

Verschillen tussen onderzoekers en behandelaars in het inschatten van het risico van gewelddadig gedrag

Vivienne de Vogel · Corine de Ruiter

Abstract In this article we present the results of a study into the interrater reliability of a structured instrument for violence risk assessment. Two independent researchers and two clinicians (treatment supervisors and group leaders) coded the Dutch version of the HCR-20 for sixty patients in a forensic psychiatric hospital. We examined differences between researchers and clinicians in coding the HCR-20, and possible interference of clinicians' feelings towards their patients. The interrater reliability of the HCR-20 was high. Group leaders scored significantly lower on the HCR-20 than researchers did. There were no significant differences between the mean HCR-20 scores of treatment supervisors and researchers. There was, however, a significant difference in the interpretation of the scores: treatment supervisors gave more 'low risk' judgments than researchers. Furthermore, feelings of clinicians towards their patients influenced their risk assessments. Feelings of being controlled and manipulated by the patient resulted in higher HCR-20 scores, whereas positive feelings (helpful, happy, relaxed) predicted lower risk ratings.

Samenvatting Verschillen onderzoekers en behandelaars in hun inschatting van het risico van gewelddadige recidive bij forensisch psychiatrische patiënten? In dit artikel

proberen we deze vraag te beantwoorden. De HCR-20, een risicotaxatie-instrument voor gewelddadig gedrag, werd door behandelaars (groepsleiders en hoofden behandeling) en onderzoekers gecodeerd voor zestig patiënten uit de Dr. Henri van der Hoeven Kliniek, forensisch psychiatrisch instituut te Utrecht. Dit prospectieve onderzoek had drie doelen: het in kaart brengen van verschillen tussen behandelaars en onderzoekers in het coderen van de HCR-20; het onderzoeken of gevoelens van behandelaars ten opzichte van hun patiënten invloed hebben op de risicotaxatie; en het vaststellen van de interbeoordelaarsbetrouwbaarheid van de HCR-20. De groepsleiders gaven gemiddeld significant lagere HCR-20-scores dan de onderzoekers. Tussen de hoofden behandeling en de onderzoekers werden geen significante verschillen gevonden in gemiddelde HCR-20-scores, maar wel in de interpretatie hiervan: de hoofden behandeling kwamen vaker tot de conclusie 'laag risico' dan de onderzoekers. Als behandelaars het gevoel hadden te worden overheerst of gemanipuleerd door de patiënt, resulteerde dit in hogere HCR-20-scores. Positieve gevoelens (willen helpen, blijdschap, ontspanning) hingen daarentegen samen met lage HCR-20-scores. De interbeoordelaarsbetrouwbaarheid van de HCR-20 was hoog.

Vivienne de Vogel, en, (✉)

DRS. V. DE VOGEL is psycholoog en verricht promotieonderzoek naar risicotaxatie van (seksueel) gewelddadig gedrag in de Dr. Henri van der Hoeven Kliniek te Utrecht.

Correspondentieadres: Vivienne de Vogel, Dr. Henri van der Hoeven Kliniek, afdeling onderzoek, Postbus 174, 3500 AD Utrecht. E-mail: vdevogel@hoevenstichting.nl. De auteurs bedanken alle behandelaars en onderzoekers die aan dit onderzoek hebben meegewerkt. Een speciaal dankwoord gaat uit naar Cécile Vandeputte-Van de Vijver, die samen met de eerste auteur de trainingen verzorgde en tevens als onderzoeker fungeerde.

Inleiding

Het inschatten van het risico van (seksueel) gewelddadig gedrag is een belangrijke taak voor deskundigen die werkzaam zijn in forensisch psychiatrische instellingen. Een accurate risicotaxatie voorafgaand aan een (proef) verlop of bij het opstellen van een verlengingsadvies voor de rechtbank bij terbeschikkinggestelden is essentieel en kan gewelddadige recidive helpen voorkomen (Douglas

& Webster, 1999a). In de meeste forensisch psychiatrische instellingen werd tot nu toe overwegend gebruik gemaakt van de zogenaamde klinische risicotaxatie, dat wil zeggen, van de kennis en ervaring van klinici. Maar uit onderzoek is gebleken dat aan deze inschatting nadelen zijn verbonden, zoals de vatbaarheid van de beoordelaars voor cognitieve vertekeningen en de lage betrouwbaarheid en validiteit van deze methode (zie voor een bespreking De Ruiters, 1999 ; Quinsey, Harris, Rice, & Cormier, 1998, p. 55-72). Gebruik van gestandaardiseerde risicotaxatie-instrumenten wordt sterk aanbevolen om de betrouwbaarheid en validiteit van de risicotaxatie te vergroten (Borum, 1996 ; Douglas, Cox, & Webster, 1999).

Een van de veelbelovende risicotaxatie-instrumenten voor gewelddadig gedrag is de *Historical, Clinical, Risk Management-20* (HCR-20; Webster, Douglas, Eaves, & Hart, 1997 ; Nederlandse geautoriseerde vertaling: Philipse, De Ruiters, Hildebrand, & Bouman, 2000). Dit instrument bestaat uit twintig items die betrekking hebben op risicofactoren voor gewelddadig gedrag in het verleden, het heden en de toekomst. Onderzoek in psychiatrische en justitiële instellingen in verschillende landen toonde aan dat de HCR-20 een goede interbeoordelaarsbetrouwbaarheid en predictieve validiteit heeft (Belfrage, 1998 ; Belfrage, Fransson, & Strand, 2000 ; Douglas, Ogloff, Nicholls, & Grant, 1999 ; Strand, Belfrage, Fransson, & Levander, 1999) (Zie voor een meer gedetailleerde bespreking van studies met de HCR-20: De Vogel & Hildebrand, 2001 .) Zo vonden Douglas et al. (1999) dat in een groep van 193 onvrijwillig opgenomen psychiatrische patiënten de HCR-20-scores een sterk verband hadden met verschillende vormen van gewelddadig gedrag in de maatschappij. Bovendien toonden zij aan dat met de HCR-20 verandering in het risico van gewelddadig gedrag vóór en na klinische interventie gemeten kan worden (Belfrage & Douglas, 2002). Een belangrijke beperking van veel onderzoeken naar de HCR-20 is hun retrospectieve opzet. In retrospectieve onderzoeken wordt het risico ‘achteraf’ ingeschat: de HCR-20 wordt alleen op basis van informatie uit het verleden gecodeerd en vervolgens gerelateerd aan recidivegevens. In prospectieve onderzoeken wordt het risico daarentegen ‘vooruit’ ingeschat. Tot nu toe werden slechts enkele prospectieve onderzoeken uitgevoerd (bijvoorbeeld Belfrage et al., 2000 ; Douglas et al., 1999), maar nog geheel niet in Nederland.

