

Cahier 2007-5

Psychometrische kwaliteiten van de Recidive Inschattingsschalen (RISc)

**Interbeoordelaarsbetrouwbaarheid, interne
consistentie en congruente validiteit**

L. M. van der Knaap

L. E. W. Leenarts

L. T. J. Nijssen

Exemplaren van deze publicatie kunnen schriftelijk worden besteld bij

Bibliotheek WODC, kamer KO 14
Postbus 20301, 2500 EH Den Haag

Fax: (070) 370 45 07

E-mail: wodc@minjus.nl

Cahiers worden in beperkte mate gratis verspreid zolang de voorraad strekt

Alle nadere informatie over WODC-publicaties is te vinden op Justweb en op www.wodc.nl

Dankwoord

Voor u ligt het eindrapport van het onderzoek naar de interbeoordelaarsbetrouwbaarheid, interne consistentie en de congruente validiteit van de RISC. De uitvoering van dit onderzoek is mogelijk gemaakt door de inspanningen van velen. Op de eerste plaats zijn wij dank verschuldigd aan de reclasseringscliënten die hun medewerking hebben willen verlenen aan dit onderzoek. Zij ‘ondergingen’ een tweede afname van de RISC of vulden na afloop van de RISC-afname een vragenlijst in. Ook de reclasseringswerkers die de werving van de deelnemers, de afname van de tweede RISC’s en het verzamelen van ingevulde vragenlijsten voor hun rekening namen, hebben een onschatbare bijdrage geleverd aan dit onderzoek.

Veel dank zijn wij verschuldigd aan Eefje de Ruijter en Deborah Thompson voor hun werk als landelijk onderzoekscoördinator van het onderzoek naar de interbeoordelaarsbetrouwbaarheid respectievelijk de congruente validiteit van de RISC. Dankzij hun grote inzet is het gelukt om de dataverzameling van beide deelonderzoeken te laten slagen. Eefje de Ruijter kreeg daarbij de hulp van Agnes van Gent, Remco Roersch en Manja Sjardin, die als regiocoördinatoren een onmisbare rol speelden. Ook Deborah Thompson werd ondersteund door regiocoördinatoren: Ruud van Diemen, Greet Doornbos, Josien Leurdijk, Wim Veldhof, Ank Verdonk, Christine Wijshake en Rens Zijlstra. Medewerkers van de afdeling I&A van de RN zorgden steeds voor Excelbestanden met de RISC-data. Collega’s van het WODC — Ad Essers en Gijs Weijters — zijn behulpzaam geweest bij de verwerking van de onderzoeksdata.

De leden van de klankbordgroep (zie bijlage 1), ten slotte, hebben een zeer waardevolle bijdrage geleverd. Wij zijn hen veel dank verschuldigd voor hun bijzonder opbouwende wijze van advisering en begeleiding, maar zeker ook vanwege hun grote enthousiasme voor dit onderzoek.

De onderzoekers

Inhoud

Samenvatting	1
1 Inleiding	11
1.1 Achtergrond	11
1.2 RISC	12
1.3 Probleemstelling, doelstelling en onderzoeksvragen	14
1.4 Leeswijzer	16
2 Methoden van onderzoek	19
2.1 Inleiding	19
2.2 Het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISC	19
2.3 Het onderzoek naar de schaalstructuur en interne consistentie van de RISC	21
2.4 Het onderzoek naar de congruente validiteit van de RISC	22
2.4.1 Twee keer congruente validiteit	22
2.4.2 Congruente validiteit: RISC en de voorspelling van recidive	22
2.4.3 Congruente validiteit: de leefgebieden die door de RISC-schalen in kaart worden gebracht	24
3 Interbeoordelaarsbetrouwbaarheid	29
3.1 Inleiding	29
3.2 Beschrijving van de onderzoeksgroep	29
3.3 Kwaliteit van de data	32
3.4 Het analyseren van de interbeoordelaarsbetrouwbaarheid	33
3.5 Resultaten van de analyses van de interbeoordelaarsbetrouwbaarheid	35
3.5.1 Schaal 1&2 Delictgegevens	35
3.5.2 Schaal 3 Huisvesting en wonen	37
3.5.3 Schaal 4 Opleiding, werk en leren	38
3.5.4 Schaal 5 Inkomen en omgaan met geld	39
3.5.5 Schaal 6 Relaties met partner, gezin en familie	40
3.5.6 Schaal 7 Relaties met vrienden en kennissen	41
3.5.7 Schaal 8 Druggebruik	42
3.5.8 Schaal 9 Alcoholgebruik	43
3.5.9 Schaal 10 Emotioneel welzijn	43
3.5.10 Schaal 11 Denkpatronen, gedrag en vaardigheden	44
3.5.11 Schaal 12 Houding	45
3.5.12 Schaal- en totaalscores	47
4 Schaalstructuur en interne consistentie van de RISC	49
4.1 Het beoordelen van structuur en de interne consistentie van de RISC	49
4.2 Resultaten: de schaalstructuur en interne consistentie van de RISC	51
5 Congruente validiteit	57
5.1 Inleiding	57
5.2 Resultaten van de validering met behulp van een recidivevoorspellingsmodel	57

5.2.1	Selectie van de onderzoeksgroep	57
5.2.2	De relatie tussen de RISC-scores en StatRec: vergelijking van vier risicogroepen	59
5.2.3	Samenhang tussen de RISC-scores en StatRec: correlaties tussen RISC en StatRec	60
5.2.4	Multivariate analyses van de samenhang tussen de RISC-scores en StatRec	67
5.3	Congruente validiteit van de RISC-schalen 10, 11 en 12: schaalstructuur en interne consistentie van de voor validering gebruikte instrumenten	78
5.3.1	Onderzoeksgroep	78
5.3.2	Controle van de schaalstructuur van NPV, UCL en BDHI-D	82
5.4	Resultaten van de validering van RISC met behulp van NPV, UCL en DHI-D	83
5.4.1	Verwachte samenhangen tussen RISC-schalen en schalen van NPV, UCL en BDHI-D	83
5.4.2	Resultaten congruente validiteit schalen 10, 11 en 12	84
6	RISC en specifieke doelgroepen	89
6.1	De relatie tussen de RISC en StatRec bij specifieke doelgroepen	89
6.2	Correlaties tussen RISC en StatRec bij specifieke doelgroepen	91
6.3	Multivariate analyses van de samenhang tussen de RISC-scores en StatRec bij specifieke doelgroepen	95
7	Conclusies en aanbevelingen	99
7.1	De interbeoordelaarsbetrouwbaarheid, interne consistentie en congruente validiteit van de RISC	99
7.2	Aanbevelingen met betrekking tot de interbeoordelaarsbetrouwbaarheid van de RISC	101
7.3	Aanbevelingen met betrekking tot de structuur van de RISC	103
7.4	Conclusies naar aanleiding van het validiteitsonderzoek van de RISC	104
7.4.1	Congruente validiteit in termen van recidivevoorspelling	104
7.4.2	Congruente validiteit van de afzonderlijke RISC-schalen: meten ze wat ze beogen te meten?	105
7.4.3	Congruente validiteit in een aantal specifieke doelgroepen	105
7.4.4	Enkele opmerkingen naar aanleiding van het onderzoek naar de congruente validiteit in termen van recidivevoorspelling	106
7.5	Tot slot	107
	Summary	111
	Literatuur	121
Bijlage 1	Klankbordgroep	125
Bijlage 2	Factorladingen RISC-schalen	127

Samenvatting

1 Aanleiding, doel en onderzoeksvragen

Het instrument 'Recidive InschattingsSchalen' (RISc) is het diagnose-instrument van de Nederlandse reclasseringsorganisaties dat in 2002-2003 door Adviesbureau Van Montfoort is ontwikkeld in opdracht van de drie reclasseringsorganisaties — Reclassering Nederland (RN), Stichting Verslavingsreclassering GGZ Nederland (SVG) en Leger des Heils Jeugdzorg en Reclassering (LJ&R) — en het beleidsprogramma Terugdringen Recidive (TR). Voor de ontwikkeling van de RISc heeft de What Works benadering als uitgangspunt gediend (zie bijvoorbeeld McGuire, 1995). Deze benadering gaat ervan uit dat gedragsinterventies die erop gericht zijn om recidive terug te dringen, aangepast moeten zijn op het recidiverisico van een dader en dat zij gericht moeten zijn op het beïnvloeden van factoren die het delictgedrag bevorderen. In overeenstemming met deze uitgangspunten is het doel van de RISc om te komen tot een inschatting van de kans op recidive (gedefinieerd als een nieuwe veroordeling) en van de statische — niet te veranderen factoren zoals leeftijd, geslacht en justitiële voorgeschiedenis — en dynamische — in principe veranderbare en beïnvloedbare factoren — criminogene factoren die aan dit risico ten grondslag liggen. De RISc bestaat uit twaalf schalen die elk één van de criminogene factoren in kaart beogen te brengen: (1) Delictgeschiedenis; (2) Huidig delict en delictpatroon; (3) Huisvesting en wonen; (4) Opleiding, werk en leren; (5) Inkomen en omgaan met geld; (6) Relaties met partner, gezin en familie; (7) Relaties met vrienden en kennissen; (8) Druggebruik; (9) Alcoholgebruik; (10) Emotioneel welzijn; (11) Denkpatronen, gedrag en vaardigheden; en (12) Houding. Samen vormen deze schalen de totaalscore die het risico op recidive aangeeft.

