

RIJKSUNIVERSITEIT GRONINGEN

The Disability Assessment Structured Interview

Its reliability and validity in work disability assessment

Proefschrift

ter verkrijging van het doctoraat in de
Medische Wetenschappen
aan de Rijksuniversiteit Groningen
op gezag van de
Rector Magnificus, dr. F. Zwarts,
in het openbaar te verdedigen op
woensdag 17 maart 2010
om 14.45 uur

door

Jerry Spanjer

geboren op 29 november 1958
te Groningen

| | |
|------------------------|--|
| Promotor: | Prof. dr. J. W. Groothoff |
| Copromotores: | Dr. B. Krol Dr. S. Brouwer |
| Beoordelingscommissie: | Prof. dr. M. H. W. Frings-Dresen Prof. dr. J. H. B. Geertzen Prof. dr. J. H. B. M. Willems |

Samenvatting

Als werknemers door ziekte hun werk niet meer kunnen doen betaalt de werkgever het loon twee jaar door. Hierna kan een WIA uitkering bij de UWV worden aangevraagd. De beoordeling of de werknemer voor deze uitkering in aanmerking komt begint bij de verzekeringsarts, die de functionele beperkingen en mogelijkheden beoordeelt. Verzekeringsartsen in Nederland baseren hun oordeel voornamelijk op het gesprek dat ze met de cliënt hebben en daarnaast op het lichamelijk onderzoek en op de informatie van de bedrijfsarts en behandeld arts.

Deze belastbaarheidbeoordeling kan grote gevolgen hebben voor zowel de cliënt als de maatschappij. Het is daarom van belang dat verschillende verzekeringsartsen tot hetzelfde oordeel komen (betrouwbaarheid) en dat de juiste beperkingen en mogelijkheden worden vastgesteld (validiteit). In de literatuur zijn echter nauwelijks gegevens bekend over de betrouwbaarheid en validiteit van deze belastbaarheidbeoordelingen.

Het gesprek dat de verzekeringsarts heeft met de cliënt speelt een belangrijke rol bij de beoordeling van de belastbaarheid. In dit proefschrift wordt het Belastbaarheid Gericht Beoordelingsgesprek (BGB), nader onderzocht. De BGB is een semi-gestructureerd interview protocol met vaste gespreksonderwerpen, die grotendeels gebaseerd zijn op het model van de International Classification of Functioning, Disability and Health (ICF). De belangrijkste onderwerpen zijn: werk, stoornissen, ervaren belemmeringen, participatie, visie van de cliënt en de mening van de verzekeringsarts. Karakteristiek voor de BGB is de semi-gestructureerde manier van uitvragen en het doorvragen naar specifieke en gedetailleerde voorbeelden van beperkingen en concrete activiteiten die nog wel worden ondernomen.

In het eerste deel van dit proefschrift wordt in een model beschreven welke instrumenten gebruikt kunnen worden om de belastbaarheid te beoordelen en welke mogelijke oorzaken er zijn voor de variatie tussen artsen in belastbaarheidbeoordeling. Verder werd in de literatuur gezocht naar de betrouwbaarheid en validiteit van instrumenten om de beperkingen te kunnen beoordelen bij mensen die een uitkering aanvragen.

In het tweede deel worden vier studies beschreven waarin de betrouwbaarheid en validiteit werden onderzocht van belastbaarheidbeoordelingen, waarbij gebruik werd gemaakt van de BGB. Ook werd de mening van cliënten en verzekeringsartsen over de BGB gemeten.

Hoofdstuk 1 bevat een algemene introductie waarin het doel en de inhoud van dit proefschrift worden beschreven, uitmondend in de volgende onderzoeksvragen:

1. Wat zijn mogelijke bronnen van variatie in de beoordeling van belastbaarheid?
2. Welke instrumenten worden beschreven om de beperkingen te meten of beoordelen bij mensen die een uitkering aanvragen en wat zijn de psychometrische kwaliteiten?
3. Wat is, naast de medische anamnese, het effect van gedetailleerde informatie over functioneren op de beoordeling van beperkingen en de betrouwbaarheid?
4. Zijn artsen van mening dat ze beperkingen kunnen beoordelen op basis van een schriftelijk BGB rapport?