Bij bestudering van de literatuur valt op dat in vrijwel alle onderzoeken de HCR-20 werd gecodeerd door onafhankelijke onderzoekers en niet door diagnostici of behandelaars uit de klinische praktijk. Deze onderzoekers waren meestal niet bekend met de patiënt en codeerden de HCR-20 op basis van dossierinformatie voor onderzoeksdoeleinden. Maar voor het gebruik van de

HCR-20 voor klinische doelen, namelijk het inschatten van het risico van gewelddadig gedrag en besluitvorming omtrent verlop, wordt codering door gekwalificeerde en ervaren klinici vereist (Philipse et al., 2000, p. 30). In de dagelijkse praktijk zijn het bovendien doorgaans de directe behandelaars die verantwoordelijk zijn voor verlopbeslissingen (Dernevik, Falkheim, Holmqvist, & Sandell, 2001). Nu is discussie gaande over de objectiviteit van klinici, met name van klinici die direct betrokken zijn bij de behandeling van de patiënt (Dernevik et al., 2001 ; Litwack & Schlesinger, 1999). Zijn behandelaars die de patiënt goed kennen en veel geïnvesteerd hebben in diens behandeling, in staat hun persoonlijke gevoelens buiten beschouwing te laten wanneer zij het recidiverisico inschatten? Verschillende auteurs zijn van mening dat dergelijke gevoelens wel degelijk meespelen. Ze leiden tot subjectiviteit en maken daardoor de rollen van forensisch beoordelaar en behandelaar onverenigbaar. Zij adviseren om forensische evaluaties, zoals risicotaxaties te laten uitvoeren door een beoordelaar die niet direct bij de behandeling is betrokken (Ackerman, 1999 ; Greenberg & Shuman, 1997 ; Litwack & Schlesinger, 1999). Anderzijds kan worden beargumenteerd dat de behandelaar de meeste informatie over de patiënt heeft en dat gevoelens van tegenoverdracht minder ruimte krijgen door een gestructureerde wijze van risico inschatten, zoals met de HCR-20. Ook kan men de vraag stellen of deze gevoelens werkelijk irrelevant of irrationeel zijn en een versturende invloed hebben op de risicotaxatie, of dat ze juist bijdragen aan een accurate risicotaxatie. Dernevik et al. (2001) onderzochten de relatie tussen gevoelens van behandelaars, gemeten met de *Feeling Word Checklist* (FWC; Whyte, Constantopoulos, & Bevans, 1982), en hun coderingen op de HCR-20 in een onderzoek in een forensisch psychiatrische instelling in Zweden. Veertig ervaren verpleegkundigen, die een eendaagse training in risicotaxatie hadden gevolgd, codeerden ieder de HCR-20 voor één van in totaal acht patiënten. Iedere patiënt werd dus door vijf verpleegkundigen beoordeeld. Het risico van deze acht patiënten werd tevens beoordeeld door onafhankelijke codeurs met een ruime ervaring in het coderen van de HCR-20, *experts* genoemd. De verpleegkundigen bleken significant hogere HCR-20-totaalscores te geven dan de experts. In dit onderzoek werd een sterk verband gevonden tussen de scores op de FWC en de totaalscore van de HCR-20 (R (multipel correlatie) = .66, R^2 (percentage verklaarde variantie) = .43). Recentelijk bleek dat na een follow-up tijd van twee jaar de verpleegkundigen institutioneel geweld goed hadden voorspeld, terwijl de voorspellende waarde voor gewelddadige recidive laag was. De experts waren juist goed in het voorspellen van gewelddadige recidive na ontslag (Dernevik & Douglas, 2002).

In dit artikel bespreken we een deelstudie van een in januari 2001 gestart prospectief onderzoek naar risicotaxatie van seksueel gewelddadig gedrag. Voor zestig patiënten uit de Dr. Henri van der Hoeven Kliniek werd de HCR-20 gecodeerd door onafhankelijke onderzoekers en behandelaars en deze coderingen werden met elkaar vergeleken. Het doel van deze studie was, naast het bepalen van de interbeoordelaarsbetrouwbaarheid van de HCR-20, inzicht te verkrijgen in verschillen tussen behandelaars en onderzoekers in het coderen van de HCR-20 en de invloed van gevoelens van behandelaars ten opzichte van hun patiënten (gemeten met de FWC) op de risicotaxatie.

Methodie

Instelling

De Dr. Henri van der Hoeven Kliniek is een forensisch psychiatrisch instituut met 114 intramurale en 30 transmurale plaatsen. De meeste patiënten zijn opgenomen in het kader van de strafrechtelijke maatregel Terbeschikkingstelling (TBS). De TBS is door de rechter uitgesproken op grond van de ernst van de gepleegde delicten, de mate van ontoerekeningsvatbaarheid (op basis van ernstige psychische toestands- of persoonlijkheidspathologie) en het als hoog ingeschatte recidiverisico.

Sinds de jaren vijftig biedt de kliniek een breed opgezette behandeling, waarin sociotherapeutische uitgangspunten en werkwijzen een centrale rol spelen en waarbij de patiënten in de gelegenheid worden gesteld deel te nemen aan een gevarieerd ontwikkelingsprogramma op het gebied van arbeidstraining, onderwijs, sport, creatieve vakken en psychotherapie (Dankers & Van der Linden, 1995). Tevens is in de behandeling aandacht voor de ontwikkeling van een steunend netwerk. De focus van de behandeling is op direct delictgerelateerde risicofactoren op basis van een grondige analyse van de aanloop naar het delict, het zogenaamde delictscenario. Het uitgangspunt van deze vorm van behandeling, *no cure but control*, komt voort uit het terugvalpreventie-model (Laws, Hudson, & Ward, 2000). Dit wil zeggen dat de pretentie de patiënt te genezen bij de meesten een te hoge verwachting is en dat betere hantering van risicofactoren recidive kan beperken.

Een belangrijke fase in de behandeling is de transmurale fase, waarin de patiënt buiten de muren van de kliniek woont, maar nog valt onder de verantwoordelijkheid van de kliniek. In deze resocialisatiefase kan de patiënt oefenen met zijn of haar nieuwe inzichten en gedrag, zoals die zijn opgenomen in het terugvalpreventieplan. De patiënt wordt gedurende de transmurale fase begeleid door een behandelteam van de kliniek dat erop

toeziet of het de patiënt daadwerkelijk lukt om zich in de maatschappij staande te houden ondanks zijn of haar kwetsbaarheden.

Onderzoeksgroep

De onderzoeksgroep bestond uit 53 mannen en 7 vrouwen. De gemiddelde leeftijd was 36.6 jaar ($SD = 8.0$). Zeventig procent was voor het TBS-delict al eerder veroordeeld, met een gemiddelde van 4.9 ($SD = 13.0$) eerdere veroordelingen. De TBS-delicten waren: 57 procent (poging tot) moord of doodslag, 20 procent seksuele delicten, 17 procent overige geweldsdelicten en 7 procent brandstichting. Bij 8 procent van de patiënten werd een DSM-IV as-I-stoornis (APA, 1994) vastgesteld, 52 procent had een as-II-stoornis en bij 35 procent was er comorbiditeit van as-I- en as-II-stoornissen. Misbruik van middelen werd niet onder as-I-stoornissen meegerekend. De meerderheid van de groep had problemen met middelengebruik: 7 procent alcohol, 13 procent drugs en 45 procent meerdere middelen.

