

De betrouwbaarheid van risicotaxatie in de pro Justitia rapportage

Een onderzoek met behulp van de HKT-30

W.J. CANTON, T.S. VAN DER VEER, P.J.A. VAN PANHUIS, R. VERHEUL, W. VAN DEN BRINK

ACHTERGROND Het inschatten van de kans op herhaling van gewelddadig gedrag neemt een belangrijke plaats in in de forensische psychiatrie. Met een gestandaardiseerd risicotaxatie-instrument voor de pro Justitia-rapportages zou deze kans betrouwbaar ingeschat kunnen worden.

DOEL Onderzoek naar de interbeoordelaarsovereenstemming van risicotaxatie met behulp van een gestandaardiseerd risicotaxatie-instrument.

METHODE De HKT-30 (historisch-klinisch-toekomst), een gestandaardiseerd risicotaxatie-instrument, werd door 3 ervaren rapporteurs onafhankelijk van elkaar ingevuld aan de hand van 74 pro Justitia rapportages.

RESULTATEN De betrouwbaarheid van de historische items en de toekomstitems was in het algemeen goed tot zeer goed, van de klinische items redelijk tot goed.

CONCLUSIE Met behulp van een gestandaardiseerd instrument (HKT-30) kan een redelijk betrouwbare risicotaxatie plaatsvinden. Belangrijke aanwijzingen voor mogelijke verbeteringen in het gebruikte risicotaxatie-instrument worden beschreven.

[TIJDSCHRIFT VOOR PSYCHIATRIE 46(2004)8, 537-542]

TREFWOORDEN betrouwbaarheid, forensische psychiatrie, risicotaxatie

Het inschatten van de kans op herhaling van gewelddadig gedrag neemt een belangrijke plaats in in de forensische psychiatrie en speelt een rol gedurende het hele proces van de rechtspleging.

In de pro Justitia rapportage (PJ-rapportage) wordt aan de gedragskundige een mening gevraagd over het recidiverisico en het risicomangement. De behandeling en/of begeleiding, die door de rechtbank kunnen worden opgelegd, zijn gericht op het verminderen van het recidiverisico. Ook bij de vraagstelling of een gedwongen maatregel (bv. terbeschikkingstelling, tbs) beëindigd kan worden en bij het al dan niet toekennen van

proefverlof speelt risicotaxatie een grote rol.

Risicotaxatie vond tot voor kort uitsluitend plaats door middel van een 'klinische inschatting'. Om tot een meer betrouwbare inschatting te komen met een betere voorspellende waarde is tussen september 1999 en juni 2000 gewerkt aan het ontwikkelen van een risicotaxatie-instrument (de HKT-30), dat gebaseerd is op de reeds bestaande instrumenten en waarin resultaten van recidiveonderzoek na ontslag of proefverlof uit de tbs-klinieken zijn meegenomen. In dit artikel worden gegevens gepresenteerd over de betrouwbaarheid van de HKT-30.

METHODE

Brongegevens

In het kader van het onderzoek is door 3 ervaren forensisch psychiaters onafhankelijk van elkaar bij 74 PJ-rapportages, afkomstig uit de arrondissementen Den Bosch, Arnhem en Zutphen, de HKT-30 ingevuld. Het gaat daarbij zowel om enkelvoudige psychiatrische of psychologische rapportages als om dubbelrapportages, dat wil zeggen rapportages waaraan zowel een psycholoog als een psychiater een bijdrage heeft geleverd. De rapportages maken deel uit van een bestand van 755 rapporten, dat in het kader van een groter onderzoeksproject (gerandomiseerd) geselecteerd is.

Instrument

Het onderzochte risicotaxatie-instrument, de HKT-30, is opgebouwd uit een H(istorisch) en een K(linisch) cluster, die beide bestaande uit 12 items, en uit een T(oekomst) cluster, dat bestaat uit 6 items. De verschillende items worden elders in dit tijdschriftnummer uitgebreid beschreven in een artikel over de validiteit van risicotaxatie (Canton e.a. 2004).

Aan ieder item van de HKT-30 is aan de hand van de handleiding een score toegekend met een cijfer tussen 0 en 6. Dit gebeurde door 3 rapporteurs, onafhankelijk van elkaar. Voorafgaand hieraan was door de rapporteurs de handleiding bestudeerd en waren onderling een aantal afspraken gemaakt.

De meeste items van de HKT-30 waren op basis van de bestaande PJ-rapportages in te vullen.

