

# Psychometrische stand van zaken van risicotaxatie-instrumenten voor volwassenen in Nederland

G.T. BLOK, E. DE BEURS, A.G.S. DE RANITZ, T. RINNE

**ACHTERGROND** Hoewel er in het forensisch-psychiatrische veld steeds meer gebruikgemaakt wordt van risicotaxatie-instrumenten, is de discussie over de waarde en bruikbaarheid van deze instrumenten nog volop gaande.

**DOEL** Een bijdrage aan deze discussie leveren door een overzicht te geven van de psychometrische kwaliteiten van de instrumenten die het meest worden gebruikt in Nederland voor risicotaxatie bij volwassenen.

**METHODE** Gegevens verzamelen over de betrouwbaarheid en de validiteit van de Historical, Clinical, and Risk Management (HCR-20), de Historische, Klinische en Toekomstige Risico-indicatoren-30 (HKT-30), de Sexual Violence Risk-20 (SVR-20), de Forensisch Psychiatrische Profielen (FP-40) en de Psychopathy Checklist-Revised (PCL-R) en deze vergelijken met gegevens over het ongestructureerde klinische oordeel.

**RESULTATEN** De interbeoordelaarsbetrouwbaarheid bleek over het algemeen voldoende, maar de interne consistentie was vaak onvoldoende. De predictieve validiteit was met uitzondering van enkele studies over het algemeen redelijk.

**CONCLUSIE** De huidige stand van zaken dwingt tot voorzichtigheid wanneer men voor het inschatten van de kans op recidive risicotaxatie-instrumenten of enkel het ongestructureerde klinische oordeel gebruikt. Wellicht is het niet mogelijk om recidive nauwkeuriger te voorspellen. In afwachting van nieuwe ontwikkelingen is het raadzaam zoveel mogelijk gegevens over een onderzochte met elkaar te combineren en de uitkomsten daarvan te toetsen aan het oordeel van andere professionals.

[TIJDSCHRIFT VOOR PSYCHIATRIE 52(2010)5, 331-341]

**TREFWOORDEN** betrouwbaarheid, forensische psychiatrie, risicotaxatie, validiteit

Wanneer iemand een (ernstig) delict heeft gepleegd, willen we graag zekerheid over de kans dat diegene in de toekomst opnieuw een delict zal plegen. Ook van psychiaters wordt steeds vaker een inschatting van potentieel gevaarlijk gedrag gevraagd. Steeds meer psychiaters zijn werkzaam in het forensisch-psychiatrische veld en in de reguliere psychiatrie worden we vaker geconfron-

teerd met patiënten die (soms ernstig) agressief gedrag laten zien.

Tot een jaar of 10 geleden werden dergelijke voorspellingen uitsluitend gedaan op basis van het klinische oordeel, maar vooral door onderzoek in het buitenland is duidelijk geworden dat dit onvoldoende betrouwbaar en valide is (Quinsey e.a. 2006a, b). Het menselijk beoordelingsvermo-

gen is te beperkt en we kunnen bij de beoordeling van een situatie of een persoon niet meer dan een beperkt aantal variabelen tegelijk in ogenschouw nemen.

Vanwege deze beperkingen ontstonden de eerste risicotaxatie-instrumenten. Deze waren gebaseerd op biografische of historische factoren waarvan in onderzoek was aangetoond dat deze een statistisch significant voorspellend karakter hadden wat betreft het gevaar op recidive. Deze instrumenten worden dan ook wel actuariële instrumenten genoemd, omdat ze historische onveranderlijke factoren betreffen, zoals een criminele jeugd en eerdere delicten.

Recidiverisico is echter niet een statisch, maar een potentieel veranderlijk gegeven. Voorbeelden van veranderbare of dynamische risicofactoren zijn 'gebrek aan zelfinzicht', 'impulsiviteit' of 'geringe bereidheid om aan een behandeling mee te werken'. Deze factoren kunnen aangrijpingspunten zijn voor behandeling en zijn verdisconteerd in nieuwere risicotaxatie-instrumenten, zoals de *Historical, Clinical, and Risk Management (HCR-20)* in Noord-Amerika en de *Historische, Klinische en Toekomstige Risico-indicatoren-30 (HKT-30)* in Nederland. Bij deze instrumenten dient bovendien een forensisch expert (psychiater of psycholoog) afsluitend een klinisch oordeel te geven over het door hem of haar ingeschatte risico van herhaling. Dit oordeel is niet alleen afhankelijk van de hoogte van de eindscore, maar ook van de klinische weging van verschillende combinaties van items door de beoordelaar. Dit 'gestructureerd klinisch oordeel' deelt men meestal onder in 3 categorieën: hoog, matig en laag risico van recidive.

Verschillende risicotaxatie-instrumenten worden inmiddels veelvuldig gebruikt in de forensisch-psychiatrische praktijk. Het is daarom belangrijk om goed op de hoogte te zijn van de psychometrische kenmerken van veelgebruikte risicotaxatie-instrumenten in Nederland. Wij geven in dit artikel een overzicht van de huidige kennis.

