	3,20	3,00	2,33	2,50	3,00	2,00	3,00	2,94
	Gebruiksgemak AGREE (30,31,32)	**3,17**	**2,80**	**3,17**	**3,33**	**3,08**	**3,25**	**2,67**	**3,08**	**2,96**
30	*Mijn deskundigheid was voldoende*	3,50	3,00	2,75	3,67	3,25	3,25	3,00	3,25	3,36
31	*beantwoorden vragen was eenvoudig.*	3,25	2,20	3,25	3,00	3,00	3,00	2,75	3,00	2,99
32	*informatie was goed te vinden in document.*	2,75	3,20	3,50	3,33	3,00	3,50	2,25	3,00	3,22
33	*aanvullende informatie gebruikt**	4,00	3,40	4,00	4,00	2,75	4,00	3,50	4,00	2,32
	Waardering AGREE (36,37,38,39,40,43,45,46)	**3,06**	**2,58**	**3,22**	**3,25**	**2,94**	**3,03**	**2,94**	**3,17**	**3,19**
34	*Mijn oordeel is door invullen AGREE veranderd*	1,75	1,40	1,50	2,33	1,50	1,75	1,25	2,25	2,16
36	*AGREE vragenlijst zinvol instrument*	3,50	2,40	3,00	3,33	3,25	3,25	3,00	3,33	3,11
37	*vragen onvoldoende toegespitst op arbocontext**	2,50	2,00	2,75	3,00	2,25	3,00	3,50	3,00	1,80
38	*AGREE is te gedetailleerd**	2,75	2,60	2,75	3,33	3,00	3,50	2,50	3,25	1,82
39	*niet als richtlijn bedoeld, dus AGREE niet relevant**	3,00	2,60	3,50	3,67	2,50	3,00	2,75	3,50	2,13
40	*kwaliteitsoordeel eenvoudiger zonder AGREE**	3,00	2,00	3,25	3,00	3,25	2,75	2,75	2,50	1,81
41	*in AGREE ontbreken noodzakelijke vragen**	2,50	2,60	2,75	3,00	2,75	3,00	3,50	3,00	1,95
43	*bij brede verspreiding (AGREE) keurmerk wenselijk*	4,00	3,20	3,75	3,33	3,00	2,75	3,25	3,50	3,09
44	*arboprofessionals hebben onvoldoende kennis voor beoordeling inhoudelijke kwaliteit**	2,00	3,40	3,00	2,00	2,75	2,50	2,50	1,75	2,26
45	*kennis over mate van bewijs beïnvloedt vrijheids-graden*	3,00	3,20	3,00	2,67	3,25	3,00	3,50	2,75	3,05
46	*AGREE hulpmiddel bij zelf beoordelen arbokennis-product*	2,75	2,60	3,75	3,67	3,00	3,00	2,25	2,75	2,88
	Aantal beoordelaars	2	5	4	3	4	3	4	3	3,7

*: antwoordscores omgepoold

Bijlage 12 Scores toegevoegde vragen per arbokennisproduct: NVAB

		9	10	11	12	13	14	
Voor feitelijke formulering vragen zie vragenformulier		Richtlijn Lage Rugklachten	Richtlijn Psychische Klachten	Richtlijn Astma en COPD	Richtlijn Klachten van Arm Nek en Schouder	Leidraad Verwijzen	LESA Overspanning	Alle arbokennisproducten gemiddeld
	Richtlijnigheid (26,27,28,29)	**3,94**	**3,88**	**3,69**	**3,56**	**3,88**	**3,45**	**3,14**
26	*bevat aanbevelingen goed handelen bij probleem*	4,00	4,00	3,75	3,25	4,00	3,17	3,18
27	*handelen veel arboprofessionals wordt beïnvloed*	3,75	4,00	3,75	3,75	3,75	3,17	3,21
28	*uitgevende organisatie heeft vertrouwen praktijk.*	4,00	4,00	3,75	3,75	4,00	3,67	3,18
29	*wordt gezien als richtinggevend*	4,00	3,50	3,50	3,50	3,75	3,60	2,94
	Gebruiksgemak AGREE (30,31,32)	**3,67**	**2,83**	**3,33**	**3,50**	**3,50**	**3,28**	**2,96**
30	*Mijn deskundigheid was voldoende*	3,75	3,00	3,50	3,75	3,50	3,33	3,36
31	*beantwoorden vragen was eenvoudig.*	3,50	2,50	3,00	3,25	3,25	3,00	2,99
32	*informatie was goed te vinden in document.*	3,75	3,00	3,50	3,50	3,75	3,50	3,22
33	*aanvullende informatie gebruikt**	3,00	3,00	2,75	3,50	2,50	3,00	2,32
	Waardering AGREE (36,37,38,39,40,43,45,46)	**3,69**	**3,19**	**3,29**	**3,06**	**2,91**	**2,79**	**3,19**
34	*Mijn oordeel is door invullen AGREE veranderd*	1,50	2,00	1,50	1,00	1,75	2,17	2,16
36	*AGREE vragenlijst zinvol instrument*	3,75	3,50	3,50	3,25	3,50	2,67	3,11
37	*vragen onvoldoende toegespitst op arbocontext**	3,75	3,00	2,75	3,25	3,50	2,83	1,80
38	*AGREE is te gedetailleerd**	3,75	3,00	3,00	2,75	2,75	2,67	1,82
39	*niet als richtlijn bedoeld, dus AGREE niet relevant**	4,00	4,00	4,00	3,75	2,75	3,17	2,13
40	*kwaliteitsoordeel eenvoudiger zonder AGREE**	3,25	3,00	3,00	3,00	2,50	2,50	1,81
41	*in AGREE ontbreken noodzakelijke vragen**	3,75	3,00	2,25	3,25	2,75	2,67	1,95
43	*bij brede verspreiding (AGREE) keurmerk wenselijk*	3,50	3,00	3,25	2,50	2,50	2,83	3,09
44	*arboprofessionals hebben onvoldoende kennis voor beoordeling inhoudelijke kwaliteit**	3,50	2,00	2,75	3,00	3,00	2,83	2,26
45	*kennis over mate van bewijs beïnvloedt vrijheids-graden*	3,75	3,50	3,25	2,75	3,25	3,00	3,05
46	*AGREE hulpmiddel bij zelf beoordelen arbokennis-product*	3,75	2,50	3,00	3,25	2,50	2,67	2,88
	Aantal beoordelaars	4	2	4	4	4	6	3,7

