

Onderzoek naar interdoktervariatie bij beoordelingen van verzekeringsartsen

Ton Schellart, Jan Besseling, Romy Steenbeek, Herman Kroneman

In Nederland is de interdoktervariatie bij verzekeringsartsen, c.q. de betrouwbaarheid van de toepassing van de functionele mogelijkhedenlijst (FML)¹ bij arbeidsongeschiktheidsbeoordelingen het onderwerp geweest van een aantal studies.²⁻⁶ Deze studies hebben plaatsgevonden in een 'laboratorium'-situatie of in een kleine setting. Er zijn bij ons geen grootschalige studies bekend naar interdoktervariatie bij arbeidsongeschiktheidsbeoordelingen.

In 2004 heeft de Raad voor Gezondheidsonderzoek het belang van onderzoek naar interdoktervariatie bij verzekeringsartsen onderstreept.⁷ Het centrale thema van ons onderzoek sluit hierbij aan. We zijn hierbij vooral geïnteresseerd in de verzekeringsartsen, hun werkomgeving (kantoor) en de cliënten als bronnen van variatie. Zo'n studie is niet alleen nationaal, maar ook internationaal van belang, omdat in veel landen arbeidsongeschiktheidsbeoordeling plaatsvindt.⁸⁻¹⁰

ONDERZOEKSVRAGEN

In een eerdere studie ontwikkelden we een conceptueel kader conform het 'Attitude-Social norm-self-Efficacy'-model (ASE-model) voor de variabelen die een belangrijke bron kunnen vormen van interdoktervariatie. We hebben die

SAMENVATTING

Doel is inzicht te krijgen in de mate en mogelijk daarmee samenhangende factoren van systematische verschillen tussen beoordelingen door verzekeringsartsen in de praktijk.

Variabelen conform het 'Attitude - Sociale Norm - Eigen Effectiviteit'-model uit een vragenlijstonderzoek onder verzekeringsartsen, zijn gekoppeld aan uitkomsten van hun beoordelingen van cliënten. Multilevelanalyses zijn uitgevoerd op 196 verzekeringsartsen en 83.755 beoordelingen voor WAO, Herbo of WIA in de periode 2003-2008. 'Intraclass'-correlatiecoëfficiënten zijn berekend. Kenmerken, meningen en zelfgerapporteerd beoordelingsgedrag van verzekeringsartsen hangen in zeer geringe mate samen met de uitkomst van de arbeidsongeschiktheidsbeoordeling. Er blijken wel kleine tot matige systematische verschillen, met name voor de mate van urenbeperkingen en de ernst van beperkingen. Onze bevindingen leveren geen indicaties op van andere oorzaken voor deze verschillen. Zij kunnen te maken hebben met werkrouines of opvattingen over de 'normale' werklust van een werknemer. Onderzoek moet uitwijzen of protocollering deze verschillen vermindert.

variabelen op hun betrouwbaarheid onderzocht¹¹ en gevalideerd in een structuurmodel.¹² Hierover is gerapporteerd in het eerste onderzoeksartikel in dit nummer van TBV. In dit artikel willen we vier vragen beantwoorden:

1. Wat is de variatie in relevante uitkomstmaten van arbeidsongeschiktheidsbeoordelingen op verzekeringsartsniveau (VA-niveau) en op (regionaal) kantoonniveau?
2. Hoe sterk is de relatie tussen de achtergrondvariabelen van de cliënt en de uitkomstmaten?
3. Welke achtergrondvariabelen, meningen en overwegingen van de verzekeringsartsen zelf hangen samen met de uitkomstmaten van arbeidsongeschiktheidsbeoordelingen?
4. Hoe sterk is deze samenhang van deze uitkomstmaten met het VA-niveau, respectievelijk het kantoonniveau?

METHODE

Opzet van de studie en procedure

We verzamelden informatie over sociale verzekeringsartsen en hun cliënten uit twee bronnen: een eenmalige enquête in maart 2008 en de data uit het Claim Beoordelings- en Borgingssysteem

Dr. A.J.M. Schellart is sociaal wetenschapper en senior onderzoeker bij de afdeling Sociale Geneeskunde, EMGO Instituut voor onderzoek naar gezondheid en zorg, VU medisch centrum en het Kenniscentrum Verzekeringsgeneeskunde van AMC-UMCG-UWV-VUmc (KCVG).

Drs. J.J.M. Besseling is sociaal wetenschapper en was ten tijde van het onderzoek senior onderzoeker bij TNO.

Dr. R. Steenbeek is onderzoeker bij TNO.

Dr. H. Kroneman, MHCM, is verzekeringsarts, medisch adviseur UWV.