## METHODE

Er zijn in Nederland vooral studies gedaan naar de *HCR-20*, de *HKT-30*, de *Forensisch Psychiatrische Profielen (FP-40)* en de *Sexual Violence Risk-20 (SVR-20)*. In dit artikel laten wij buiten beschouwing: onderzoeken naar de *Static-99*, een actuariële instrument betreffende seksuele recidive, en onderzoeken naar de *RISC* (Van der Knaap e.a. 2007) en de *Quickscan* (Von Bergh e.a. 2006); dit zijn beide risicotaxatie-instrumenten die door reclasseringsinstellingen worden gebruikt. Ook instrumenten die ontwikkeld zijn voor jeugdigen laten we buiten beschouwing, zoals de *Structured Assessment of Violence Risk in Youth (SAVRY)* (Lodewijks e.a. 2008) en een jeugdversie van de *Psychopathy Checklist-Revised (PCL-R)* (Das e.a. 2008), of een ambulante forensisch risicotaxatie-instrument, zoals momenteel binnen De Waag Nederland wordt ontwikkeld (de Waagschaal; Van Horn e.a. 2008).

Tevens is er in Nederland een aantal studies gedaan naar de predictieve validiteit van de *PCL-R*; feitelijk is dit geen risicotaxatie-instrument, maar een instrument bedoeld om de mate van psychopathie te meten. Overigens wordt de score op de *PCL-R* ook meegenomen bij instrumenten zoals de *HCR-20*.

## RESULTATEN

### Algemene psychometrische kenmerken

De interbeoordelaarsbetrouwbaarheid wordt meestal bepaald met de intraklassecorrelatiecoëfficiënt (ICC). Een ICC-waarde  $> 0,75$  wordt over het algemeen als zeer goed beoordeeld, een ICC-waarde van 0,6 tot 0,75 als goed, van 0,4 tot 0,6 als redelijk en een waarde  $< 0,4$  als matig of onvoldoende (Fleisch 1986).

De interne consistentie van een schaal of de mate waarin de samenstellende items van een schaal bij elkaar horen, wordt uitgedrukt met Cronbachs alfa. De Commissie Test Aangelegenheden Nederland (COTAN) stelt de volgende grenzen voor: een alfa  $< 0,80$  is onvoldoende,

van 0,80 tot 0,90 is voldoende en  $> 0,90$  is goed.

De predictieve validiteit wordt in de meeste studies bepaald door de sensitiviteit (de proportie juist voorspelde recidivisten) en de specificiteit (de proportie juist voorspelde niet-recidivisten) van een instrument vast te stellen. Een handzame maat om de predictieve validiteit uit te drukken is de oppervlakte onder de curve (AUC) van de *receiver operating characteristics*. In een curve worden de sensitiviteit en 1-de specificiteit afgezet voor alle mogelijke grenswaarden op het instrument. Bij een lage grenswaarde is het instrument optimaal sensitief (alle recidivisten worden als zodanig voorspeld), maar weinig specifiek (ook vele niet-recidivisten worden als recidivist aangemerkt). Bij een hoge grenswaarde wordt voor bijna niemand recidive voorspeld, wat resulteert in optimale specificiteit, maar een lage sensitiviteit. De kunst is nu het afkappunt te kiezen met een optimale balans tussen sensitiviteit en specificiteit. Een goed instrument, dat wil zeggen een instrument met een goede predictieve validiteit, detecteert voldoende recidivisten zonder te veel niet-recidivisten als recidivist aan te merken. Met de AUC is deze predictieve validiteit van het instrument uit te drukken. Bij een AUC-waarde van 0,5 loopt de curve recht en is bij elke grenswaarde de sensitiviteit en specificiteit 50%, wat gelijk is aan kans. Een AUC-waarde tussen 0,70 en 0,75 is redelijk (De Vogel e.a. 2005) en een waarde  $> 0,75$  of 0,80 wordt als goed beoordeeld (Brand 2005a, b; Fawcett 2006).

#### Uitkomsten voor onderzochte instrumenten

**HCR-20** De HCR-20 is als risicotaxatie-instrument ontwikkeld in Noord-Amerika. Deze is ook in Nederland meerdere keren onderzocht en wordt in de klinische praktijk inmiddels frequent gebruikt. De lijst bestaat uit 20 items, onderverdeeld in historische items ('H'), klinische items ('C') en risicohanteringsitems ('R'), die een score tussen 0 en 2 krijgen. De historische items zoals 'eerder gewelddadig gedrag' en 'problemen in de kindertijd' verwijzen naar recidivevoorspellende factoren in het verleden (zogenoemde actuariële

factoren), terwijl de klinische factoren zoals 'gebrek aan zelfinzicht' en 'impulsiviteit' en de risicohanteringsitems zoals 'blootstelling aan destabiliserende factoren' en 'geringe beschikbaarheid steun' betrekking hebben op factoren die in de toekomst de kans op recidive zouden kunnen verhogen. In de HCR-20 is ook de mate van psychopathie meegenomen, waarvoor de PCL-R (zie hieronder) ook nog gescoord moet worden. Daarnaast wordt er nog een 'gestructureerd klinisch eindoordeel' naar de kans op recidive gevraagd.

Er werd in Nederland een aantal studies verricht naar de betrouwbaarheid en de validiteit van de HCR-20 (De Vogel e.a. 2004; 2005; 2006; Hildebrand e.a. 2005; Philipse e.a. 2002).

De interbeoordelaarsbetrouwbaarheid van de HCR-20 bleek voor vrijwel alle onderdelen goed tot zeer goed (ICC van 0,55 tot 0,89). Alleen de subschaal 'risicohanteringsitems' haalde in alle onderzoeken niet meer dan een redelijke interbeoordelaarsbetrouwbaarheid (ICC van 0,49 tot 0,58).