*: antwoordscores omgepoold

Bijlage 13 Scores toegevoegde vragen per arbokennisproduct: Overige

		15	16	17	18	19	20	21	22	
	Voor feitelijke formulering vragen zie vragenformulier	Geen benutbare mogelijkheden UWV	Zwangerschap en arbeids-ongeschiktheid UWV	Registratierichtlijn Burnout NCvB	Interventies Acute Lage Rugklachten NKAB	Arbeidshulpverlening Psychische Klachten NKAP	Diabetes Handboek Arbeid en Belastbaarheid BSL	PAGO Huisvuilbeladers Coronel	SMB Richtlijn Rugklachten Arbouw	Alle arbokennisproducten gemiddeld
	Richtlijnigheid (26,27,28,29)	**2,75**	**2,81**	**3,25**	**2,50**	**3,25**	**2,67**	**3,50**	**2,83**	**3,14**
26	*bevat aanbevelingen goed handelen bij probleem*	4,00	3,50	3,25	2,50	3,75	2,33	3,50	3,00	3,18
27	*handelen veel arboprofessionals wordt beïnvloed*	2,67	2,25	3,50	2,25	3,25	2,67	3,50	3,00	3,21
28	*uitgevende organisatie heeft vertrouwen praktijk.*	2,00	2,50	3,00	3,00	3,00	3,00	3,50	3,33	3,18
29	*wordt gezien als richtinggevend*	2,33	3,00	3,25	1,50	3,00	2,67	3,50	2,00	2,94
	Gebruiksgemak AGREE (30,31,32)	**3,33**	**3,42**	**3,17**	**2,92**	**3,00**	**3,11**	**3,25**	**3,33**	**2,96**
30	*Mijn deskundigheid was voldoende*	3,67	3,50	3,50	3,25	3,50	3,33	3,50	3,33	3,36
31	*beantwoorden vragen was eenvoudig.*	3,00	3,25	3,00	3,25	2,50	2,67	3,00	3,00	2,99
32	*informatie was goed te vinden in document.*	3,33	3,50	3,00	2,25	3,00	3,33	3,25	3,67	3,22
33	*ik hebaanvullende informatie gebruikt**	3,00	2,75	3,75	2,25	3,50	3,67	3,25	3,67	2,32
	Waardering AGREE (36,37,38,39,40,43,45,46)	**2,83**	**2,84**	**2,91**	**3,19**	**2,50**	**2,33**	**2,81**	**3,00**	**3,19**
34	*Mijn oordeel is door invullen AGREE veranderd*	2,00	2,00	1,50	3,00	2,50	1,33	2,00	2,67	2,16
36	*AGREE vragenlijst zinvol instrument*	3,00	2,75	3,25	3,00	3,00	2,33	3,00	3,33	3,11
37	*vragen onvoldoende toegespitst op arbocontext**	2,67	2,25	2,75	3,00	1,75	3,00	2,50	3,00	1,80
38	*AGREE is te gedetailleerd**	2,67	3,00	1,75	3,25	2,25	3,00	2,25	2,67	1,82
39	*niet als richtlijn bedoeld, dus AGREE niet relevant**	3,33	3,25	3,25	3,00	2,75	1,67	2,75	3,00	2,13
40	*kwaliteitsoordeel eenvoudiger zonder AGREE**	3,00	2,75	3,25	3,50	2,25	1,67	3,00	3,00	1,81
41	*in AGREE ontbreken noodzakelijke vragen*	3,00	2,75	3,50	3,00	3,00	2,33	3,25	2,33	1,95
43	*bij brede verspreiding (AGREE) keurmerk wenselijk*	2,33	2,75	3,50	3,50	2,75	2,33	3,50	2,67	3,09
44	*arboprofessionals hebben onvoldoende kennis voor beoordeling inhoudelijke kwaliteit**	3,00	2,75	2,50	3,50	2,75	3,00	2,50	2,67	2,26
45	*kennis over mate van bewijs beïnvloedt vrijheids-graden*	3,00	3,50	2,75	3,25	2,25	2,33	3,00	3,00	3,05
46	*AGREE hulpmiddel bij zelf beoordelen arbokennis-product*	2,67	2,50	2,75	3,00	3,00	2,33	2,50	3,33	2,88
	Aantal beoordelaars:	3	4	4	4	4	3	4	3	3,7

*: antwoordscores omgepoold