

Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. De Psycholoog, september 2014, 29-39.

Woorden: (voorblad=46, artikel=4.659, referenties=1.535, figuren/tabellen: 2

Gestructureerde Risicotaxatie: Noodzakelijker Wijs

Wineke J. Smid, MSc¹

Evelyn Klein Haneveld MSc¹

Vivienne de Vogel PhD¹

Joan van Horn PhD¹

1) De Forensische Zorgspecialisten, Utrecht

Adres voor correspondentie:

Wineke Smid

De Forensische Zorgspecialisten

Afdeling Onderzoek

Postbus Box 174

3500 AD Utrecht

Tel: 030-275 8 275

wsmid@hoevenkliniek.nl

Gestructureerde Risicotaxatie: Noodzakelijker Wijs

De kwaliteit van risicotaxatie-instrumenten is onvoldoende, zo suggereren Fazel, Singh, Doll en Grann in hun artikel in the British Medical Journal van juli 2012 en ze stellen dat de forensische praktijk maar beter niet teveel kan leunen op dit type instrumenten bij het nemen van belangrijke beslissingen omtrent straf, behandeling of verlop. Naar aanleiding van deze publicatie laaide in Nederland opnieuw de discussie op waarbij de bruikbaarheid van gestructureerde risicotaxatie in diverse krantenberichten en opiniestukken in twijfel werd getrokken. Een terugkerend argument daarin is dat risicotaxatie-instrumenten weliswaar voor grote groepen een goede inschatting van de kans op recidive kunnen geven, maar dat we daarmee niets kunnen zeggen over individuele gevallen. Betoogd wordt dat in het individuele geval een klinische analyse van de persoon en de context waarin het delictgedrag plaats vindt meer recht zou doen aan de dader en dat risicotaxatie daarin géén of slechts een kleine rol zou moeten spelen. De richtlijnen voor de pro justitia rapportages van het Nederlands Instituut voor Forensische Psychiatrie en Psychologie (NIFP) weerspiegelen bijvoorbeeld deze gedachtegang. Gestructureerde risicotaxatie wordt bij deze rapportages weliswaar inmiddels aanbevolen (Nederlandse Vereniging voor Psychiatrie, 2012) en steeds vaker gebruikt, maar maakt er nog geen niet standaard/verplicht onderdeel van uit en is zeker niet leidend in de uiteindelijke conclusies en het daaruit volgend advies.

Wij beamen met Fazel et al. (2012) dat risicotaxatie-instrumenten niet zaligmakend zijn, maar stellen daar tegenover dat 1) de kwaliteit van de risicotaxatie-instrumenten veel beter is dan Fazel e.a. suggereren 2) de alternatieven voor het gebruik van risicotaxatie-instrumenten slechtere resultaten opleveren en 3) het nut of gevaar van het gebruik van deze instrumenten afhangt van de manier waarop ze gebruikt worden. We zullen in deze bijdrage betogen dat zorgvuldig en verantwoord gebruik van risicotaxatie-instrumenten een zeer waardevolle bijdrage kan leveren aan het terugdringen van recidive. Eerst zullen de meta-analyse van Fazel e.a. en de conclusies die daaraan worden verbonden kritisch worden besproken.

De vals positieven van Fazel

Fazel e.a. (2012) zochten data uit 251 onafhankelijke risicotaxatie studies (341 datasets) waarvan ze er uiteindelijk 68 (73 datasets) daadwerkelijk bemachtigden en gebruikten in hun meta-analyse/review. In een eerder artikel, dat niet zoveel stof deed opwaaien, onderzochten zij hoe goed de diverse instrumenten ten opzichte van elkaar recidive voorspelden (Singh, Grann & Fazel, 2011). In het artikel uit 2012 lijken Fazel e.a. op zoek naar een hogere waarheid, namelijk hoe goed de voorspelling van de instrumenten is ten opzichte van 'de werkelijkheid'. De focus ligt op de kritische beschouwing van de beslissingen die voort zouden vloeien uit het gebruik van deze instrumenten: hoe vaak zou een dergelijke beslissing goed of fout zijn geweest? Er is natuurlijk niets tegen een kritische beschouwing (zie bijvoorbeeld Harte & Breukink, 2010), maar er kleven grote bezwaren aan de benadering die Fazel e.a. kiezen en risicotaxatie komt daardoor in een onnodig kwaad daglicht te staan.

De meest prominente conclusie van Fazel e.a. is dat de risicotaxatie-instrumenten tekort schieten in hun discriminerend vermogen met als gevolg dat er met name teveel 'vals positieven' ontstaan. Dat wil zeggen dat er teveel delinquenten als risicovol werden aangemerkt terwijl ze niet recidiveerden en dus volgens Fazel e.a. onterecht zouden zijn vastgehouden op basis van hun risicotaxatie. Dat zou vooral voor zedendelinquenten gelden, bij wie een detentie ratio van 5 werd gevonden. Dat wil zeggen dat er maar liefst 5 delinquenten vastgehouden zouden moeten worden om 1 delict te voorkomen, met andere woorden: zo'n 80% van de delinquenten in de 'vasthouden' categorie recidiveerde niet. Dat klinkt inderdaad ernstig, maar hoe komen Fazel e.a. tot deze conclusie?

De in hun meta-analyse verzamelde studies bieden geen specifieke insteek met betrekking tot een bepaalde subgroep van delinquenten, instrumenten of een bepaalde toepassing van risicotaxatie, maar vormen een willekeurige verzameling van risicotaxaties die werden gedaan bij zeer uiteenlopende doelgroepen, voor diverse doeleinden (behandeltoewijzing, evaluatie, validering, etc.), op verschillende meetmomenten (voor, tijdens of na behandeling), met behulp van diverse instrumenten,

gekoppeld aan verschillende uitkomstmaten (recidive, interne incidenten etc.). De instrumenten werden in de geïncludeerde studies niet gebruikt om beslissingen te nemen omtrent vasthouden of vrijlaten en zeker niet op de door Fazel e.a. gesuggereerde manier. Hun conclusies zijn dan ook zuiver hypothetisch/theoretisch en verwijzen op geen enkele wijze naar de forensische realiteit. Om van de bonte verzameling aan studies tot de hogere waarheid van goede en verkeerde beslissingen te komen, doen Fazel e.a. onderweg een aantal vergaande aannames en maken ze drastische keuzes. Keuzes die allen desastreus zijn voor het aantal vals positieven. De uiteindelijke conclusie dat het aantal vals positieven te hoog is, berust dan ook vooral op een *self fulfilling prophecy*.