Instrumenten

Historical, Clinical, Risk Management-20 (HCR-20)

De HCR-20 is een klinische beoordelingsschaal (checklist) voor het inschatten van het risico op toekomstig gewelddadig gedrag bij volwassenen met een gewelddadige voorgeschiedenis of een psychische of persoonlijkheidsstoornis (DSM-as-I of -II). De checklist is gebaseerd op een uitgebreid literatuuroverzicht en de deskundigheid van ervaren klinici. De HCR-20 bestaat uit twintig items, verdeeld over drie subschalen. De Historische schaal, Klinische schaal en Risicohanteringsschaal hebben betrekking op risicofactoren in het verleden, het heden en de toekomst (tabel 1). De Historische items zijn statische, onveranderbare factoren, de Klinische items en Risicohanteringsitems zijn dynamische, door klinische interventie te beïnvloeden risicofactoren.

De items worden gecodeerd op een driepuntsschaal: '0' betekent dat het item volgens de beschikbare informatie niet aanwezig is, '1' dat het item wellicht of gedeeltelijk aanwezig is en '2' dat het item duidelijk aanwezig is. Bij het scoren van de Risicohanteringsitems moet men rekening houden met de context waarvoor de risicotaxatie geldt: een inrichting of de samenleving. De HCR-20 biedt ook ruimte voor 'Overige overwegingen', dat wil zeggen, casusspecifieke risicofactoren die niet onder de definitie van een van de twintig items vallen. De items dienen door een forensisch geschoold clinicus gecodeerd te worden op basis van alle beschikbare informatie, bij voorkeur afkomstig uit verschillende bronnen en met

Tabel 1 Items van de HCR-20 (overgenomen uit Philipse, De Ruiter, Hildebrand, & Bouman, 2000).

Historische items	Klinische items
1 Eerder geweld	11 Gebrek aan zelfinzicht
2 Jonge leeftijd bij eerste gewelddadigincident	12 Negatieve opvattingen
3 Instabiliteit van relaties	13 Actieve psychotische symptomen
4 Problemen in het arbeidsverleden	14 Impulsiviteit
5 Problemen met middelengebruik	15 Reageert niet op behandeling
6 Psychische stoornis	
7 Psychopathie	Risicohanteringsitems
8 Problemen in de kindertijd	16 Geringe kans dat plannen zullen slagen
9 Persoonlijkheidsstoornissen	17 Blootstelling aan destabiliserendefactoren
10 Eerdere onttrekking aan toezicht	18 Geringe beschikbaarheid vanpersoonlijke steun
	19 Werkt niet mee aan behandelmogelijkheden
	20 Hoog niveau van ervaren stress

verschillende methoden verzameld, zoals psychologische rapportages, interviews met derden, processen-verbaal en observaties. Bovendien moet de *Psychopathie Checklist-Revised* (PCL-R; Hare, 1991 ; Nederlandse bewerking: Vertommen, Verheul, De Ruiter, & Hildebrand, 2002) worden gecodeerd, omdat de diagnose Psychopathie, zoals gesteld op basis van de PCL-R, als item 7 is opgenomen in de HCR-20.

Het eindoordeel van het risico op gewelddadig gedrag wordt door de beoordelaar aangegeven in termen van waarschijnlijkheid (laag, matig of hoog) en geldt voor een specifieke periode (bijvoorbeeld gedurende een bepaalde behandelfase). Het eindoordeel is niet een optelsom van de coderingen van de items, maar hangt mede af van combinaties van risicofactoren. Het is mogelijk dat de aanwezigheid van één risicofactor voldoende is om te concluderen dat er een hoog risico is, bijvoorbeeld bij actieve psychotische symptomen (zoals auditieve hallucinaties die ertoe aanzetten iemand te vermoorden), of als de score op de PCL-R hoog is (meer dan dertig). Op het uitgebreide codeerblad (Hart, De Ruiter, Hildebrand, & De Vogel, 2001) worden inhoudelijk relevante vragen gesteld over het risico van gewelddadig gedrag, bijvoorbeeld: ‘Wat zal de aard en ernst van het geweld zijn?’, ‘Wanneer en hoe vaak zal het zich voordoen?’, ‘Wie is potentieel slachtoffer?’ en ‘Welke gebeurtenissen kunnen het risico vergroten of juist verkleinen?’ Deze vragen kunnen behulpzaam zijn bij het komen tot het eindoordeel. In feite is het eindoordeel een *gestructureerd klinisch oordeel* dat is gebaseerd op het proces van het coderen van het instrument, waarbij alle verzamelde informatie over de patiënt door de codeur wordt geïntegreerd in het eindoordeel.

Feeling Word Checklist (FWC)

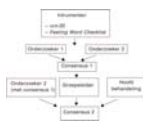
De FWC werd afgenomen om gevoelens van behandelaars voor hun patiënt in kaart te brengen. Dit instrument

werd ontwikkeld door Whyte et al. (1982) en door Holmqvist en Armelius (1994) en Holmqvist en Fogelstam (1996) bewerkt. De lijst bestaat uit dertig items (tabel 2), die verschillende gevoelens jegens de patiënt beschrijven. De items moeten worden gecodeerd op een vierpuntsschaal (0 = helemaal niet, 1 = enigszins, 2 = behoorlijk, 3 = heel erg). De dertig items zijn onderverdeeld in acht subschalen die in een circumplexmodel met vier dimensies worden weergegeven: *Positief* versus *Negatief*; *Nabij* versus *Afstandelijk*; *Accepterend* versus *Afwijzend*; en *Autonoom* versus *Overheerst*.

Tabel 2 Items van de Feeling Word Checklist (overgenomen uit Whyte, Constantopoulos, & Bevans, 1982; Nederlandse vertaling door De Vogel en De Ruiter, 2001)*.

Wanneer ik denk aan (naam patiënt) voel ik me:			
1	Behulpzaam	16	Verrast
2	Blij	17	Moe
3	Boos	18	Bedreigd
4	Enthousiast	19	Ontvankelijk
5	Angstig	20	Objectief
6	Sterk	21	Overweldigd
7	Gemanipuleerd	22	Verveeld
8	Ontspannen	23	Moederlijk
9	Voorzichtig	24	Verward
10	Teleurgesteld	25	Beschaamd
11	Onverschillig	26	Geïnteresseerd
12	Hartelijk / warm	27	Afstandelijk
13	Wantrouwend	28	Verdrietig
14	Sympathie voelend	29	Inadequaat
15	Afkerig	30	Gefrustreerd

* Items horende bij de subschalen: Positief: 1, 2, 8; Negatief: 3, 10, 15, 18, 24; Nabij: 12, 16, 21, 23; Afstandelijk: 9, 11, 27; Accepterend: 4, 14, 19, 26; Afwijzend: 13, 17, 22, 30; Autonoom: 6, 20; Overheerst: 5, 7, 25, 28, 29.