5. Wat zijn de kenmerken van de BGB in de dagelijkse praktijk?
6. Wat is de cliënttevredenheid over artsen die de BGB gebruiken?
7. Welk commentaar hebben verzekeringsartsen op de BGB?
8. Wat vinden verzekeringsartsen van de BGB als ze die zelf toepassen?
9. Wat is de intra- en interbeoordelaar betrouwbaarheid van beoordelingsgesprekken bij gebruik van de BGB?
10. Wat is de validiteit van beoordelingsgesprekken bij gebruik van de BGB?

In **Hoofdstuk 2** wordt in een model beschreven welke instrumenten gebruikt kunnen worden om de belastbaarheid te beoordelen en welke mogelijke oorzaken er zijn voor de variatie tussen artsen in hun oordeel over de belastbaarheid. Variatie tussen artsen in het beoordelen van belastbaarheid kan veroorzaakt worden op drie verschillende niveaus, met op elk niveau verschillende instrumenten die gebruikt kunnen worden. Ten eerste moeten gegevens worden verzameld. Instrumenten die hierbij gebruikt kunnen worden zijn vragenlijsten, functionele capaciteit evaluatie (FCE) en medische beoordelingen. Ten tweede dienen deze gegevens te worden geïnterpreteerd, waarvoor instrumenten als richtlijnen en protocollen gebruikt kunnen worden. Tenslotte dient het uiteindelijke oordeel te worden vastgelegd, waarvoor in Nederland vaak de Functionele Mogelijkheden Lijst (FML) wordt gebruikt. In alle drie stappen spelen persoonlijke eigenschappen van de beoordelaar een belangrijke rol, waaronder culturele achtergrond, normen en opleiding van de beoordelaar.

In **Hoofdstuk 3** wordt besproken dat er relatief veel onderzoek is verricht naar betrouwbaarheid en validiteit van instrumenten om de beperkingen te beoordelen bij patiënten in een revalidatie setting en minder bij mensen die een arbeidsongeschiktheidsuitkering aanvragen. Daarom werd een systematisch literatuuronderzoek verricht naar instrumenten om de beperkingen te beoordelen bij mensen die een uitkering aanvragen. Er werd gezocht in de bibliografische databestanden van Medline, Embase, CINAHL en PsycINFO naar studies die psychometrische kwaliteiten beschrijven van deze instrumenten. Twee beoordelaars selecteerden onafhankelijk van elkaar relevante artikelen en beoordeelden de psychometrische kwaliteiten van de gevonden instrumenten. Er werden vier instrumenten gevonden: de Roland-Morris Disability Questionnaire (RDQ), de Patient-Specific Functional Scale (PSFS), de Isernhagen Work System (IWS) en de Multiperspective Multidimensional Pain Assessment Protocol (MMPAP). De vragenlijsten (RDQ en PSFS) waren niet specifiek op de werksituatie georiënteerd en beperkingen werden slechts oppervlakkig gemeten. De psychometrische kwaliteiten van de IWS in de claimsituatie waren slecht tot matig. Van de MMPAP werd alleen de predictieve validiteit gemeten. De instrumenten beoordeelden drie tot 34 lichamelijke beperkingen en er werden geen instrumenten gevonden om mentale beperkingen te beoordelen. Er werden geen instrumenten gevonden met aangetoonde acceptabele psychometrische kwaliteiten.

In **Hoofdstuk 4** wordt een studie beschreven waarin cliënten gedetailleerde informatie over hun functioneren verstrekten. Het effect van deze informatie, naast de medische anamnese, op de beoordeelde beperkingen en de interbeoordelaar betrouwbaarheid werd onderzocht. Drie verschillende groepen van elk negen verzekeringsartsen kregen verschillende informatie over 30 cliënten: de eerste groep kreeg alleen medische informatie, de tweede groep kreeg gedetailleerde informatie over het functioneren van de cliënt en de derde groep kreeg beide.