CORRESPONDENTIEADRES

E-mail: ton.schellart@vumc.nl

INTERDOKTER-
VARIATIE,
VERZEKERINGS-
ARTSEN, ARBEIDS-
ONGESCHIKTHEIDS-
BEOORDELINGEN,
FUNCTIONELE
MOGELIJKHEDEN
EN BEPERKINGEN,
MULTILEVELANALYSE,
INTRACLASS-
CORRELATIE-
COËFFICIËNTEN

SUMMARY

A study of the assessment behaviour of insurance physicians

The aim of this study is to determine the extent and the associated factors of systematic variations in the disability assessments of long-term sick-listed employees by Dutch insurance physicians (IP's). Variables from a questionnaire survey among IPs, based on the 'Attitude - Social norm - self-Efficacy' (ASE) model, were linked to the outcomes of the disability assessments. We performed multilevel analyses on the data provided by 196 IPs and 83,755 assessments in the period 2003-2008 and we calculated Intra Class Correlations (ICCs).

The characteristics, opinions and self-reported assessment behaviour of IPs showed very little association with the outcomes of their disability assessments. Our study uncovered small to moderate systematic variations, comparable to those in other medical professions, especially for working hour restrictions and the severity of the functional incapacity. We found no indications of other reasons for these variations. They may be related to work routines or to differing views on the 'normal' employee's workload. Further research is needed to determine whether guidelines could reduce these systematic variations.

Note: for an English publication on this subject see: Schellart et al.¹³

(CBBS)¹ van UWV met de beoordelingsresultaten en achtergrondvariabelen van cliënten. In augustus 2008 zijn de gegevens uit het CBBS en de resultaten van de vragenlijst gekoppeld en geanonimiseerd. Een uitvoeriger beschrijving van opzet, kenmerken van variabelen en analyse-methode is elders gepubliceerd.¹³

Populatie

De doelgroep bestond uit verzekeringsartsen in actieve dienst van UWV op mei 2008 die beoordelingen deden voor WAO, Herbo en WIA. De beoordelingen van de cliënten moesten liggen in de periode van juli 2003 tot maart 2005 voor de WAO, van november 2004 tot december 2006 voor de Herbo, en van januari 2006 tot april 2008 voor de WIA. Van de 231 verzekeringsartsen die

reageerden op de enquête gaven er 200 toestemming om hun antwoorden te koppelen aan de CBBS-gegevens. Uiteindelijk werden de gegevens van 196 verzekeringsartsen werkend op 27 kantoren met 83.755 cliënten in de analyse betrokken.

Uitkomstmaten

Zes uitkomstmaten zijn in onze studie opgenomen:

1. GDBM: geen duurzame beschikbare mogelijkheden (nee/ja);
2. TOM: het totaal van onbenutbare mogelijkheden, dat wil zeggen GDBM maar nu uitgebreid met het niet dan wel hebben van urenbeperkingen van 10 uur of meer (nee/ja);
3. MUB: de mate van urenbeperkingen (score 1-4, + is meer urenbeperkingen);
4. FML's: de beperkingenscore op de FML (score 0-6, + is meer beperkingen): de som van de drie categorische variabelen (score 0-2), die de ernst van de beperkingen op drie functionele gebieden weerspiegelt (mentaal, fysiek en autonomie);
5. AO1: 80-100% arbeidsongeschiktheid (nee/ja);
6. AO2: 35-100% arbeidsongeschiktheid (nee/ja).

Achtergrondvariabelen cliënten

De achtergrondvariabelen die voor alle cliënten beschikbaar waren, zijn: geslacht, leeftijd, sector van beroep, wettelijk regime, diagnosegroep, objectiveerbaarheid van de klachten. Deze laatste variabele is opgesteld op basis van inzichten van (regio)stafverzekeringsartsen in klachten, die naar hun mening moeilijk te objectiveren zijn in de praktijk. Daarnaast zijn beschikbaar voor de uitkomstmaten FML's, AO1 en AO2: maatmanloon, maatmanuren en opleidingsniveau; en voor AO1 en AO2: ernst van de beperkingen respectievelijk mentaal, fysiek en op gebied van auto-noom functioneren.

Achtergrondvariabelen en ASE-variabelen van verzekeringsartsen uit vragenlijst

De achtergrondvariabelen zijn: geslacht: man (referentie), vrouw; leeftijd (continue variabele); jaren werken als verzekeringsarts (continue variabele); curatief arts zijn (geweest): neen (referentie), ja; opleiding tot verzekeringsarts gevolgd: nee (referentie), ja; aantal werkuren per week: <25 uur (referentie), 25-31 uur, >32 uur; wettelijk regime waarvoor de meeste cliënten worden beoordeeld, respectievelijk WIA, WAO en Wajong: neen (referentie), ja; gemiddeld aantal beoordelingen per week (continue variabele). Verder zijn in de analyse de variabelen opgeno-

I Gegevens van 196 verzekeringsartsen en 83.755 beoordelingen (WAO/Herbo/WIA) uit de jaren 2003-2008 zijn gekoppeld.