Het aantal ontbrekende waarden was beperkt. Bij 5 van de 74 rapporten (7%) ontbrak op 1 of meer clusters meer dan een derde deel van de items, waardoor deze rapportages niet bij het betrouwbaarheidsonderzoek betrokken konden worden. De resultaten hebben dan ook betrekking op 69 rapporten.

Statistische analyse

De gemiddelden en de spreiding zijn bepaald van ieder item, van de totaalscores per cluster en van de drie clusters gezamenlijk. Om een (voor toeval gecorrigeerde) overeenstemmingscoëfficiënt te bepalen tussen de 3 onderzoekers gelijktijdig (simultaan), werd met behulp van het statistische programma Agree voor ieder item de kwadratisch gewogen (*square weighted*, *sw*) kappa bepaald. Om de overeenstemming op clusterniveau te bepalen werd een intraklassecorrelatiecoëfficiënt (ICC) bepaald – een voor toeval gecorrigeerde maat voor de overeenkomst tussen de 3 onderzoekers.

Bij de interpretatie van de diverse kappawaarden en van de ICC wordt een waarde boven de 0,80 als een zeer goede interbeoordelaarsovereenstemming beschouwd, een waarde tussen de 0,60 en 0,80 als goed, een waarde tussen de 0,40 en 0,60 nog als redelijk en een waarde onder de 0,40 als matig of onvoldoende (Landis & Koch 1977).

Onderzoekspopulatie

Onder de 74 gerapporteerden waren 71 mannen en 3 vrouwen. Zevenendertig procent was jon-

TABEL 1 De gemiddelde door de 3 beoordelaars gegeven scores op de HKT-30 (7-puntsschaal) en de 3 clusters daarvan, aan de hand van bestaande rapporten (n=69)

Beoordelaar	Historisch cluster			Klinisch cluster			Toekomstcluster			HKT totaal		
	1	2	3	1	2	3	1	2	3	1	2	3
Gemiddelde totaal	27	28	29	26	30	30	21	17	20	78	75	74
Gemiddelde per item	2,25	2,33	2,42	2,17	2,5	2,5	3,50	2,83	3,33	2,60	2,50	2,47
Spreiding	1-56	1-56	13-49	2-62	8-47	4-52	1-35	4-28	5-30	4-140	20-131	19-129

ger dan 26 jaar. De delicten die aanleiding vormden tot de rapportage, waren als volgt verdeeld: 25% van de respondenten werd gerapporteerd voor een levensdelict, 25% voor een zedendelict en 25% voor geweldsdelicten. Voorts pleegde 16% een vermogensdelict zonder geweld, 5% brandstichting en was er bij 4% sprake van overige delicten. Volgens de rapporteurs was er bij 23% van de personen sprake van verslavingsproblematiek, bij 43% van een psychiatrisch toestandsbeeld en bij 88% van een persoonlijkheidsstoornis.

RESULTATEN

HKT-30 scores volgens de verschillende beoordelaars

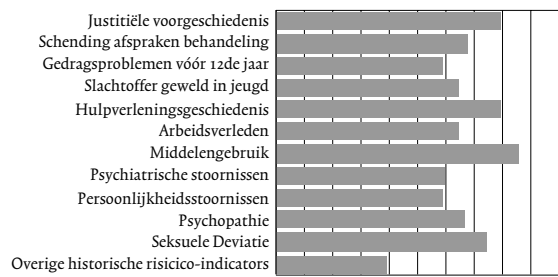
Tabel 1 toont de totaalscores per cluster, de gemiddelde itemscores per cluster en de spreiding van de scores per cluster voor de verschillende beoordelaars.

Uit tabel 1 blijkt dat voor de verschillende onderzoekers de gemiddelde somscores dicht bij elkaar lagen en dus ook de gemiddelde itemwaarde per cluster. De toekomstitems kregen gemiddeld een hogere score dan de historische en de klinische items, wat betekent dat de te verwachten sociaal-economische omstandigheden in het algemeen als zorgelijk werden ingeschat. De spreiding was groot en voor ieder item gold dat alle scoringsmogelijkheden van 0 tot en met 6 gebruikt werden.

Overeenstemming op itemniveau¹

Figuur 1 toont de overeenstemming tussen de beoordelaars op de historische items. Met uitzondering van H12 (‘overige’), varieerde de betrouwbaarheid tussen redelijk en zeer goed: kwadratisch gewogen kappa 0,38 tot 0,86 (mediaan 0,77).