In het onderzoek van Hildebrand e.a. (2005) werd ook gekeken naar de interne consistentie van de HCR-20. Hierbij valt op dat alleen de subschaal 'risicohanteringsitems' voldoende scoorde. Alle andere (sub)schalen scoorden onvoldoende.

De predictieve validiteit van de HCR-20 is terug te vinden in tabel 1. De meeste onderzoeken lieten een matige tot goede validiteit zien. Vermelenswaardig is dat de risico-inschatting bij de onderzoeken van De Vogel e.a. gebeurde op basis van zogeheten 'consensusbesprekingen'. Daarbij vergeleken verschillende onderzoekers hun scores met elkaar en moesten tot een consensus komen over de weging van de verschillende items.

**HKT-30** De HKT-30 is een risicotaxatie-instrument dat is ontwikkeld voor de Nederlandse situatie. Het instrument heeft veel overeenkomsten met de HCR-20, maar bevat een groter aantal (dynamische) risicofactoren. De HKT-30 bestaat uit 30 items en is net als de HCR-20 onderverdeeld in 3 categorieën: 'historisch', 'klinisch' en 'toekomst'. Ook bij de HKT-30 wordt vaak na het scoren van de items een gestructureerd

TABEL 1		Onderzoeken naar de predictieve validiteit van de Historical, Clinical, and Risk Management (HCR-20) voor geweldadige recidive				
Variabele	Philipse e.a. (2002)	De Vogel e.a. (2004)	De Vogel e.a. (2005)		De Vogel e.a. (2004/2006)	Hildebrand e.a. (2005)
Onderzoeksopzet	R	R	R/P		P	R
Groeps grootte	45	120	42	42	127	153
Gemiddelde time at risk; in jaren	4,3	6	6,2 (R)/ 0,9 (P)	6,8 (R)/ 1,6 (P)	1,8	7,2
AUC totaalscore	0,78*	0,82***	0,59 (ns)	0,88***	0,85***	0,67**
AUC gestructureerd klinisch oordeel	0,82*	0,79***	0,86***	0,91***	0,86***	0,64**
AUC historische items	0,89**	0,80***	0,63 (ns)	0,83***	0,77***	0,71***
AUC klinische items	0,67	0,77***	0,61 (ns)	0,75***	0,80***	0,62*
AUC risicohanterings-items	0,64	0,79***	0,52 (ns)	0,88***	0,79***	0,62*

R = retrospectief onderzoek; P = prospectief onderzoek; AUC = area under the curve; ns = niet significant.  
 \*p < 0,05.  
 \*\*p < 0,01.  
 \*\*\*p < 0,001.

klinisch oordeel gevraagd over het recidiverisico volgens de score. Het instrument is in de loop van de tijd aangepast en er zijn inmiddels een tweede en sinds kort een derde, gemodificeerde, editie van de HKT-30 in omloop.

Wij vonden twee publicaties over studies van de eerste editie van de HKT-30 (Canton e.a. 2004a, b) op basis van Pro Justitia-rapportages en twee publicaties (Hildebrand e.a. 2005; Schönberger e.a. 2008) over de tweede editie van de HKT-30 op basis van tbs-dossiers. De interbeoordelaarsbetrouwbaarheid van de HKT-30 van beide edities was goed tot zeer goed (ICC varieerde van 0,61 tot 0,87). De interne consistentie voor de verschillende subschalen varieerde van onvoldoende tot goed (Hildebrand e.a. 2005).

De predictieve validiteit van de HKT-30 voor ernstige of geweldadige recidive (zie tabel 2) was in twee van de drie studies redelijk wat betreft de totaalscore en het gestructureerde klinische oordeel (Canton e.a. 2004b; Hildebrand e.a. 2005). In het onderzoek van Schönberger e.a. (2008) waren de uitkomsten niet of onvoldoende significant ten opzichte van kansniveau.

**SVR-20** De SVR-20 is in Noord-Amerika ontwikkeld om specifiek het seksueel recidiverisico in kaart te brengen. De SVR-20 bestaat uit 20 items, onderverdeeld in 3 categorieën: psychosociale aanpassing (bijvoorbeeld 'seksuele deviatie' en 'psychopathie'), seksuele delicten (bijvoorbeeld 'meerdere typen seksuele delicten en

TABEL 2		Retrospectieve dossieronderzoeken naar de predictieve validiteit van de Historische, Klinische en Toekomstige Risico-indicatoren-30 (HKT-30)		
Variabele	Canton e.a. (2004)	Hildebrand e.a. (2005)	Schönberger e.a. (2008)	
Groepskenmerk	PJ-rapportages	tbs na vrijlating	tbs na vrijlating	
Groeps grootte	123	153	83	
Gemiddelde time at risk; in jaren	5,9	7,2	10	
AUC totaalscore	0,72*	0,72***	0,61 (ns)	
AUC gestructureerd klinisch oordeel	onbekend	0,73***	0,59 (ns)	
AUC historische items	0,69*	0,67**	0,62 (ns)	
AUC klinische items	0,68*	0,69***	0,55 (ns)	
AUC toekomstitems	0,73*	0,68**	0,65*	

PJ = Pro Justitia; AUC = area under the curve; ns = niet significant.  
 \*p < 0,05.  
 \*\*p < 0,01.  
 \*\*\*p < 0,001.