Het grootste probleem schuilt in de tweedeling van de classificatie. Om de uitkomst van de gebruikte instrumenten te laten aansluiten bij de virtuele beslissing 'vasthouden' of 'laten gaan', worden de uitkomstmaten van alle instrumenten in tweeën gedeeld. Volgens Fazel e.a. betekent daarbij alleen de kwalificatie 'laag risico' dat geen recidive wordt verwacht en dat iemand zou mogen uitstromen, zowel 'matig risico' als 'hoog risico' betekent dat wel recidive wordt verwacht en dat iemand *dus* zou worden vastgehouden. Als voorbeeld: de meeste zedendelinquenten in het onderzoek zijn beoordeeld met behulp van de Static-99 (Hanson & Thornton, 2000) (gebruikt in 13 van de 20 geïncludeerde zedenstudies). Zie Tabel 1 voor voorbeelden van items uit de Static-99. Bij de Static-99 wordt alleen de classificatie 'laag' gekoppeld aan het oordeel 'laten gaan', de categorieën 'laag-matig', 'matig-hoog' en 'hoog' worden allen gekoppeld aan het oordeel 'vasthouden'. In werkelijkheid wordt de Static-99 nooit op die manier gebruikt. Tabel 2 bevat ter illustratie percentages afkomstig uit de omschrijving van de internationale Static-99 normgroep¹ (Phenix, Hanson, Harris & Thornton, 2008). In deze tabel is te zien dat bij de tweedeling 'vasthouden'/'vrijlaten' volgens de Fazel methode de detentie ratio maar liefst 7 zou bedragen. De originele Static-99 hoog-risico categorie levert echter een detentie ratio op van 3.

¹ In de studie van Fazel e.a. ontbreken duidelijke cijfers om als vergelijkingmateriaal te kunnen dienen

Naast de grove tweedeling van de categorieën wordt door Fazel e.a. de meest beperkte variant als uitkomstmaat voor de zedendelinquenten gekozen: alleen een nieuw zedendelict telt als recidive. Dat is een prima uitkomstmaat wanneer het er om gaat te bepalen welk van de bestudeerde instrumenten het best seksuele recidive voorspelt. Maar als het er om gaat een uitspraak te doen over wie er onterecht zou zijn vastgehouden dan had een uitkomstmaat 'ernstige recidive' (i.c. geweld en/of zeden) meer voor de hand gelegen. Het is immers bekend dat zedendelinquenten relatief vaker met andersoortige delicten dan zedendelicten recidiveren (Hanson & Bussière, 1998). De gebruikte zeden-instrumenten zijn dan ook niet speciaal gemaakt om zedenrecidive te voorspellen, maar om recidive bij zedendelinquenten te voorspellen, een subtiel maar belangrijk verschil. Het is met name niet reëel om geweldsrecidive door een zedendelinquent als 'geen recidive' te beschouwen. Bij de door Fazel e.a. gehanteerde criteria zou bijvoorbeeld een hoog-risico verkrachter die recidiveert met een poging tot moord aangemerkt worden als een 'vals positief'. In Tabel 2 is te zien dat de inclusie van geweldsdelicten in de uitkomstmaat de detentie ratio in de Static-99 hoog-risico groep verder terugbrengt van 3 naar 2, (in de matig-hoog risico groep van 7 naar 3).

Tot slot is de gemiddelde follow up tijd van de geïncludeerde studies met een duur van 50 maanden (4,1 jaar) aan de korte kant om te concluderen dat mensen onterecht zouden zijn vastgehouden. Het kan alleen al maanden tot jaren duren voordat nieuwe zaken juridisch helemaal afgehandeld zijn en als veroordeling op het strafblad verschijnen (Stuiveling & Van Schoten, 2012). Het is niet denkbeeldig dat een deel van de 'vals positieven' in de jaren na de follow up termijn alsnog gerecidiveerd zou zijn of een eventuele recidive alsnog op het strafblad verschenen zou zijn. Andermaal, om de kwaliteit van de voorspelling van verschillende instrumenten te vergelijken, is een gemiddelde follow up periode van vier jaar natuurlijk prima. Maar om zwaarwegende uitspraken te doen over het onterecht vasthouden van mensen ligt een lange follow up meer voor de hand. In Tabel 2 is te zien dat bij een langere follow up periode de detentie ratio in de Static-99 hoog-risico groep nog verder terugloopt naar minder dan 2, (in de matig-hoog risico groep van 3 naar 2,5).

De grove indeling, de geamputeerde uitkomstmaat, de krappe follow up periode en de daaraan gerelateerde recidivepercentages hebben nadrukkelijk invloed op het aantal vals positieven; ze doen deze toenemen. Voorspellen wordt immers moeilijker naarmate datgene wat je voorspelt minder vaak voorkomt. Het percentage seksuele recidive voor de totale groep zedendelinquenten (de base rate voor seksuele recidive) uit de Static-99 normtabel (zie Tabel 2) is 12% in 5 jaar. Ze recidiveren dus veel vaker niet dan wel. Wanneer we van iedereen zouden gokken dat ze *niet* recidiveren, hebben we het alleen al op grond van kans in 88% van de gevallen goed. De 95% goede voorspellingen in de 'laag risico' categorie is dus 1,1 keer beter dan kans. Wanneer we van iedereen zouden voorspellen dat ze *wel* recidiveren dan hebben we het op grond van kans in slechts 12% van de gevallen goed. De 28% recidive in de hoog risico categorie maakt die schatting 2,3 keer beter dan kans. Wanneer we dus corrigeren voor kans dan is het resultaat in de hoog risico categorie eigenlijk beter dan dat in de laag risico categorie. De stelling van Fazel e.a. dat de instrumenten vooral het *aanwezige* risico slecht voorspellen, is dan ook discutabel. Deze risicotaxatie-instrumenten beogen iets te voorspellen wat we ernstig vinden en wat weinig voorkomt, ze zijn gemaakt om sensitief te zijn. Een hoge sensitiviteit gaat altijd tot op zeker hoogte ten koste van de specificiteit. Een verhoging van de specificiteit zal op zijn beurt weer ten koste gaan van de sensitiviteit.