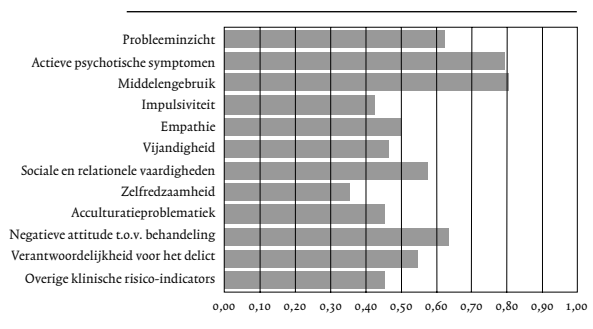
FIGUUR 1 Interbeoordelaarovereenstemming* van de historische items van de HKT-30, zoals toegekend door 3 beoordelaars op basis van 69 bestaande rapporten



*Kwadratisch gewogen kappa

Figuur 2 toont de interbeoordelaarovereenstemming van de klinische items. De betrouwbaarheid van deze items was in het algemeen wat lager dan van de historische items en er waren grote verschillen tussen de items: de kwadratisch gewogen kappa varieerde van 0,36 tot 0,81 (mediaan 0,53).

FIGUUR 2 Interbeoordelaarovereenstemming* van de klinische items van de HKT-30, zoals toegekend door 3 beoordelaars op basis van 69 bestaande rapporten

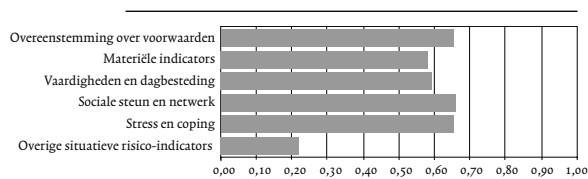


*Kwadratisch gewogen kappa

Figuur 3 toont een overzicht van de interbeoordelaarovereenstemming van de toekomstitems. De betrouwbaarheid van deze items was, met uit-

zondering van het item 'overig' (T6), redelijk tot goed te noemen: de kwadratisch gewogen kappa varieerde van 0,22 tot 0,66 (mediaan 0,62).

FIGUUR 3 Interbeoordelaarsovereenstemming* van de historische items van de HKT-30, zoals toegekend door 3 beoordelaars op basis van 69 bestaande rapporten



* Kwadratisch gewogen kappa

Overeenstemming op clusterniveau De overeenstemming op clusterniveau (ICC) was voor de historische items 0,85, voor de klinische items 0,81 en voor de toekomstitems 0,78. Voor de totaalscore was de ICC 0,85. De betrouwbaarheid van deze somscores is dus goed tot zeer goed.

DISCUSSIE

Onderzoek heeft plaatsgevonden naar de betrouwbaarheid van risicotaxatie bij de pro Justitia rapportages met behulp van de HKT-30. De betrouwbaarheid van de verschillende items was sterk wisselend. Allereerst werd duidelijk dat de items 'overig' (H12, K12, T6) matig tot slecht scoren, wat niet vreemd is vanwege het open karakter ervan. De historische items en de toekomstitems waren overwegend met een goede betrouwbaarheid te scoren. De grootste mate van onbetrouwbaarheid bleek bij de klinische items te zitten. Van slechts 4 items was de betrouwbaarheid goed te noemen, 7 items scoorden redelijk en 1 matig. De bevindingen komen in grote lijnen overeen met de resultaten betreffende de interbeoordelaarsovereenstemming van de HKT-30, zoals die beschreven staan in het rapport van de Werkgroep Pilotstudy Risicotaxatie in de Forensische Psychiatrie (2002). Bij laatstgenoemd onderzoek is echter gebruikgemaakt van een gemodificeerde versie van de HKT-30, waarin de items 'overig' zijn vervallen en er sprake is van 11 historische items, 13 klinische

items en 6 toekomstitems.

De oorzaken van de geringe betrouwbaarheid van een aantal items moeten gezocht worden op verschillende gebieden. Een belangrijke bron van onbetrouwbaarheid is gelegen in het PJ-rapport als informatiebron voor het invullen van de HKT-30. Zelden was informatie over een item geheel niet terug te vinden in de rapportage. Vaker kwam het voor dat de informatie onduidelijk of onvolledig was. Dat liet veel ruimte voor interpretatieverschillen tussen de beoordelaars en leidde dus tot een verminderde betrouwbaarheid. Een volgende bron van verminderde betrouwbaarheid is gelegen in de toelichting op de scores in de HKT-30, die bij een aantal items te veel ruimte laat voor interpretatieverschillen. Ook het feit dat er een zekere overlap bestaat tussen de inhoud van een aantal items, is een bron van verminderde betrouwbaarheid. Dezelfde informatie kan (door verschillende beoordelaars) bij verschillende items worden weergegeven.