De RISc is gebaseerd op het Britse Offender Assessment System (OASys; Howard, Clark & Garnham, 2003), het instrument dat door de reclassering en het gevangeniswezen in Engeland en Wales wordt gebruikt als een systeem van diagnostiek, indicatiestelling en sanctieplanning. OASys is aangepast aan de Nederlandse situatie en tussen april en september 2003 bij 465 delinquenten afgenomen. Op basis van de op deze manier verzamelde data is de eerste gebruikersversie van het instrument ontwikkeld (Adviesbureau Van Montfoort & Reclassering Nederland, 2004). Tijdens de ontwikkeling van de RISc heeft men zich geconcentreerd op de kwaliteit van de items en is aandacht besteed aan de interne consistentie van de schalen van het instrument. Verder onderzoek naar de betrouwbaarheid en de validiteit was echter, gezien het karakter van het instrument en het doel waarvoor het wordt ingezet, van het grootste belang. De RISc is immers een beoordelingsinstrument op basis waarvan belangrijke beslissingen over individuen worden genomen en het mag daarom voor de resultaten op de RISc niet uitmaken of het instrument door reclasseringswerker A of B wordt afgenomen. Anders geformuleerd moeten beoordelingen generaliseerbaar zijn over beoordelaars. Dit is de vraag naar de interbeoordelaarsbetrouwbaarheid van het instrument. Omdat op basis van de RISc verstrekkende beslissingen kunnen worden genomen over een delinquent, is het van het grootste belang dat het instrument daadwerkelijk meet wat het pretendeert te meten, namelijk recidiverisico en factoren die daaraan ten grondslag liggen. De vraag die hiermee wordt gesteld is de vraag naar de begripsvaliditeit van het in-

strument. Daarnaast is voor elk deugdelijk instrument een heldere, goed te interpreteren schaalstructuur essentieel. In de ontwikkelfase van de RISC is uitgegaan van een aantal a priori opgestelde schalen die tezamen een totaalscore vormen. Deze structuur diende in een grote onderzoeksgroep op basis van statistische analyses te worden getoetst. Dergelijke analyses lenen zich er tegelijkertijd voor om de interne consistentie van de RISC-totaalscore en de RISC-schalen in een grote groep te onderzoeken. Deze vorm van betrouwbaarheid geeft aan in hoeverre per schaal de score replicerbaar is over de items die deel uitmaken van die schaal

Het WODC is in 2005 gestart met een onderzoek naar de interbeoordelaarsbetrouwbaarheid, structuur, interne consistentie en de begripsvaliditeit van de RISC. Het doel van het onderzoek is om inzicht te krijgen in deze psychometrische kenmerken van de RISC, maar voornamelijk om aanbevelingen te doen om het instrument psychometrisch te optimaliseren. De volgende onderzoeksvragen stonden centraal:

- 1 Wat is de interbeoordelaarsbetrouwbaarheid van de RISC?
- 2 Hoe is de begripsvaliditeit van de RISC?
De vraag naar de begripsvaliditeit van het instrument wordt in dit onderzoek toegespitst op de congruente validiteit ofwel soortgenootvaliditeit: de mate waarin een test correleert met een andere test waarvan wordt aangenomen dat die dezelfde eigenschap meet (Drenth & Sijtsma, 1990).
- 3 Welke aanbevelingen ter verbetering van het instrument zijn op grond van de antwoorden op vraag 1 en 2 te doen?
- 4 Hoe is de schaalstructuur van de RISC en wat is de interne consistentie van de RISC-schalen en de RISC-totaalscore en welke aanbevelingen ter verbetering zijn met betrekking tot de structuur van de RISC te doen?

Op basis van de ervaringen die in het eerste jaar met de RISC zijn opgedaan, is door de drie reclasseringsorganisaties geconstateerd dat bij reclasseringswerkers de indruk bestaat dat de RISC bij een aantal specifieke doelgroepen tot een lagere inschatting van het recidiverisico leidt dan zij in de praktijk aannemelijk achten. Als dit daadwerkelijk het geval zou zijn, betekent dit dat de validiteit van het instrument voor deze doelgroepen onvoldoende zou zijn. Naar aanleiding hiervan is het WODC gevraagd het onderzoek naar de congruente validiteit uit te breiden met de volgende onderzoeksvraag:

- 5 Wat is de relatie tussen de RISC-totaalscore en de score op het recidivevoorspellingsmodel StatRec voor de volgende subgroepen:
 - a Plegers van huiselijk geweld
 - b Zedendelinquenten
 - c Veelplegers
 - d Oplichters
 - e Oudere ouders
 - f Bestuurders onder invloed

Na het bespreken van de voor de beantwoording van de onderzoeksvragen gehanteerde onderzoekopzet en een weergave van de resultaten per deelonderzoek, wordt aan het eind van deze samenvatting een overzicht gegeven van de belangrijkste conclusies en aanbevelingen.

2 Onderzoeksopzet

Het onderzoek naar de interbeoordelaarsbetrouwbaarheid, schaalstructuur, interne consistentie en de congruente validiteit van de RISc is in drie deelonderzoeken uitgevoerd.

De interbeoordelaarsbetrouwbaarheid van de RISc

Om de interbeoordelaarsbetrouwbaarheid van de RISc te onderzoeken is het instrument in de periode van november 2005 tot half mei 2006 bij een groep van 75 reclasseringscliënten twee maal afgenomen door twee verschillende reclasseringswerkers. Naast een reguliere (eerste) afname is de RISc nogmaals afgenomen door een tweede reclasseringswerker, die dit onafhankelijk van de eerste werker deed. De reclasseringswerkers werkten in vaste koppels, maar hadden onderling geen contact over de cliënt bij wie zij de RISc afnamen. De aan het onderzoek deelnemende reclasseringswerkers, die random waren geselecteerd door de onderzoekers, vroegen de cliënten bij wie zij een RISc afnamen of zij mee wilden werken aan het onderzoek. In totaal werkten negentien koppels van reclasseringswerkers mee aan het onderzoek: elf RN-koppels, vijf SVG-koppels en drie LJ&R-koppels. Wanneer een cliënt aangaf mee te willen werken aan het onderzoek, werd de collega-werker gewaarschuwd waarna die een afspraak maakte met de cliënt voor een tweede afname. Deze tweede afspraak diende ongeveer twee tot drie weken na de eerste afname te worden gepland. Cliënten kregen €25,- voor hun deelname aan het onderzoek.

De schaalstructuur en interne consistentie

Het onderzoek naar de schaalstructuur en de interne consistentie van de RISc is uitgevoerd op basis van een van de Reclassering Nederland (RN) verkregen database waarin alle RISc's zijn opgenomen die tussen november 2004 en mei 2006 door de drie reclasseringsorganisaties zijn gestart. Na de essentiële datacleaning konden de analyses uitgevoerd worden op 11.666 RISc's.

De congruente validiteit van de RISc

Aan het onderzoek naar de congruente validiteit van de RISc zijn twee onderdelen onderscheiden. Het belangrijkste doel van de RISc is om een inschatting van het risico op recidive te bepalen. De vraag of de RISc in voldoende mate correleert met een instrument dat eveneens het recidiverisico beoogt in te schatten, vormt het eerste deel van het onderzoek naar de congruente validiteit. Voor dit deel van het onderzoek is gebruik gemaakt van dezelfde database als bij de analyse van de schaalstructuur en de interne consistentie van de RISc. Voor 9.985 van de 11.666 RISc's uit deze database kon een score worden berekend op de StatRec-schaal. Dit gevalideerde instrument voorspelt op basis van een aantal statische achtergrondkenmerken de recidivekans van daders. Recidive is hierin gedefinieerd als het opnieuw in contact komen met justitie. Door de samenhang tussen beide instrumenten te onderzoeken, kon de congruente validiteit van de RISc-recidivevoorspelling worden vastgesteld. Voor het onderzoek naar de specifieke doelgroepen is volgens dezelfde onderzoeksopzet gewerkt.

Het tweede deel van het onderzoek naar de congruente validiteit van de RISc heeft betrekking op de inhoud van de RISc-schalen. Met behulp van deze schalen worden specifieke leefgebieden in kaart gebracht en de aan- of afwezigheid van criminogene problemen op deze gebieden wordt gebruikt om een indicatie te geven

voor het aanbod van de reclassering, bijvoorbeeld in de vorm van gedragsinterventies. De vraag of deze schalen daadwerkelijk de concepten meten die ze beogen te meten, is in het tweede deel van het validiteitsonderzoek bestudeerd. Het valideren van alle RISC-schalen bleek een te grote belasting te zijn voor de reclasseringsorganisaties en bovendien bestonden er inhoudelijke problemen met betrekking tot het valideren van de meer feitelijke RISC-schalen zoals huisvesting, opleiding / werk en financiën. Het lag namelijk het meest voor de hand om dergelijke schalen te valideren met behulp van dossierinformatie, maar omdat de reclasseringswerker bij het afnemen en invullen van de RISC gebruik maakt van dezelfde dossierinformatie, was dat geen mogelijkheid. Uiteindelijk is in overleg met de drie reclasseringsorganisaties besloten de drie minst feitelijke en meest subjectieve schalen uit het instrument te valideren. Het gaat daarbij om de schalen 10 *Emotioneel welzijn*, 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding*. Dit zijn bovendien schalen die een belangrijke rol spelen in de beoordeling van de vraag of een cliënt een gedragsinterventie zou moeten volgen. Om de congruente validiteit van deze schalen te bestuderen, is op zoek gegaan naar betrouwbare en gevalideerde instrumenten die (delen van) dezelfde concepten meten. Drie vragenlijsten — de Nederlandse Persoonlijkheidsvragenlijst (NPV), Utrechtse Copinglijst (UCL) en de Buss-Durkee Hostility Inventory - Dutch (BDHI-D) — werden voor het onderzoek geselecteerd. Reclasseringswerkers vroegen hun cliënten na afloop van de afname van de RISC om medewerking aan het onderzoek naar de congruente validiteit. Iedere cliënt vulde slechts één van de drie geselecteerde vragenlijsten in. Tijdens het invullen van de vragenlijst was de reclasseringswerker aanwezig voor eventuele uitleg. Als de cliënt klaar was met invullen, controleerde de werker of alle vragen waren ingevuld en of geen bladzijden waren overgeslagen. Voor hun medewerking aan het onderzoek ontvingen cliënten €10,-. De vragenlijsten zijn tussen april en half september 2006 door 204 reclasseringscliënten ingevuld. In een aantal gevallen was de RISC van de cliënt onvindbaar, onbruikbaar of op de sluitingsdatum van de dataverzamelingsperiode nog niet afgesloten. In de totale validiteitsonderzoeksgroep resteren daarom 185 personen.