Tot slot includeren Fazel et al. een interessante vergelijking tussen forensische- en medische risicotaxatie-instrumenten. Hoewel de vergelijking wordt aangegrepen om de tekortschietende kwaliteit van de forensische instrumenten te onderstrepen, valt eigenlijk op dat ze juist niet slecht uit de vergelijking naar voren komen (AUC's tussen de .66-.74 voor de forensische risicotaxatie versus AUC's tussen de .57-.86 voor de medische taxaties die cardiovasculaire incidenten (CVA) voorspellen; Cooney, Dubina & Graham, 2009). Dat geldt zeker wanneer rekening gehouden wordt met de lastige uitkomstmaat van de forensische instrumenten. Een nieuwe veroordeling vereist immers niet alleen dat er een nieuw delict moet zijn gepleegd (datgene wat we eigenlijk willen meten), maar dat delict moet ook worden ontdekt en er moet een strafrechtelijk traject in gang worden

gezet van aangifte, arrestatie, aanklacht en uiteindelijk een veroordeling. Voordat een gepleegd delict uiteindelijk op het strafblad verschijnt, kan onderweg een hoop ruis ontstaan. De medische instrumenten richten zich op de voorspelling van relatief duidelijk waarneembare fysieke aandoeningen waar patiënten last van hebben en waar ze zichzelf mee komen melden. En toch doen de forensische instrumenten het, zelfs met een relatief korte follow up periode van 4 jaar, bijna net zo goed. Dat is indrukwekkend te noemen.

Naast de vergelijking tussen 'medische risicotaxatie' en 'forensische risicotaxatie' was een vergelijking tussen 'medische risicotaxatie' en 'medisch klinische voorspelling' hier ook informatief geweest evenals de eigenlijk onontbeerlijke vergelijking tussen 'forensische risicotaxatie' en 'forensisch klinische voorspelling'. Uit die vergelijkingen zou namelijk naar voren komen dat in beide gevallen risicotaxatie beter voorspelt dan de klinische inschatting. Risicotaxatie is niet volmaakt, maar het alternatief is beduidend slechter.

Klinisch oordeel beter dan risicotaxatie?

Er zijn vele studies en meta-analyses waaruit in de afgelopen decennia een *body of evidence* is ontstaan met de algemene conclusie dat een klinisch oordeel zonder instrumenten leidt tot een aanzienlijk slechtere inschatting van het recidiverisico dan wanneer deze inschatting met instrumenten gemaakt wordt. In tegenstelling tot wat Fazel e.a. beweren is er dienaangaande geen sprake van grote verdeeldheid, maar bestaat hierover in het wetenschappelijk veld wel degelijk brede consensus (Andrews, Bonta & Wormith, 2006; Janus & Prentky, 2003; Monahan, 2007; Otto & Douglas, 2010; Quinsey, Harris, Rice & Cormier, 2006). Deze meerwaarde van gestructureerde taxatie is niet nieuw (Meehl, 1954) en niet uniek voor het forensisch veld, maar geldt met betrekking tot het voorspellen van allerlei vormen van menselijk gedrag en gezondheid (Aegisdottir et al., 2006, Dawes, 2002; Grove et al., 2000). Het veelgehoorde argument dat het klinisch oordeel wel beter zou zijn 'in het individuele geval' snijdt geen hout, een groep is immers niets anders dan een verzameling individuele gevallen en als het klinisch oordeel in het individuele geval werkelijk beter zou zijn dan zou dat bij het bestuderen

van een groep van 1000 individuele gevallen ook naar voren moeten komen. Dat is niet zo.

Met de stelling dat "risicotaxatie *alleen* niet voldoende is" suggereren Fazel e.a. dat de combinatie van een instrument én een klinisch oordeel betere resultaten op zou leveren dan risicotaxatie alleen. Dat is niet het geval. Zeker niet wanneer het gaat om het inschatten van het basis recidiverisico, voorafgaande aan een eventuele behandeling. In sommige individuele gevallen zal de extra klinische afweging bij de taxatie iets aan de voorspelling toevoegen, in andere gevallen zal het de voorspelling slechter maken. Het grote probleem is dat we niet weten in welke gevallen het positief bijdraagt en in welke gevallen negatief. Het netto resultaat over de hele groep wordt er doorgaans slechter van, de voorspelling zit er dus vaker naast (Hanson & Morton-Bourgon, 2009; Michel et al., 2013; Storey, Watt, Jackson & Hart, 2012; Wormith, Hogg & Guzzo, 2012). Voor gebruik in instellingen waar intensieve behandeling wordt geboden (zoals in tbs-instellingen) ligt dit genuanceerder, maar ook dan geldt dat wetenschappelijk onderbouwde risicofactoren de leidraad moeten vormen (De Vogel, De Vries Robbé, De Spa & Wever, 2013).

Een deel van het probleem lijkt te zitten in de (on)betrouwbaarheid. Betrokken beoordelaars worden het niet eens over de klinische toevoeging: wanneer moet er wel en wanneer niet een extra klinische afweging gemaakt worden bij de eindscore van het instrument en wat moet er dan precies worden meegewogen? Is bijvoorbeeld het feit dat het een 'heel naar' delict betrof een reden voor een extra afweging en wat is dan precies 'heel naar' en hoeveel hoger wordt het risico daar dan van? Uiteindelijk neemt elke individuele beoordelaar daar een andere beslissing over. Hanson, Harris, Scott en Helmus (2007) berekenden dat beoordelaars het meestal niet eens waren ($ICC = .15$) wat betreft het wel of niet toevoegen van een extra klinische afweging. Factoren die met zo weinig overeenstemming vastgesteld kunnen worden, kunnen niet tot valide voorspellingen leiden omdat de score dan meer afhankelijk wordt van wie de beoordeling heeft gedaan dan van wie er beoordeeld is.