TABEL 3 Retrospectief dossieronderzoek naar de predictieve validiteit van de Sexual Violence Risk-20 (SVR-20) voor seksuele recidive

Variabele	De Vogel e.a. (2004)	Koster e.a. (2006)	Schönberger e.a. (2008)
Groepsgrootte	121	30	86
Gemiddelde time at risk; in jaren	11,7	8,8	10
AUC totaalscore seksuele recidive	0,80*	0,63 ns	0,52 ns
AUC finale risicobeoordeling	0,83*	0,69 ns	0,57 ns
AUC 'psychosociale aanpassing'	0,68*	0,62 ns	0,51 ns
AUC 'seksuele recidive'	0,79*	0,54 ns	0,49 ns
AUC 'toekomstplannen'	0,76*	0,64 ns	0,59 ns

AUC = area under the curve; ns = niet significant.

\*p < 0,001.

'lichamelijk letsel bij slachtoffer(s) van seksuele delicten') en toekomstplannen ('ontbreken van realistische toekomstplannen' en 'negatieve houding ten opzichte van interventies'). Ook bij de SVR-20 wordt een gestructureerd klinisch eindoordeel gevraagd.

Wij vonden 3 studies over de SVR-20 in Nederland, namelijk een studie van De Vogel e.a. (2004), van Koster e.a. (2006) en van Schönberger e.a. (2008). De interbeoordelaarsbetrouwbaarheid van de verschillende onderdelen varieerde van redelijk tot zeer goed in de studies (ICC: van 0,42 tot 0,78). De interne consistentie (Koster e.a. 2006) was onvoldoende, met uitzondering van de subschaal 'toekomst'.

De predictieve validiteit (zie tabel 3) wat betreft seksuele recidive was in slechts een van de drie studies voor de meeste onderdelen goed.

FP-40 Sinds 1990 inventariseert men alle relevante kenmerken van terbeschikkingstelling(tbs)-patiënten die van belang zijn voor recidive. Dit heeft uiteindelijk geresulteerd in een set van vragenlijsten die moeten worden ingevuld door verschillende professionals die bij behandeling betrokken zijn, zoals psychotherapeuten, psychologen en psychiaters. De verschillende items worden ondergebracht in 40 'schalen', zoals milieufactoren, persoonlijkheidsfactoren en psychiatrische symptomen. Deze schalen worden onderverdeeld in 4 categorieën, te weten: 'ontwikkeling en maatschappelijke inbedding', 'psychiatrisch beeld', 'persoonlijkheid en vaar-

digheden' en 'situationele en overige invloeden'.

De laatste versie van de FP-40 dateert van 2006, maar veel onderzoeksgegevens zijn gebaseerd op eerdere versies. De meeste resultaten daarvan staan vermeld in de handleiding FP40, versie 2006.

Voor alle schalen werd de interne consistentie bepaald (Brand 2003). Van de 40 schalen hadden er 23 een redelijke tot goede interne consistentie. De interbeoordelaarsbetrouwbaarheid werd alleen vermeld voor de 5 schalen van de Lijst Sociotherapie en varieerde van goed tot zeer goed (ICC varieerde van 0,64 tot 0,86). De predictieve validiteit van de FP40-schalen voor middelzware tot zeer zware delictrecidive werd op meerdere manieren bepaald. Voor de totale populatie was de predictieve validiteit redelijk (AUC 0,72). Wanneer men de lijsten en scores echter uitsplitste naar verschillende subpopulaties, dan bleek de predictieve validiteit goed tot zeer goed (AUC varieerde van 0,80 tot 0,93).

PCL-R De Psychopathy Checklist Revised (PCL-R) is in Noord-Amerika door Hare (Ver-tommen e.a. 2002) ontwikkeld op basis van de criteria van Cleckley met betrekking tot psychopathie, om op een meer gestructureerde wijze te komen tot de diagnose psychopathie. De PCL-R is primair een diagnostisch instrument met een onderliggend nosologisch concept, maar wordt regelmatig als risicotaxatie-instrument gebruikt en past daarom in dit overzicht van instrumenten. In eerste instantie werden de 20 items onder-

verdeeld in twee factoren. Factor 1 betrof de meer berekenende psychopathietrekken, zoals 'gladde prater' en 'ontbreken emotionele diepgang' en factor 2 betrof de meer impulsieve, antisociale gedragingen, zoals 'prikkelhongerig/neiging tot verveling' en 'impulsiviteit'. Tegenwoordig wordt de PCL-R veelal onderverdeeld in 3 of 4 factoren. De items 'seksueel gedrag' en 'partnerrelaties' vallen niet onder een bepaalde factor.

Er werden 4 studies in Nederland uitgevoerd, waarin gekeken werd naar de interbeoordelaarsbetrouwbaarheid van de PCL-R (Hildebrand e.a. 2002; 2005; Koster e.a. 2006; Schönberger e.a. 2008). De interbeoordelaarsbetrouwbaarheid van de totaalscore van de PCL-R en van de twee factoren was goed tot zeer goed (ICC varieerde van 0,71 tot 0,88) en die van de vier facetten was redelijk tot goed (ICC varieerde van 0,54 tot 0,84).

In drie studies (Hildebrand e.a. 2002; 2005; Spreen e.a. 2008) was de interne consistentie van de totaalscore voldoende, in één studie was deze onvoldoende (Koster e.a. 2006). De interne consistentie van factor 1 en factor 2 was wisselend voldoende of onvoldoende.

Een publicatie in 2004 liet zien dat er een significante correlatie was tussen de score op de PCL-R en ontwrichtend gedrag binnen de kliniek (Hildebrand e.a. 2004). Er werden daarnaast verschillende studies gedaan naar de predictieve validiteit van de PCL-R in Nederland middels een

ROC-analyse (zie tabel 4).