3 Resultaten van het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISC

Negentien koppels van reclasseringswerkers hebben de RISC onafhankelijk van elkaar twee maal afgenomen bij 75 reclasseringscliënten. De eerste RISC-afname werd gevolgd door de tweede RISC-afname met een tussenpoos van gemiddeld 29 dagen. Nadat is vastgesteld dat de onderzoeksgroep met betrekking tot hun achtergrondkenmerken en hun scores op de RISC in grote mate overeenkomt met de reguliere reclasseringspopulatie is de kwaliteit van de data in termen van missende waarden onderzocht. Deze bleek goed te zijn. De interbeoordelaarsbetrouwbaarheid van de RISC is zowel op itemniveau als op schaalniveau onderzocht. Ook de interbeoordelaarsbetrouwbaarheid van de RISC-totaalscore is bestudeerd. De gebruikte analysetechnieken om de interbeoordelaarsbetrouwbaarheid van de RISC-items en —schalen vast te stellen, zijn afgestemd op het meetniveau van items en schalen. Om de mate van overeenstemming tussen de reclasseringswerkers op *nominale* items te onderzoeken, is coëfficiënt κ berekend (Cohen, 1960). Voor *ordina-*le items, de schaalcores en de totaalscore is de volgende strategie gebruikt (cf. Born, 1995, p. 130-132).

- 1 Het berekenen van de proportie overeenstemming. Deze maat is de meest gebruikte en eenvoudigste maat, maar hiermee kon niet worden volstaan. Het ontbreekt namelijk zowel aan een controle voor kansovereenstemming als aan een formele toetsing van de mate van overeenstemming.
- 2 Het berekenen van de χ^2 -maat van Lawlis en Lu (1972). Deze maat geeft aan of de overeenstemming significant groter is dan op basis van toeval te verwachten is (kansovereenstemming).
- 3 T-maat van Tinsley en Weiss (1975) berekenen. Deze maat is een afgeleide van de χ^2 -maat van Lawlis en Lu en geeft een indicatie van de grootte van de overeenstemming (0=de overeenstemming is niet groter dan op basis van toeval is te verwachten, 1=perfecte overeenstemming).

In alle gevallen is op basis van de resultaten een uitspraak gedaan over de mate van overeenstemming tussen de reclasseringswerkers. Met betrekking tot de items van de RISc is de mate van overeenstemming tussen reclasseringswerkers over het algemeen redelijk tot sterk en ook op schaalniveau is de overeenstemming tussen de beoordelaars in alle gevallen redelijk tot sterk. Met betrekking tot de totaalscore van de RISc komen reclasseringswerkers in sterke mate tot hetzelfde oordeel. De interbeoordelaarsbetrouwbaarheid van de RISc is op basis van deze resultaten goed te noemen. Wel valt op dat de interbeoordelaarsbetrouwbaarheid van items afneemt naarmate minder naar feitelijkheden wordt gevraagd en dat een groot aantal items van de schalen 11 en 12 slechts een matige overeenstemming laat zien.

4 Resultaten van het onderzoek naar de schaalstructuur en de interne consistentie van de RISc

Om de schaalstructuur van het instrument te onderzoeken, is steeds voor elke schaal en voor de totaalscore een principale componentenanalyse (pca) uitgevoerd op de bijbehorende items respectievelijk schalen. Voor elke schaal en voor de totaalscore is nagegaan in hoeverre de pca één factor opleverde en zijn de factorloadingen bestudeerd. Op deze manier werd nagegaan of de items respectievelijk schalen eenzelfde achterliggend begrip in kaart brengen. Daarnaast is voor elke schaal en voor de totaalscore een betrouwbaarheidsanalyse uitgevoerd. De interne consistentie van elke schaal en van de totaalscore is berekend met behulp van Cronbach's coëfficiënt alfa.

De RISc-schalen vormen over het algemeen goede schalen: de meeste items van elke schaal meten ieder een deel van hetzelfde achterliggende concept dat de schaal in kaart beoogt te brengen en de scores zijn in de meeste schalen voldoende replicerbaar over de items. Dat wil zeggen dat met betrekking tot de meeste schalen sprake is van een betrouwbare meting. Dit blijkt uit de interne consistentie van de schalen, die voor de meeste schalen voldoende tot goed is. De uitzonderingen worden gevormd door de schalen 6 *Relaties met partner, gezin en familie* en 9 *Alcoholgebruik*. Deze schalen bestaan uit items die in onvoldoende mate eenzelfde achterliggend begrip in kaart brengen en die onvoldoende betrouwbaar zijn in hun meting (onvoldoende interne consistentie).

De analyse van de RISc-totaalscore toont aan dat ook de RISc-schalen samen een goede schaal vormen: ze laden over het algemeen goed op de geëxtraheerde factor en meten dus elk een deel van hetzelfde achterliggende begrip. De enige schaal die

weinig bijdraagt aan de totaalscore is schaal 9 *Alcoholgebruik*. Aan de interne consistentie van de totaalscore worden strengere eisen gesteld dan aan de interne consistentie van de afzonderlijke RISC-schalen omdat op basis van deze score belangrijkere beslissingen over individuele daders worden genomen dan op basis van de afzonderlijke schaalcores. De betrouwbaarheidsanalyse van de totaalscore toont aan dat aan deze strengere eisen wordt voldaan.

5 Resultaten van het onderzoek naar de congruente validiteit van de RISC

Ten behoeve van het onderzoek naar de congruente validiteit van de RISC in termen van recidivevoorspelling is de samenhang tussen RISC en de StatRec-schaal bestudeerd. Dit is op verschillende manieren gedaan. Als eerste stap is de onderzoeksgroep op basis van hun RISC-totaalscore in vier risicogroepen verdeeld en is nagegaan in hoeverre de StatRec-scores van de vier groepen significant van elkaar verschillen. Vervolgens zijn correlaties berekend tussen RISC en StatRec en tot slot zijn regressieanalyses uitgevoerd.

De resultaten van het onderzoek naar de congruente validiteit van de RISC in termen van de voorspelling van de recidivekans zijn gunstig. Er werd, conform de verwachting, een sterk verband gevonden tussen de RISC-totaalscore, die een inschatting beoogt te geven van de recidivekans, en de gevalideerde voorspelling van de recidive zoals deze met StatRec wordt berekend. Op basis van het uitgangspunt van de RISC — elke schaal houdt verband met (het risico op) delictgedrag — werd verwacht dat alle correlaties tussen de afzonderlijke RISC-schalen en de StatRec-recidivekans middelsterk positief zouden zijn. In de meeste gevallen werd deze verwachting in de totale onderzoeksgroep ondersteund. De uitzonderingen worden gevormd door de schalen 6 *Relaties met partner, gezin en familie*, 9 *Alcoholgebruik* en 10 *Emotioneel welzijn*. De samenhang tussen deze schalen en de StatRec-voorspelling van recidive is zwak. De totale onderzoeksgroep is op basis van haar achtergrondkenmerken — geslacht, leeftijd, herkomst en uitgangsdelict — verdeeld in een aantal subgroepen waarvoor de samenhang tussen RISC en StatRec eveneens is bestudeerd. Het beeld dat uit deze analyses naar voren komt, is vergelijkbaar met dat uit de analyse van de totale onderzoeksgroep.

Regressieanalyses zijn uitgevoerd om na te gaan in hoeverre de score op de StatRec te verklaren is op basis van de RISC-schalen. De resultaten laten zien dat de RISC-schalen samen 48% van de variantie van de StatRec-recidivekans verklaren. Dit is een zeer aanzienlijk deel en deze bevinding ondersteunt dan ook de congruente validiteit van de RISC. Ook in vrijwel alle subgroepen wordt een aanzienlijk deel van de StatRec-recidivekans verklaard door de RISC-schalen. De groep vrouwen vormt hierop een uitzondering. De mate waarin de StatRec-score van deze groep wordt verklaard door hun scores op de RISC is weliswaar groot, maar aanzienlijk minder groot dan bij mannen het geval is. Hoewel het, gezien de onderlinge samenhang tussen de afzonderlijke RISC-schalen, niet reëel was om te verwachten dat van elke afzonderlijke schaal de unieke bijdrage aan de verklaring van de StatRec-recidivekans kon worden bepaald, zijn enkele opmerkelijke resultaten gevonden. De schalen 4 *Opleiding, werk en leren* en 7 *Relaties met vrienden en kennissen* leveren, naast schaal 1&2 *Delictgegevens*, in de totale onderzoeksgroep een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-score. Dit betekent dat deze leefgebieden, ondanks de samenhang die ze vertonen met de andere leefgebieden, een eigen invloed hebben op de verklaring van de StatRec-score. De

invloed van opleiding en relaties met vrienden en kennissen op de verklaring van de op basis van StatRec berekende recidivekans is ook in veel van de onderzochte subgroepen gevonden. Ook schaal 8 *Druggebruik* blijkt in een aantal subgroepen een eigen bijdrage te leveren aan de verklaring van de StatRec-recidivekans. Daarnaast bestaan er verschillen tussen de bestudeerde subgroepen in de aard van de dynamische factoren die, ondanks de onderlinge samenhang tussen de RISC-schalen, een unieke bijdrage leveren aan de verklaring van de StatRec-voorspelling van de recidivekans. Ook is er onderscheid te maken tussen de diverse groepen in de sterkte van de effecten van RISC-schalen op de verklaring van de StatRec-score.

Ook in het kader van het onderzoek naar de congruente validiteit van de RISC is de samenhang bestudeerd tussen scores op de schalen 10 *Emotioneel welzijn*, 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding* van de RISC en de scores op de Nederlandse Persoonlijkheidsvragenlijst (NPV), Utrechtse Copinglijst (UCL) en de Buss-Durkee Hostility Inventory-Dutch (BDHI-D). Voorafgaand aan de analyses is een aantal verwachtingen opgesteld met betrekking tot de samenhang tussen de RISC-schalen en bepaalde, geselecteerde schalen van de NPV, UCL en BDHI-D. De uitkomsten van het onderzoek naar de congruente validiteit van de schalen 10 *Emotioneel welzijn*, 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding* laten een gunstig beeld zien van de congruente validiteit van de twee eerstgenoemde schalen. De verwachte, middelsterke samenhang wordt gevonden tussen Schaal 10 *Emotioneel welzijn* en alle schalen die een soortgelijk begrip meten. Ditzelfde geldt ook voor schaal 11 *Denkpatronen, gedrag en vaardigheden*. Dit vormt een onderbouwing voor de aanname dat deze schalen de begrippen in kaart brengen die ze beogen te meten. Schaal 12 *Houding* vertoont daarentegen met twee van de drie onderzochte schalen niet de verwachte, middelsterke samenhang. Ondanks de kanttekeningen die bij dit deelonderzoek zijn te plaatsten en ondanks het feit dat tussen schaal 12 en de derde van de onderzochte schalen wel de verwachte, middelsterke samenhang werd gevonden, kunnen de resultaten van dit onderzoek twijfels opwekken over de vraag of de RISC-schaal 12 er voldoende in slaagt de houding van de reclasseringscliënt ten opzichte van andere mensen, de maatschappij, het delict en criminaliteit in het algemeen in kaart te brengen.