Het beperkte aantal onderzoeken dat er tot nu toe gedaan is naar de inhoud van de extra afwegingen laat zien dat het meewegen van extra factoren die niet in het instrument staan er bijvoorbeeld toe kan leiden dat factoren worden meegewogen die weliswaar van belang zijn in het leven van de patiënt, maar die niet direct samenhangen met zijn of haar recidiverisico; factoren die als zodanig dus terecht ontbraken in het instrument, bijvoorbeeld of iemand verlegen of teruggetrokken is (Wormith, Hogg & Guzzo, 2012). Daarnaast gebeurt het dat risicofactoren die al in het instrument verdisconteerd zijn, op een iets andere manier geformuleerd, als extra risicofactor worden toegevoegd, waardoor deze als het ware dubbel gewogen worden. Bijvoorbeeld de toevoeging dat iemand erg anti sociaal is, terwijl daarvoor de eerdere veroordelingen al gescoord zijn. Een toegevoegd klinisch oordeel lijkt er vooral toe te leiden dat het risico voor veel delinquenten aanzienlijk hoger wordt ingeschat dan zonder deze aanpassing. Het voorspellende vermogen van de score neemt daardoor af (Wormith et al., 2012).

Dat is natuurlijk onbevredigend en teleurstellend en geeft zeker aanleiding voor verder onderzoek naar mogelijkheden tot verbetering. Het is echter anno 2013 niet meer mogelijk om de kop in het zand te steken. De constatering dat risicotaxatie niet volmaakt is, is geen reden om er van af te zien en te volharden in iets wat nog veel minder volmaakt is. We zijn het zowel aan de samenleving als aan de delinquenten verplicht om het best mogelijke te doen. En de conclusie kan op dit moment niet anders zijn dan dat gestructureerde risicotaxatie het best mogelijke is. De onvolmaaktheid van de gebruikte methoden dient vervolgens natuurlijk wel meegenomen te worden in de bepaling van het vervolgtraject. Er zitten duidelijke grenzen aan wat op basis van risicotaxatie geconcludeerd en verantwoord kan worden (zie ook Harte & Breukink, 2010). En de consequenties van risicotaxaties dienen binnen die grenzen te passen.

(On)verantwoord gebruik van risicotaxatie in strafzaken

Enerzijds zijn er toepassingen waar risicotaxatie-instrumenten duidelijk niet geschikt voor zijn. Veel staten in de VS kennen bijvoorbeeld *Sexual Violent Predator Acts*, wetten die

Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. *De Psycholoog*, september 2014, 29-39.

bepalen dat elke zedendelinquent die 'recidivegevaarlijk' is, voor onbepaalde tijd preventief mag worden opgesloten (zie bijvoorbeeld http://kansasstatutes.lesterama.org/Chapter_59/Article_29a/59-29a02.html). Een risicotaxatie-instrument (vaak de Static-99) wordt dan gebruikt met het doel om iedereen die er te hoog op scoort voor eeuwig op te sluiten. Overigens ligt de grenswaarde dan bij een Static-99-score die veel hoger ligt dan de waarde die Fazel e.a. in hun onderzoek hanteren. Er zijn in het wetenschappelijk veld maar weinig voorstanders van het gebruik van risicotaxatie voor dit doeleinde. Er bestaan immers geen instrumenten die kunnen aantonen dat iemand zeker zal recidiveren en die daarmee levenslange preventieve opsluiting eventueel zouden kunnen legitimeren.

Anderzijds zijn er duidelijk gevallen waarin wel verantwoord gebruik gemaakt kan worden van risicotaxatie en het zijn juist die gevallen die bij toepassing in Nederland een enorme verbetering op zouden leveren ten opzichte van de status-quo. Het meest duidelijke voorbeeld is het gebruik bij de pro justitia rapportage. Levenslange opsluiting voorkomt natuurlijk recidive, maar er zijn efficiëntere manieren om de recidive terug te dringen, bijvoorbeeld door middel van interventie/behandeling. Onderzoek laat keer op keer zien (Andrews & Bonta, 2006; French & Gendreau, 2006; Hanson, Bourgon, Helmus & Hodgson 2009; Landenberger & Lipsey, 2005; Olver, Wong & Nicholaichuk, 2009; Wilson, Bouffard & Mackenzie, 2005) dat interventies die het recidiverisico omlaag brengen het meest effectief zijn wanneer ze voldoen aan de drie basisprincipes van het *Risk Need Responsivity* model (Andrews & Bonta, 2010). Het RNR-model schrijft voor dat 1) de plegers met het hoogste risico de meest intensieve interventie dienen te krijgen (het risico principe), 2) de interventie gericht moet zijn op die (dynamische) factoren die daadwerkelijk samenhangen met het delictgedrag (het behoefte principe) en 3) de interventie dient te worden gegeven in een vorm die aansluit bij de leerstijl en mogelijkheden van de pleger (het responsiviteits principe). Met name het risico principe is belangrijk voor de discussie. Wanneer delinquenten met het hoogste risico de meest intensieve interventie moeten krijgen, dan is het zaak om deze mensen op betrouwbare

wijze uit de totale groep delinquenten te selecteren. Het gaat hier om de inschatting van het basisrisico, dus de kans op recidive zonder interventie.