Figuur 1 Onderzoeksopzet

interbeoordelaarsbetrouwbaarheid
HCR-20.

Procedure

Figuur 1 geeft de opzet van de interbeoordelaarsbetrouwbaarheidstudie met de HCR-20 weer. Eerst werden alle twaalf behandelteams, bestaande uit een hoofd behandeling en een aantal groepsleiders, gedurende één dag getraind in het coderen van de HCR-20. Deze training voorzag in een overzicht van relevante literatuur over risicotaxatie. Aan de hand van dossierinformatie en videomateriaal ging men oefenen met cases uit de praktijk. De training werd verzorgd door een senior klinisch psycholoog en de eerste auteur.

In de behandeling zijn enkele momenten te onderscheiden waarin de vrijheden van de patiënt worden vergroot. Het is belangrijk om het risico van gewelddadig gedrag in te schatten als patiënten voor het eerst onbegeleid naar buiten gaan en op het punt staan te beginnen met de transmurale fase. Vanaf januari 2001 werd voor alle patiënten die zich in een van beide bovenstaande situaties bevonden of die reeds in de transmurale fase verkeerden, de HCR-20 gecodeerd. Daarnaast werd de HCR-20 gecodeerd voor patiënten die pas waren opgenomen, om het risico van gewelddadig gedrag binnen de kliniek te kunnen inschatten. Twee onderzoekers, een groepsleider en het hoofd behandeling, codeerden onafhankelijk van elkaar de HCR-20. De twee onderzoekers kwamen samen tot consensus, waarna een van de onderzoekers met deze consensus een volgende consensusbespreking voerde met de groepsleider en het hoofd behandeling. De laatste twee codeerden voor aanvang van de consensusbespreking en onafhankelijk van elkaar de FWC. De consensusbesprekingen namen gemiddeld één uur in beslag en werden door de betrokkenen als zinvol ervaren. Vooral het scherpen van de geest over wat risicofactoren precies zijn en wat ze betekenen voor de behandeling vond men waardevol. Overigens werden de resultaten van de risicotaxatie niet aan de patiënten meegedeeld, omdat we de HCR-20 op dit moment nog beschouwen als onderzoeksinstrument en er nog onvoldoende bekend is over de psychometrische eigenschappen (in het bijzonder de voorspellende waarde) van de Nederlandse HCR-20.

Codeurs

Aan het huidige onderzoek deden 44 codeurs mee: vijf onderzoekers, zeven hoofden behandeling en 32

groepsleiders. De vijf onderzoekers waren allen psycholoog en getraind in het coderen van de HCR-20 en de PCL-R en in het uitvoeren van psychodiagnostiek. De zeven hoofden behandeling waren ervaren klinici – (klinisch) psycholoog of psychotherapeut. De achtergrond van de groepsleiders liep sterk uiteen, de meesten hadden een HBO- of universitaire opleiding gevolgd.

Data-analyse

De interbeoordelaarsbetrouwbaarheid van de HCR-20 werd onderzocht met de *Intraklasse Correlatie Coëfficiënt* (ICC) (McGraw & Wong, 1996). Om verschillen te toetsen tussen de HCR-20-scores van onderzoekers, hoofden behandeling en groepsleiders werd de *F*-toets gebruikt, voor verschillen in de eindoordeelen werd de chi-kwadraat-toets gebruikt. Om de samenhang te onderzoeken tussen de FWC-subschalen en de HCR-20-totaalscore en eindoordeel werden Pearson's product-momentcorrelaties berekend. De invloed van gevoelens van behandelaars voor hun patiënt op de risicotaxaties (HCR-20-totaalscore en eindoordeel) werd onderzocht met een multiële regressieanalyse.

Resultaten

Risicobeoordelingen

In tabel 3 zijn de gemiddelde HCR-20-scores te lezen van de verschillende groepen codeurs en de scores zoals overeengekomen tijdens de consensusbesprekingen tussen onderzoekers en behandelaars. Wat betreft de eindoordeelen werd het risico van 17 patiënten ingeschat als laag (gemiddelde totaalscore = 20.8, range = 10-28), van 24 als matig (gemiddelde totaalscore = 24.7, range = 16-34), en van 19 als hoog (gemiddelde totaalscore = 32.4, range = 23-37). De verschillen in gemiddelde totaalscores van de laag-, matig- en hoog-risicocategorieën waren significant ($F = 31.6, p < .01$). Opvallend was de overlap in ranges van totaalscores voor de eindoordeelen laag, matig en hoog. Dat betekent dat er geen gebruik kan worden gemaakt van vaste *cut off* HCR-20-totaalscores voor deze eindoordeelen.

Verschillen tussen codeurs in risicobeoordelingen

In tabel 3 is te lezen dat de groepsleiders, vergeleken met de onderzoekers, lager scoorden op de Historische items, Risicohanteringsitems, Overige overwegingen en totaalscore van de HCR-20. Er werden geen significante verschillen gevonden in de gemiddelde scores van de onderzoekers en de hoofden behandeling, behalve voor

Tabel 3 Risicobeoordelingen voor zestig patiënten.

	Gemiddelde score			totaal	overig	Eindoordeel		
	H	C	R			laag	matig	hoog
Onderzoekers	14.5 ^a	5.2	6.2 ^c	25.9 ^e	2.4 ^g	8 ^j	28	24
Hoofden behandeling	13.8	5.0	5.8	24.6	0.8 ^h	18 ^k	21	21
Groepsleiders	13.2 ^b	4.5	5.0 ^d	22.7 ^f	0.5 ⁱ	12	28	20
Consensus	14.6	5.3	6.1	26.1	3.3	17	24	19

^a > ^b, $p < .05$. ^c > ^d, $p < .01$. ^e > ^f, $p < .01$. ^g > ^h > ⁱ, $p < .05$. ^k > ^j, $p < .05$.

H = Historische items; C = Klinische items; R = Risicohanteringsitems; Overig = Overige overwegingen.

de Overige overwegingen. Maar in de eindoordelen bestond wel een significant verschil: de hoofden behandeling kwamen vaker tot de conclusie 'laag risico' dan de onderzoekers.