6 Resultaten van het onderzoek naar de samenhang tussen RISC en StatRec in een aantal specifieke doelgroepen

Omdat onder reclasseringswerkers de indruk bestond dat de RISC in het geval van een aantal specifieke doelgroepen een te laag recidiverisico inschat, is de samenhang tussen RISC en StatRec in deze groepen afzonderlijk bestudeerd. Zowel RISC als StatRec beoogt de kans in te schatten dat een dader opnieuw met justitie in aanraking komt. De verwachting is dat de RISC-totaalscore en de StatRec-voorspelling van de recidivekans ook in de bestudeerde specifieke doelgroepen een sterke samenhang vertonen. Dit blijkt voor alle groepen, behalve voor de groep veelplegers (daders met elf of meer eerdere strafzaken) te worden bevestigd. In deze groep is de correlatie tussen RISC en StatRec middelsterk. De afzonderlijke RISC-schalen die dynamische criminogene factoren in kaart beogen te brengen, worden verwacht een middelsterke samenhang met de StatRec-score te vertonen. Deze verwachting wordt over het algemeen bevestigd. Net als in de totale onderzoeksgroep vormen de schalen 6 *Relaties met partner, gezin en familie*, 9 *Alcoholge-*

bruik en 10 *Emotioneel welzijn* de belangrijkste uitzonderingen. Deze schalen vertonen in alle subgroepen zwakke correlaties met de door StatRec voorspelde recidivekans.

Regressieanalyses wijzen uit dat de StatRec-score in de groep bestuurders onder invloed en de groep oplichters het best door de RISC wordt verklaard. Ook in de groep plegers van huiselijk geweld en de groep zedendelinquenten verklaart de RISC een aanzienlijk deel van de variantie van de StatRec-voorspelling van het recidiverisico. De RISC-schalen verklaren de minste variantie van de StatRec-score bij de groep veelplegers. Hoewel dit overeenkomt met een sterke correlatie, slaagt RISC er beduidend minder goed dan in de andere groepen in om de StatRec-voorspelling van de recidivekans te verklaren.

Van de dynamische RISC-schalen levert, ondanks de samenhang met alle andere RISC-schalen, in de meeste groepen de schaal 4 *Opleiding, werk en leren* een noemenswaardig unieke bijdrage aan de verklaring van de StatRec-score. De schalen 5 *Inkomen en omgaan met geld* en 9 *Alcoholgebruik* leveren elk in drie van de zes bestudeerde groepen een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-voorspelling van de recidivekans. Tussen de specifieke subgroepen worden ook enkele verschillen gevonden in de aard van de dynamische factoren die een unieke bijdrage leveren aan de verklaring van de StatRec-voorspelling van de recidivekans.

Over het algemeen leveren de analyses geen aanwijzingen op dat de RISC voor de onderzochte doelgroepen een te laag recidiverisico inschatten. Dit geldt echter niet voor de groep veelplegers. Afgaande op het op basis van StatRec ingeschatte recidiverisico van deze groep lijkt het erop dat reclasseringswerkers gelijk hebben wanneer zij stellen dat RISC bij sommige veelplegers tot een te lage inschatting van de recidivekans komt. Veelplegers die volgens RISC een laag of een middelhoog recidiverisico hebben, hebben namelijk op basis van StatRec een recidivekans van 66% respectievelijk 76%.

7 Conclusies en aanbevelingen

De resultaten van de drie uitgevoerde studies naar de interbeoordelaarsbetrouwbaarheid, de structuur en de congruente validiteit van de RISC in ogenschouw nemend, luidt de algemene conclusie dat de RISC in termen van de onderzochte kenmerken over gunstige psychometrische kwaliteiten beschikt. Om het instrument verder te verbeteren zijn op verschillende punten aanpassingen mogelijk en vervolgonderzoek, met name naar de predictieve validiteit van het instrument, is nodig. Voor het vergroten van de interbeoordelaarsbetrouwbaarheid van de RISC worden de volgende aanbevelingen gedaan:

- Formuleer de invulinstructies van de RISC nog eenduidiger en voorkom op die manier mogelijk bestaande ambivalentie en verschillen in interpretatie door reclasseringswerkers.
- Overwogen dient te worden de antwoordschalen aan te passen aan de inhoud van het item. De antwoordschaal van moeilijk in te schatten items dient ruimte te bieden voor nuance en antwoordschalen met slechts twee antwoordcategorieën dienen in dat soort gevallen te worden vermeden.
- Geadviseerd wordt om nog meer aandacht te besteden aan opleiding en deskundigheidsbevordering voor reclasseringswerkers.

- De matige interbeoordelaarsbetrouwbaarheid van de items van schalen 11 en 12 dienen aangepakt te worden. Op de eerste plaats door middel van deskundigheidsbevordering, maar daarnaast kan worden overwogen de scoringsinstructies te onderwerpen aan een onderzoek door een ervaren gedragsdeskundige.

De kwaliteit van de structuur van de RISC kan op een aantal manieren worden verbeterd. Door de items 2.10 *Verantwoordelijkheid nemen voor het delictgedrag*, 6.4 *Familie- of gezinslid heeft een justitieel dossier*, 8.5 *Motivatie om druggebruik aan te pakken* en 9.5 *Motivatie om alcoholgebruik aan te pakken* niet meer te betrekken in de berekening van de score op de schalen waar ze bij horen, kan de kwaliteit van deze schalen worden verbeterd. Hiermee wordt niet aangeraden deze items uit de RISC te verwijderen, maar te zoeken naar andere mogelijkheden om de informatie die met deze items wordt verzameld, nuttiger in te zetten. Aangeraden wordt om item 2.10 bij schaal 12 *Houding* in te delen. De items 8.5 en 9.5 zouden behouden kunnen worden als aparte indicatoren voor motivatie en zouden opgenomen kunnen worden in een aparte 'motivatieschaal'. Onderzoek naar de samenhang tussen item 6.4 en daadwerkelijke recidive zal uitsluitend moeten geven over de feitelijke waarde van dit item. Met betrekking tot de totaalscore worden geen aanbevelingen gedaan. Hoewel verbeteringen mogelijk zijn, wordt aangeraden eerst op schaalniveau aanpassingen te doen en vervolgens na te gaan welk effect dit op de psychometrische kenmerken van de totaalscore heeft.

De congruente validiteit van de RISC in termen van de voorspelling van recidive is goed. De RISC-totaalscore, die nieuwe contacten met justitie beoogt te voorspellen, hangt zowel in de totale onderzoeksgroep als in vrijwel alle bestudeerde subgroepen in sterke mate samen met de StatRec-voorspelling van de recidivekans. Uit de resultaten van de regressieanalyses blijkt dat de score op de StatRec in de totale onderzoeksgroep en in vrijwel alle onderzochte subgroepen goed te verklaren is op basis van de scores op de RISC-schalen tezamen. Er zijn voorts op basis van de resultaten in dit onderzoek, met uitzondering van de groep veelplegers, geen aanwijzingen dat de RISC in bepaalde groepen tot een te lage inschatting komt van de recidivekans. Deze resultaten vormen een overtuigende onderbouwing voor de congruente validiteit van de totaalscore van de RISC. Hier moet echter wel bij worden aangetekend dat de RISC is gevalideerd met behulp van een *voorspelling van recidive*. De samenhang tussen RISC en StatRec, bestudeerd met behulp van correlaties en regressieanalyses, geeft dan ook geen uitsluitend over de daadwerkelijke voorspellingskracht van de RISC en onderzoek naar de mate waarin de RISC samenhangt met daadwerkelijke recidive is daarom nodig. Speciale aandacht zal hierbij uit moeten gaan naar de voorspelling van recidive bij vrouwen en veelplegers. De congruente validiteit van de RISC is bij vrouwen niet in het geding, maar de RISC lijkt er toch wat minder goed in te slagen de recidivekans te verklaren. Voor veelplegers zijn er aanwijzingen dat de RISC ten onrechte een te laag recidive-risico inschat.

De vraag of de schalen 10, 11 en 12 meten wat ze beogen te meten, kan in het geval van de schalen 10 *Emotioneel welzijn* en 11 *Denkpatronen, gedrag en vaardigheden* vrij gunstig worden beantwoord. Vervolgonderzoek naar de congruente validiteit van schaal 12 *Houding* is echter aan te bevelen omdat de resultaten uit het onderhavige onderzoek niet eenduidig waren. Daarnaast zal de congruente validiteit

van de overige schalen van de RISC in een of meerdere vervolgstudies moeten worden onderzocht.

Wanneer de resultaten van de drie deelonderzoeken tezamen worden overzien, kan worden geconcludeerd dat er twee RISC-schalen zijn waarbij vraagtekens kunnen worden geplaatst met betrekking tot hun waarde voor het instrument: schaal 6 *Relaties met partner, gezin en familie* en schaal 10 *Emotioneel welzijn*. Van schaal 6 kon zowel de betrouwbaarheid (interne consistentie) als de congruente validiteit in dit onderzoek niet worden ondersteund. De congruente validiteit van schaal 10 in termen van *inhoud* (emotioneel welzijn) wordt weliswaar ondersteund, maar de schaal hangt zowel in de totale onderzoeksgroep als in alle onderzochte subgroepen slechts zwak samen met de voorspelling van recidive. De congruente validiteit van deze schaal in termen van de samenhang met de recidivekans kan in dit onderzoek dan ook niet worden onderbouwd.

