

Risicotaxatie en behandelevaaluatie met twee forensische observatielijsten

E. BRAND, H. NIJMAN

ACHTERGROND Behandelevaaluatie en risicotaxatie gaan hand in hand in de terbeschikkingstellingsector (tbs-sector).

DOEL Het testen van twee observatielijsten die zijn ontwikkeld voor de evaluatie van tbs-behandeling.

METHODE Twee observatielijsten voor sociotherapeuten dan wel groepsleiders werden onderzocht op basis van metingen die zijn verricht in het voormalige Meijers (tbs-selectie) Instituut ($n = 556$) en in De Kijvelanden ($n = 100$), een tbs-behandelinginstelling voor forensische psychiatrie. Deze lijsten, de Observatielijst voor sociotherapeutisch medewerkers (STM-lijst) en de Copinglijst, beogen onder meer de sociale vaardigheden, coöperatie, zelfredzaamheid en copingvaardigheden van forensisch psychiatrische patiënten te meten.

RESULTATEN De twee lijsten laten een goede interne consistentie zien. Bevindingen in De Kijvelanden wijzen erop dat de STM-lijst over adequate test-hertestbetrouwbaarheid beschikt. Daarnaast blijken de STM-totaalscores en de afzonderlijke schalen ervan significant samen te hangen met criminele recidive. Vooral op de afdeling waargenomen antisociaal gedrag, zelfredzaamheid en sociale vaardigheden van de patiënt, en in mindere mate coöperatief gedrag en negatieve copingstijlen, bleken significant geassocieerd te zijn met latere terugval in crimineel gedrag.

CONCLUSIE De lijsten lijken geschikt om behandeffecten binnen forensisch psychiatrische settings in kaart te brengen en bieden mogelijk een aanvulling op gangbare risicotaxatiemethoden.

[TIJDSCHRIFT VOOR PSYCHIATRIE 49(2007)4, 221-232]

TREFWOORDEN behandelevaaluatie, forensische psychiatrie, risicotaxatie, terbeschikkingstelling

De afdeling Monitoring en Research TBS in het Meijers Instituut (nu Oostvaarderskliniek) heeft een set van tien vragenlijsten ontwikkeld die voor forensische instellingen bedoeld is. Deze vragenlijsten hebben samen de naam *Forensische Profiellijsten 40* (FP40) gekregen (Brand 2003; Brand 2006; Brand & Van Emmerik 2001).

Dit artikel heeft betrekking op een selectie uit deze set, namelijk de *Observatielijst voor sociothera-*

peutisch medewerkers (STM-lijst) en de Copinglijst. Deze twee onderdelen van de FP40 zijn bedoeld om door sociotherapeuten, ook wel groepsleiders genoemd, te worden ingevuld naar aanleiding van observaties van het gedrag van patiënten op de afdeling. Het zijn dus geen zelfbeoordelingslijsten die door tbs-patiënten worden ingevuld, maar observatielijsten die door de afdelingsstaf worden gescoord. Andere onderdelen van de FP40 moeten

door psychologen, psychiaters of maatschappelijk werkers te worden ingevuld op basis van hun specifieke deskundigheid.

Met deze twee vragenlijsten werd in het verleden op grote schaal in het Meijers Instituut en in tbs-kliniek De Kijvelanden gewerkt (Brand 2003; Brand 2006). In dit artikel komt de interne consistentie van de twee lijsten aan bod zoals reeds eerder op basis van de gegevens van het Meijers Instituut werd onderzocht (Brand 2003). Daarna wordt ingegaan op de test-hertestbetrouwbaarheid van de STM-lijst. Deze is bepaald aan de hand van herhaalde metingen die werden uitgevoerd in De Kijvelanden. Het accent zal echter liggen op de voorspellende waarde van de observatielijsten wat betreft later optredende criminele recidive: in hoeverre blijken de scores op de twee lijsten samen te hangen met latere terugval in crimineel gedrag?

Eerst wordt hier echter beschreven met welk doel de twee nieuwe lijsten en de bijhorende subschalen werden ontwikkeld. Er is een aantal redenen te noemen waarom diagnostische instrumenten die ontwikkeld zijn voor de reguliere psychiatrie, minder geschikt zouden kunnen zijn voor het forensische veld. Op de eerste plaats heeft zelfbeoordeling van de eigen symptomatologie door patiënten, zoals vaak in de reguliere ggz wordt gebruikt, haar beperkingen. Zelfbeoordeling is vooral problematisch bij een beoordeling over het al dan niet verlengen van een tbs; de patiënt zal zich hiertegen vaak verzetten. Op de tweede plaats kunnen veel gedragingen niet gescoord worden omdat de lange vrijheidsbeneming waarmee tbs doorgaans gepaard gaat, bepaald gedrag (grotendeels) uitsluit zoals reizen of alcoholgebruik. Als derde kan genoemd worden dat forensische patiënten op diverse kenmerken (bv. het vertonen van antisociaal gedrag) anders kunnen scoren, waardoor specifieke normgegevens en/of extra items nodig kunnen zijn voor forensische populaties. Om deze redenen werd in 1990 gestart met de ontwikkeling van de set van 10 vragenlijsten, die werden onderverdeeld in 40 subschalen, met de naam FP40.

Bij risicotaxatie en forensische diagnostiek

worden variabelen vaak ingedeeld in statische en dynamische variabelen; dynamische variabelen zijn daarbij (in theorie) veranderbaar. De twee lijsten voor sociotherapeuten die in dit artikel onderzocht worden, beogen dynamische variabelen te meten. Diverse onderzoekers melden dat het vooral belangrijk is om te zoeken welke dynamische variabelen een rol spelen in het ontstaan van agressief of gewelddadig gedrag (Langton & Barbaree 2001). Dit kan zowel van belang zijn voor behandeladviezen (Serin & Kuriychuk 1994) als voor het schatten van het risico op delictrecidive (Dernevik 2000).