Wanneer de codeurs na de consensusbespreking gevraagd werd hoeveel tijd zij aan de risicotaxatie hadden besteed en waar zij hun taxatie op baseerden, bleken grote verschillen te bestaan. Onderzoekers besteedden gemiddeld 120 minuten aan het coderen van de HCR-20, groepsleiders 30 minuten en hoofden behandeling 15 minuten. Onderzoekers baseerden zich met name op dossierinformatie, terwijl groepsleiders en hoofden behandeling voornamelijk afgingen op persoonlijke ervaringen met een patiënt.

Overige overwegingen

Overige overwegingen die regelmatig werden genoemd, zijn: sadistische fantasieën, schijnaanpassing/ongrijpbaarheid, sociale isolatie, financiële problemen en gebrek aan toekomstperspectief.

Interbeoordelaarsbetrouwbaarheid

De interbeoordelaarsbetrouwbaarheid van de HCR-20 was hoog (tabel 4). Vooral voor de Historische items werd een zeer hoge mate van overeenstemming gevonden. Voor twee Historische items werd echter een lage betrouwbaarheid gevonden, te weten *Eerder geweld* en *Problemen in de kindertijd*. De interbeoordelaarsbetrouwbaarheid voor de Klinische items was hoog. Voor het Klinische item *Impulsiviteit* werd een lage overeenstemming gevonden. Voor de

Risicohanteringsitems was de interbeoordelaarsbetrouwbaarheid matig. De items *Blootstelling aan destabiliserende factoren*, *Werkt niet mee aan behandelmogelijkheden* en *Hoog niveau van ervaren stress* hadden een lage betrouwbaarheid.

De groepen codeurs verschilden in de mate van overeenstemming. Tussen de onderzoekers was de interbeoordelaarsbetrouwbaarheid over het geheel genomen zeer hoog. Voor zowel de onderzoekers en groepsleiders als de onderzoekers en hoofden behandeling werd een hoge overeenstemming gevonden, behalve voor de Risicohanteringsitems. Maar tussen de hoofden behandeling en groepsleiders werd een matige betrouwbaarheid gevonden voor de Klinische en de Risicohanteringsitems.

Invloed van gevoelens voor de patiënt op de risicotaxatie

Om de samenhang te onderzoeken tussen de FWC-subschalen, ingevuld door de behandelaars (hoofden behandeling en groepsleiders), en hun HCR-20-totaalscores en eindoordelen, werden eerst Pearson's product-momentcorrelaties berekend. De FWC-subschalen *Negatief*, *Afstandelijk*, *Afwijzend* en *Overheerst* hingen samen met hoge HCR-20-totaalscores en eindoordelen (tabel 5). De subschalen *Positief* en *Accepterend* hingen daarentegen samen met lage HCR-20-totaalscores en eindoordelen. Vervolgens werd een stapsgewijze multiële regressieanalyse uitgevoerd om de samenhang nader te onderzoeken. Met een Pearson's correlatie wordt de statistische samenhang tussen twee variabelen aangetoond. Bij multiële regressie wordt onderzocht of met één variabele (de zogenaamde predictorvariabele, in dit geval de FWC-subschalen) een

Tabel 4 Intraklasse Correlatie Coëfficiënten (ICCS)*.

	H	C	R	Totaal-score	Eind-oordeel
Onderzoeker 1 en 2	.85	.75	.70	.85	.76
Onderzoekers en hoofden behandeling	.83	.67	.56	.79	.65
Onderzoekers en groepsleiders	.78	.70	.58	.82	.68
Hoofden behandeling en groepsleiders	.86	.54	.58	.77	.63
Onderzoekers, hoofden behandeling en groepsleiders	.82	.64	.57	.79	.65

*Alle iccs zijn groter dan 0 ($p < .001$). Kritische waarden die aangehouden werden bij de iccs (*single measure*) zijn: $icc \geq .75$ = zeer hoog; $.60 < icc < .75$ = hoog; $.40 \geq icc < .60$ = matig; $icc < .40$ = laag (Fleiss, 1986). Onderzoekers = consensus van onderzoeker 1 en 2. H = Historische items. C = Klinische items. R = Risicohanteringsitems.

Tabel 5 Pearson's product-momentcorrelaties FWC-subschalen en HCR-20-totaalscore en eindoordeel.

	Positief	Negatief	Nabij	Afstandelijk	Accepterend	Afwijzend	Autonoom	Overheerst	HCR-20-totaalscore
Positief	-								
Negatief	-.45*	-							
Nabij	.19*	.07	-						
Afstandelijk	-.48**	.50**	-.10	-					
Accepterend	.72**	-.49**	.36**	-.42**	-				
Afwijzend	-.35**	.69**	.07	.55**	-.44**	-			
Autonoom	.24**	.05	.21	-.03	.11	-.02	-		
Overheerst	-.40**	.63**	.27**	.42**	-.24**	.56**	-.06	-	
HCR-20-totaalscore	-.28**	.38**	.12	.20*	-.19*	.34**	.04	.46**	-
HCR-20-eindoordeel	-.34**	.33**	.19*	.40**	-.23*	.34**	-.12	.37**	.56**

* = $p < .05$ (tweezijdig getoetst). ** $p < .01$ (tweezijdig getoetst).

andere variabele (HCR-20-totaalscore of eindoordeel) kan worden voorspeld. In dit onderzoek werd een stapsgewijze multiële regressie uitgevoerd, waarbij alle predictorvariabelen stap voor stap werden getoetst op hun voorspellende waarde. Uiteindelijk blijven alleen de variabelen met een significante predictie over.

Uit deze analyse bleek dat gevoelens van overheerst en gemanipuleerd worden door de patiënt, verwoord in de subschaal *Overheerst*, hoge HCR-20-totaalscores voorspelden; 21 procent van de variantie in de HCR-20-totaalscores werd door deze subschaal verklaard (tabel 6). Geen van de andere FWC-subschalen bleek in staat de HCR-20-totaalscore te voorspellen. Voor het HCR-20-eindoordeel hadden drie subschalen een significante voorspellende waarde. *Afstandelijk* en *Nabij* voorspelden hogere eindoordeelen, terwijl de subschaal *Positief* een lager eindoordeel voorspelde. Drieëntwintig procent van de variantie in de HCR-20-eindoordeelen werd verklaard door deze drie subschalen tezamen (tabel 7).

Beschouwing

In dit onderzoek werd aangetoond dat de HCR-20 betrouwbaar te coderen is, dat verschillen bestaan tussen onderzoekers en behandelaars in het coderen van de HCR-20 en dat gevoelens van behandelaars ten opzichte van hun patiënten van invloed zijn op de HCR-20-coderingen.