De predictieve validiteit van de totaalscore van de PCL-R voor gewelddadige recidive varieerde van onvoldoende tot goed. Voor algemene (gewelddadige en niet-gewelddadige) recidive was deze redelijk of onvoldoende. De predictieve validiteit van factor 2 voor gewelddadige recidive kwam in twee studies als onvoldoende naar voren, terwijl deze in de overige studies goed was.

**Ongestructureerd klinisch oordeel** Binnen de gevonden publicaties waren er drie waarbij het ongestructureerde klinische oordeel werd meegenomen bij het bepalen van de predictieve validiteit van een instrument.

In het onderzoek van Philipse e.a. (2002) was de predictieve validiteit voor ernstige recidive goed (AUC 0,76). De Vogel e.a. (2004) extrapoleerden retrospectief het ongestructureerde klinische oordeel uit de adviezen die aan de rechtbank waren gegeven. De predictieve validiteit voor gewelddadige (AUC 0,68) en algemene recidive (AUC 0,63) was significant beter dan kans, maar was onvoldoende voor een adequate inschatting van het recidiverisico. In het retrospectieve onderzoek van Canton e.a. (2004b) naar de predictieve validiteit van de HKT-30 was het ongestructureerde klinische oordeel redelijk als inschatting van het risico voor gewelddadige recidive (AUC 0,70) en voor minder ernstige recidive (AUC 0,72).

TABEL 4 Retrospectief dossieronderzoek naar de predictieve validiteit van de Psychopathy Checklist-Revised (PCL-R)

Variabele	Hildebrand e.a. (2003)	De Vogel e.a. (2004)	Hildebrand e.a. (2005)	Koster e.a. (2006)	Schönberger e.a. (2008)
Groepsgrootte	94	120	153	58	84
Gemiddelde time at risk; in jaren	11,8	6	7,2	8,8	10
AUC totaalscore gewelddadige recidive	0,70**	0,75***	0,72***	0,82**	0,61 (ns)
AUC factor 1 gewelddadige recidive	0,62 (ns)	0,63**	0,61 (ns)	0,69 (ns)	0,50 (ns)
AUC factor 2 gewelddadige recidive	0,69**	0,79***	0,77***	0,83***	0,63*
AUC totaalscore algemene recidive	0,74**	0,68**	0,69***		
AUC factor 1 algemene recidive	0,67**	0,63*	0,57 (ns)		
AUC factor 2 algemene recidive	0,71**	0,70***	0,74***		

AUC = area under the curve; ns = niet significant; factor 1 betreft berekenende psychopathietrekken; factor 2 betreft meer impulsieve, antisociale gedragingen.

\*p < 0,05.

\*\*p < 0,01.

\*\*\*p < 0,001.

## DISCUSSIE

Met dit artikel willen wij een overzicht bieden van de uitkomsten van onderzoeken naar de psychometrische kwaliteit van verschillende, in Nederland toegepaste, forensische risicotaxatie-instrumenten bij volwassenen. Wij denken dat de gevonden artikelen een goed beeld van de stand van zaken geven.

**Goede betrouwbaarheid** Vrijwel alle publicaties zijn gebaseerd op retrospectief uitgevoerd onderzoek binnen de populatie van - voornamelijk mannelijke - tbs-patiënten en voormalig tbs-patiënten.

Over het geheel genomen, bleek de interbeoordelaarsbetrouwbaarheid van de genoemde instrumenten goed. De interne consistentie van de HCR-20, HKT-30 en SVR-20 is over het geheel genomen onvoldoende. De interne consistentie van de PCL-R is globaal beter. Bij de FP-40 heeft ruim de helft van de schalen een goede intraschaalbetrouwbaarheid.

**Vereisten** Voor instrumenten en oordelen op grond waarvan men belangrijke beslissingen moet nemen, dient de predictieve validiteit meerdere keren te zijn bepaald binnen een populatie die dezelfde samenstelling heeft als waarin men het instrument wil gebruiken (Brand 2005b; Grisso e.a. 1992). Tot op heden zijn er alleen over de HCR-20 meerdere studies gepubliceerd waarin een AUC-waarde boven de 0,75 werd gevonden. Deze studies zijn echter merendeels uitgevoerd door dezelfde onderzoeksgroep (De Vogel 2005). Een latere studie naar de HCR-20 laat aanzienlijk minder goede uitkomsten zien. Mogelijk dat het vaststellen van het risico op recidive bij een onderzochte middels consensusbesprekingen, zoals De Vogel e.a. deden, tot een verhoging van de predictieve validiteit leidt.

De PCL-R lijkt minder geschikt voor risicotaxatie. Voor dit instrument is alleen voor factor 2 bij verschillende onderzoeken een AUC-waarde boven de 0,75 gevonden.

Hoewel veel instrumenten meestal een significant hogere predictieve validiteit hebben dan kansniveau, hebben ze onvoldoende voorspellende waarde om doorslaggevend te zijn bij zeer ingrijpende beslissingen. Er zijn in de laagrisicogroepen relatief veel fout-negatieve uitslagen en van de gewelddelictplegers in de hoogrisicogroep is de uitslag bij ongeveer de helft fout-positief (Hildebrand e.a. 2005).