1 Inleiding

1.1 Achtergrond

Het instrument ‘Recidive InschattingsSchalen’ (RISc) is het diagnose-instrument van de Nederlandse reclasseringsorganisaties waarmee reclasseringswerkers het recidiverisico van delinquenten inschatten en in kaart brengen welke zogeheten criminogene factoren (kenmerken en omstandigheden van personen die samenhangen met delictgedrag) aan dit risico ten grondslag liggen. Het is ontwikkeld in het kader van het in 2002 gestarte beleidsprogramma ‘Terugdringen Recidive’ (TR). Dit programma heeft tot doel recidive bij volwassen veroordeelden te verminderen. Het programma introduceerde hiertoe een geïntegreerde aanpak waarin verschillende methoden worden gebruikt om recidive te verminderen: een wetenschappelijk onderbouwd diagnose-instrument (RISc), gedragsinterventies die aan kwaliteitscriteria voldoen, een goede samenwerking tussen de reclasseringsorganisaties en het gevangeniswezen, nazorg en een structurele evaluatie van de effecten van de aanpak (zie voor meer informatie www.justitie.nl/recidive).

In 2002-2003 heeft Adviesbureau Van Montfoort in opdracht van TR en de drie reclasseringsorganisaties — Reclassering Nederland (RN), Stichting Verslavingsreclassering GGZ Nederland (SVG) en Leger des Heils Jeugdzorg en Reclassering (LJ&R) — de ontwikkeling van de RISc op zich genomen. Een experimentele versie van de RISc is bij 465 delinquenten afgenomen waarna met behulp van itemanalyses en betrouwbaarheidsanalyses het definitieve instrument is ontwikkeld. Tijdens de ontwikkeling van de RISc is aandacht besteed aan de kwaliteit van de items en aan de interne consistentie van de schalen van het instrument. Verder onderzoek naar de betrouwbaarheid en validiteit van de RISc was echter nodig omdat deze belangrijke psychometrische kwaliteiten van het instrument vooralsnog onbekend waren. In het najaar van 2005 is het WODC hiertoe gestart met een onderzoek naar de interbeoordelaarsbetrouwbaarheid en de begripsvaliditeit van het instrument. Dit onderzoek zal in 2008 gevolgd worden door een studie naar de predictieve validiteit van de RISc wanneer zal worden nagegaan in hoeverre het instrument erin slaagt daadwerkelijke recidive te voorspellen.

In dit rapport wordt verslag gedaan van de resultaten van het onderzoek naar de interbeoordelaarsbetrouwbaarheid en de begripsvaliditeit van de RISc. Dit eerste hoofdstuk geeft een nadere toelichting op het instrument, beschrijft de probleemstelling, doelstelling en onderzoeksvragen van deze studie en sluit af met een leeswijzer. Hoofdstuk 2 behandelt de methoden van onderzoek waarna in hoofdstuk 3 de resultaten van het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISc worden beschreven. In hoofdstuk 4 wordt ingegaan op de interne consistentie en de schaalstructuur van het instrument en hoofdstuk 5 beschrijft de resultaten van het onderzoek naar de begripsvaliditeit. Het zesde en laatste hoofdstuk gaat in op de conclusies die op basis van de resultaten getrokken kunnen worden en geeft aanbevelingen voor de doorontwikkeling van de RISc.

1.2 RISc

Voor de ontwikkeling van de RISc heeft de What Works benadering als uitgangspunt gediend (zie bijvoorbeeld McGuire, 1995). Deze benadering, die onder meer in Canada en het Verenigd Koninkrijk veel navolging vindt, gaat ervan uit dat gedragsinterventies die erop gericht zijn om recidive terug te dringen, aangepast moeten zijn op het recidiverisico van een dader en dat zij gericht moeten zijn op het beïnvloeden van factoren die het delictgedrag bevorderen. Daarnaast moeten interventies passen bij de (on)mogelijkheden van een dader om te kunnen profiteren van die interventie (responsiviteit). Met andere woorden, de interventie moet afgestemd zijn op de kenmerken van de dader en dus ‘passen’ bij de dader. In overeenstemming met deze uitgangspunten is het doel van de RISc om te komen tot een inschatting van de kans op recidive (gedefinieerd als een nieuwe veroordeling) en van de statische — niet te veranderen factoren zoals leeftijd, geslacht en justitiële voorgeschiedenis — en dynamische — in principe veranderbare en beïnvloedbare factoren — criminogene factoren die aan dit risico ten grondslag liggen¹. Tevens levert de RISc een inschatting op van het gevaarsrisico (voor de dader zelf of diens omgeving) en geeft het instrument een indruk van de mate waarin een dader beïnvloedbaar is door middel van gedragsinterventies die het recidiverisico beogen te verminderen.

Als eerste stap in de ontwikkeling van de RISc is een inventarisatie uitgevoerd van bestaand forensisch instrumentarium dat in de Nederlandse reclasseringspraktijk bruikbaar zou kunnen zijn (Adviesbureau Van Montfoort & Reclassering Nederland, 2004). Op basis van deze inventarisatie en naar aanleiding van interviews die de ontwikkelaars van de RISc in het kader van een nulmeting hebben gehouden, is ervoor gekozen de RISc te baseren op het Britse Offender Assessment System (OASys; Howard, Clark & Garnham, 2003). Dit instrument is op zijn beurt ontwikkeld op basis van de instrumenten Level of Service Inventory-Revised (LSI-R; Andrews & Bonta, 1996) en Assessment Case management and Evaluation System (ACE; Gibbs, 1999). OASys is het instrument dat door de reclassering en het gevangeniswezen in Engeland en Wales wordt gebruikt als een systeem van diagnostiek, indicatiestelling en sanctieplanning. Hoewel het doel van de RISc beperkter was — sanctieplanning behoort niet tot de doelen van de RISc — vormde OASys een werkbaar uitgangspunt. Het instrument is aangepast aan de Nederlandse situatie, waarbij scoringssysteem en opbouw zoveel mogelijk intact zijn gehouden (Adviesbureau Van Montfoort & Reclassering Nederland, 2004). In de periode april tot en met september 2003 is de testversie van de RISc bij 465 delinquenten afgenomen. Op basis van de op deze manier verzamelde data is de eerste gebruikersversie van het instrument ontwikkeld (Adviesbureau Van Montfoort & Reclassering Nederland, 2004).

De RISc is een instrument dat door een daartoe gecertificeerde reclasseringswerker wordt ingevuld. Hij of zij beoordeelt in een gesprek met de delinquent in hoeverre er problemen zijn op het gebied van een twaalfstal criminogene factoren. Naast informatie uit het gesprek, dient de beoordelaar zich ook te baseren op dossierinformatie. Het instrument bestaat uit drie onderdelen: basisdiagnostiek, verdie-

¹ Voor een uitgebreid overzicht van onderzoek naar statische en dynamische factoren in de verklaring van recidive verwijzen we naar: Gendreau, Little & Goggin, 1996; Hanson & Harris, 2000; Abracen, Mailloux, Serin et al., 2004; Hanson & Morton-Bourgon, 2005; Philipse, Koeter, van der Staak et al., 2006.

pingsdiagnostiek en indicatiestelling. De basisdiagnostiek vormt het eerste niveau van de RISC en wordt altijd doorlopen. Dit is het deel van het instrument op basis waarvan het recidiverisico wordt ingeschat en waar de criminogene factoren in kaart worden gebracht. Mocht de afname van dit onderdeel nog vragen onbeantwoord laten, dan kan de reclasseringswerker ervoor kiezen verdiepingsdiagnostiek uit te (laten) voeren. In tegenstelling tot dit tweede onderdeel van het instrument, dat niet voor alle delinquenten wordt doorlopen, wordt het derde niveau van de RISC voor iedere delinquent ingevuld. In dit deel van het instrument wordt een indicatie gegeven voor het aanbod van de reclassering, bijvoorbeeld in de vorm van gedragsinterventies. RISC is hiermee handelingsdiagnostiek, bedoeld voor de reclasseringspraktijk om tot een gestructureerd en onderbouwd oordeel te komen over de juiste invulling van een justitieel traject c.q. reclasseringstoezicht. Het onderhavige onderzoek is uitsluitend gericht op het deel dat de basisdiagnostiek omvat. Als in het vervolg over de RISC wordt gesproken, wordt dit deel van het instrument bedoeld.

Het eerste niveau van de RISC, de basisdiagnostiek, bestaat uit kwantitatieve items, die gescoord worden, en kwalitatieve items, die niet gescoord worden. De te scoren items zijn zo veel mogelijk gebaseerd op concrete kenmerken van een dader of zijn omstandigheden, en hebben alle betrekking op onderwerpen waarvan uit onderzoek is gebleken dat deze een bijdrage leveren aan de kans op recidive (Adviesbureau Van Montfoort & Reclassering Nederland, 2004). Het instrument bestaat uit 12 schalen die de volgende statische en dynamische criminogene factoren in kaart brengen:

- 1 Delictgeschiedenis
- 2 Huidig delict en delictpatroon
- 3 Huisvesting en wonen
- 4 Opleiding, werk en leren
- 5 Inkomen en omgaan met geld
- 6 Relaties met partner, gezin en familie
- 7 Relaties met vrienden en kennissen
- 8 Druggebruik
- 9 Alcoholgebruik
- 10 Emotioneel welzijn
- 11 Denkpatronen, gedrag en vaardigheden
- 12 Houding

Naast deze twaalf schalen wordt op een dertiende schaal aanvullende informatie over de situatie van de dader verzameld. Deze schaal bestaat echter volledig uit kwalitatieve items en draagt niet bij aan de berekening van de totaalscore. De kwantitatieve items van de eerste twaalf RISC-schalen worden op een drie- of tweepuntsschaal gescoord (met de categorieën 0, 1 en 2 respectievelijk 0 en 2). Score 0 duidt op de afwezigheid van problemen, de hoogste score komt overeen met de aanwezigheid van ernstige problemen. Op elke schaal wordt een ruwe schaalscore berekend door de itemscores bij elkaar op te tellen. De schalen 1 en 2 worden gecombineerd tot één schaal waardoor een RISC-profiel uit elf schaalscores bestaat. Deze ruwe schaalscore wordt op basis van dit scoringsprofiel omgezet naar een gewogen score. Deze weging is gebaseerd op het Engelse instrument OASys en is volgens de handleiding van de RISC nodig omdat uit onderzoek blijkt dat sommige leefgebieden een sterkere samenhang vertonen met de kans op recidive dan

andere. De gewogen schaalscores van de RISC worden ingedeeld in drie categorieën. Deze categorieën geven de mate weer waarin sprake is van de criminogene factor:

- 1 Geen criminogeen probleem (afwezig).
- 2 Ciminogeen probleem is aanwezig.
- 3 Criminogeen probleem is in ernstige mate aanwezig.