De interbeoordelaarsbetrouwbaarheid van de HCR-20 was over het geheel genomen hoog en kwam overeen met eerdere onderzoeksbevindingen (Belfrage, 1998 ; Douglas & Webster, 1999b). Bovendien werd het verschil in betrouwbaarheid per subschaal (zeer hoge betrouwbaarheid voor de Historische items, hoge betrouwbaarheid voor de Klinische items en matige betrouwbaarheid voor de Risicohanteringsitems) ook door anderen gevonden (Belfrage, 1998). Voor bepaalde individuele items werd een lage betrouwbaarheid gevonden. Hiervoor kunnen drie oorzaken worden genoemd. Ten eerste zou de lage betrouwbaarheid voor deze items te wijten kunnen zijn

Tabel 6 Stapsgewijze multiële regressieanalyse met FWC-subschalen als predictorvariabelen voor HCR-20-totaalscores.

Variabele FWC	B	SE B	β	sign. T	R ²	Adjusted R ²
Overheerst	5.8	1.0	.46	< .01	.21	.21

B = ongestandaardiseerde regressiecoëfficiënt; SE B = standaardmeetfout B; β = gestandaardiseerde regressiecoëfficiënt; Sign. T = significantieniveau T-toets; R² = percentage verklaarde variantie; Adjusted R² = percentage verklaarde variantie gecorrigeerd voor het aantal cases en aantal predictorvariabelen.

Tabel 7 Stapsgewijze multiële regressieanalyse met FWC-subschalen als predictorvariabelen voor HCR-20-eindoordeelen.

Variabelen FWC	B	SE B	β	sign. T
Afstand	.46	.14	.31	<.01
Nabij	.35	.11	.26	<.01
Positief	-.34	.13	-.24	<.01
Multiële R	.50			
R ²	.25			
Adjusted R ²	.23			

aan onervarenheid van de codeurs in het gebruik van de HCR-20. Een voorbeeld hiervan is de codering van het item *Eerder geweld*. Met name de groepsleiders rekenden het TBS-delict abusievelijk niet mee onder *Eerder geweld*. Voor de onderzoekers, die meer ervaring hadden in het gebruik van gestandaardiseerde instrumenten als de HCR-20 en PCL-R, werd voor het item *Eerder geweld* juist een zeer hoge betrouwbaarheid gevonden. Ten tweede zijn sommige HCR-20-items onduidelijk of te globaal gedefinieerd in de handleiding. Veel codeurs waren van mening dat de items *Impulsiviteit* en *Blootstelling aan destabiliserende factoren* te globaal waren beschreven en daardoor te veel ruimte lieten voor interpretaties. Ten derde kunnen verschillen in klinische ervaring en persoonlijke opvattingen van de codeurs hebben bijgedragen aan verminderde betrouwbaarheid. Waarschijnlijk gold dit speciaal voor de items *Problemen in de kindertijd* en *Hoog niveau van ervaren stress*. Een regelmatig terugkerend discussiepunt bij deze items was: wanneer is de problematiek van de patiënt ernstig genoeg om een '2' te coderen? Meer ervaren behandelaars, met name de hoofden behandeling, waren vaker geneigd problematiek als minder ernstig te bestempelen, omdat zij de desbetreffende patiënt vergeleken met patiënten die zij eerder behandeld hadden.

In het huidige onderzoek werden verschillen geconstateerd tussen behandelaars en onderzoekers, niet alleen in hun HCR-20-coderingen, maar ook in de wijze van coderen, namelijk de snelheid en de informatie waarop de codering gebaseerd werd. Behandelaars gingen af op hun persoonlijke ervaringen met de patiënt en gebruikten relatief weinig dossierinformatie, terwijl onderzoekers de patiënt vaak minder goed kenden en daardoor vooral gebruikmaakten van dossierinformatie. Of dit invloed heeft op de juistheid van risicotaxaties is nog onbekend, toekomstig onderzoek naar de predictieve validiteit van de HCR-20 zal dat moeten uitwijzen.

Er bestonden verschillen tussen de drie groepen codeurs. Ten eerste verschilden onderzoekers en hoofden behandeling niet qua gemiddelde HCR-20-scores, maar wel wat betreft de interpretatie hiervan. De hoofden behandeling kwamen vaker tot de conclusie 'laag risico' en leken dus optimistischer in hun interpretatie van de HCR-20-coderingen dan de onderzoekers. Dit zou te maken kunnen hebben met het feit dat hoofden behandeling sterker de druk voelen om de doorstroom binnen de forensische psychiatrie te bevorderen, dus patiënten zo snel mogelijk te laten doorstromen naar de transmurale fase. Het zou – ondanks de gestructureerde wijze van risico inschatten – kunnen zijn dat behandelaars vatbaarder waren voor cognitieve vertekeningen (*biases*) die gelden voor alle klinisch diagnostische beslissingen, bijvoorbeeld 'bevestiging zoeken voor een reeds gevormd oordeel' (zie voor een bespreking hiervan De Ruiter, 1999). De onderzoekers stonden

verder af van de behandeling van de patiënten en waren in mindere mate betrokken bij de besluitvorming over (proef)verlofplannen, waardoor zij wellicht minder vatbaar waren voor deze cognitieve vertekeningen. Een tweede verschil is dat de groepsleiders over het geheel genomen lagere scores op de Historische items, Risicohanteringsitems en totaalscore van de HCR-20 gaven vergeleken met de onderzoekers en – hoewel niet significant – ook met de hoofden behandeling. Een mogelijke oorzaak hiervoor is dat zij dagelijks met de patiënten omgaan en dit werk bijzonder lastig zou zijn wanneer zij zich voortdurend bewust zouden zijn van de aanwezige risicofactoren. Ook kan het zijn dat zij door het veelvuldige contact met patiënten in grotere mate emotioneel betrokken waren bij de patiënt en meer zicht hadden op de 'leuke kanten' van een patiënt en deze in hun risico-inschattingen verdisconteerden. Het feit dat groepsleiders lager scoorden dan onderzoekers werd niet in eerder onderzoek gevonden. Tot nu toe is slechts één onderzoek bekend dat de coderingen van onderzoekers en behandelaars vergeleek. Dernevik et al. (2001) vonden dat verpleegkundigen juist hogere HCR-20-totaalscores gaven dan onafhankelijke *experts* in risicotaxatie. Maar de verpleegkundigen uit dit onderzoek kunnen nauwelijks worden vergeleken met de groepsleiders uit het huidige onderzoek, omdat er een groot verschil bestaat in het gemiddelde aantal jaren werkzaam in de instelling (24.6 versus 6.2). Ten derde viel de matige betrouwbaarheid tussen hoofden behandeling en groepsleiders voor de Klinische en Risicohanteringsitems op. Wellicht is dit verschil te verklaren door de verschillende rollen die zij vervullen in de behandeling, waardoor zij de patiënt en zijn problematiek op een andere manier kennen en bekijken. Het hoofd behandeling zet vanaf een afstand de behandellijnen uit, terwijl de groepsleiders de dagelijkse begeleiding verzorgen en de meeste uren met de patiënten doorbrengen. Overigens leken deze verschillen in coderingen op itemniveau geen grote invloed te hebben gehad op de gehele risicotaxatie, aangezien de betrouwbaarheid voor de totaalscore zeer hoog was en ook de gemiddelde HCR-20-scores van hoofden behandeling en groepsleiders niet significant van elkaar verschilden.