Teruggrijpend op het voorbeeld uit het begin van dit betoog: van de Nederlandse zedendelinquenten die 'laag' scoorden op de Static-99R (de gereviseerde versie van de Static-99) bleek in recent onderzoek ongeveer 4% binnen 10 jaar aangeklaagd of veroordeeld te zijn voor een nieuw zedendelict, van de zedendelinquenten die 'hoog' scoorden was dat 36% (Smid, Kamphuis, Wever, Van Beek & Hoebe, 2013c). Het betekent dus niet dat iedereen die laag scoort nooit zal recidiveren en evenmin dat iedereen die hoog scoort zeker zal recidiveren. De exacte recidive percentages verschillen per onderzoek en hangen natuurlijk mede af van follow up tijd en uitkomstmaten. Maar het *relatieve* risico houdt in alle onderzoeken stand (Helmus, Hanson, Thornton, Babchishin & Harris, 2012a). Dat wil zeggen dat de kans op recidive altijd veel groter is in de hoog-risico categorie dan in de laag-risico categorie. Daarbij is het goed om te beseffen dat, over alle veroordeelde delinquenten bekeken, het absolute aantal delinquenten in de categorie 'laag' veel groter is dan in de categorie 'hoog'. In de laagste categorie van de Static-99R zit bijvoorbeeld zo'n 40% van alle veroordeelde zedendelinquenten, in de hoogste categorie een kleine 10%. Dat kleine groepje hoog-risico delinquenten recidiveert dus in het Nederlandse geval negen keer zo vaak als de laag-risico delinquenten wanneer we niet ingrijpen.

Om recidive effectief te kunnen bestrijden is het van belang om in elk geval het relatief kleine groepje hoog-risico plegers goed te selecteren en hen de meest intensieve interventie te laten ondergaan. Onderzoek van Smid, Kamphuis, Wever en Van Beek (2013a) laat zien dat dit ook in Nederland zonder gestructureerde risicotaxatie niet goed gaat. De ongestructureerde klinische inschatting, die nog steeds leidend is in de Nederlandse pro justitia rapportage, leidt er bijvoorbeeld toe dat meer dan de helft van de hoog-risico verkrachters en meer dan een vijfde van de hoog-risico misbruikplegers zonder enige vorm van interventie na hun straf de samenleving weer in gaat. Vervolgonderzoek laat zien dat deze onbehandelde hoog-risico delinquenten ook daadwerkelijk veel vaker recidiveren (Smid, Kamphuis, Wever & Van Beek, 2013b).

Vergelijkbaar onderzoek voor geweldsdelinquenten ontbreekt vooralsnog, maar er zijn weinig redenen om aan te nemen dat de resultaten daar anders zouden zijn.

Naast vals negatieven, levert de ongestructureerde klinische inschatting ook vele 'vals positieven' op: meer dan de helft van de misbruik plegers met het laagste risico wordt doorverwezen naar soms jarenlange (ambulante) behandelingen (Smid et al., 2013a). Dat duidt op een weinig efficiënt gebruik van beschikbare resources, en draagt bovendien ook een ander risico in zich. Uit onderzoek blijkt namelijk dat (intensieve) behandeling van laag-risico delinquenten juist *risicoverhogend* kan werken (Bonta, Wallace-Capretta & Rooney, 2000; Lowenkamp & Latessa, 2004; Lowenkamp, Latessa & Holsinger, 2006; Wakeling, Mann & Carter, 2012). Lowenkamp et al. (2006) voeren een aantal mogelijke redenen aan voor dat effect, waaronder het contact met hoog-risico plegers die in dezelfde groepen behandeld worden. Die vermenging van risico-niveaus binnen behandelgroepen komt ook in Nederland voor (Smid, Kamphuis, Wever & Verbruggen, 2013d). Daarnaast hebben laag risico plegers vaak een aanzienlijk 'normaal' deel in hun leven zoals werk, gezin, hobby's, vrienden etc. Dat deel heeft een beschermende werking en door intensieve vormen van behandeling worden ze daar langer en vaker van weggehouden.

Goed selecteren is dus belangrijk en dat kan relatief eenvoudig door ten tijde van de veroordeling standaard gebruik te maken van gestandaardiseerde instrumenten waarmee op basis van veelal feitelijke informatie het recidiverisico kan worden ingeschat. Voor geweldsdelinquenten kan men bijvoorbeeld denken aan de Violence Risk Appraisal Guide (VRAG of VRAG-R) (Quinsey et al., 2006; Rice, Harris & Lang 2013) of de LSI-R (Andrews, Bonta & Wormith, 2004), voor zedendelinquenten zijn de eerder genoemde Static-99R (Hanson & Thornton, 2000; Helmus, Thornton, Hanson & Babchishin, 2012b) of de Static-2002R (Hanson & Thornton, 2003; Helmus et al., 2012b) goed voor dit doeleinde te gebruiken (Smid et al., 2013c). Gebruik van deze instrumenten is niet alleen beter voor de slachtoffers, omdat het een bijdrage levert aan een betere bestrijding van recidive, maar het is ook beter voor de delinquenten. Hoewel het professionals soms onmenselijk toeschijnt om een verdachte aan de hand van een 'simpel lijstje' te

beoordelen, gaat achter de klinische beoordeling in wezen een onbedoelde, maar veel grotere willekeur schuil. Uit onderzoek van Kordelaar (2002) bleek bijvoorbeeld dat gerechtelijke arrondissementen verschilden in de mate waarin pro justitia rapportages werden aangevraagd en dat er duidelijke regionale verschillen bestonden in het aantal tbs-opleggingen. Individuele opvattingen van rechters en officieren van justitie over het nut van behandeling speelden daarbij mogelijk een rol. Door middel van het instrument BooG (Kordelaar, 2002) wordt sinds 2007 geprobeerd om in ieder geval voor het aanvragen van een rapportage landelijk dezelfde criteria aan te houden. De criteria van BooG zijn echter niet gerelateerd aan risicofactoren. Hoewel er langzamerhand meer aandacht komt voor de rol van risicotaxatie in de pro justitia rapportage, is de gestructureerde risicotaxatie, voor zover aanwezig, niet sturend of doorslaggevend en is het uiteindelijke oordeel van de individuele rapporteurs nog steeds overwegend klinisch van aard. Dit leidt tot behandelopleggingen die te weinig samenhangen met het recidiverisico. Het standaard inzetten van risicotaxatie instrumenten bij elke pleger kan ervoor zorgen dat iedereen langs dezelfde lat wordt gelegd en dat beslissingen over het inzetten van een intensieve en dure behandeling op een zo betrouwbaar mogelijke manier worden gemaakt, in overeenstemming zijn met het risico principe.