I Individuele meningen en kenmerken van verzekeringsartsen spelen vrijwel geen rol bij arbeidsongeschiktheidsbeoordelingen.

I Er zijn kleine tot matige systematische verschillen, vergelijkbaar met andere medische beroepsgroepen, met name bij urenbeperking en ernst van de beperkingen.

I Mogelijk kan protocollering de systematische verschillen verkleinen.

men die op basis van het ASE-model zijn geconstrueerd (zie deel 1 van dit artikel): Attitude, Sociale Norm, Eigen Effectiviteit, Barrières, Kennis en Informatie, Intentie, Gedrag-proces, Gedrag-beoordeling.

Analyse

De representativiteit van de verzekeringsartsen in de onderzoeksgroep is in het vorige artikel in dit nummer van TBV beschreven. We analyseerden aanvullend het verschil tussen de verzekeringsartsen van wie de vragenlijst wel of niet kon worden gekoppeld met CBBS. We vergeleken de gegevens over de cliënten in onze analyse met de gegevens van alle cliënten in CBBS over dezelfde beoordelingperiodes voor dezelfde drie wettelijke regimes.

Met behulp van kruistabellen analyseerden we de uitkomstmaten univariaat tussen verzekeringsartsen en kantoren met SPSS 15.0. We berekenden het (rekenkundige) gemiddelde, de standaarddeviatie, het minimum en maximum, en de variatiecoëfficiënt (standaarddeviatie gedeeld door het gemiddelde) tussen verzekeringsartsen respectievelijk tussen kantoren met Excel 2003.

We gebruikten multilevelanalyse met MLwin versie 2.02,^{14,15} voor de zes uitkomstmaten: gewone regressie voor MUB, logistische regressie voor GDBM, TOM, AO1 en AO2, en Poissonregressie voor FML's. Met multilevelanalyse kunnen naast invloeden op het laagste niveau (hier: het cliënt-niveau) ook de invloeden van hogere niveaus (hier: het VA-niveau en het kantoor-niveau) op de uitkomstmaat worden geschat. Deze invloeden op hogere niveaus wijzen op systematische verschillen in een uitkomstmaat tussen, in ons geval, verzekeringsartsen, respectievelijk kantoren.

Omdat niet van alle cliënten alle achtergrondvariabelen beschikbaar waren voor alle uitkomstmaten, hebben we deze uitkomstmaten geanalyseerd met verschillende sets van de achtergrondvariabelen van cliënten. De regressieresultaten hiervan worden kortheidshalve niet gepresenteerd, omdat ze niet de kern vormen van ons onderzoek.

Intraclasscorrelatiecoëfficiënten

Om de samenhangen van de uitkomstmaten te bepalen op VA-niveau, respectievelijk op kantoor-niveau, berekenden we 'intraclasscorrelatiecoëfficiënten' (ICC's). Die worden in onderzoek vaak gebruikt om de mate van systematische variatie te bepalen tussen elementen op een hoger niveau.¹⁶⁻¹⁸ Een ICC meer richting 0 bete-

kent dat er weinig systematisch verschil is op VA-niveau respectievelijk kantoor-niveau, en meer richting 1 betekent dat er veel verschil is. Meestal, in 'real life' cross-sectionele studies onder artsen, is de ICC niet hoger dan 0,20.¹⁵ Op basis van een ICC-database op gebied van gezondheidszorg in het Verenigd Koninkrijk,^{19,20} kwalificeren wij een ICC ≤ 0,10 als laag, 0,10 < ICC ≤ 0,20 als matig, en een ICC > 0,20 als hoog.

De bijdrage van de achtergrondvariabelen van de cliënten aan het model berekenden we als het verschil tussen de totale verklaarde variantie enerzijds en de som van de ICC's voor VA-niveau en kantoor-niveau anderzijds.

RESULTATEN

Representativiteit

De verzekeringsartsen in onze analyse zijn representatief op de onderzochte aspecten (leeftijd, geslacht, aantal werkuren) voor de totale populatie bij UWV. De verzekeringsartsen die niet gekoppeld werden, verschilden op een aantal aspecten van hen die wel gekoppeld konden worden. De groep die gekoppeld kon worden, ervaart minder Barrières ($p=0,003$), heeft een hogere Eigen Effectiviteit ($p=0,027$) en scoort hoger op het hebben van voldoende Kennis en Informatie ($p=0,035$).

Tussen de cliënten in onze analyse en alle cliënten in CBBS in dezelfde periode bleken slechts kleine verschillen (plus of min 3%-punt) te bestaan: er waren wat meer cliënten met een WIA-beoordeling, wat minder met een WAO-beoordeling en wat meer cliënten in de leeftijdsgroep 45-54 jaar in onze analysegroep dan in de totale groep.