Het gevoel overheerst of gemanipuleerd te worden door de patiënt vertoonde een significant verband met een hoge HCR-20-totaalscore. Dit biedt een interessante invalshoek voor toekomstig onderzoek. Bekend is dat patiënten die hoog scoren op psychopathie, dergelijke gevoelens bij anderen kunnen oproepen. Verder gingen negatieve gevoelens die behandelaars hadden ten opzichte van hun patiënten, zoals frustratie, teleurstelling, wantrouwen en afwijzing samen met hoge HCR-20-scores. Interessant is dat zowel gevoelens van nabijheid of (te) dicht bij een patiënt betrokken zijn (bijvoorbeeld gevoelens van moederlijkheid) als gevoelens van afstandelijkheid en

onverschilligheid een hoog HCR-20-eendoordeel voorspelden. Positieve gevoelens voor de patiënt als willen helpen, blijdschap en ontspanning voorspelden daarentegen een laag eendoordeel. De bevinding dat gevoelens van behandelaars ten opzichte van de patiënten die ze behandelen invloed hebben op hun risicotaxaties, werd eerder aangetoond in het onderzoek van Dernevik et al. (2001).

Verder dient nog vermeld te worden dat de hoogte van de gemiddelde HCR-20-scores van de gehele groep codeurs overeenkwam met bevindingen uit eerdere onderzoeken die werden uitgevoerd in buitenlandse forensisch psychiatrische instellingen (Belfrage, 1998; Strand et al., 1999). Het feit dat er een grote overlap bestond in de ranges van totaalscores behorende bij de eendoordelen laag, matig en hoog bevestigt de aanbeveling van Webster et al. (1997) om niet op basis van een totaalscore tot een eendoordeel te komen, maar om klinisch en inhoudelijk naar (combinaties van) risicofactoren te kijken.

Er zijn beperkingen aan het huidige onderzoek. De belangrijkste is dat we nog niets weten over de voorspellende waarde (predictieve validiteit) van de Nederlandse versie van de HCR-20. Ook de betekenis en eventuele implicaties van de invloed van gevoelens die behandelaars hebben ten opzichte van hun patiënten op de risicotaxatie is vooralsnog onbekend. Een belangrijke vraag hierbij is of deze gevoelens van invloed zijn op de juistheid van de risicotaxaties. Dernevik en Douglas (2002) vonden dat behandelaars institutioneel geweld op korte termijn goed voorspelden, terwijl de onderzoekers gewelddadige recidive op langere termijn goed voorspelden. Toekomstig onderzoek zal moeten uitwijzen of er in het huidige onderzoek verschil bestaat in de voorspellende waarde van HCR-20-coderingen van onderzoekers en behandelaars, op korte en op lange termijn. Interessant is ook of de consensus tussen onderzoekers en behandelaars wellicht leidt tot de beste voorspelling. De vraag wie het meest geschikt is om risicotaxaties uit te voeren – de behandelend clinicus of de meer op afstand staande onderzoeker – kan op dit moment nog niet beantwoord worden. Een andere beperking is het grote aantal beoordelaars dat aan dit onderzoek meedeed. Vooral in de groepsleiding waren veel wisselingen, waardoor maar door enkele groepsleiders meer dan twee keer een risicotaxatie werd uitgevoerd en ervaring werd opgedaan. Doordat de risicotaxatie door de groepsleiders meestal werd uitgevoerd voor een eerste of tweede keer, werden telkens dezelfde fouten gemaakt, zoals eerder beschreven over het item *Eerder geweld*. Dit is echter op dit moment de dagelijkse praktijk waarin gewerkt wordt, gezien de relatief snelle wisseling van groepsleiders in de Dr. Henri van der Hoeven Kliniek (in 2001: 28,2 in dienst getreden en 17,6 uit dienst getreden, op 31-12 2001 totaal in dienst: 119,1; Dr. Henri van der Hoeven Stichting, 2002).

Op basis van de resultaten en ervaringen uit het huidige onderzoek willen wij enkele aanbevelingen doen voor de implementatie van gestandaardiseerde risicotaxatie-instrumenten in de forensische praktijk. Ten eerste is een goede training, waarin relevante literatuur besproken wordt en uitgebreid wordt geoefend met cases uit de praktijk, van groot belang. De te trainen groep bestaat bij voorkeur uit maximaal tien personen, omdat dit het oefenen en de discussie vergemakkelijkt en er hierbij voldoende individuele aandacht geboden kan worden. Idealiter volgen een of meerdere terugkomdagen om vragen te beantwoorden en de getrainde vaardigheden op peil te houden. De tijd tussen de training en de eerste risicotaxatie mag niet langer dan zes maanden bedragen, omdat de in de training opgedane kennis en vaardigheden dan nog vers in het geheugen liggen. In de training dient benadrukt te worden dat men voldoende tijd moet nemen voor het coderen en bij iedere risicotaxatie de handleiding van de HCR-20 te lezen. Daarnaast is het te allen tijde aan te bevelen om dossierinformatie te gebruiken bij het coderen van de HCR-20; ook behandelaars die de patiënt goed kennen kunnen deze informatie gebruiken om hun persoonlijke ervaring te verifiëren. Het is tevens van belang dat de kwaliteit van de dossierinformatie en verslaglegging van het behandelverloop goed is, omdat deze mede de basis vormen waarop de HCR-20 wordt gecodeerd.

Ten tweede is het belangrijk om voldoende tijd te nemen voor de consensusbesprekingen met collega's. In het huidige onderzoek vonden alle partijen de consensusbesprekingen bijzonder zinvol. De verschillende invalshoeken – de onafhankelijke, op afstand staande onderzoekers en de behandelende klinici die de patiënt goed kennen – leken elkaar goed aan te vullen. Het is aan te bevelen in deze consensusbesprekingen niet alleen stil te staan bij de risicofactoren en het eendoordeel, maar tevens te discussiëren over hoe deze factoren in de behandeling kunnen worden verminderd en of eventueel beschermende factoren bestaan. Ook is het belangrijk om na te denken en te bespreken of er bij een patiënt relevante risicofactoren aanwezig zijn die niet in de HCR-20 zijn opgenomen. Deze factoren kunnen onder *Overige overwegingen* worden gecodeerd, maar het is mogelijk dat een bepaalde risicofactor bij zoveel patiënten voorkomt, dat hij wellicht opgenomen dient te worden in een herziene versie van de (Nederlandse) HCR-20. De HCR-20 is een instrument in ontwikkeling (Philipse et al., 1997, p. 17), hetgeen betekent dat verfijning van items, verwijdering van irrelevante items of toevoeging van nieuwe risicofactoren tot de mogelijkheden behoort.