De groep artsen die gedetailleerde informatie over het functioneren van de cliënt had gekregen concludeerde dat er zwaardere beperkingen waren dan de groep artsen die alleen van medische informatie was voorzien. De gedetailleerde informatie over het functioneren had geen duidelijke invloed op de interbeoordelaar betrouwbaarheid vergeleken met medische of beide soorten informatie.

In **Hoofdstuk 5** wordt een onderzoek beschreven waarin werd onderzocht of artsen op basis van een schriftelijk rapport een belastbaarheid beoordeling kunnen verrichten en wat in dat geval de interbeoordelaar betrouwbaarheid is. In totaal 12 artsen beoordeelden op basis van een schriftelijk rapport de belastbaarheid van 12 cliënten. De rapporten deden verslag van BGB gesprekken die gevoerd en vastgelegd waren door speciaal daarvoor opgeleide verpleegkundigen. De verzekeringsartsen werden gevraagd of ze van mening waren dat ze op basis van deze rapporten een betrouwbare beoordeling konden verrichten. Daarnaast werd de betrouwbaarheid tussen de artsen gemeten door het percentage overeenstemming op de items van het Functie Informatie Systeem (FIS) en de Psychische Mogelijkheden Lijst (PML) te berekenen. De kwaliteit van de rapporten werd door de artsen als redelijk tot goed beoordeeld. De helft van de artsen was van mening dat op basis van deze rapporten een redelijk betrouwbare beoordeling kon worden verricht, een kwart vond dat niet en een kwart had geen mening. De overeenkomst tussen de verzekeringsartsen was redelijk tot goed met een gemiddeld overeenstemmingspercentage van 76% (spreiding 64 – 88%). Wat betreft de urenbeperkingen (als de arts van mening is dat de cliënt geen hele dag kan functioneren) was de overeenstemming tussen de artsen laag.

In **Hoofdstuk 6** wordt een onderzoek beschreven waarin de inhoud van de BGB werd geanalyseerd, artsen naar hun commentaar op de BGB werd gevraagd en waarin de intra- en interbeoordelaars betrouwbaarheid van beoordelingen gebaseerd op de BGB werd bepaald. Totaal 14 BGB gesprekken van vijf verzekeringsartsen werden op de video opgenomen. Deze video opnames werden getoond aan 22 verzekeringsartsen, die hun commentaar gaven op de gesprekken. De gesprekken werden geanalyseerd door de duur van de verschillende gespreksonderwerpen te meten. De interbeoordelaar betrouwbaarheid tussen de 22 artsen werd gemeten door het percentage overeenstemming op de items van het Functie Informatie Systeem (FIS) en de Psychische Mogelijkheden Lijst (PML) te berekenen. De intra-beoordelaar betrouwbaarheid werd gemeten door de vijf verzekeringsartsen, die de opnames hadden gemaakt, te vragen de belastbaarheid te beoordelen vlak na de opname

en na het terugzien van de video na zes maanden.

De gemiddelde duur van de gesprekken was 33 minuten (spreiding 19 – 77 min). Er werd evenveel tijd besteed aan stoornissen als aan ervaren belemmeringen en participatie (beide 30% van het totale gesprek). De gemiddelde duur van de introductie was 9% van het totale gesprek, 9% ging over het werk, de visie van de cliënt duurde 8%, de mening van de verzekeringsarts 13% en de daarop volgende reactie van de cliënt duurde gemiddeld 4% van het hele gesprek. In het algemeen waren de artsen van mening dat de beoordelingsgesprekken gestructureerd, functioneel en efficiënt waren. In hun eigen beoordelingsgesprekken werd meer aandacht aan medische zaken besteed en minder aan de door de cliënten ervaren belemmeringen. Het gemiddelde overeenstemmingspercentage tussen de 22 artsen op de items van de FIS en PML was 74% (spreiding 56 – 85%).