De som van alle gewogen schaalscores vormt de totaalscore, die zicht geeft op de kans op recidive. Ook de totaalscore wordt in drie categorieën ingedeeld. Deze geven de ernst van het recidiverisico weer: 1) laag risico, 2) midden risico, en 3) hoog risico. Wanneer de RISC wordt afgenomen bij een ontkenkende, nog niet veroordeelde cliënt kan schaal 2 *Huidig delict en delictpatroon* niet ingevuld worden, wat op basis van de scoringsinstructies van de RISC tot gevolg heeft dat geen totaalscore berekend wordt.

In de reclasseringspraktijk wordt de RISC voor verschillende doelen gebruikt.

- 1 Voor voorlichtingsrapporten aan de Rechterlijke Macht of het Openbaar Ministerie.
- 2 Voor het ontwikkelen van een plan van aanpak voor reclasseringstoezicht. Reclasseringstoezicht is verplichte reclasseringsbegeleiding in het kader van een voorwaardelijke veroordeling, proefverlof TBS, Penitentiair Programma, Elektronisch Toezicht en dergelijke.
- 3 Voor het opzetten van een reïntegratieplan voor gedetineerden die na veroordeling in eerste aanleg een strafresistant hebben van minimaal vier maanden.

In alle gevallen wordt de RISC gebruikt om te komen tot een inschatting van enerzijds de kans op recidive, anderzijds het gevaar voor zichzelf of anderen. Tevens dient het instrument om inzicht te krijgen in de vraag op welke leefgebieden criminogene factoren zich manifesteren. Tot slot wordt een inschatting gemaakt van de (on)mogelijkheden van de delinquent om deel te nemen aan bepaalde interventies (denk aan motivatie of persoonlijkheidskenmerken).

1.3 Probleemstelling, doelstelling en onderzoeksvragen

De RISC wordt sinds november 2004 gebruikt door de reclasseringsorganisaties in het kader van riskassessment en riskmanagement bij delinquenten binnen het gevangeniswezen en de reclassering. Tijdens de ontwikkeling van het instrument is, zoals eerder gesteld, aandacht besteed aan de kwaliteit van de items en de interne consistentie van de schalen. Gezien het karakter van het instrument en het doel waarvoor het wordt ingezet, is onderzoek naar de interbeoordelaarsbetrouwbaarheid en de begripsvaliditeit echter van het grootste belang. Immers, het gaat hier om een beoordelingsinstrument op basis waarvan belangrijke beslissingen over individuen worden genomen. De RISC is richtinggevend voor het handelen: als de werker weet wat aan het delictgedrag ten grondslag ligt, welke in hoge mate aanwezige criminogene factoren er zijn, welke leerstijl, motivatie en leeromgeving het meest geschikt zijn voor deze delinquent, kan de werker (in overleg met werkbegeleider en delinquent) kiezen of indiceren voor bijvoorbeeld een specifieke justitiële interventie (Adviesbureau Van Montfoort & Reclassering Nederland, 2004). Het mag daarom voor de resultaten op de RISC niet uitmaken of het instrument door

reclasseringswerker A of B wordt afgenomen: beoordelingen moeten generaliseerbaar zijn over beoordelaars. Dit is de vraag naar de interbeoordelaarsbetrouwbaarheid van het instrument: wat is de mate van overeenstemming tussen verschillende beoordelaars? Omdat op basis van de RISC vérstrekkende beslissingen kunnen worden genomen over een delinquent, is het van het grootste belang dat het instrument daadwerkelijk meet wat het pretendeert te meten, namelijk recidiverisico en factoren die daaraan ten grondslag liggen. Omdat op basis van deze factoren beslissingen worden genomen over het type interventie dat voor een delinquent geschikt is, is het niet voldoende om uitsluitend naar de samenhang met recidive-risico te kijken, maar moeten de RISC-schalen tevens inhoudelijk overeenkomen met de begrippen die ze beogen te meten.

Het doel van het onderhavige onderzoek is om inzicht te krijgen in de interbeoordelaarsbetrouwbaarheid en de begripsvaliditeit van de RISC, maar voornamelijk om aanbevelingen te doen om het instrument psychometrisch te optimaliseren. De volgende onderzoeksvragen dienen hiervoor in dit onderzoek beantwoord te worden:

- 1 Wat is de interbeoordelaarsbetrouwbaarheid van de RISC?
- 2 Hoe is de begripsvaliditeit van de RISC?
- 3 Welke aanbevelingen ter verbetering van het instrument zijn op grond van de antwoorden op vraag 1 en 2 te doen?

Bij de eerste twee vragen gaat het om de psychometrische kwaliteiten van zowel de individuele schalen van het instrument als van de totaalscore. Het antwoord op de derde vraag wordt geformuleerd op basis van de uitkomsten van de eerste twee vragen. Het onderzoek naar de begripsvaliditeit van de RISC betreft in de onderhavige studie onderzoek naar de congruente validiteit ofwel soortgenootvaliditeit. Dit is een onderdeel van het proces van begripsvalidering en heeft betrekking op de mate waarin een test correleert met een andere test waarvan wordt aangenomen dat die dezelfde eigenschap meet (Drenth & Sijtsma, 1990). Een andere manier om de begripsvaliditeit van een test vast te stellen is bijvoorbeeld om na te gaan in hoeverre een test negatief samenhangt met een andere test waarvan wordt aangenomen dat het een tegengesteld concept meet (discriminante validiteit). In het volgende hoofdstuk wordt uitgebreid ingegaan op de gebruikte methoden van onderzoek.

Hoewel tijdens de ontwikkeling van de RISC al aandacht is besteed aan de interne consistentie van het instrument, is ervoor gekozen om dit, samen met de schaalstructuur van de RISC, ook tot onderwerp van het onderhavige onderzoek te maken. Voor het onderhavige onderzoek kon immers gebruik worden gemaakt van een omvangrijke database met meer dan 10.000 RISC's. Een zodanig groot onderzoeksbestand biedt de mogelijkheid om de psychometrische kwaliteiten van de RISC ook in termen van interne consistentie optimaal vast te stellen. De onderzoeksvraag die hiermee beantwoord dient te worden is:

- 4 Hoe is de schaalstructuur van de RISC en wat is de interne consistentie van de RISC-schalen en de RISC-totaalscore en welke aanbevelingen ter verbetering zijn met betrekking tot de structuur van de RISC te doen?

RISc en specifieke doelgroepen

Op basis van de ervaringen die in het eerste jaar met de RISc zijn opgedaan, is door de drie reclasseringsorganisaties een knelpuntenevaluatie uitgevoerd (Bosker, 2006). Hieruit kwam naar voren dat de RISc bij een aantal specifieke doelgroepen tot een lagere inschatting van het recidiverisico lijkt te leiden dan reclasseringswerkers in de praktijk aannemelijk achten. Als dit daadwerkelijk het geval zou zijn, betekent dit dat de validiteit van het instrument voor deze doelgroepen onvoldoende zou zijn. Om de congruente validiteit van de risico-inschatting van de RISc te bestuderen, was ten tijde van het verschijnen van de genoemde knelpuntenevaluatie reeds besloten de RISc-totaalscore in het onderhavige onderzoek te relateren aan het recidiverisico zoals dat uit het gevalideerde recidivevoorspellingsmodel StatRec blijkt (zie hoofdstuk 2 voor een uitgebreide toelichting). Naar aanleiding van de knelpuntenevaluatie is het WODC gevraagd dit deel van het onderzoek uit te breiden met een extra onderzoeksvraag:

- 5 Wat is de relatie tussen de RISc-totaalscore en de score op het recidivevoorspellingsmodel StatRec voor de volgende subgroepen:
 - a Plegers van huiselijk geweld
 - b Zedendelinquenten
 - c Veelplegers
 - d Oplichters
 - e Oudere daders
 - f Bestuurders onder invloed

In de oorspronkelijke vraag om uitbreiding van het onderzoek naar de congruente validiteit van de RISc stonden ook mannen en vrouwen en groepen met verschillende etnische achtergronden vermeld. Omdat het bestuderen van eventuele verschillen tussen deze groepen al binnen het oorspronkelijke onderzoek gepland was, zijn ze niet afzonderlijk in deze extra, vijfde onderzoeksvraag opgenomen.

1.4 Leeswijzer

Een onderzoek naar de psychometrische kwaliteiten van een instrument is van nature een vrij ‘technisch’ onderzoek dat logischerwijs ook een vrij technisch rapport oplevert. Hoewel er niet aan valt te ontkomen dat de beschrijving van de methoden van onderzoek ook ingaat op de gebruikte statistische methoden, is bij het schrijven van hoofdstuk 2 — waarin de onderzoeksmethoden worden uiteengezet — getracht het geheel zo leesbaar mogelijk te houden. De resultatenhoofdstukken zijn daarentegen per definitie zeer technisch van aard. Hoofdstuk 3 behandelt de resultaten van het onderzoek naar de interbeoordelaarsbetrouwbaarheid, hoofdstuk 4 beschrijft de schaalstructuur van de RISc en de interne consistentie van de schalen en de totaalscore en hoofdstuk 5 behandelt de resultaten van het onderzoek naar de congruente validiteit. In hoofdstuk 6 worden ten slotte de resultaten beschreven van de analyses op specifieke doelgroepen. De geïnteresseerde lezer wordt van harte aanbevolen deze hoofdstukken te lezen, maar wie liever alleen wil weten wat de resultaten van het onderzoek betekenen voor de bruikbaarheid van de RISc en op welke manier het instrument psychometrisch kan worden verbeterd, wordt naar hoofdstuk 7 verwezen. Hierin wordt op een heldere en begrijpelijke manier uitgebreid ingegaan op de conclusies die op basis van dit onderzoek kun-

nen worden getrokken over de kwaliteit van de RISc en op de vraag welke consequenties aan de resultaten kunnen worden verbonden.