METHODE

Proefpersonen

De resultaten van enkele onderzoeken die hebben plaatsgevonden binnen het tbs-veld worden hier beschreven. De gepresenteerde data zijn verzameld in het voormalige tbs-selectie-instituut, het Meijers Instituut ($n = 556$) en in De Kijvelanden ($n = 100$), een tbs-behandelinginstelling voor forensische psychiatrie. Voor tbs-patiënten in algemene zin geldt dat er, zoals het Wetboek van Strafrecht het formuleert, ten tijde van hun delict sprake was van een gebrekkige ontwikkeling of ziekelijke stoornis van de geestvermogens waardoor hun delictgedrag hun niet dan wel verminderd kan worden toegerekend. Bij instroom in de tbs (wat destijds altijd via het Meijers Instituut geschiedde) is de leeftijd van patiënten gemiddeld 31 jaar, heeft 86% een as-II-stoornis en 73% een as-I-stoornis volgens de criteria van de DSM. Bij de as-I-problematiek is dan ook verslaving meegenomen. De overgrote meerderheid van tbs'ers zijn mannen (zo'n 95%). Voor de steekproef van de Kijvelanden ($n = 100$) geldt dat het uitsluitend om mannelijke tbs-patiënten ging, aangezien in de Kijvelanden alleen mannen worden opgenomen.

In het onderhavige onderzoek zijn de tbs-patiënten opgenomen die eind jaren negentig (vanaf juli 1996) in het toenmalige Meijers Instituut instroomden en dus (bijna zonder uitzondering) aan

het begin van hun tbs-behandeling stonden. In totaal werden in die jaren bij 556 tbs-patiënten van het Meijers Instituut STM-lijsten afgenomen. Aangezien de Copinglijst tijdens deze onderzoeksperiode nog inhoudelijk werd aangepast op psychometrische gronden, ligt het aantal hiervan (in de laatste versie) in dit onderzoek lager ($n = 376$). Voor wat betreft de dataset van De Kijvelanden gaat het om STM-gegevens van 100 personen die tussen februari 1998 en juni 2001 in De Kijvelanden opgenomen waren en waarvan minimaal 2 herhaalde metingen met de STM-lijst voorhanden waren. In de genoemde periode werd deze lijst namelijk halfjaarlijks gescoord in de Kijvelanden.

Instrumenten

De FP40 Observatielijst voor sociotherapeutisch medewerkers (STM-lijst) De STM-lijst bestaat uit 67 items die door sociotherapeuten op basis van observaties van het gedrag van de patiënten dienen te worden gescoord. De STM bevat de volgende subschalen: sociale vaardigheden (SOCV), coöperatie met de behandeling (COOP), het vertonen van antisociaal gedrag op de afdeling (ASOC) en zelfredzaamheid van de patiënten (REDZ).

FP40 Copinglijst De Copinglijst is een lijst waarin zeven typen probleemsituaties beschreven staan die zich in een tbs-kliniek kunnen voordoen. Achter elke situatie staan acht mogelijke manieren van reageren. Op basis van resultaten van factoranalyses zijn de subschalen positieve copingvaardigheden (COPP) en vermijdend gedrag (COPA) ontstaan. Daarnaast worden ook (frequent) voorkomende negatieve reacties op probleemsituaties onder de tbs-patiënten vastgelegd, waaruit de subschaal negatieve copingmanieren (COPN) wordt berekend.

Statistische analyses

Op basis van de in het Meijers Instituut verzamelde gegevens werd de interne consistentie

van de STM- en Copinglijst bestudeerd. De interne consistentie werd hierbij uitgedrukt in Cronbachs alfa voor de afzonderlijke subschalen van de twee lijsten.

Aan de hand van de in de Kijvelanden verzamelde herhaalde metingen werd de test-hertestbetrouwbaarheid van de STM-lijst bepaald. Zoals eerder beschreven werden hiervoor de STM-gegevens van personen gebruikt die tussen februari 1998 en juni 2001 in de Kijvelanden waren opgenomen en waarvan minimaal 2 herhaalde (halfjaarlijkse) metingen voorhanden waren. Het ging in totaal om 100 patiënten waarvoor dit het geval was. De test-hertestbetrouwbaarheid van de subschalen van de STM-lijst wordt uitgedrukt in pearsoncorrelaties tussen de herhaalde metingen.

De predictieve validiteit van latere criminaliteit werd onderzocht aan de hand van de STM- en Copinglijstgegevens die bij instroom in het Meijers Instituut eind jaren negentig waren verzameld. Hiervoor werd later op gezette tijden informatie over nieuwe veroordelingen opgevraagd bij de Centrale Justitiële Documentatie over alle patiënten die in het selectiecentrum opgenomen waren geweest. Deze delictrecidivegegevens werden gekoppeld aan de instroommetingen met de STM- ($n = 556$) en de voorhanden zijnde Copingscores ($n = 376$) zoals uitgevoerd in het Meijers Instituut. Hierbij dient opgemerkt te worden dat de 'time at risk' voor recidive in dit onderzoek dus de periode vanaf instroom in tbs-behandeling betreft. Empirisch onderzoek (bv. Hilterman 2002), maar in feite ook recente incidenten die in de media veel aandacht kregen, laten zien dat criminele recidives kunnen optreden terwijl de tbs-maatregel van kracht is. Vooral naarmate de patiënt meer vrijheden en verlofmogelijkheden krijgt gedurende de tbs-behandeling, stijgt het aantal potentiële risicosituaties, maar ook intramuraal in de tbs-kliniek kunnen ernstige incidenten optreden. Op basis van ervaringen lijkt het relatief vaak om gewelddadige delicten te gaan die intramuraal optreden, terwijl recidive in zedendelicten, indien ze optreden, doorgaans later (in situaties met minder

toezicht) optreden, en soms vele jaren na beëindiging van de tbs-maatregel (Hildebrand e.a. 2003).