Univariate resultaten

In tabel 1 zijn de centrale tendentie en de spreiding van percentages op arts-niveau en op kantoor-niveau gegeven. Wij zijn vooral geïnteresseerd in de variatiecoëfficiënten, omdat die de ongecorrigeerde variatie in de uitkomstmaten voor de twee niveaus weergeven. Op VA-niveau waren de hoogste drie variatiecoëfficiënten die voor cliënten met een hogere beperkingscore op de FML (FMLs456), daarna voor cliënten met GDBM en vervolgens voor cliënten met meer dan 10 uur beperkingen (MUB234). Op kantoor-niveau zijn deze drie variatiecoëfficiënten ook het hoogst, maar daar is de variatiecoëfficiënt van MUB234 nu hoger dan die voor GDBM. De variatiecoëfficiënten op het kantoor-niveau zijn vanzelfsprekend lager dan die op VA-niveau, omdat er meer cliënten per kantoor zijn dan per verzekeringsarts.

Tabel 1Centrale tendentie en spreiding van percentages 'ja' van uitkomstmaten voor cliënten, over verzekeringsartsen ($n=196$) en over kantoren ($n=27$)[#]

Uitkomstmaten	GDBM	TOM	MUB234	FMLs123	FMLs456	AO1	AO2
VA-niveau							
gemiddelde	12,6	35,1	22,5	27,3	4,1	42,9	57,5
mediaan	11,7	34,1	20,8	25,9	2,0	42,9	57,1
minimum	2,5	8,6	3,3	4,8	0,5	10,5	21,1
maximum	77,8	83,3	45,6	61,2	19,3	86,1	91,7
standaarddeviatie	7,0	11,9	9,6	9,0	3,0	9,8	9,0
variatiecoëfficiënt	0,56	0,34	0,43	0,33	0,75	0,23	0,16
Kantoonniveau							
gemiddelde	12,1	34,7	22,6	28,1	3,3	43,9	58,3
mediaan	11,2	33,3	21,7	28,3	3,2	42,7	57,7
minimum	8,7	18,2	9,5	20,7	0,9	35,1	51,6
maximum	18,4	49,9	35,4	47,3	8,8	53,0	67,3
standaarddeviatie	2,5	8,3	6,8	5,1	2,0	4,4	4,4
variatiecoëfficiënt	0,21	0,24	0,30	0,18	0,60	0,10	0,08

[#] Op basis van rijpercentages van aantal cliënten over verzekeringsartsen (VA), respectievelijk over kantoren per 'ja'-categorie van uitkomstmaten; $n=83.755$ cliënten voor GDBM, TOM, AO1 en AO2; $n=22.917$ cliënten voor MUB; $n=64.190$ cliënten voor FML's.

Legenda:

GDBM = geen duurzaam beschikbare mogelijkheden (categorie = ja);

TOM = geen duurzaam beschikbare mogelijkheden of urenbeperking van 10 uur of meer (categorie = ja);

MUB234 = urenbeperking van >10 werkuren (categorie = ja);

FMLs123 = functionele mogelijkheden score (score 1 tot en met 3 = ja);

FMLs456 = functionele mogelijkheden score (score 4 tot en met 6 = ja);

AO1 = arbeidsongeschiktheid 80-100% (categorie = ja);

AO2 = arbeidsongeschiktheid 35-100% (categorie = ja).

Resultaten multilevelanalyses

In tabel 2 is te zien dat er slechts een paar samenhangen zijn tussen de diverse uitkomstmaten en de variabelen van verzekeringsartsen. Hierbij is dus gecorrigeerd voor achtergrondvariabelen op cliëntniveau. Bij AO1 is sprake van een zwak positief interactie-effect tussen een hogere leeftijd van de verzekeringsarts en een hogere leeftijd van de cliënt: in die gevallen is dus sprake van wat meer cliënten met 80-100% arbeidsongeschiktheid. Vroeger opgeleid zijn als curatief arts hangt ook samen met meer 80-100% arbeidsongeschiktheid van de cliënten. Gespecialiseerd zijn als verzekeringsarts hangt samen met meer cliënten die GDBM zijn beoordeeld. Een langere carrière als verzekeringsarts hangt samen met cliënten met een hogere score voor beperkingen

volgens de FML. Er zijn geen samenhangen tussen de zes uitkomstmaten en de ASE-variabelen Sociale Norm, Barrières, Intentie en Gedrag-proces; met de andere ASE-variabelen wel. Een positieve Attitude hangt samen met minder cliënten die GDBM, dan wel 80-100% arbeidsongeschikt (AO1) zijn beoordeeld. Meer Kennis en Informatie van de verzekeringsarts hangt samen met meer cliënten die 35-100% dan wel 80-100% arbeidsongeschikt zijn beoordeeld. Gedrag-beoordeling – een strengere, meer de regels volgende verzekeringsarts – hangt samen met meer cliënten die 80-100% arbeidsongeschikt zijn beoordeeld. Bij de uitkomstmaat AO1 zijn de meeste samenhangen met verzekeringsartsvariabelen te vinden, bij de uitkomstmaat mate van urenbeperking (MUB) geen enkele. Er zijn bij de verze-