2 Methoden van onderzoek

2.1 Inleiding

Het onderzoek naar de interbeoordelaarsbetrouwbaarheid, de interne consistentie en de congruente validiteit van de RISc is in drie deelonderzoeken uitgevoerd. De resultaten van deze onderzoeken worden in de navolgende hoofdstukken beschreven. In dit hoofdstuk wordt de opzet van de drie deelstudies beschreven en wordt besproken hoe de data voor de studies zijn verzameld. De onderzoeksopzetten van de drie deelonderzoeken zijn door het WODC opgesteld en in overleg met de drie reclasseringsorganisaties nader uitgewerkt. Voor de opzet van de deelonderzoeken naar de interbeoordelaarsbetrouwbaarheid en de congruente validiteit van de RISc is nagegaan of geput kon worden uit ervaringen in buitenlands onderzoek naar de psychometrische kwaliteiten van OASys en LS/CMI (de opvolger van de LSI-R). Zoals in hoofdstuk 1 is opgemerkt, is de RISc gebaseerd op het Britse OASys en daarmee indirect op de Canadese LSI-R. Helaas bleken studies naar de interbeoordelaarsbetrouwbaarheid van de LS/CMI bij nadere bestudering in feite test-hertestbetrouwbaarheid in kaart te brengen (Andrews, Bonta & Wormith, 2004). De kwaliteit van de opzet van de studie naar de interbeoordelaarsbetrouwbaarheid van OASys (Howard, Clark & Garnham, 2003) was onvoldoende om te kunnen gebruiken als uitgangspunt voor het onderhavige onderzoek. Onderzoek naar de congruente validiteit van OASys en LS/CMI ontbreekt of is kwalitatief onvoldoende (Andrews, Bonta & Wormith, 2004; Howard, Clark & Garnham, 2003). Hiermee is overigens niet gezegd dat de instrumenten in het geheel niet zijn onderzocht op hun psychometrische kenmerken. De interne consistentie van beide instrumenten is onderzocht evenals de predictieve validiteit. Er bestaat dus inzicht in de mate waarin daadwerkelijke recidive kan worden voorspeld op basis van OASys en de LS/CMI.

2.2 Het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISc

Om de interbeoordelaarsbetrouwbaarheid van de RISc te onderzoeken is het instrument in de periode van november 2005 tot half mei 2006 bij een groep reclasseringscliënten twee maal afgenomen door twee verschillende reclasseringswerkers. Naast een reguliere (eerste) afname is de RISc nogmaals afgenomen door een tweede reclasseringswerker, die dit onafhankelijk van de eerste werker deed. Dat wil zeggen dat de twee reclasseringswerkers onderling geen contact hadden over de cliënt bij wie zij de RISc afnamen. Voorafgaand aan het onderzoek was met de drie reclasseringsorganisaties — Reclassering Nederland (RN), Stichting Verslavingsreclassering GGZ Nederland (SVG) en Leger des Heils Jeugdzorg en Reclassering (LJ&R) — overeengekomen dat zij alledrie zouden participeren in het onderzoek. Het onderzoek is uitgevoerd in de reclasseringsregio's Alkmaar-Haarlem, Rotterdam en Den Bosch. In de regio's Alkmaar-Haarlem en Rotterdam deden de reclasseringsorganisaties alledrie mee, in de regio Den Bosch participeerden alleen RN en LJ&R. Om organisatorische redenen besloot de SVG in die regio niet aan het onderzoek mee te werken.

Bij de start van het onderzoek is in elke regio een aantal reclasseringswerkers random uitgekozen om aan dit onderzoek mee te werken. Deze selectie werd door de onderzoekers uitgevoerd aan de hand van door de drie reclasseringsorganisaties aangeleverde overzichten met werkers die over een geldige RISC-licentie beschikten en die in de onderzoeksperiode beschikbaar waren (dat wil zeggen, niet met zwangerschapsverlof, langdurig ziek of langere tijd op vakantie). De drie reclasseringsorganisaties zijn niet alledrie even groot; de RN is veruit de grootste (ongeveer 67% van de delinquenten die met de reclassering te maken krijgen, is een RN-cliënt), het LJ&R de kleinste (8%) en de SVG neemt een middenpositie in (25%). Om deze verhoudingen in de groep van aan het onderzoek deelnemende reclasseringswerkers te benaderen, is in overleg met de reclasseringsorganisaties besloten om tweentwintig RN-medewerkers te selecteren, tien SVG-medewerkers en zes LJ&R-medewerkers. Dit leverde een verdeling op van 58% RN, 26% SVG en 16% LJ&R. Nadat de onderzoekers de random selectie hadden uitgevoerd, hebben de drie reclasseringsorganisaties de gelegenheid gekregen werkers, die in het uitvoeringsproces absoluut niet gemist konden worden, te laten vervangen door een van de vooraf (eveneens random) geselecteerde reservewerkers. De overige reservewerkers konden ingezet worden in het geval dat tijdens het onderzoek een van de deelnemende reclasseringswerkers uitviel.

De geselecteerde reclasseringswerkers dienden de dubbele afname van de RISC in vaste koppels uit te voeren. Hiertoe zijn de geselecteerde reclasseringswerkers op regioniveau en per reclasseringsorganisatie random door de onderzoekers tot koppels gevormd. Deze random koppeling van werkers leverde in een aantal gevallen problemen op omdat mensen weliswaar in dezelfde regio werkten, maar niet binnen dezelfde unit. Hoewel hier zoveel mogelijk rekening mee is gehouden, bleek dit in de regio Rotterdam voor de RN niet tot een werkbaar situatie te leiden. Vlak na het vormen van koppels werden de units in deze regio gereorganiseerd waardoor werkers die een koppel vormden niet meer in dezelfde unit werkten. Uiteindelijk is daarom op die locatie in een aantal gevallen afgeweken van de random koppeling van werkers.

De werving van cliënten voor dit onderzoek is door de reclasseringswerkers uitgevoerd. De onderzoekers hebben in twee bijeenkomsten een toelichting op het onderzoek verzorgd voor de werkers en er is een instructie voor hen opgesteld. De reclasseringswerkers is gevraagd om alle cliënten te benaderen bij wie zij in de onderzoeksperiode een reguliere RISC afnamen. Wanneer een cliënt aangaf mee te willen werken aan het onderzoek, werd de collega-werker gewaarschuwd waarna die een afspraak maakte met de cliënt voor een tweede afname. Deze tweede afspraak diende ongeveer twee tot drie weken na de eerste afname te worden gepland. Cliënten kregen €25,- voor hun deelname aan het onderzoek. De reclasseringswerkers hielden tijdens het onderzoek bij welke cliënten hun medewerking toezegden, wanneer de tweede afspraak gepland stond en of de tweede RISC daadwerkelijk kon worden afgenomen. De reclasseringswerkers is tevens gevraagd een overzicht bij te houden van cliënten die geen deel wisten te nemen aan het onderzoek. Met behulp van deze overzichten zou nagegaan kunnen worden in hoeverre de deelnemende cliënten verschillen van de cliënten die een tweede RISC-afname weigerden. De verwerking van de data door de onderzoekers is geheel anoniem verlopen: van de reclasseringscliënten zijn uitsluitend de cliëntnummers bekend. De RN heeft zorggedragen voor de koppeling van RISC-data aan de cliëntnummers.

Om de dataverzameling in goede banen te leiden, is er een coördinatiegroep opgericht. Hierin zat een landelijk onderzoekscoördinator, werkzaam bij de RN en door de reclasseringsorganisaties gevraagd de landelijke coördinatie van het onderzoek voor haar rekening te nemen. Verder bestond de coördinatiegroep uit drie regionale onderzoekscoördinatoren, elk werkzaam bij de RN en belast met de bewaking van de dataverzameling in de drie onderzoeksregio's. Met zijn vieren overzagen en controleerden zij de voortgang van de dataverzameling. De landelijke coördinator had zeer regelmatig contact met de onderzoekers en ongeveer eens per maand kwamen onderzoekers en coördinatiegroep bijeen om de voortgang van de dataverzameling te bespreken. In de oorspronkelijke onderzoeksopzet is een streefaantal van 80 dubbele RISC-afnames vastgesteld. De negentien reclasseringskoppels dienden daarom elk vier à vijf dubbele RISC's af te nemen. In overleg met de reclasseringsorganisaties was de tijd die hiervoor nodig zou zijn, geschat op drie maanden. Dit bleek echter een onhaalbare planning en het was nodig de dataverzamelingsperiode met ruim drie maanden te verlengen.

In de periode van november 2005 tot half mei 2006 hebben in totaal 90 reclasseringscliënten toegezegd mee te werken aan het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISC. In vier gevallen is een van beide RISC's of zijn beide RISC's nooit afgerond. Bij zeven cliënten is om uiteenlopende redenen geen tweede RISC afgenomen. In drie gevallen lijkt een te oude RISC te zijn opgevoerd als eerste RISC; deze is (ruim) voor de start van het onderzoek afgenomen. Uiteindelijk is de RISC bij 75 reclasseringscliënten daadwerkelijk twee keer afgenomen. De achtergrondkenmerken van deze groep worden in hoofdstuk 3 beschreven.

2.3 Het onderzoek naar de schaalstructuur en interne consistentie van de RISC

Ten behoeve van het onderzoek naar de schaalstructuur en de interne consistentie van de RISC kon gebruik worden gemaakt van een van de Reclassering Nederland (RN) verkregen database waarin alle RISC's zijn opgenomen die tussen november 2004 en mei 2006 door de drie reclasseringsorganisaties zijn gestart. Het bestand bevatte in totaal 16.717 RISC's. De RISC's die nog in behandeling waren, zijn hieruit verwijderd waarna 12.868 RISC's overbleven. Vervolgens is besloten de RISC's waarvan vier of meer schalen onvoldoende of niet waren ingevuld niet te betrekken in de analyses. Wanneer zoveel schalen missen, betekent dat vrijwel altijd dat de RISC om wat voor reden dan ook nooit is afgerond (persoonlijke communicatie J. Bosker (RN), 6-6-2006). Dit had tot gevolg dat 986 RISC's niet zijn betrokken in het onderzoek, zodat 11.882 RISC's overbleven. Vervolgens is een aantal RISC's dat op dezelfde dag is afgenomen uit de data verwijderd evenals de RISC's die in het kader van het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISC dubbel zijn afgenomen (de reguliere RISC's zijn in het bestand bewaard gebleven). Het aantal RISC's waarop analyses konden worden uitgevoerd, bedroeg na deze datacleaning 11.666 RISC's.