**Vraagtekens en beperkingen** Gezien de geringe voorspellende waarde van de instrumenten kan men vanuit ethisch oogpunt vraagtekens plaatsen bij het belang dat zeker ook vanuit het ministerie van Justitie wordt gehecht aan de totaalscores op risicotaxatie-instrumenten bij het toekennen van verlof of beëindiging van intramurale zorg in het kader van een tbs-maatregel. We sluiten ons dan ook graag aan bij een uitspraak van Brand (2005b): *‘Zo lang de predictieve validiteit van instrumenten voor risicotaxatie niet hoger is dan het niveau wat recent gevonden is, lijkt het af te raden om zeer ingrijpende beslissingen alleen, of vrijwel alleen, van een testscore af te laten hangen. De scores op een risicotaxatielijst, zowel de somscore als losse scores, kunnen zeker wel als nuttige extra informatie gebruikt worden’*. Daarnaast lijkt er meer aandacht nodig voor de ‘houdbaarheidsdatum’ van taxaties, aangezien de huidige risicotaxaties slechts gaan over recidive, zonder dat de periode waarbinnen deze op zou moeten treden is gedefinieerd (Nijman e.a. 2006).

Er zijn ook aanwijzingen dat alleen al de juridische fase waarin iemand zich bevindt, van invloed is op de uitkomsten van een risicotaxatie-instrument (Urbaniok e.a. 2007). Ook voor het ongestructureerde klinische oordeel geldt dat er onvoldoende wetenschappelijke gegevens zijn om louter op basis hiervan ingrijpende beslissingen te nemen.

Men kan betwijfelen in hoeverre de tegenvalende voorspellende waarde van risicotaxatie voor daadwerkelijke recidive aan onvolkomenheden van het meetinstrumentarium te wijten is. Waarschijnlijk wordt recidive in grote mate mede

bepaald door toevallige, niet te voorziene omstandigheden. Een verwante vraag is in hoeverre het überhaupt mogelijk is om betrouwbaar de uitkomsten van een bepaalde groep te vertalen naar een individu uit deze groep (Hart e.a. 2007). Bij de inschatting van de kans op recidive en de daaruit volgende adviezen en behandelingen zal de beoordelend psychiater of psycholoog zich dus bewust moeten blijven van de grote mate van onzekerheid die er is wat betreft de daadwerkelijke kans op recidive.

**Nieuwe ontwikkelingen** Om te komen tot een uniforme wijze van scoren geldt voor verschillende instrumenten, zoals de HCR-20, de SVR-20 en de PCL-R, dat hiervoor het volgen van een training vereist is. Om de kwaliteit van het voorspellen van recidive verder te verbeteren, zijn er op dit moment meerdere ontwikkelingen gaande. Zo loopt er op dit moment een internationaal onderzoek naar de HCR-20 (versie 3), waarbij de mate van psychopathie niet meer als apart item is opgenomen (K.S. Douglas, schriftelijke mededeling, 2009).

Ook is het uitsplitsen van de totale populatie naar specifieke subpopulaties (Brand e.a. 2006) met bijbehorende specifieke risicotaxatie-instrumenten wellicht een manier om de voorspelling te verbeteren. Daarnaast worden nieuwe instrumenten ontwikkeld die zich vooral richten op beschermende factoren, zoals de *Structured Assessment of Projective Factors for violence risk* (SAPROF; De Vogel e.a. 2007). Deze brengen belangrijke aspecten in kaart die tot nu toe buiten beschouwing blijven, terwijl die juist het recidiverisico vermindern.

Een laatste ontwikkeling die gesignaleerd wordt, is het ontwerpen van zogeheten expertsystemen die met behulp van meer factoren dan tot nog toe gebruikelijk is een op het individu toegespitst risicoprofiel voor de herhaling van een specifiek delict proberen te voorspellen. Daarmee is potentieel beter maatwerk te leveren. Een voorbeeld van zo'n systeem is het instrument *Forensisch Operationalisierter Therapie- und Risiko-Evaluations-*

*System* (FOTRES; Urbaniok e.a. 2007). Dit in Zwitserland ontwikkelde systeem wordt momenteel binnen het Nederlands Instituut voor Forensische Psychiatrie en Psychologie (NIFP) onderzocht.

## BESLUIT

Toekomstig onderzoek naar de betrouwbaarheid en validiteit van deze nieuwe instrumenten en ontwikkelingen zal moeten uitwijzen in hoeverre de voorspelling van recidive kan verbeteren. In afwachting hiervan zullen we klinische indrukken, de uitkomsten op verschillende onderdelen van de instrumenten en alle bijkomende informatie moeten meewegen - zoals in het gestructureerd klinisch eindoordeel ook gebeurt. De uitkomst daarvan moeten we toetsen aan de inschattingen van andere professionals. Dit laatste kan bijvoorbeeld gebeuren in de vorm van consensusbesprekingen, zoals nu al gebeurt binnen de Van Der Hoeven kliniek (centrum voor klinische forensische psychiatrie in Utrecht) of in het Pieter Baan Centrum tijdens de stafvergadering van een Pro Justitia-onderzoek.