Tabel 2Coëfficiënten (multilevelanalyse) van variabelen van verzekeringsartsen op uitkomstmaten[#]

Variabelen/categorieën	GDBM	TOM	MUB	FML	AO1	AO2
Soort coëfficiënt	OR	OR	B	OR	OR	OR
Leeftijd_VA (+)					0,819**	
Leeftijd_VA *leeftijd_CL (+)					1,039*	
Curatief gespecialiseerd (ja)					1,150**	
Gespecialiseerd als VA (ja)	1,320**					
Jaren als VA (+)				1,009**		
Attitude (+)	0,930**	0,950*				0,950**
Eigen Effectiviteit (+)					0,988*	
Kennis en Informatie (+)					1,057**	1,057**
Gedrag-beoordeling (+)					1,070**	

[#] Gecorrigeerd voor achtergrondvariabelen van cliënten; $n=83.755$ cliënten voor GDBM, TOM, AO1 en AO2; $n=22.917$ cliënten voor MUB; $n=64.190$ cliënten voor FML's; OR = odds ratio, >1 is positieve (+) relatie, <1 is negatieve (-) relatie; B = lineaire regressiecoëfficiënt; VA = verzekeringsarts; CL = cliënt.

* = $0,05 < p < 0,10$; ** = $0,01 < p < 0,05$.

Legenda: zie legenda bij tabel 1.

keringsartsvariabelen geen invloeden te vinden van het kantoonniveau.

Resultaten ICC's

De ICC's voor de samenhang tussen elk van de drie uitkomstmaten enerzijds en het VA-niveau en het kantoonniveau anderzijds zijn samen met de verklaarde variantie van het gehele model (R^2) en de bijdrage van de cliëntvariabelen in tabel 3 weergegeven. De hoogste ICC's op VA-niveau zijn die voor MUB (17,2%) en FML's (11,8%). Voor de andere uitkomstmaten (GDBM, TOM, AO1 en AO2) zijn de ICC's op VA-niveau beneden de 10%. De bijdrage van achtergrondvariabelen van verzekeringsartsen en de ASE-variabelen aan de ICC's op VA-niveau is heel gering, ongeveer 1%-punt (niet in tabel 3). Op kantoonniveau is alleen de ICC voor TOM nog van enige betekenis (1,6%). Voor de andere uitkomstmaten zijn de ICC's op kantoonniveau (vrijwel) verwaarloosbaar (0,0-0,5%). De R^2 is het hoogst (37,2%) voor AO1, daarna voor AO2 (32,6%). De bijdrage van de achtergrondvariabelen van cliënten hieraan is bijna 31%, respectievelijk 28%: dit komt met name (ruim 20%-punt) door de drie variabelen voor de mate van beperkingen (fysiek, mentaal en autonoom) in het model (niet in tabel 3).

DISCUSSIE

Beantwoording onderzoeksvragen

De vier onderzoeksvragen van ons onderzoek kunnen nu worden beantwoord.

Ten eerste, de variatie in de uitkomstmaatregelen van de arbeidsongeschiktheidsbeoordelingen op VA-niveau is univariaat het hoogst bij de zwaardere beperkingen volgens de FML (FMLs456). Hetzelfde resultaat geldt voor het kantoonniveau. Ten tweede, de sterkte van de samenhang (bijdrage aan de verklaarde variantie) tussen de achtergrondvariabelen van cliënten en de uitkomstmaten is ongeveer 5-6% voor GDBM, MUB en FML's, en ongeveer 10% voor TOM. Voor AO1 en AO2 is dit rond de 30%, met name ten gevolge van de drie variabelen ontleend aan het CBBS voor de mate van beperkingen in die modellen.

Ten derde, achtergrondvariabelen van verzekeringsartsen die geassocieerd zijn met uitkomst-

maten, zijn: leeftijd van de verzekeringsarts in interactie met de leeftijd van de cliënt (zwak), opgeleid zijn als curatief arts, opgeleid zijn als verzekeringsarts en aantal jaren gewerkt als verzekeringsarts. Vier ASE-variabelen hebben een verband met uitkomstmaten: Attitude (negatief met GDBM en AO2, zwak negatief met TOM), Eigen Effectiviteit (zwak negatief met AO1), Kennis en Informatie (positief met AO1 en AO2) en Gedrag-beoordeling (positief met AO1). De bijdrage van de achtergrondvariabelen en de ASE-variabelen van verzekeringsartsen aan de ICC's op VA-niveau is zeer klein, ongeveer 1%-punt.

Tot slot, op basis van de eerder genoemde evaluatiecriteria^{19,20} is de ICC voor MUB, 17%, het hoogst, maar valt nog in de categorie 'matig'; die van FML's, 12%, is ook 'matig'. Voor de overige uitkomstmaten (GDBM, TOM, AO1 en AO2) zijn de ICC's (5-7%) als 'laag' te kwalificeren. De ICC's op het kantoonniveau zijn zeer laag: alleen de ICC voor TOM, 1,6%, is nog van enige betekenis.