De onderzoekers zijn niet getraind in het werken met de HKT-30. Ze hebben voor het onderzoek alleen uitgebreid de handleiding bestudeerd en vooraf onderling overlegd. Door een goede training in het werken met een onderzoeksinstrument zal de betrouwbaarheid verder kunnen toenemen.

De laatste bron van verminderde betrouwbaarheid is gelegen in het feit dat de toelichting op een aantal items is toegespitst op mensen die al in behandeling of veroordeeld zijn. Deze items zijn moeilijk te interpreteren aan de hand van PJ-rapporten. Ten slotte geldt voor een aantal items dat er gelijktijdig verschillende bronnen van onbetrouwbaarheid aanwezig zijn.

De betrouwbaarheid van de clusterscores en de totaalscore waren goed tot zeer goed. Dit roept direct de vraag op of deze somscores ook als zodanig gebruikt kunnen worden als een kwantitatieve maat voor het bestaande recidiverisico. Toekomstig onderzoek zal duidelijk moeten maken of dat het geval is. Vooralsnog moet een dergelijk gebruik van de HKT-30 nog worden ontraden

CONCLUSIE

Op basis van bestaande PJ-rapportages kan met behulp van een gestandaardiseerd instrument (HKT-30) een redelijk betrouwbare risicotaxatie plaatsvinden. Belangrijke aanwijzingen voor mogelijke verbeteringen in het gebruikte risicotaxatie-instrument zijn in bovenstaande beschreven.

NOOT

1. Een uitgebreide bespreking op itemniveau is te verkrijgen bij de eerste auteur.

LITERATUUR

- Canton, W.J., van der Veer, T.S., van Panhuis, P.J.A., e.a. (2004). De voorspellende waarde van risicotaxatie bij de rapportage pro Justitia. Onderzoek naar de HKT-30 en de klinische inschatting. *Tijdschrift voor Psychiatrie*, 46, 525-535.
- Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.
- Werkgroep Pilotstudy Risicotaxatie Forensische Psychiatrie. (2002). *Bevindingen van een landelijke pilotstudy naar de HKT-30*. Den Haag: Ministerie van Justitie.

AUTEURS

W.J. CANTON is als psychiater verbonden aan de Forensisch Psychiatrische Dienst te Den Bosch.

T.S. VAN DER VEER is als psychiater en hoofd van dienst verbonden aan de Forensisch Psychiatrische Dienst te Arnhem.

P.J.A. VAN PANHUIS is als psychiater en hofressortscoördinator verbonden aan de Forensisch Psychiatrische Dienst te Den Bosch.

R. VERHEUL is als Manager Academisering verbonden aan het Psychotherapeutisch Centrum de Viersprong te Halsteren.

W. VAN DEN BRINK is als hoogleraar verslavingszorg verbonden aan het Academisch Medisch Centrum, Universiteit van Amsterdam, Afdeling Psychiatrie.

Correspondentieadres: W.J. Canton, FPD Den Bosch, Postbus 26, 5201 AA Den Bosch.

Geen strijdige belangen meegegeeld.

Het artikel werd voor publicatie geaccepteerd op 13-4-2004.

SUMMARY

The reliability of risk assessment in 'pro Justitia' reports. An investigation on the basis of the hkt-30 – W.J. Canton, T.S. van der Veer, P.J.A. van Panhuis, R. Verheul, W. van den Brink –

BACKGROUND Estimating the chance that offenders will relapse into violent behaviour is an important concern in forensic psychiatry. It should be possible to reliably estimate the chance of relapse in 'pro Justitia' reports by awarding scores on the basis of a standardised instrument for risk assessment.

AIM To investigate the inter-rater reliability of risk assessment with the help of a standardised instrument for risk assessment.

METHOD Three experienced reporters reviewed the same 74 'pro Justitia' reports independently and awarded their own scores for the information contained therein. These scores were entered on the HKT-30 (historical-clinical-future), which is a standardised instrument for risk assessment.

RESULTS The reliability of the historical items and the future items was generally good to very good; the reliability of the clinical items was reasonable to good.

CONCLUSION A standardised instrument such as the HKT-30 can be used to produce a reasonably reliable risk assessment. The discussion gives useful advice about improvements that can be made in this risk assessment instrument.

[TIJDSCHRIFT VOOR PSYCHIATRIE 46(2004)8, 537-542]

KEY WORDS forensic psychiatry, reliability, risk assessment