In dit onderzoek wordt in de statistische vergelijkingen de time at risk voor recidivisten en niet-recidivisten dus gerekend vanaf hun instroommeting aan het begin van hun tbs-behandeling, en zijn in alle berekeningen de follow-upperiodes voor beide groepen steeds gelijk gehouden. Meer specifiek worden de recidivecijfers na twee, drie, vier, vijf en zes jaar na instroom geanalyseerd (het eerste jaar wordt wegens een te laag aantal recidivisten voor statistische analyses buiten beschouwing gelaten). Verschillen in scores op STM- en Copinglijst tussen recidivisten en niet-recidivisten worden over deze follow-upperiodes met t-testen onderzocht. De predictieve validiteit wordt tevens onderzocht met de methode die hiervoor de laatste jaren als meest geschikte methode wordt gezien, namelijk de zogenaamde *area under the curve* (AUC)-waarde van een ROC-curve.

Voor een ROC-curve worden getallenparen geplotted op een X-Y-plot, in 2 percentages. Voor alle mogelijke scores wordt berekend wat het percentage patiënten is dat terecht als hoog risico wordt beschouwd (percentage detectie) en het percentage dat bij die gekozen grenswaarde onterecht als hoog risico wordt beschouwd (percentage vals alarm). Bij een goede test is bij elke gekozen grenswaarde het percentage detectie hoger dan het percentage vals alarm. Wanneer men geen voorkennis

heeft en volslagen willekeurig tot beslissingen komt over risico's, zal de AUC-waarde 0,50 zijn. Een meer grondige uitleg van ROC-curves en de AUC-waarde staat in een recent verschenen statistiekboek voor strafrechtjuristen (Brand 2005). Bij een AUC-waarde boven de 0,65 kunnen op groepsniveau statistisch al zeer significante groepsverschillen bestaan. Om individuele diagnostiek te kunnen uitvoeren zijn echter AUC-waarden nodig die ruim boven de 0,70 liggen.

RESULTATEN

Intraschaalbetrouwbaarheid (interne consistentie)

Op basis van de in het Meijers Instituut verzamelde STM- en Copinglijsten werd de interne consistentie – ofwel intraschaalbetrouwbaarheid – van de subschalen van de STM- en van de Copinglijst in Cronbachs alfa uitgedrukt. De resultaten in tabel 1 tonen dat de items binnen de schalen ruim voldoende samenhangen (Cronbachs alfa 0,84 – 0,94).

Test-hertestbetrouwbaarheid van de STM-lijst

De test-hertestbetrouwbaarheid van de STM-lijst werd berekend op de data die verzameld werden in De Kijvelanden. De resultaten van de test-hertestanalyses staan in tabel 2. De test-hertestbe-

TABEL 1 Intraschaalbetrouwbaarheid (interne consistentie) van de subschalen van de Observatielijst voor sociotherapeutisch medewerkers (STM-lijst) en de Copinglijst

Lijst	Aantal items	Cronbachs alfa
STM-lijst		
ASOC	8	0,87
COOP	9	0,85
REDZ	12	0,88
SOCV	9	0,85
Copinglijst		
COPA	14*	0,92
COPN	14*	0,94
COPP	21**	0,89

ASOC = antisociaal gedrag op de afdeling; COOP = coöperatief gedrag op de afdeling; REDZ = REDzaamheid; SOCV = sociale vaardigheden; COPA = vermijgend gedrag; COPN = negatieve copingmanieren; COPP = positieve copingvaardigheden;

* Zeven probleemsituaties keer twee manieren van omgaan met het probleem

** Zeven probleemsituaties keer drie manieren van omgaan met het probleem

TABEL 2 Test-hertestcorrelaties van de subschalen van de Observatielijst voor sociotherapeutisch medewerkers (STM-lijst) in tbs-kliniek De Kijvelanden 1998-2001

Schaal STM-lijst	Dataselectie A*			Dataselectie B**		
	r	n	p	r	n	p
ASOC	0,499	100	0,000	0,795	45	0,000
REDZ	0,617	99	0,000	0,706	44	0,000
SOCV	0,488	98	0,000	0,678	45	0,000
COOP	0,472	100	0,000	0,544	45	0,000

ASOC = antisociaal gedrag op de afdeling; REDZ = redzaamheid; SOCV = sociale vaardigheden; COOP = coöperatief gedrag op de afdeling; r = pearsons correlatiecoëfficiënt
 *Dataselectie A: tussenperiode een half jaar, hertest vaak door andere beoordelaar
 **Dataselectie B: tussenperiode een half jaar, hertest door zelfde beoordelaar

trouwbaarheid is berekend op basis van 100 eerste en tweede metingen die halfjaarlijks in de Kijvelanden werden uitgevoerd (dataselectie A genoemd in tabel 2). Hierbij kon het bij de 2 metingen gaan om 2 verschillende medewerkers van de afdeling. Om zicht te krijgen op de invloed die scoring door verschillende beoordelaars heeft, worden ook de correlaties van herhaalde, halfjaarlijkse metingen getoond in geval de beoordelaar een half jaar later dezelfde was ($n = 45$; in de tabel aangegeven als dataselectie B).

De tabel laat zien dat bij beoordeling door dezelfde beoordelaar de test-hertestbetrouwbaarheid van antisociaal gedrag (ASOC) en zelfredzaamheid (REDZ) op de afdeling en sociale vaardigheden (SOCV) redelijk zijn, zeker wanneer de behoorlijk lange tussenliggende (test-hertest) periode (een half jaar) in ogenschouw wordt genomen. De minste overeenstemming wordt gevonden voor de beoordeling van coöperatief gedrag (COOP), ook indien dezelfde beoordelaar de schaal tweemaal had gescoord. Coöperatief gedrag is mogelijk een meer veranderlijk, situatieafhankelijk kenmerk (*state*) terwijl het vertonen van sociale en antisociale gedragingen op de afdeling (ASOC) wellicht meer een stabiel persoonskenmerk (*trait*) is. Ook de groep rondom Barbaree (Langton & Barbaree 2001; Peacock e.a. 2001) vond overeenkomstige resultaten. Deze onderzoekers, en ook Hanson (Hanson & Harris 2000), splitsen sindsdien dynamische variabelen in fluctuerende dynamische variabelen zoals coöperatie en iets meer stabiele dynamische variabelen zoals sociale vaardig-

heden.