## LITERATUUR

- Bergh, M. von, van Poppel, J., & Römkens, R. (2006). Evaluatie bruikbaarheid Quick Scan Reclassering. Tilburg: IVA Beleidsonderzoek en advies.
- Brand, E.F.J.M. (2005a). Risicotaxatie van delictrecidive: methoden, statistieken adviezen. In M.F. Sjerps & J.A. Coster van Voorhout (Red.) *Het onzekere bewijs. Gebruik van statistiek en kansrekening in het strafrecht.* (pp. 399-428). Deventer: Kluwer.
- Brand, E.F.J.M. (2005b). Een maat voor de kwaliteit van instrumenten voor risicotaxatie. In M.F. Sjerps & J.A. Coster van Voorhout (Red.) *Het onzekere bewijs. Gebruik van statistiek en kansrekening in het strafrecht.* (pp. 429-455). Deventer: Kluwer.
- Brand E. (2003). Forensische Profiellijsten (FP40), Psychodiagnostisch gereedschap. *De Psycholoog*, 38, 239-245
- Brand, E., & van Emmerik, J.L. (2006). FP40 Handleiding V2006 Forensische profiellijsten. Den Haag: Dienst Justitiële Inrichtingen.
- Brand, E., & Nijman, H. (2007). Risicotaxatie en behandelbeoordeling met twee forensische observatielijsten. *Tijdschrift voor Psychiatrie*, 49, 221-232.



- Canton, W.J., van der Veer, T.S., van Panhuis, P.J.A., e.a. (2004a). De betrouwbaarheid van risicotaxatie in de pro Justitia rapportage, een onderzoek met behulp van de HKT-30. *Tijdschrift voor Psychiatrie*, 46, 537-542.
- Canton, W.J., van der Veer, T.S., van Panhuis, P.J.A., e.a. (2004b). 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.
- Das, J., de Ruiter, C., Doreleijers, T., e.a. (2009). Reliability and construct validity of the Dutch Psychopathy Checklist: Youth Version: Findings from a sample of male adolescents in a juvenile justice treatment institution. *Assessment*, 16, 88-102.
- Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874.
- Fleisch, J.L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
- Grisso, Th., & Appelbaum, T.S. (1992). Is it unethical to offer predictions of future violence? *Law and Human Behavior*, 16, 621-633.
- Hart, S.D., Michie, C., & Cooke, D.J. (2007). Precision of actuarial risk assessment instruments, evaluating the 'margins of error' of group v. individual predictions of violence. *British Journal of Psychiatry*, 190, s60-s65.
- Hildebrand, M., de Ruiter, C., & van Beek, D.J. (2001). *SVR-20; Richtlijnen voor het beoordelen van het risico van seksueel gewelddadig gedrag*, Nederlandse vertaling. Utrecht: Forum.
- Hildebrand, M., de Ruiter, C., de Vogel, V., e.a. (2002). Reliability and factor structure of the Dutch language version of Hare's Psychopathy Checklist-Revised. *International Journal of Forensic Mental Health*, 1, 139-154.
- Hildebrand, M., de Ruiter, C., & de Vogel, V. (2003). Recidive van verkrachters en aanranders na tbs; de relatie met psychopathie en seksuele deviatie. *De Psycholoog*, 38, 114-122.
- Hildebrand, M., de Ruiter, C., & Nijman, H. (2004). PCL-R psychopathy predicts disruptive behavior among male offenders in a Dutch forensic psychiatric hospital. *Journal of Interpersonal Violence*, 19, 13-29.
- Hildebrand, M., Hesper, B.L., Spreen, M., e.a. (2005). De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie; een onderzoek naar de betrouwbaarheid en predictieve validiteit van de HCR-20, HKT-30 en PCL-R. Utrecht: Expertisecentrum Forensische Psychiatrie.
- Horn, J. van, Wilpert, J., Scholing, A., e.a. (2008). *Waagschaal volwassenen, risicotaxatie-instrument voor de Ambulante Forensische Psychiatrie*. Utrecht: De Waag.
- Knaap, L.M. van der, Leenarts, L.E.W., & Nijssen, L.T.J. (2007). *Psychometrische kwaliteiten van de Recidive Inschattingsschalen (RISC); interbeoordelaarsbetrouwbaarheid, interne consistentie en congruente validiteit*. Den Haag: WODC.
- Koster, K., van Lankveld, J., & Spreen, M. (2006). Voorspelling van recidive bij zedendelinquenten met behulp van retrospectief gebruik van de PCL-R en SVR-20. *Tijdschrift voor Seksuologie*, 30, 204-214.
- Lodewijks, H.P.B., Doreleijers, T.A.H., & de Ruiter, C. (2008). SAVRY risk assessment in violent Dutch adolescents: Relation to sentencing and recidivism. *Criminal Justice and Behavior*, 35, 696-709.
- Nijman, H., & Bulten, E. (2006). Risicotaxatie en behandelbeoordeling in de forensische psychiatrie - een overzicht van recente promoties, *Tijdschrift voor Criminologie*, 48, 78-85.
- Philipse, M., van Erven, T., & Peters, J. (2002). Risicotaxatie in de tbs, van geloof naar empirie. *Justitiële verkenningen*, 28, 77-93.
- Philipse, M., & de Ruiter, C., & Bouman, Y. (2000). *HCR-20: Beoordelen van het risico van gewelddadig gedrag versie 2*, Nederlandse vertaling. Nijmegen/Utrecht: Pompestichting/Van der Hoevenstichting.
- Quinsey, V.L., Harris, G.T., Rice, M.E., e.a. (2006a). *Clinical judgment*. In V.L. Quinsey, G.T. Harris, M.E. Rice, e.a. (Red.) *Violent offenders: appraising and managing risk* (2de druk) (pp. 61-81). Washington: American Psychological Association.
- Quinsey, V.L., Harris, G.T., Rice, M.E., e.a. (2006b). Criticism of actuarial risk assessment. In V.L. Quinsey, G.T. Harris, M.E. Rice, C.A. Cormier (Red.) *Violent offenders: appraising and managing risk* (2de druk) (pp. 197-223). Washington: American Psychological Association.
- Schönberger, H.J.M., Hildebrand, M., Spreen, M., e.a. (2008). De waarde van gestructureerde risicotaxatie en van de diagnose psychopathie bij seksuele delinquenten; een onderzoek naar de betrouwbaarheid en predictieve validiteit van de SVR-20, Static-99, HKT-30 en PCL-R. Utrecht: Expertisecentrum Forensische Psychiatrie.
- Spreen, M., ter Horst, P., Lutjehuis, B., e.a. (2008). De kwaliteit en de rol van de PCL-R in het Nederlands TBS-systeem. *De Psycholoog*, 43, 78-86.
- Urbaniok, F. (2007). *FOTRES, Forensisches Operationalisiertes Therapie-Risiko-Evaluations-System*. (2de druk). Bern: Zytglogge.
- Urbaniok, F., Endrass, J., Rossegger A., e.a. (2007). The prediction of criminal recidivism, The implication of sampling in prognostic models. *European Archives of Psychiatry and Clinical Neuroscience*, 257, 129-134.
- Vertommen, H., Verheul, R., de Ruiter, C., e.a. (2002). *Hare's Psychopathie Checklist, handleiding*. Utrecht: Harcourt Test Publishers.
- Vogel, V. de, & de Ruiter, C. (2004). Differences between clinicians and