Risicotaxatie tijdens behandeling

Eenmaal in die behandeling, is de bruikbaarheid van de genoemde, relatief statische instrumenten een stuk beperkter. Hoewel het goed is om ook binnen verschillende behandelsettings goed zicht te hebben het basisrisico van de patiënten, gaat het in de behandeling vooral om het vaststellen van risicofactoren die zich lenen voor beïnvloeding, de dynamische risicotaxatie. Volgens het behoefte principe vloeien de behandeldoelen logischerwijs voort uit de dynamische risicofactoren. Deze dynamische vorm van risicotaxatie kan ook worden gebruikt om het effect van de behandeling na enige tijd te evalueren en uiteindelijk om te bepalen of en wanneer het risico voldoende afgenomen is om een patiënt uit te laten stromen, hoewel aan de empirische onderbouwing van dit gebruik nog hard gewerkt moet worden (Wong, Olver &

Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. *De Psycholoog*, september 2014, 29-39.

Nicholaichuk, 2013; De Vries Robbé, De Vogel, Douglas & Nijman, 2013). Het inventariseren van dynamische risicofactoren is een bewerkelijkere vorm van risicotaxatie en de instrumenten zijn veel moeilijker in te vullen. Het scoren van de items vergt vaak een (klinische) inschatting van de beoordelaar. De vraag: 'Heeft de patiënt het afgelopen half jaar veel impulsief gedrag laten zien?' is van een heel andere orde dan de vraag: 'Hoeveel eerdere veroordelingen heeft de patiënt op zijn strafblad staan?' Het gebruik van dynamische risicotaxatie-instrumenten vereist dan ook specifieke deskundigheid, training en grote consciëntieusheid van de beoordelaars. Zie voor een uitgebreidere beschrijving en richtlijnen voor zorgvuldig gebruik van risicotaxatie in de forensische behandelpraktijk De Vogel et al. (2013).

De dynamische risicotaxatie-instrumenten die gebruikt worden bij de forensische behandeling zijn veel moeilijker op hun waarde te toetsen dan eerder beschreven instrumenten. De dynamische items zijn vaak moeilijk met terugwerkende kracht te scoren, hetgeen het gebruikelijke retrospectieve onderzoek moeilijk maakt. Prospectief onderzoek is problematisch omdat de uitkomstmaat, recidive, zoveel mogelijk voorkomen moet worden. Het doel van het gebruik van deze instrumenten is immers het voorkomen van recidive door het bieden van adequaat risicomanagement (beveiliging en/of behandeling). Dat wil zeggen dat een hoog-risico patiënt doorgaans niet zal uitstromen en daardoor niet recidiveert en zijn hoge risico dus niet kan 'waarmaken'. Het is erg moeilijk om de voorspellende waarde van je instrument vast te stellen terwijl je datzelfde instrument gebruikt als richtlijn voor je handelen. Prospectief onderzoek heeft dan ook vaak niet recidive als uitkomstmaat, maar richt zich bijvoorbeeld op incidenten in de behandelinstelling, die waarschijnlijk niet één op één samenhangen met recidive. Maar alle beperkingen terzijde wijst het beschikbare onderzoek naar dit soort risicotaxatie er op dat het in elk geval duidelijk beter werkt dan *geen* risicotaxatie (Douglas & Reeves, 2010; De Vogel, 2005).

Bij het gebruik van deze instrumenten in behandelinstellingen bestaat natuurlijk wel het door Fazel e.a. opgemerkte risico dat mensen eindeloos vast blijven zitten vanwege hun indeling in de categorie 'hoog' terwijl dat absoluut geen 100% kans op

recidive inhoudt. Om dat te voorkomen is het noodzakelijk te zorgen dat bijvoorbeeld tbs behandelingen niet eindeloos voortduren. De wens van staatssecretaris Teeven om de behandelduur in de tbs met 20% terug te dringen lijkt hierbij aan te sluiten. Slechts in zeer uitzonderlijke gevallen mag een tbs uitmonden in een zogenaamde long stay, de levenslange preventieve opsluiting die vergelijkbaar is met wat in de VS gebeurt onder de Sexual Violent Predator Act. De focus kan beter verlegd worden naar een goede selectie van de instroom. De effectiviteit van de tbs is immers veel groter als alle hoog-risico delinquenten instromen en na een redelijk aantal jaren met een veel lager risico weer uitstromen. In plaats van dat, zoals nu, slechts een beperkt aantal hoog-risico delinquenten min of meer toevallig in de tbs terecht komen en er vervolgens bijna niet meer uit komen, terwijl een groot deel van de hoog-risico delinquenten volledig aan de aandacht ontsnapt.

Risicotaxatie: ook in het individuele geval!

Concluderend kan gesteld worden dat, hoewel risicotaxatie niet volmaakt is en verder onderzoek naar mogelijke verbetering noodzakelijk blijft, risicotaxatie wel bewezen waardevol is, zowel bij individuele strafzaken als tijdens behandeling. Het draagt bij aan rationele besluitvorming over welke plegers (intensieve) behandeling behoeven, en aan een gefundeerd oordeel over het effect van die behandelingen. Bij goed en verantwoord gebruik kan gestructureerde risicotaxatie een belangrijke bijdrage leveren aan een veiliger samenleving.

Tabel 1. Voorbeelden van items uit het risicotaxatie-instrument Static-99R

Jonge leeftijd
Geen duurzame intieme relaties
Eerdere veroordelingsdata
Eerdere seksuele delicten
Enig onbekend slachtoffer

Tabel 2. Recidivepercentages behorende bij Static-99 classificaties bij verschillende uitkomstmaten en follow up tijden.