Tabel 2 toont dat een herhaalde beoordeling na een half jaar door dezelfde beoordelaar (dataselectie B in tabel 2) bevredigende test-hertestbetrouwbaarheden oplevert voor de verschillende subschalen. De correlaties dienen namelijk ook weer niet te hoog te zijn aangezien het gedrag betreft waarvan verondersteld wordt dat het aan verandering onderhevig is tijdens de behandeling. In dat licht lijken de significante correlaties na een half jaar tussen 2 verschillende beoordelaren (dataselectie A in tabel 2) ook nog redelijk te zijn.

Predictieve validiteit van de STM- en de Copinglijst

Voor de 556 tbs-patiënten waarvoor de STM-lijsten in het Meijers Instituut bij instroom waren gescoord, werd onderzocht wie gerecidiveerd was binnen 2, 3, 4, 5 en 6 jaar na de instroommeting. Het gaat bij de gepresenteerde recidivecijfers om geweldsdelicten, vermogensdelicten en zedendelicten. De recidivecijfers werden tot en met 1 december 2004 opgevraagd. De follow-upduur tussen instroommeting in het Meijers Instituut en deze peildatum bedroeg gemiddeld voor de 556 patiënten 6,8 jaar (bereik 5,2 tot 8,3 jaar). Alle patiënten waren dus minimaal 5 jaar 'gevolgd' en voor de recidivemeting over 6 jaar gold dat 80% reeds 6 of meer jaren 'gevolgd' had kunnen worden na de instroommeting.

Tot 1 december 2004 bleken van de 556 patiënten bij wie de STM-lijst was afgenomen, 65 patiënten opnieuw veroordeeld te zijn voor een delict. De

TABEL 3		De samenhang van de subschalen van de Observatielijst voor sociotherapeutisch medewerkers (STM-lijst) en de Copinglijst met recidive (veroordeling voor gewelds-, vermogens- of zedendelict) per tijdseenheid in jaren; weergegeven in p-waarden van t-testen en gemiddelde ruwe score				
		Na 2 jaar	Na 3 jaar	Na 4 jaar	Na 5 jaar	Na 6 jaar
Aantal recidivisten op totale n van 556 (STM-lijst)		n = 12	n = 22	n = 33	n = 46	n = 65
ASOC						
	p-waarde	0,068	0,033*	0,011*	0,005*	0,002*
	gemiddelde recidive vs. gemiddelde niet-recidive**	1,55 vs. 1,68	1,56 vs. 1,68	1,56 vs. 1,68	1,56 vs. 1,68	1,57 vs. 1,69
COOP						
	p-waarde	-	-	-	-	-
	gemiddelde recidive vs. gemiddelde niet-recidive**	-	-	-	-	-
REDZ						
	p-waarde	-	0,099	0,021*	0,002*	0,008*
	gemiddelde recidive vs. gemiddelde niet-recidive**	-	1,61 vs. 1,70	1,58 vs. 1,71	1,56 vs. 1,71	1,60 vs. 1,71
SOCV						
	p-waarde	-	-	0,033*	0,013*	0,060
	gemiddelde recidive vs. gemiddelde niet-recidive**	-	-	1,36 vs. 1,50	1,36 vs. 1,50	1,41 vs. 1,50
STM totaal						
	p-waarde	-	0,045*	0,018*	0,004*	0,017*
	gemiddelde recidive vs. gemiddelde niet-recidive**	-	-	1,41 vs. 1,52	1,41 vs. 1,52	1,44 vs. 1,52
Aantal recidivisten op het aantal van n = 376 waarvoor een copinglijst was ingevuld		n = 7	n = 14	n = 23	n = 31	n = 44
COPA						
	p-waarde	0,076***	-	-	-	-
	gemiddelde recidive vs. gemiddelde niet-recidive**	1,64 vs. 1,42	-	-	-	-
COPN						
	p-waarde	0,039*	-	0,030*	0,039*	0,023*
	gemiddelde recidive vs. gemiddelde niet-recidive**	0,82 vs. 1,22	-	1,00 vs. 1,24	1,04 vs. 1,24	1,05 vs. 1,24
COPP						
	p-waarde	-	-	-	-	-
	gemiddelde recidive vs. gemiddelde niet-recidive**	-	-	-	-	-
COPING totaal						
	p-waarde	-	-	-	0,054	0,071
	gemiddelde recidive vs. gemiddelde niet-recidive**	-	-	-	1,13 vs. 1,23	1,15 vs. 1,23

*p < 0,05 (eenzijdig getoetst). De waarden p 0,05 ≤ p < 0,10 zijn ook weergegeven en worden beschouwd als trends.

** lagere scores reflecteren gunstigere scores op de STM-lijst en de Copinglijst

*** let op: randsignificante trend in omgekeerde richting

ASOC = antisociaal gedrag op de afdeling; COOP = coöperatief gedrag op de afdeling; REDZ = redzaamheid; SOCV = sociale vaardigheden; COPP = positieve copingvaardigheden; COPN = negatieve copingmanieren; COPA = vermijdend gedrag

grootste groep betrof recidives in gewelddadige delicten (n = 45). In 15 gevallen ging het om veroordelingen voor vermogensdelicten (zonder geweld) en in 5 gevallen waren de tbs-patiënten binnen de periode van 6 jaar na instroom voor een zedendelict veroordeeld. Met t-testen werd onderzocht welke van de subschalen van de STM- en Copinglijst samenhangen met later optredende recidive (ge-

welds-, vermogens- en zedendelicten samen; zie tabel 3).