- researchers in assessing risk of violence in forensic psychiatric patients. *The journal of forensic psychiatry and psychology*, 15, 145-164.
- Vogel, V. de, de Ruiter, C., van Beek, D., e.a. (2004). Predictive validity of the SVR-20 and Static-99 in a Dutch sample of treated sex offenders. *Law and human behaviour*, 28, 235-251.
- Vogel, V. de, & de Ruiter, C. (2005). The HCR-20 in Personality disordered female offenders: A comparison with a matched sample of males. *Clinical psychology and psychotherapy*, 12, 226-240.
- Vogel, V. de, de Ruiter, C., Hildebrand, M., e.a. (2004). Type of discharge and risk of recidivism measured by the HCR-20: a retrospective study in a Dutch sample of treated forensic psychiatric patients. *International journal of forensic mental health*, 3, 149-165.
- Vogel, V. de. (2005). Structured risk assessment of (sexual) violence in forensic clinical practice. *The HCR-20 and SVR-20 in Dutch forensic psychiatric patients* (proefschrift). Amsterdam: Dutch University Press.
- Vogel, V. de, & de Ruiter, C. (2006). Structured professional judgement of violence risk in forensic clinical practice: A prospective study into the predictive validity of the dutch HCR-20. *Psychology, Crime & Law*, 12, 321-336.
- Vogel, V. de, de Ruiter, C., Bouman, Y., e.a. (2007). *SAPROF, Structured Assessment of PROjective Factors for violence risk; versie 1; richtlijnen voor het beoordelen van beschermende factoren voor gewelddadig gedrag*. Utrecht: Forum.

## AUTEURS

G.T. BLOK was ten tijde van het schrijven van dit artikel arts in opleiding tot psychiater en is thans als psychiater werkzaam bij De Waag Zeeland/Emergis, Middelburg.

E. DE BEURS is psycholoog en werkzaam als hoofd Onderzoek & Ontwikkeling binnen het Nederlands Instituut voor Forensische Psychiatrie en Psychologie (NIFP), Utrecht.

A.G.S. DE RANITZ is forensisch psychiater en werkzaam binnen het NIFP, locatie Pieter Baan Centrum, Utrecht.

T. RINNE is forensisch psychiater, lid van de raad van bestuur van het NIFP en medisch directeur van het Pieter Baan Centrum, Utrecht.

Correspondentieadres: G.T. Blok, De Waag Zeeland, Postbus 72, 4330 AB Middelburg.

E-mail: gblok@dewaag-nederland.nl.

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 29-9-2009.

## SUMMARY

The current psychometric state of risk assessment scales for adults in the Netherlands – G.T. Blok, E. de Beurs, A.G.S. de Ranitz, T. Rinne –

**BACKGROUND** Although Dutch forensic psychiatry is making increasing use of structural risk assessment scales, the controversy about the value and usefulness of these instruments continues unabated.

**AIM** To provide an overview of the psychometric qualities of the instruments used most often in the Netherlands for risk assessment in adults.

**METHOD** Dutch data about the Historical, Clinical, and Risk Management (HCR-20), the Sexual Violence Risk-20 (SVR-20), the Psychopathy Checklist-Revised (PCL-R) and two Dutch instruments, the ‘Historische, Klinische en Toekomstige Risico-indicatoren-30’ (HKT-30) and the ‘Forensisch Psychiatrische Profielen’ (FP-40) were reviewed. In addition, data relating to the unstructured clinical judgement were studied.

**RESULTS** The inter-rater reliability values of the instruments discussed were in general satisfactory, but the internal consistency was often unsatisfactory. Except in some studies, the predictive validity was in general reasonable.

**CONCLUSION** At present, caution is called for with regard to the assessment of the risk of recidivism when this is based purely on risk assessment scales or purely on the unstructured judgement. Perhaps it is simply not possible to predict recidivism more accurately. Until there are some new developments in this area, it seems advisable to combine as many data as possible about a person under investigation derived from assessment scales and clinical judgement and to compare the outcome with the conclusions of the other professionals.

[TIJDSCHRIFT VOOR PSYCHIATRIE 52(2010)5, 331-341]

**KEY WORDS** forensic psychiatry, risk assessment reliability, validity