Er was een aanzienlijk verschil tussen de artsen bij beoordeling van de urenbeperkingen. De intrabeoordelaar overeenkomst was 80% (spreiding 52 – 100%). Als de artsen hun eigen video opnames na 6 maanden weer terug zagen werden minder beperkingen gescoord in vergelijking met de beperkingen op basis van een zojuist verricht gesprek.

In **Hoofdstuk 7** wordt de betrouwbaarheid en validiteit van de BGB in een gerandomiseerd onderzoek met controlegroep onderzocht. Cliënten die na 21 maanden ziekte een WIA uitkering aanvroegen, werden geïnterviewd en onderzocht door twee artsen die een BGB training hadden ondergaan (n=32) of door twee artsen uit de controle groep zonder BGB training (n=30). De twee artsen beoordeelden de cliënten onafhankelijk van elkaar. Overeenstemmingspercentages, wel of niet in aanmerking komen voor een WIA uitkering en verschillen in scores op de FML items werden gemeten om de betrouwbaarheid en concurrente validiteit te onderzoeken. De inhoudsvaliditeit werd bepaald door de verzekeringsartsen, die een BGB training hadden ondergaan (n=8), een vragenlijst te laten invullen met betrekking tot hun mening over de BGB. Daarnaast vulden de cliënten een vragenlijst in om hun tevredenheid met betrekking tot gedragsaspecten van de artsen te meten.

Er werden geen duidelijke verschillen in overeenstemmingspercentages gevonden tussen de BGB en de controle groep, het gemiddelde overeenstemmingspercentage was 81 % (spreiding 59 - 91%). De inhoudsvaliditeit was goed: alle artsen vonden dat in de BGB aandacht wordt besteed aan alle relevante aspecten en dat de BGB voldoende informatie geeft om de beperkingen te beoordelen. De artsen waren van mening dat de BGB een realistisch beeld geeft van de cliënt en een acceptabele gespreksmethodiek is in de dagelijkse praktijk. Zeven van de acht verzekeringsartsen waren van mening dat de BGB een betere basis was om de FML in te vullen dan hun gebruikelijke beoordelingsgesprek. De artsen van de controlegroep gaven in 9 van de 21 FML items minder beperkingen aan dan de artsen van de BGB groep. Dit leidde echter niet tot meer uitkeringen in de BGB groep. De cliënttevredenheid met betrekking tot de artsen was in zowel de controle als in de BGB groep 7,7 op een schaal van 1 tot 10.

Hoofdstuk 8 geeft een samenvatting en discussie van de belangrijkste bevindingen van dit proefschrift. Verder worden de methodologische aspecten besproken en aanbevelingen gedaan voor verder onderzoek, de praktijk en beleid. De belangrijkste bevindingen kunnen worden samengevat op geleide van de 10 onderzoeksvragen:

1. Variatie in de beoordeling van belastbaarheid kan worden veroorzaakt op drie niveaus: gegevensverzameling, interpretatie en vastleggen van het oordeel.
2. In de literatuur werden geen instrumenten met voldoende psychometrische kwaliteiten gevonden om de belastbaarheid in een uitkeringssituatie te beoordelen.
3. Door de cliënt verstrekte gedetailleerde informatie over diens functioneren, naast de medische anamnese, had als gevolg dat artsen meer beperkingen aanwezig achtten. Deze informatie over het functioneren had geen invloed op de interbeoordelaar betrouwbaarheid.
4. De helft van de artsen was van mening dat een betrouwbare beoordeling verricht kan worden op basis van een schriftelijke rapportage, 25% vond dat niet en 25% had geen mening.
5. Een BGB gesprek zonder lichamelijk onderzoek duurde gemiddeld 30-45 minuten. Ongeveer 30% van het totale gesprek ging over medische zaken en 30% over beperkingen en activiteiten.
6. De cliënttevredenheid over artsen die de BGB gebruikten was met een score van 7,7 op een schaal van 1 tot 10 gelijk aan die van een controlegroep.
7. Verzekeringsartsen die video opnames van BGB interviews beoordeelden gaven aan dat de BGB functioneel en efficiënt was, doch dat de puur medische anamnese aan de magere kant was.
8. De artsen die de BGB zelf toepasten na een training waren van mening dat de BGB een realistisch beeld geeft van de beperkingen en mogelijkheden van de cliënt en een acceptabele gespreksmethodiek is in de dagelijkse praktijk. Ook de medische anamnese werd voldoende bevonden.
9. De intra- en interbeoordelaar betrouwbaarheid van beoordelingsgesprekken bij gebruik van de BGB was redelijk tot goed, behalve op het item "urenbeperking". De interbeoordelaar betrouwbaarheid was vergelijkbaar met een controlegroep zonder BGB training.
10. De inhoudsvaliditeit van de BGB was goed. Artsen die de BGB toepasten gaven meer beperkingen aan in hun beoordeling dan artsen zonder BGB training. Omdat een "gouden standaard" ontbreekt, is niet bekend wie de "juiste" beperkingen beoordeelde.

In de discussie over de methodologie worden het gebrek aan een gouden standaard en het gebrek aan voldoende psychometrische kwaliteiten van de uitkomstmaten, waaronder de FML, besproken. Het interview in het algemeen en de BGB in het bijzonder lijkt een waardevol instrument, maar onderzocht zou moeten worden of aanvullende instrumenten een rol in de beoordeling van functionele mogelijkheden kunnen spelen. Zo kan een combinatie van functionele capaciteitstesten en/of vragenlijsten naast het onderzoek van de verzekeringsarts overwogen worden; of

beoordelingen door meerdere verzekeringsartsen van dezelfde cliënt.

Om de variatie tussen verzekeringsartsen te reduceren lijkt het eenduidig interpreteren van informatie meer bij te dragen dan eenduidige gegevensverzameling in het beoordelingsgesprek. Het is daarom nuttig aandacht te besteden aan onderzoek naar - en ontwikkeling van duidelijke richtlijnen en protocollen, bijvoorbeeld de standaard verminderde arbeidsduur. Aanbevolen wordt om richtlijnen en protocollen eerst in de praktijk uit te testen alvorens ze te implementeren in de dagelijkse praktijk. Verder kunnen de controle op de uitvoering en evaluatie na implementatie nog verbeterd worden.

Het is te vroeg om te adviseren de BGB landelijk als voorkeurs gespreksmethodiek in te voeren, omdat het wenselijk is eerst verder onderzoek naar de validiteit van de BGB te verrichten. Wel zijn er voldoende argumenten om de BGB training aan te bieden aan verzekeringsartsen in opleiding en aan geregistreerde verzekeringsartsen in het kader van nascholing.

Algemene conclusie

In de literatuur werden geen instrumenten met voldoende psychometrische kwaliteiten gevonden om de belastbaarheid in een uitkeringssituatie te beoordelen. Het Belastbaarheid Gerichte Beoordelingsgesprek (BGB) lijkt dit manco op te kunnen vullen, omdat het een bruikbaar instrument is met acceptabele betrouwbaarheid en tevredenheid bij zowel verzekeringsartsen als cliënten. Er werd een goede inhoudsvaliditeit aangetoond, echter verder onderzoek naar de validiteit is nodig omdat verzekeringsartsen die de BGB toepasten meer beperkingen aangaven dan een controlegroep die de BGB niet toepaste. Om de overeenkomst in beoordelingen tussen verzekeringsartsen te verbeteren lijkt onderzoek naar interpretatie van informatie (richtlijnen en protocollen) meer zin te hebben dan verder onderzoek naar de gegevensverzameling (FCE, interview en vragenlijsten).