Sterke en zwakke punten

Het sterke punt van deze studie is dat de verzekeringsartsen representatief zijn voor alle verzekeringsartsen bij UWV. Bovendien zijn er slechts kleine verschillen tussen de cliënten in onze analyse en het totale aantal cliënten. Een ander sterk punt is dat de uitkomstmaten voor het werk van de verzekeringsartsen bij UWV en voor de Nederlandse samenleving zeer relevant zijn. Een derde sterkte ligt in de koppeling tussen de uitkomstmaten op cliëntniveau en de vragenlijstgegevens van de verzekeringsartsen, in samenhang met de getoetste betrouwbaarheid en de validiteit van de ASE-variabelen.^{11,12}

Zwak punt in onze studie is het cross-sectionele ontwerp van de studie: dit laat niet toe dat we harde conclusies kunnen trekken over causale relaties. Verder weten we niet of de scores voor de ASE-variabelen op basis van de enquêtegegevens stabiel in de tijd zijn en geldig voor de voorgaande jaren. Een ander zwak punt is dat we geen cliëntgegevens hebben over comorbiditeit en de ernst van de aandoeningen. De afwezigheid van dergelijke gegevens is verantwoordelijk voor de

Tabel 3

Verklaarde variantie (R^2) en intraclasscorrelatiecoëfficiënten (ICC's) voor uitkomstmaten (in procenten variantie)[#]

	GDBM	TOM	MUB	FML	AO1	AO2
ICC-VA	5,1%	7,1%	17,2%	11,8%	6,5%	4,7%
ICC-kantoor	0,4%	1,6%	0,0%	0,0%	0,0%	0,2%
R^2 totale model	11,3%	19,1%	21,0%	18,5%	37,2%	32,6%
bijdrage R^2 door cliëntvariabelen	5,8%	10,4%	4,8%	6,7%	30,7%	27,7%

[#] $n=83.755$ cliënten voor GDBM, TOM, AO1 en AO2; $n=22.917$ cliënten voor MUB; $n=64.190$ cliënten voor FML's.

Legenda: zie legenda bij tabel 1.

relatief lage verklaarde variantie van de modellen voor GDBM, TOM, MUB en FML's.

Interpretaties

De samenhangen tussen de ASE-variabelen en de uitkomstmaten kunnen naar onze mening alleen zinvol worden geïnterpreteerd als causale verbanden. Ten eerste, een positievere houding ten opzichte van het werk als verzekeringarts, de kwaliteit ervan en het sociaal zekerheidssysteem, waaronder de WIA, (Attitude) hangt samen met minder beoordelingen GDBM en met meer oordelen <35% AO. Misschien dat deze verzekeringsartsen hun beroep, de kwaliteit van hun werk en de regelgeving – waaronder de bewezen opvatting dat werk goed is voor iemands gezondheid – (nog) meer serieus nemen. Ten tweede, het naleven van regels (Gedrag-beoordeling) hangt samen met meer 80-100% AO. Daarnaast hangt een hogere score op Kennis en Informatie over de medische status van de cliënt en over de inspanningen van de werkgever en de cliënt – dat een positieve relatie heeft met Gedrag-beoordeling in het structuurmodel¹² – ook samen met meer 80-100% AO. Onze interpretatie is dat het hebben van meer informatie (bijvoorbeeld van de bedrijfsarts over de gezondheid of uit het re-integratieverslag) juist de meer arbeidsongeschikte cliënten betreft. Het naleven van regels, in het geval van deze cliënten, kan dan een hogere kans betekenen op een 80-100% AO-beoordeling. Deze interpretatie lijkt in overeenstemming met een bevinding van Spanjer.⁴ Die vindt in zijn studie dat *aanvullende* informatie van de cliënt over participatie en beperkingen in activiteiten tot meer beperkingenscores leidt ten opzichte van de scores op basis van informatie over de medische anamnese alleen.

De bevinding dat ASE-variabelen slechts een zwakke samenhang met de uitkomstmaten blijken te hebben, kan worden verklaard door de aanname dat zelfgerapporteerde attitudes, intenties en gedrag nou eenmaal zwak samenhangen met geregistreerde uitkomstmaten omdat ze verschillend van aard zijn.^{21,22} Naar onze mening is deze uitleg hier niet van toepassing, omdat de achtergrondvariabelen van de verzekeringsartsen ook zwak geassocieerd zijn met deze uitkomstmaten. Wij zijn dan ook geneigd om de zwakke relatie tussen de achtergrondvariabelen en ASE-variabelen en de uitkomstmaten als 'echt' en 'waar' beschouwen, met andere woorden, de individuele kenmerken en meningen van de verzekeringsartsen hebben zeer weinig invloed op de uitkomst van een arbeidsongeschiktheidsbeoordeling in de Nederlandse sociale verzekering.