Static-99 Classificatie	Bij de indeling volgens Fazel e.a.		Bij de gebruikelijke indeling			
	Percentage delinquenten	Seksuele recidive 5 jr	Percentage delinquenten	Seksuele recidive 5 jr	Geweld/ Seksuele recidive 5 jr	Geweld/ Seksuele recidive 10 jr
Laag	31%	5%	31%	5%	10%	16%
Laag-matig	69%	15%	42%	9%	17%	26%
Matig-hoog			18%	15%	27%	39%
Hoog			9%	28%	45%	57%
Totaal	100%	12%	100%	12%	22%	33%

Referenties

- Ægisdóttir, S., White, M. J., Spengler, P. M., Maugherman, A. S., Anderson, L. A., Cook, R. S.,...Rush, J. D. (2006). The meta-analysis of clinical judgment project: fifty-six years of accumulated research on clinical versus statistical prediction. *The Counseling Psychologist*, 34, 341-82.
- Andrews, D. A., & Bonta, J. (2006). *The psychology of criminal conduct (4th ed.)*. Newark, NJ: LexisNexis/Anderson.
- Andrews, D. A., & Bonta, J. (2010). *The psychology of criminal conduct (5th ed.)*. Newark, NJ: LexisNexis/Anderson.
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2004). *The Level of Service/Case Management Inventory (LS/CMI)*. Toronto, Canada: Multi-Health Systems.
- Andrews, D. A., Bonta, J., & Wormith, S. J. (2006). The recent past and near future of risk/need assessment. *Crime and Delinquency*, 52, 7–27.
- Bonta, J., Wallace-Capretta, S., & Rooney, J. (2000). A quasi-experimental evaluation of an intensive rehabilitation supervision program. *Criminal Justice and Behavior*, 27, 312–329.
- Cooney, M. T., Dubina, A., & Graham, I. (2009). Value and limitations of existing scores for the assessment of cardiovascular risk. *Journal of the American College of Cardiology*, 54:1209-27.
- Dawes, R. M. (2002). The ethics of using or not using statistical prediction rules in psychological practice and related consulting activities. *Philosophy of Science*, 69, S178-S184.
- Douglas, K.S. & Reeves, K.A. (2010). Historical-Clinical-Risk Management-20 (HCR-20) Violence risk assessment scheme: rationale, Application, and Empirical Overview. In R.K. Otto & K.S. Douglas (Eds.), *Handbook of violence risk assessment. International perspectives on forensic mental health* (p. 147-185). New York: Routledge.

- Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. *De Psycholoog*, september 2014, 29-39.
- Fazel, S., Singh, J.P., Doll, H., & Grann, M. (2012). Use of risk assessment instruments to predict violence and antisocial behaviour in 73 samples involving 24.827 people: Systematic review and meta-analysis. *British Medical Journal*, 345, e4692.
- French, S. A., & Gendreau, P. (2006). Reducing prison misconducts: What works! *Criminal Justice and Behavior*, 33, 185-218.
- Grove, W. M., Zald, D. H., Lebow, B. S., Snits, B. E., & Nelson, C. E. (2000). Clinical vs. mechanical pre-diction: A meta-analysis. *Psychological Assessment*, 12, 19-30.
- Hanson, R. K., Bourgon, G., Helmus, L., & Hodgson, S. (2009). The principles of effective correctional treatment also apply to sexual offenders: A meta-analysis. *Criminal Justice and Behavior*, 36, 865-891.
- Hanson, R. K., & Bussière, M. T. (1998). Predicting relapse: A meta-analysis of sexual offender recidivism studies. *Journal of Consulting and Clinical Psychology*, 66, 348–362.
- Hanson, R. K., Harris, A. J. R., Scott, T.-L., & Helmus, L. (2007). *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project* (User Report 2007-05). Ottawa, ON: Public Safety Canada.
- Hanson, R. K., & Morton-Bourgon, K. E. (2009). The accuracy of recidivism risk assessments for sexual offenders: A meta-analysis of 118 prediction studies. *Psychological Assessment*, 21, 1–21.
- Hanson, R.K., & Thornton, D. (2000). Improving risk assessment for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119-136.
- Hanson, R. K., & Thornton, D. (2003). *Notes on the development of Static-2002*. (Corrections Research User Report No. 2003-01). Ottawa, ON: Department of the Solicitor General of Canada.

- Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. *De Psycholoog*, september 2014, 29-39.
- Harte, J. & Breukink, M. (2010). Objectiviteit of schijnzekerheid? Kwaliteit, mogelijkheden en beperkingen van instrumenten voor risicotaxatie. *Tijdschrift voor Criminologie*, 52, 52-72.
- Helmus, L., Hanson, R. K., Thornton, D., Babchishin, K. M., & Harris, A. J. R. (2012a). Absolute recidivism rates predicted by Static-99R and Static-2002R sex offender risk assessment tools vary across samples: A meta-analysis. *Criminal Justice and Behavior*, 39, 1148-1171. doi:10.1177/0093854812443648
- Helmus, L., Thornton, D., Hanson, R. K., & Babchishin, K. M. (2012b). Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights. *Sexual Abuse: A Journal of Research and Treatment*, 24, 64-101.
- Janus, E. S., & Prentky, R. (2003). Forensic use of actuarial risk assessment with sex offenders: Accuracy, admissibility, and accountability. *American Criminal Law Review*, 40, 1443-1499.
- Kordelaar, W. F. J. M. van (2002). *BooG: Beslissingsondersteuning Onderzoek Geestvermogens in het strafrecht voor volwassenen*. Deventer, Nederland: Kluwer.
- Kordelaar, W. F. J. M. van. (2006). BooG 2.0, Eenduidigheid in het besluit over het onderzoek naar de persoon van de verdachte. In T.I. Oei & M.S. Groenhuisen (Eds.) *Capita Selecta forensische Psychiatrie 2006*. Deventer, Nederland: Kluwer.
- Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A metaanalysis of factors associated with effective treatment. *Journal of Experimental Criminology*, 1, 451-476.
- Lowenkamp, C. & Latessa, E. (2004). Understanding the risk principle: How and why correctional interventions can harm low-risk offenders. *Topics in Community Corrections*. Washington, D.C.: National Institute of Corrections, Department of Justice.

- Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. *De Psycholoog*, september 2014, 29-39.
- Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime & Delinquency*, 52, 77–93.
- Meehl, P. E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota Press.
- Michel, S. F., Riaz, M., Webster, C., Hart, S. D., Levander, S., Müller-Isberner, R., ... & Hodgins, S. (2013). Using the HCR-20 to Predict Aggressive Behavior among Men with Schizophrenia Living in the Community: Accuracy of Prediction, General and Forensic Settings, and Dynamic Risk Factors. *International Journal of Forensic Mental Health*, 12, 1-13. doi: 10.1080/14999013.2012.760182.
- Monahan, J. (2007). Clinical and actuarial predictions of violence. In D. Faigman, D. Kaye, M. Saks, J. Sanders, & E. Cheng (Eds.), *Modern scientific evidence: The law and science of expert testimony* (pp. 122–147). St. Paul, MN: West Publishing.
- Nederlandse Vereniging voor Psychiatrie (2012). *Richtlijn psychiatrisch onderzoek en rapportage in strafzaken*. Utrecht, Nederland: De Tijdstroom.
- Olver, M. E., Wong, S. C. P., & Nicholaichuk, T. P. (2009). Outcome evaluation of a high-intensity inpatient sex offender treatment program. *Journal of Interpersonal Violence*, 24, 522-536.
- Otto, R. K., & Douglas, K. S. (Eds.). (2010). *Handbook of violence risk assessment tools*. Milton Park, UK: Routledge.
- Phenix, A., Hanson, R. K., Harris, A. J. R., & Thornton, D. (2008) Detailed recidivism tables Static-99 (October 2008). Retrieved online on May 14 2013 from <http://www.static99.org/pdfdocs/detailedrecidivismtablesoctober2008.pdf>

- Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. *De Psycholoog*, september 2014, 29-39.
- Quinsey, V. L., Harris, G. T., Rice, M. E., & Cormier, C. A. (2006). *Violent offenders: Appraising and managing risk (2nd ed.)*. Washington, DC: American Psychological Association.
- Rice, M. E., & Harris, G. T. (2005). Violent recidivism: Assessing predictive validity. *Journal of Consulting and Clinical Psychology*, 63, 737-748.
- Rice, M. E., Harris, G. T., & Lang, C. (2013). Validation of and revision to the VRAG and SORAG: The Violence Risk Appraisal Guide-Revised (VRAG-R). *Psychological assessment*, 25, 951-965. doi: 10.1037/a0032878.
- Singh JP, Grann M, & Fazel S. A (2011). Comparative study of violence risk assessment tools: a systematic review and meta regression analysis of 68 studies involving 25 980 participants. *Clinical Psychology Review*, 31, 499 – 513.
- Smid, W. J., Kamphuis, J. H., Wever, E. C., & Beek, D. Van (2013a). Treatment Selection for Sex Offenders Based on Clinical Judgment vs. Actuarial Risk Assessment: Match and Analysis of Mismatch. *Journal of Interpersonal Violence*, 28, 2273–2289. doi:10.1177/0886260512475313.
- Smid, W. J., Kamphuis, J. H., Wever, E. C., & Beek, D. J. Van (2013b). *A quasi experimental evaluation of sex offender treatment and an exploration of associated risk factors*. Manuscript submitted for publication.
- Smid, W. J., Kamphuis, J. H., Wever, E. C., Beek, D. J. van, & Hoebe, I. (2013c). *Validation and Comparison of Nine Risk Assessment Instruments in a Dutch National Sample of Sex Offenders*. Manuscript submitted for publication.
- Smid, W. J., Kamphuis, J. H., Wever, E. C., & Verbruggen, M. C. F. M. (2013d). *Risk levels, treatment duration, and drop out in a clinically composed outpatient sex offender treatment group*. Manuscript submitted for publication.

- Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. *De Psycholoog*, september 2014, 29-39.
- Storey, J. E., Watt, K. A., Jackson, K. J., & Hart, S. D. (2012). Utilization and implications of the Static-99 in practice. *Sexual Abuse: A Journal of Research and Treatment*, 24, 289-302. doi: 10.1177/1079063211423943.
- Stuiveling, S. J., & Schoten, J. M. A. van (2012). *Prestaties in de strafrechtketen*. (Rapport van de algemene rekenkamer). Den Haag: Algemene Rekenkamer.
- 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 (beschikbaar: <http://dare.uva.nl/document/13725>).
- Vogel, V. de, Vries Robbé, M. de, Spa, E. de, & Wever, E. (2013). Risicotaxatie van geweld. *Maandblad Geestelijke volksgezondheid*, 68, 137 (volledige versie online: mgv.boomtijdschriften.nl/artikelen/GV-0-0-7.htm).
- Vries Robbé, M. de, Vogel, V. de, Douglas, K. S., & Nijman, H. L. I. (2013). *Changes in dynamic risk and protective factors for violence: Long-term recidivism reduction*. Manuscript submitted for publication.
- Wakeling, H. C., Mann, R. E., & Carter, A. J. C. (2012). Do Low-risk Sexual Offenders Need Treatment? *The Howard Journal of Criminal Justice*, 51, 286–299.
- Webster, C. D., Douglas, K. S., Eaves, D., & Hart, S. D. (1997). *HCR-20: Assessing risk for violence (Version 2)*. Vancouver, Canada: Simon Fraser University.
- Wilson, D. B., Bouffard, L. A., & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitivebehavioral programs for offenders. *Criminal Justice and Behavior*, 32, 172-204.
- Wong, S. C., Olver, M. E., & Nicholaichuk, T. P. (2013). Assessing risk change in sexual offender treatment using the Violence Risk Scale-sexual offender version: a brief overview. *Sexual Offender Treatment*, 8. Retrieved online on November 11 2013 from <http://www.sexual-offender-treatment.org/115.html>

Smid, W.J., Klein Haneveld, E., Vogel, V. de, & Horn, J. van (2014). Gestructureerde risicotaxatie noodzakelijker wijs. *De Psycholoog*, september 2014, 29-39.

Wormith, J. S., Hogg, S., & Guzzo, L. (2012). The Predictive Validity of a General Risk/Needs Assessment Inventory on Sexual Offender Recidivism and an Exploration of the Professional Override. *Criminal Justice and Behavior*, 39, 1511-1538