De resultaten in tabel 3 wijzen erop dat van de STM-lijst vooral het vertonen van antisociaal gedrag op de afdeling (ASOC) en de mate van zelfredzaamheid (REDZ) samenhangen met later optredende recidive. In iets mindere mate blijken de sociale vaardigheden (SOCV) ook geassocieerd te

TABEL 4 De samenhang van de subschalen van de Observatielijst voor sociotherapeutisch medewerkers (STM-lijst) en de Copinglijst met recidives (veroordelingen voor geweldsdelicten) per tijdseenheid in jaren; weergegeven door p-waarden van t-testen*

	Na 2 jaar	Na 3 jaar	Na 4 jaar	Na 5 jaar	Na 6 jaar
Aantal recidivisten op totale n van 556 (STM-lijst)	n = 9	n = 17	n = 23	n = 32	n = 45
ASOC					
p-waarde	0,079	0,038*	0,016*	0,018*	0,018*
gemiddelde recidive vs. gemiddelde niet-recidive**	1,53 vs. 1,68	1,55 vs. 1,68	1,54 vs. 1,68	1,57 vs. 1,68	1,58 vs. 1,68
COOP					
p-waarde	0,029*	0,019*	0,051	0,055	-
gemiddelde recidive vs. gemiddelde niet-recidive**	0,93 vs. 1,20	0,99 vs. 1,20	1,05 vs. 1,20	1,08 vs. 1,20	-
REDZ					
p-waarde	-	0,073	0,007*	0,001*	0,006*
gemiddelde recidive vs. gemiddelde niet-recidive**	-	1,58 vs. 1,70	1,53 vs. 1,71	1,52 vs. 1,71	1,57 vs. 1,71
SOCV					
p-waarde	0,097	0,050	0,020*	0,009*	0,041*
gemiddelde recidive vs. gemiddelde niet-recidive**	1,31 vs. 1,49	1,32 vs. 1,49	1,31 vs. 1,50	1,32 vs. 1,50	1,38 vs. 1,50
STM totaal					
p-waarde	0,033*	0,012*	0,004*	0,002*	0,011*
gemiddelde recidive vs. gemiddelde niet-recidive**	1,34 vs. 1,52	1,34 vs. 1,52	1,36 vs. 1,52	1,37 vs. 1,52	1,42 vs. 1,52
Aantal recidivisten op het aantal van n = 376 waarvoor een copinglijst was ingevuld	n = 6	n = 12	n = 16	n = 21	n = 30
COPA					
p-waarde	-	-	-	-	-
gemiddelde recidive vs. gemiddelde niet-recidive**	-	-	-	-	-
COPN					
p-waarde	0,076		0,067	0,034*	0,022*
gemiddelde recidive vs. gemiddelde niet-recidive**	0,86 vs. 1,23		1,00 vs. 1,23	0,99 vs. 1,23	1,01 vs. 1,24
COPP					
p-waarde	-	-	-	-	-
gemiddelde recidive vs. gemiddelde niet-recidive**	-	-	-	-	-
COPING totaal					
p-waarde	-	-	-	-	0,064
gemiddelde recidive vs. gemiddelde niet-recidive**	-	-	-	-	1,13 vs. 1,22

* $p < 0,05$ (eenzijdig getoetst). De waarden $p > 0,05 \leq p < 0,10$ zijn ook weergegeven en worden beschouwd als trends.

** lagere scores reflecteren gunstigere scores op de STM-lijst en de Copinglijst

ASOC = antisociaal gedrag op de afdeling; COOP = coöperatief gedrag op de afdeling; REDZ = redzaamheid; SOCV = sociale vaardigheden; COPP = positieve copingvaardigheden; COPN = negatieve copingmanieren; COPA = vermijdend gedrag

zijn met recidives, hoewel dit verband na 6 jaar niet meer significant is ($p = 0,06$). De gemiddelde score op alle schalen van de STM-lijst blijkt ook significant met recidive samen te hangen. Verminderingen van de significantieniveaus na 6 jaar suggereren dat de voorspellende waarde van de verzamelde (dynamische) variabelen met het verstrijken van de tijd afnemen.

Voor de Copinglijst waren de resultaten wat betreft het voorspellen van toekomstige recidive

duidelijk ongunstiger. Alleen negatieve copingvaardigheden (COPN) bleken op de langere termijn gerelateerd aan recidives. De gemiddelde score van de 3 copingsubschalen samen bleek dan ook uiterst zwak, en niet significant, samen te hangen met recidives (p -waarden van 0,054 en 0,071 na 5 en 6 jaar). Het gaat in deze berekeningen om recidive in gewelds-, vermogens- of zedendelicten. Apart werd geanalyseerd welke subschalen samenhangen met de meest voorkomende recidivecategorie, namelijk

gewelddelicten (zie tabel 4).

De resultaten lijken in grote lijnen hetzelfde te blijven wanneer uitsluitend veroordelingen voor gewelddelicten ($n = 45$ na 6 jaar) in de analyses werden betrokken. Vooral antisociaal gedrag op de afdeling (ASOC), zelfredzaamheid (REDZ) en sociale vaardigheden (SOCV) en de totaalscore van de STM bleken over de langere termijn met recidives in gewelddelicten samen te hangen. In deze analyse werd verder een ogenschijnlijk meer kortdurende relatie tussen coöperatief gedrag (COOP) en recidive gevonden, aangezien het verband tot 3 jaar na meting significant was en daarna minder sterk werd. Van de Copinglijst bleek wederom alleen de subschaal die negatieve copingstijlen meet, samen te hangen met recidives.