Ten slotte is het belangrijk om codeurs te voorzien van feedback over de juistheid van hun risicotaxaties, bijvoorbeeld wanneer een patiënt die buiten woont opnieuw een delict heeft gepleegd.

Referenties

- Ackerman, M.J. (1999). *Essentials of forensic psychological assessment*. New York: Wiley.
- APA, American Psychiatric Association (1994). *Diagnostic and Statistical Manual of Mental Disorders (4th ed.)*. Washington, DC: American Psychiatric Association.
- Belfrage, H. (1998). Implementing the HCR-20 scheme for risk assessment in a forensic psychiatric hospital: Integrating research and clinical practice. *The Journal of Forensic Psychiatry*, 9, 328-338.
- Belfrage, H., & Douglas, K.S. (2002). Treatment effects on forensic psychiatric patients measured with the HCR-20 violence risk assessment scheme. *International Journal of Forensic Mental Health*, 1, 25-36.
- Belfrage, H., Fransson, G., & Strand, S. (2000). Prediction of violence using the HCR-20: A prospective study in two maximum-security correctional institutions. *The Journal of Forensic Psychiatry*, 11, 167-175.
- Borum, R. (1996). Improving the clinical practice of violence risk assessment: Technology, guidelines and training. *American Psychologist*, 51, 945-956.
- Dankers, J., & Linden, J. van der (1995). Van regenten en patiënten: *De geschiedenis van de Willem Arntsz Stichting: Huis en Hoeve, Van der Hoeven Kliniek en Dennendal*. Amsterdam: Boom.
- Dernevik, M., & Douglas, K.S. (2002). *Structured clinical risk assessment: What are we assessing and who should do it in clinical setting?* Paper presented at the 12th European Conference on Psychology and Law, Leuven, Belgium, September 14-17.
- Dernevik, M., Falkheim, M., Holmqvist, R., & Sandell, R. (2001). Implementing risk assessment procedures in a forensic psychiatric setting: Clinical judgement revisited. In D.P. Farrington, C.R. Hollin & M. McMurrin (Eds.), *Sex and violence: The psychology of crime and risk assessment* (pp. 83-101). London: Routledge.
- Dr. Henri van der Hoeven Stichting (2002). *Jaarverslag 2001*. Utrecht: Dr. Henri van der Hoeven Stichting.
- Douglas, K.S., & Webster, C.D. (1999a). Predicting violence in mentally and personality disordered individuals. In R. Roesch, S.D. Hart, J.R.P. Ogloff (Eds.), *Psychology and law: The state of discipline* (pp. 175-239). New York: Kluwer Academic.
- Douglas, K.S., & Webster, C.D. (1999b). The HCR-20 violence risk assessment scheme. Concurrent validity in a sample of incarcerated offenders. *Criminal Justice and Behavior*, 26, 3-19.
- Douglas, K.S., Cox, D.N., & Webster, C.D. (1999). Violence risk assessment: Science and practice. *Legal and Criminological Psychology*, 4, 149-184.
- Douglas, K.S., Ogloff, J.R.P., Nicholls, T.L., & Grant, I. (1999). Assessing risk for violence among psychiatric patients: The HCR-20 violence risk assessment scheme and the Psychopathy Checklist: Screening Version. *Journal of Consulting and Clinical Psychology*, 67, 917-930.
- Fleiss, J.L. (1986). *The design and analysis of clinical experiments*. New York: Wiley.
- Grann, M., Belfrage, H., & Tengström, A. (2000). Actuarial assessment of risk for violence. Predictive validity of the VRAG and the Historical part of the HCR-20. *Criminal Justice and Behavior*, 27, 97-114.
- Greenberg, S., & Shuman, D. (1997). Irreconcilable conflict between therapeutic and forensic roles. *Professional Psychology: Research and Practice*, 28, 50-57.
- Hare, R. (1991). *Manual for the Hare Psychopathy Checklist-Revised*. Toronto, Ontario: Multi-Health Systems.
- Hart, S.D., Ruiter, C. de, Hildebrand, M., & Vogel, V. de (2001). *HCR-20 codeerblad. Uitgebreide versie*. Utrecht: Dr. Henri van der Hoeven Stichting.
- Holmqvist, R., & Armelius, B.A. (1994). Emotional reactions to psychiatric patients. *Acta Psychiatrica Scandinavica*, 90, 204-209.
- Holmqvist, R., & Fogelstam, H. (1996). Psychological climate and countertransference in psychiatric treatment homes. *Acta Psychiatrica Scandinavica*, 93, 288-295.
- Laws, D.R., Hudson, S.M., & Ward, T. (2000). *Remaking relapse prevention: A sourcebook*. London: Sage Publications.
- Litwack, T.R., & Schlesinger, L.B. (1999). Dangerousness risk assessments: Research, legal, and clinical considerations. In A.K. Hess & I.B. Weiner (Eds.), *The handbook of forensic psychology* (pp. 171-217). New York: Wiley.
- McGraw, K.O., & Wong, S.P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1, 30-46.
- Philipse, M., Ruiter, C. de, Hildebrand, M., & Bouman, Y. (2000). *HCR-20. Beoordelen van het risico van gewelddadig gedrag. Versie 2*. Nijmegen/Utrecht: Prof. Mr. W.P.J. Pompestichting/Dr. Henri van der Hoeven Stichting.
- Quinsey, V.L., Harris, G.T., Rice, M.E., & Cormier, C.A. (1998). *Violent offenders: Appraising and managing risk*. Washington DC: American Psychological Association.
- Ruiter, C. de (1999). Risicotaxatie bij forensisch psychiatrische patiënten. In C. de Ruiter & M. Hildebrand (red.), *Behandlungsstrategieën bij forensisch psychiatrische patiënten* (pp. 9-16). Houten/Diegem: Bohn Stafleu Van Loghum.
- Strand, S., Belfrage, H., Fransson, G., & Levander, S. (1999). Clinical and risk management factors in risk prediction of mentally disordered offenders: More important than actuarial data? *Legal and Criminological Psychology*, 4, 67-76.
- Vertommen, H., Verheul, R., Ruiter, C. de, & Hildebrand, M. (2002). *De herziene versie van Hare's Psychopathie Checklist*. Lisse: Swets Test Publishers.
- Vogel, V. de, & Hildebrand, M. (2001). Beoordelen van het risico van gewelddadig gedrag: De Historical, Clinical, Risk Management-20 (HCR-20). *Gedragstherapie*, 34, 93-102.
- Webster, C.D., Douglas, K.S., Eaves, D., & Hart, S.D. (1997). *HCR-20. Assessing the risk of violence. Version 2*. Burnaby, BC, Canada: Simon Fraser University and Forensic Psychiatric Services Commission of British Columbia.
- Whyte, C.R., Constantopoulos, C., & Bevans, H.G. (1982). Types of countertransference identified by Q-analysis. *British Journal of Medical Psychology*, 55, 187-201.