De ICC's op VA-niveau zijn het hoogst voor de mate van urenbeperking (dus niet óf de verzekeringsarts een urenbeperking geeft) en voor de scores op de FML, maar kunnen nog steeds worden aangemerkt als 'matig'. Spanjer vond in zijn onderzoek^{2-4,6} over de betrouwbaarheid van de FML onder de verzekeringsartsen gemiddelde absolute overeenkomsten van 74 tot 84%. In onze studie vonden we een gemiddelde ICC van 12% bij VA-niveau voor de beperkingenscore volgens de FML, hetgeen wijst op een systematisch verschil in een minderheid van de beoordelingsresultaten volgens de FML. Dit is in overeenstemming met de absolute overeenkomst voor het grootste deel van de beoordelingen in deze studies van Spanjer. De absolute overeenkomst tussen de (verzekerings)artsen over urenbeperking van een cliënt, was in deze studies van Spanjer^{2-4,6} relatief laag. Het feit dat we in onze studie het grootste systematische verschil vinden bij de mate van urenbeperking, komt hiermee ook overeen. Mogelijk speelt hierbij een rol dat richtlijnen en protocollen over *hoe* de mate en ernst van (uren)beperkingen te schatten ontbreken in de praktijk.

De ICC's op het kantoonniveau zijn te verwaarlozen, hetgeen kan betekenen dat het regiomanagement geen invloed heeft op de AO-beoordelingen op een andere manier dan via de centrale richtlijnen vanuit het UWV-hoofdkantoor. Onze resultaten tonen aan dat de schijnbare, univariante verschillen in uitkomstmaten tussen kantoren in belangrijke mate worden verklaard door verschillen op cliënt- en VA-niveau.

De gevonden verschillen op VA-niveau kunnen het gevolg zijn van andere factoren dan de individuele kenmerken en meningen van verzekeringsartsen. Een plausibele verklaring is dat de manier waarop verzekeringsartsen routinematig hun (fysieke en psychologische) onderzoeken uitvoeren van de gezondheidstoestand van cliënten, kan leiden tot verschillende uitkomsten. Een andere plausibele verklaring is dat verzekeringsartsen verschillen in hun oordeel, bewust of onbewust, wat de werklast van een 'normale' werknemer is. Als deze verklaringen geldig zijn, kunnen systematische verschillen worden teruggebracht door goed ontwikkelde en goed geïmplementeerde richtlijnen en protocollen over de wijze van beoordeling van de ernst van beperkingen.^{23,24} De huidige trend in de verzekeringsgeneeskunde in Nederland om meer te werken volgens richtlijnen, protocollen en mediprudentie sluit daarbij aan. Verder onderzoek is nodig of vermindering van de interdoktervariatie bij verzekeringsartsen ook daadwerkelijk plaatsvindt

in de praktijk sinds 2008, na de implementatie bij UWV van de verzekeringsgeneeskundige protocollen.

CONCLUSIE

Kenmerken, meningen en zelfgerapporteerd beoordelingsgedrag van verzekeringsartsen hangen in zeer geringe mate samen met de uitkomsten van de arbeidsongeschiktheidsbeoordeling. Dat gezegd hebbende, zijn er wel systematische verschillen gebleken in ons onderzoek. Die kunnen echter worden gekwalificeerd als 'normaal', dat wil zeggen gelijk aan die bij andere medische beroepsgroepen. Onze bevindingen leveren geen indicatie op van andere oorzaken voor deze verschillen. Zij kunnen te maken hebben met andere werkrouines of met verschillende opvattingen over de 'normale' werklust van een werknemer. Als dit zo is, kunnen deze verschillen – met name die bij de mate van urenbeperking en bij de ernst van de beperking – worden verminderd door goed ontwikkelde en goed geïmplementeerde richtlijnen, protocollen en mediprudentie. Verder onderzoek is nodig om na te gaan of vermindering van deze systematische verschillen daadwerkelijk in de praktijk plaatsvindt.

NOOT

Het onderzoek is uitgevoerd door Romy Steenbeek en Jan Besseling (TNO), Ton Schellart en Han Anema (VUmc) en Henny Mulders en Herman Kroneman (UWV) met subsidie van Stichting Instituut GAK.

Dit artikel is een Nederlandse bewerking en qua uitkomsten uitgebreidere versie van een eerder verschenen Engelstalig artikel (zie literatuurverwijzing 13).

Het onderzoeksproject werd gefinancierd door Stichting Instituut Gak.

Aanvullende financiering kwam van TNO Werk en Gezondheid en van het Kenniscentrum Verzekeringsgeneeskunde AMC-UMCG-UWV-VUmc.

Er zijn door de auteurs geen belangenconflicten gemeld.