Om meer zicht te krijgen op de sterkte van de verbanden wat betreft het voorspellen van recidives, werden AUC-waarden berekend over de periode van 2 tot 6 jaar na meting voor de schalen die volgens de t-testen met recidive een samenhang vertoonden (ASOC, REDZ, SOCV, COOP, STM-totaalscore en COPN). Ook werd op basis van deze 5 schalen een gemiddelde score berekend om te onderzoeken in welke mate de combinatie van deze 5 maten recidives kon voorspellen. In tabel 5 zijn de AUC-waarden weergegeven wat betreft het voorspellen van gewelddelicten.

De AUC-waarden van de totale STM-lijst lagen gedurende de eerste 5 jaar na de afname tussen

0,71 en 0,65. Na verloop van tijd lijkt de voorspellende waarde van de dynamische STM-variabelen, in lijn met de t-testresultaten, af te nemen, waarbij de AUC-waarde na 6 jaar 0,59 is. Vooral de subschaal die coöperatief gedrag meet (COOP) blijkt in de beginmetingen wel gerelateerd te zijn aan recidives (AUC-waarden van 0,70 en 0,65), maar naarmate de periode waarover voorspeld moet worden langer wordt, neemt de voorspellende waarde van deze dynamische variabele snel af (AUC = 0,54 na 6 jaar). De subschaal die negatieve copingstijlen (COPN) beoogt te meten blijkt in beperkte mate (AUC-waarde na 5 jaar = 0,64) recidives te voorspellen, en de AUC-waarden lijken redelijk stabiel rond dit niveau naarmate de tijd verstrijkt. Het samenvoegen van de 4 STM-subschaalscores met de schaal die negatieve copingstijlen tracht te meten leverde in de eerste 5 jaar AUC-waarden van 0,73 tot 0,67 op. Hierna werd wederom een vermindering van de voorspellende waarde (AUC = 0,61) waargenomen wanneer de recidives uit het zesde jaar na instroommeting in de analyses werden toegevoegd.

DISCUSSIE

De significante verbanden met de subschalen van de STM-lijst wijzen erop dat zij voor risicotaxatie wezenlijke kenmerken zouden kunnen meten, en mogelijk ook, aangezien het dynamische variabelen betreft, zouden kunnen bijdragen

TABEL 5 AUC-waarden voor het voorspellen van gewelddelicten (veroordeling voor gewelddelicten) van de scores op de Observatielijst voor sociotherapeutisch medewerkers (STM-lijst) en van de negatieve-copingscores per tijdseenheid in jaren

	Na 2 jaar	Na 3 jaar	Na 4 jaar	Na 5 jaar	Na 6 jaar
Aantal recidivisten op totale n van 556 (STM-lijst)	$n = 9$	$n = 17$	$n = 23$	$n = 32$	$n = 45$
ASOC	0,64	0,65	0,64	0,62	0,59
COOP	0,70	0,64	0,60	0,58	0,54
REDZ	0,60	0,59	0,62	0,60	0,56
SOCV	0,66	0,62	0,63	0,62	0,56
Gemiddelde STM-score	0,71	0,66	0,66	0,65	0,59
COPN	0,69	0,58	0,62	0,64	0,62
Gemiddelde score ASOC, COOP, REDZ, SOCV en COPN	0,73	0,68	0,68	0,67	0,61

ASOC = antisociaal gedrag op de afdeling; COOP = coöperatief gedrag op de afdeling; REDZ = redzaamheid; SOCV = sociale vaardigheden; COPN = negatieve copingmanieren

aan de evaluatie van behandelprogressie. Echter, de gevonden AUC-waarden zijn niet van dien aard dat er individuele diagnostiek mee uitgevoerd kan worden. Om risicotaxatie niet alleen theoretisch, op groepsniveau, te gebruiken maar ook praktisch op individueel niveau, is een grotere nauwkeurigheid nodig dan de gevonden AUC's die tussen de 0,65 en 0,70 liggen. Ter vergelijking: in een landelijke steekproef onder gewelddadige ex-tbs-patiënten (Hildebrand e.a. 2005) werd recentelijk voor de *Historie Kliniek Toekomst schaal-30* (HKT-30)-totaalscore een AUC-waarde van 0,72 gevonden en voor de *Historical Clinical Risk Assessment-20* (HCR-20) een AUC-waarde van 0,67 voor het voorspellen van hernieuwd ernstig delictgedrag na opheffing van de tbs-maatregel. Er zijn echter ook beduidend hogere AUC-waarden gevonden. In het zojuist aangehaalde landelijke onderzoek van Hildebrand en collega's (2005) werd bijvoorbeeld voor factor 2 van de *Psychopathy Check List Revised* (PCL-R) - waarin juist veel zaken in kaart worden gebracht die zich in het verleden van de delictpleger hebben voorgedaan - een AUC-waarde van 0,77 gevonden. Verder werd in een eerder onderzoek van De Vogel (2005) onder plegers van seksuele delicten een AUC-waarde van 0,82 voor de HCR-20 gevonden.

Opgemerkt dient te worden dat het in het huidige onderzoek metingen betrof die aan het begin van de tbs-behandeling werden verricht. Het gaat in die zin om een prospectief uitgevoerd onderzoek, maar de vraag rijst of metingen meer aan het einde van tbs-behandeling niet tot een betere voorspelling zouden hebben geleid. De meeste eerdere risicotaxatieonderzoeken richtten zich op de periode na beëindiging van de tbs-maatregel, of in ieder geval op behandelperioden waarin sprake is van meer vrijheden en minder toezicht. Registratie van delictgedrag gedurende de eerste jaren van de tbs-behandeling zal leiden tot het vinden van lagere recidiveaantallen, zeker wat betreft bepaalde categorieën van delicten. Wat betreft gewelddadige incidenten/delicten die zich wel ten tijde van (intramurale) tbs-behandeling worden gepleegd, is het de vraag of dergelijke inci-

denten altijd tot aangifte dan wel een veroordeling leiden. Dit kan tot een minder scherp contrast tussen de onderzochte groepen hebben geleid wat betreft hun STM- en Copinglijstscores, aangezien patiënten die in gewelddadigheid recidiveerden bij het uitblijven van een veroordeling niet als zodanig herkend zouden zijn (en daarmee dus in de 'controlegroep' van niet-recidivisten terecht zijn gekomen).

Verder dient opgemerkt te worden dat in de meest gangbare risicotaxatie-instrumenten zoals de HCR-20 en de HKT-30, de dynamische variabelen aangevuld zijn met een substantieel aantal historische of statische (niet ten goede veranderbare) variabelen. Ook de STM-scores dienen ondersteund dan wel aangevuld te worden met bepalingen van de historische, onveranderbare risicotaxatie-items. Verder spelen bij het ontstaan van delicten diverse situationele factoren een rol. Naast gedrag dat observeerbaar is op de afdeling en persoonskenmerken die door een psychologisch onderzoek of interview te meten zijn, zullen een sociaal netwerk, huisvesting, een betaalde baan, middelenmisbruik enzovoorts een grote invloed hebben op het later wel of niet tot stand komen van een hernieuwd delict tijdens resocialisatie of na behandeling. Dat soort toekomstige risicofactoren zijn variabelen die in bijvoorbeeld de HCR-20 en de HKT-30 terug te vinden zijn, terwijl de huidige onderzochte schalen zich uitsluitend richten op geobserveerd gedrag op de afdeling.

Met al deze beperkingen en aanmerkingen in het achterhoofd, lijkt het toch gerechtvaardigd verder te onderzoeken of de intramuraal gemeten gedragskenmerken een zodanige samenhang met gewelddadige recidive laten zien dat de schaalscores wellicht een rol kunnen hebben als aanvullende informatie op de gangbare taxatiemethoden. Dit geldt dan vooral voor de subschalen van de STM-lijst. De tegenwoordig veelgebruikte taxatie-instrumenten zoals de HKT-30, HCR-20 en PCL-R dienen als sterk gestructureerde checklists op basis waarvan tot een oordeel over recidivegevaar wordt gekomen. Dit gebeurt echter doorgaans niet op basis van één somscore. De risicotaxatie-instru-

menten worden ingezet als gestandaardiseerde basis voor het schatten van recidivegevaar, waarbij de AUC-waarden de kracht van deze basis voor de verschillende instrumenten weergeven. Nader onderzocht zou kunnen worden of de gedragsobservatiescores verkregen via de STM-lijst statistisch gezien aanvullend kunnen zijn op scores die al met de gangbare taxatie-instrumenten worden verkregen.

Zoals eerder opgemerkt is een verschil met diverse eerdere recidiveonderzoeken dat de predictie van geweld in het huidige onderzoek niet gebaseerd was op dossieronderzoek of historische feiten maar volledig op gedragsobservaties die aan het begin van de tbs werden uitgevoerd. Een van de weinige andere prospectief uitgevoerde recidiveonderzoeken in Nederland, waarin ook het accent lag op onderzoek naar de predictieve validiteit van dynamische variabelen, is het grootschalige onderzoek van Philipse (2005). In dit onderzoek werd voor de onderzochte dynamische variabelen zoals gemeten met de *Clinical Inventory of Dynamic Reoffending Risk Indicators* (CIDRRI) geen voorspellende waarde gevonden. Het gebrek aan significanties in het huidige onderzoek met positieve copingvaardigheden (COPP) en vermijdende copingvaardigheden (COPA) sluit bij die bevindingen aan. Echter, in het huidige onderzoek bleek men via observatie van het gedrag op de afdeling op andere gedragskenmerken wel tot een zekere mate van nauwkeurigheid van voorspellen te kunnen komen. Het scoren van de STM-lijst en het scoren van negatieve copingstijlen zou daarmee nut kunnen hebben voor behandelplanning en behandel-evaluatie. Omdat de maatregel tbs met dwangverpleging gericht is op het veranderen dan wel reduceren van 'delictgevaar' is het van belang dat gedurende de behandeling door middel van herhaalde metingen veranderbare variabelen in kaart worden gebracht, waarvan bekend is dat ze met latere delictrecidive samenhangen. In theorie zijn sociale vaardigheden en zelfredzaamheid te behandelen gedurende het verblijf in een tbs-kliniek. De schalen coöperatie en antisociaal gedrag op de afdeling kunnen daarbij ook een indicatie geven

over het verloop van de behandeling. Deze kenmerken worden dimensionaal gescoord, dat wil zeggen mogelijk sensitief genoeg om veranderingen te kunnen registreren, terwijl het invullen van de desbetreffende schalen niet veel tijd in beslag neemt.

CONCLUSIE

De intraschaalbetrouwbaarheden uitgedrukt in Cronbachs alfa van de voor de tbs ontwikkelde STM- en Copinglijst zijn bevredigend. Daarnaast blijkt uit de herhaalde metingen uit De Kijvelanden dat de STM-lijst wat betreft test-hertestbetrouwbaarheid zich goed kan meten met andere gedragsobservatiematen wanneer deze in forensisch psychiatrische populaties ingezet worden (zoals de Nederlandse versie van de REHAB van Van der Gaag & Wilken (1994); Nijman e.a. 2002). Verder werd een aantal significante verbanden gevonden tussen een aantal dynamische (mogelijk onder invloed van behandeling veranderbare) variabelen en latere terugval in (gewelddadig) delictgedrag. Personen met de hoogste scores op de STM-lijst zijn dus later daadwerkelijk vaker gaan recidiveren in delictgedrag, en mogelijk dus dat behandeling die tot lagere STM-scores leidt, ook een vermindering van het recidivegevaar kan bewerkstelligen. De predictieve validiteit uitgedrukt in AUC-waarden van STM-totaalscores is echter niet van dien aard dat aan de hand van deze lijst individuele risicobeoordelingen kunnen worden uitgevoerd.