LITERATUUR

1. UWV Uitvoering Werknemersverzekeringen. Het Claimbeoordelings- en Borgingssysteem. Een introductie voor belangstellenden. Amsterdam: UWV, 2003; http://cba.uwv.nl/literatuur-claimbeoordeling/adlib_view?id=57771, benaderd op 20 januari 2010.
2. Spanjer J. De inter- en intra-beoordelaarsbetrouwbaarheid van WAO-beoordelingen. Tijdschr Bedrijfs Verzekeringsgeneeskde 2001; 9: 234-241.
3. Spanjer J, Krol B, Groothoff JW. Inter-rater reliability in disability assessment based on a semi-structured interview report. Disabil Rehabil 2008; 30: 1885-1890.
4. Spanjer J, Krol B, Popping R, et al. Disability assessment interview: the role of detailed information on functioning in addition to medical history-taking. Rehabil Med 2009; 41: 267-272.
5. Brouwer S. In disability in chronic low back pain: psychometric properties of ADL- and work-related instruments. Chapters 4 and 5. PhD thesis. University of Groningen, 2004, <http://irs.ub.rug.nl/ppn/267302193>, benaderd op 6 december 2009.
6. Spanjer J, Krol B, Brouwer S, Popping R, Groothoff JW, van der Klink JJ. Reliability and Validity of the Disability Assessment Structured Interview (DASI): a tool for assessing functional limitations in claimants. J Occup Rehabil 2010; 20: 33-40.
7. Raad voor Gezondheidsonderzoek. Advies Onderzoek Verzekeringsgeneeskunde. Den Haag: RGO, 2004.
8. Boer WEL de, Besseling MJ, Willems HHBM. Organisation of disability evaluation in 15 countries. Pratiques et organisation des soins 2007; 38: 205-217.
9. Boer WE de, Bruinvels DJ, Rijkenberg AM, et al. Evidence based guidelines in the evaluation of work disability: an international survey and a comparison of quality of development. BMC Public Health 2009; 9: 349.
10. Boer WE de, Rijkenberg AM, Donceel P. Guidelines for assessment of work disability: an international survey. Gesundheitswesen 2011; 73: e103-e110.
11. Steenbeek R, Schellart AJM, Mulders HPG, et al. The development of instruments to measure the work disability assessment behaviour of insurance physicians. BMC Public Health 2011; 11: 1.
12. Schellart AJM, Steenbeek R, Mulders HPG, et al. Can the disability assessment behaviour of insurance physicians be explained? Applying the ASE-model. BMC Public Health 2011; 11: 576.
13. Schellart AJM, Mulders HPG, Steenbeek R, et al. Inter-doctor variations in the assessment of functional incapacities by insurance physicians. BMC Public Health 2011; 11: 864.
14. Rasbash J, Steele F, Browne W, Prosser B. A User's Guide to MLwiN Version 2.0. Bristol, UK: Centre for Multilevel Modelling, University of Bristol, 2004.
15. Twisk JWR. Applied Multilevel Analysis. New York: Cambridge University Press, 2006.
16. Merlo J. Changing analytical approaches in European epidemiology – a short comment on a recent article. Eur J Epidemiol 2005; 20: 737.
17. Merlo J, Gerdttham UG, Eckerlund I, et al. Hospital level of care and neonatal mortality in low and high-risk deliveries – reassessing the question in Sweden by multilevel analysis. Med Care 2005; 43: 1092-1100.
18. Damman OC, Stubbe JH, Hendriks M, et al. Using multilevel modeling to assess casemix adjusters in consumer experience surveys in health care. Med Care 2009; 47: 496-503.
19. University of Aberdeen: Empirical estimates of ICCs from changing professional practice study. [Final version.] [www.abdn.ac.uk/hsrc/documents/iccs-web.xls], benaderd op 26 januari 2010.
20. Campbell M, Grimshaw J, Steen N. Sample size calculations for cluster randomised trials. Changing professional practice in Europe group (EU BIOMED II Concerted Action). J Health Serv Res Policy 2000; 5: 12-16.
21. Godin G, Bélanger-Gravel A, Eccles M, Grimshaw J. Healthcare professionals' intentions and behaviours: a systematic review of studies based on social cognitive theories. Implement Sci 2008; 3: 36.
22. Rebergen D, Hoenen J, Heinemans A, et al. Adherence to mental health guidelines by Dutch occupational physicians. Occup Med (London) 2006; 5: 461-468.
23. Zwerver F, Schellart AJM, Anema JR, et al. Intervention mapping for development of an implementation strategy for the insurance medicine guideline on depression. BMC Public Health 2011; 11: 9.
24. Schellart AJM, Zwerver F, Anema JR, Beek AJ van der. Training in toepassing van het protocol Depressieve stoornis en veranderingen in de uitkomsten van de beoordeling van de belastbaarheid. Tijdschr Bedrijfs Verzekeringsgeneeskde 2012; 8: 357-364.