Verder werden in dit onderzoek aanwijzingen gevonden dat de voorspellende waarde van de significant samenhangende dynamische factoren na verloop van tijd vermindert. Voor de eerste 5 jaar na meting werden voor de totaalscore van de STM-lijst AUC-waarden gevonden van 0,71 tot 0,65, terwijl deze na 6 jaar 0,59 was. Het lijkt er op dat de onderzochte dynamische voorspellers een bepaalde houdbaarheidsdatum hebben. Vooral het vertonen van coöperatief gedrag (COOP) leek zijn voorspellende karakter snel te verliezen (na meer dan 3 jaar bleken de verbanden met recidives

niet meer significant te zijn). Verder bleken de totaalscores van de lijst die copingstijlen tracht te meten geen voorspellende waarde voor latere recidive te hebben, met een uitzondering voor de sub-schaal waarmee beoogd wordt negatieve copingstijlen in kaart te brengen. De STM-lijst lijkt echter een dynamische lijst te zijn die nuttig kan zijn in het kader van behandel-evaluatie.

LITERATUUR

- Brand, E.F.J.M. (2003). Forensische Profiellijsten (FP40). *De Psycholoog*, 38, 239-245.
- Brand, E.F.J.M. (2005). 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. 429-456). Deventer: Kluwer.
- Brand, E.F.J.M. (2006). Handleiding FP 40, versie 2006. Forensische profiel lijsten. Den Haag: Dienst Justitiële Inrichtingen.
- Brand, E.F.J.M., & van Emmerik, J.L. (2001). *Handboek Forensisch Psychiatrische Profielen. Handleiding FP40*. Den Haag: Ministerie van Justitie.
- Dernevik, M. (2000). Implementing risk assessment procedures in the clinical context: some experiences and some pitfalls. Presentation at symposium 'Assessing risk for violent behavior', 18 april 2000. Nijmegen.
- Gaag, M. van der, & Wilken, J. (1994). *REHAB. Een maat van psychiatrische invaliditeit*. Lisse: Zwets & Zeitinger.
- Hanson, R.K., & Harris, A. (2000). Where should we intervene? Dynamic predictors of sexual offense recidivism. *Criminal Justice and Behavior*, 27, 6-35.
- Hildebrand, M., Hesper, B., 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.
- 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, 114-124.
- Hilterman, E.L.B. (2002). Tijdens de oefening terug naar af. Een onderzoek naar de predictie van ernstige recidive door tbs-gestelden tijdens verlof. *Tijdschrift voor Criminologie*, 232 - 251.
- Langton, C., & Barbaree, H.E. (2001). Clarifying the role of moderating variables in risk assessment. Presentation at the 26th International Congress on Law and Mental Health, Montreal. International Academy of Law and Mental Health.
- Nijman, H., de Kruijk, C., & van Nieuwenhuizen, Ch. (2002). Gedragsveranderingen tijdens tbs-behandeling. *Tijdschrift voor Psychiatrie*, 44, 339-344.
- Peacock, E., Langton, C.M., & Barbaree, H.E. (2001). Sex offender risk assessment: integration of treatment performance with actuarial risk. Presentation at the 26th International Congress on Law and Mental Health, Montreal. International Academy of Law and Mental Health.
- Philipse, M. (2005). Predicting criminal recidivism. Empirical studies and clinical practice in forensic psychiatry. Nijmegen: Thesis Radboud Universiteit Nijmegen.
- Serin, R.C., & Kuriyuchuk, M. (1994). Social and cognitive processing deficits in violent offenders: implications for treatment. *International Journal of Law and Psychiatry*, 17, 431-441.
- 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. Amsterdam: Dutch University Press.

AUTEURS

E.F.J.M. BRAND is researchpsycholoog en werkzaam bij de Dienst Justitiële Inrichtingen, afdeling Informatieanalyse en Documentatie, Ministerie van Justitie; voorheen werkzaam in het Meijers Instituut, afdeling Monitoring en Research TBS.

H. NIJMAN is gezondheidswetenschapper en werkzaam als onderzoeker in De Kijvelanden te Poortugaal en als bijzonder hoogleraar forensische psychologie verbonden aan de Radboud Universiteit te Nijmegen.

Correspondentieadres: E.F.J.M. Brand, Dienst Justitiële Inrichtingen, afdeling Informatieanalyse en Documentatie, Schedeldoekshaven 101, 2511 EM Den Haag.
E-mail: e.brand@dji.minjus.nl.

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 8-1-2007.

SUMMARY

Risk assessment and treatment evaluation by means of two forensics observation instruments – E. Brand, H. Nijman –

BACKGROUND Under the Dutch TBS system (compulsory detention by government order), risk assessment and treatment evaluation are closely linked procedures.

AIM To test the psychometric properties of two observation instruments developed specifically for evaluating the treatment of persons compulsorily detained (tbs).

METHOD Two observation instruments for the use of social therapists and/or group leaders were tested on the basis of measurements obtained at the 'Meijers Instituut'(n=556) and at the 'De Kijvelanden'(n=100). The two instruments, one called the STM list (list for social therapists) and the other called the 'coping list', are questionnaires. The lists are designed to measure characteristics such as social skills, cooperation, daily living skills and coping skills in forensic psychiatric patients. These characteristics are believed to be important predictors of risk of future criminal behaviour.

RESULTS The internal consistency for the two questionnaires was found to be good. The results obtained at 'De Kijvelanden' showed that the STM questionnaire had adequate test-retest reliability. The STM total scores and subscale scores were found to be significantly associated with criminal recidivism. Anti-social behaviour on the ward, daily living skills and social skills, and to a lesser extent cooperation and negative coping styles, were found to be associated with future relapse into criminal behaviour.

CONCLUSION The questionnaires appear to be suitable for repeated assessing the effects of treatment on behaviour in forensic psychiatric settings and could possibly serve as useful supplements to other risk assessment methods that are currently available.

[TIJDSCHRIFT VOOR PSYCHIATRIE 49(2007)4, 221-232]

KEY WORDS forensic psychiatry, risk assessment, treatment evaluation