Voor elk deugdelijk instrument is een heldere, goed te interpreteren schaalstructuur essentieel. In de ontwikkelfase van de RISC is uitgegaan van een aantal a priori opgestelde schalen die tezamen een totaalscore vormen. Omdat de onderzoeksgroep tijdens de ontwikkelfase te klein was, is deze structuur toen niet in statistische analyses getoetst (Vinke et al., 2003). Met behulp van de voor het onderhavige onderzoek beschikbare database kon deze toets wel worden uitgevoerd. Er is nage-

gaan in hoeverre de RISC-schalen en de RISC-totaalscore ‘goede’ schalen zijn. Dat wil zeggen in hoeverre ze bestaan uit onderdelen (items respectievelijk schalen) die elk betrekking hebben op een aspect van het achterliggende concept dat door de schaal in kaart moet worden gebracht. Goede schalen zijn echter ook voldoende betrouwbaar. De betrouwbaarheid van een schaal heeft betrekking op de precisie en nauwkeurigheid van de uitslagen; ze dienen vrij te zijn van toevallige meetfouten (Drenth & Sijtsma, 1990). Twee vormen van betrouwbaarheid die van belang zijn voor een beoordelingsinstrument zoals de RISC — de interbeoordelaarsbetrouwbaarheid en de interne consistentie — komen in dit onderzoek aan bod. Eerder is al ingegaan op de interbeoordelaarsbetrouwbaarheid van de schalen en de totaalscore en de resultaten van het onderzoek hiernaar worden beschreven in hoofdstuk 3. De interne consistentie van een schaal geeft aan in hoeverre de score replicerbaar is over de items die deel uitmaken van de schaal en is op basis van de database met 11.666 RISC’s onderzocht. In hoofdstuk 4 worden de resultaten van het onderzoek naar de schaalstructuur en interne consistentie van de RISC beschreven.

2.4 Het onderzoek naar de congruente validiteit van de RISC

2.4.1 Twee keer congruente validiteit

Aan het onderzoek naar de congruente validiteit van de RISC kunnen twee onderdelen worden onderscheiden. Het belangrijkste doel van de RISC is om een inschatting van het risico op recidive te bepalen. De validering van de totaalscore, die deze risico-inschatting vormt, en van de RISC-schalen, die immers ook geacht worden verband te houden met delictgedrag, is het eerste deel van het onderzoek naar de congruente validiteit. Het tweede deel heeft betrekking op de inhoud van de RISC-schalen. Met behulp van deze schalen worden specifieke leefgebieden in kaart gebracht en de aan- of afwezigheid van criminogene problemen op deze gebieden wordt gebruikt om een indicatie te geven voor het aanbod van de reclasering, bijvoorbeeld in de vorm van gedragsinterventies. De vraag of deze schalen daadwerkelijk de concepten meten die ze beogen te meten, wordt in het tweede deel van het validiteitsonderzoek bestudeerd.

Omdat de validering van de leefgebieden van de RISC een andere aanpak vereiste dan het onderzoek naar de samenhang met recidive, wordt hierna de opzet van beide delen van het onderzoek naar de congruente validiteit apart besproken.

2.4.2 Congruente validiteit: RISC en de voorspelling van recidive

In onderzoek naar de congruente validiteit van een instrument is het gebruikelijk om andere, gevalideerde instrumenten te gebruiken die hetzelfde of ongeveer hetzelfde construct beogen meten (Drenth & Sijtsma, 1990). Indien het te onderzoeken instrument — in dit geval de RISC — valide is, zou het een sterke positieve samenhang moeten vertonen met de gevalideerde instrumenten die hetzelfde construct meten. Er bestaan verschillende, gevalideerde risicotaxatie-instrumenten om het risico op recidive in te schatten. Deze zijn echter vrijwel zonder uitzondering ontwikkeld voor specifieke dadergroepen of voor het voorspellen van specifieke recidivedelicten. Zo is bijvoorbeeld de SVR-20 ontwikkeld om het risico op (seksueel) gewelddadige recidive te voorspellen bij volwassen zedendelinquenten en

wordt met de HCR-20 de kans op gewelddadige recidive voorspeld. Een door het WODC ontwikkelde recidiveschaal bood uitkomst (Wartna & Tollenaar, 2006). Deze schaal, StatRec, voorspelt op basis van een aantal statistische achtergrondkenmerken de recidivekans van daders. Recidive is hierin gedefinieerd als het opnieuw in contact komen met justitie. Scores op de schaal kunnen uiteenlopen van 0 tot 1, waarbij een score 0 overeenkomt met de afwezigheid van de kans op recidive en score 1 duidt op een kans van 100%. De schaal maakt onderdeel uit van de Quick Scan, een selectie-instrument van de reclassering voor gebruik in de eerste fase van het reclasseringsproces. Het instrument is opzettelijk eenvoudig gehouden omdat het snel moet kunnen worden gescoord door reclasseringswerkers die de Quick Scan invullen.

StatRec is ontwikkeld door een statistisch model (logistische regressie) te fitten op gegevens over de strafrechtelijke voorgeschiedenis van alle personen van wie in 1999 een strafzaak naar aanleiding van een misdrijf is afgedaan. Deze data zijn met gebruikmaking van de methodiek van de WODC-Recidivemonitor (Wartna, Blom & Tollenaar, 2004) bewerkt tot analyseerbare gegevens waarna de totale groep op aselechte wijze is gesplitst in een schattings- (N=73.699) en een validatiesteekproef (N=74.301). Het zoeken naar de factoren die de hoogte van de recidive zo goed mogelijk voorspellen, is gedaan in de schattingssteekproef waarna kruisvalidatie van de gevonden verbanden op de validatiesteekproef is uitgevoerd. Dit houdt in dat met het model dat geschat is op basis van de schattingssteekproef, voorspellingen zijn gedaan voor de validatiesteekproef, waarna deze voorspellingen zijn vergeleken met de werkelijke recidive. Het gebruik van een logistisch regressiemodel houdt in dat de kans om na een *vaste* periode te recidiveren wordt voorspeld. In het geval van StatRec is dat vier jaar. Op deze manier is een model ontwikkeld dat een valide voorspelling van recidive na vier jaar oplevert².

De StatRec-voorspelling van de kans op recidive wordt gebaseerd op de volgende kenmerken:

- geslacht,
- leeftijd,
- geboorteland,
- delictcategorie, dat wil zeggen het soort delict waarvoor de dader wordt vervolgd,
- het aantal strafzaken in jeugd en volwassenheid, ongeacht de afdoening,
- de veroordelingsdichtheid (het aantal strafzaken waarvoor een dader is vervolgd, gedeeld door de duur in jaren van zijn of haar criminele (justitiële) carrière).

In het kader van het onderzoek naar de congruente validiteit van de RISC is de RISC-totaalscore van de daders uit de door de RN verstrekte database met 11.666 RISC's gerelateerd aan de voorspelling van de recidivekans op basis van StatRec. Ook de samenhang tussen de StatRec en de scores op de afzonderlijke schalen van de RISC is onderzocht. Het uitgangspunt van de RISC is immers dat de schalen van

² Gelet op de primaire functie van de schaal – het aangeven van de kans op recidive – is het vooral van belang dat de voorspelling overeenkomt met de feitelijke recidivepercentages die men in diverse dadergroepen aantreft. Uit het resultaat van de vergelijking van de voorspelde en geobserveerde recidive in de andere helft van de populatie van volwassen verdachten uit 1999 (N=74.301) kan geconcludeerd worden dat op elk niveau van het voorspelde risico de voorspelde recidive nauwelijks afwijkt van de recidive die feitelijk is opgetreden. Zowel bij laag als bij hoog risico wordt het aantal recidivisten vrij nauwkeurig voorspeld (Wartna en Tollenaar, 2006).

het instrument criminogene factoren in kaart brengen die herhaling van het delictgedrag voorspellen (Adviesbureau Van Montfoort & Reclassering Nederland, 2004). Hoewel de verwachting hierdoor is dat alle RISC-schalen afzonderlijk samenhangen met de voorspelling van de recidivekans op basis van StatRec, is het niet reëel om te verwachten dat elke schaal, rekening houdend met de invloed van de andere RISC-schalen, een unieke bijdrage levert aan de verklaring van de StatRec-score. Een dergelijke verwachting is in strijd met de eis dat de RISC-schalen samen een intern consistente RISC-totaalscore vormen. Dat houdt immers in dat de verschillende onderdelen van de RISC-totaalscore elk een deel van hetzelfde achterliggende construct meten. Dit verhoudt zich slecht met de verwachting dat iedere schaal een unieke bijdrage kan leveren aan het recidiverisico. Bovendien is het voorspellen van recidive niet het enige doel van de RISC. Het is ook een instrument voor handelingsdiagnostiek: welke dynamische criminogene factoren kunnen met behulp van gedragsinterventies worden beïnvloed opdat het recidiverisico afneemt? Voor het voorspellen van recidive zijn gegevens over de delictgeschiedenis van essentieel belang — statische criminogene factoren zijn zeer sterke voorspellers van recidive — terwijl voor het stellen van een diagnose en het indiceren van geschikte interventies de dynamische criminogene factoren van belang zijn. Het is niet reëel te veronderstellen dat dynamische factoren een (redelijk grote) unieke bijdrage kunnen leveren aan het voorspellen van recidive als ook informatie over delictgeschiedenis bij deze voorspelling wordt betrokken. De dynamische factoren zouden dan immers informatief moeten zijn over een toekomstig delict zonder dat ze dat zijn over het huidige en vroegere delictgedrag.

Voor de congruente validiteit van de RISC in termen van de voorspelling van recidive is het het belangrijkste dat de RISC-schaalscores en de RISC-totaalscore een (middel)sterke positieve samenhang vertonen met de voorspelling van het recidiverisico zoals bepaald met StatRec. Ook is het van belang dat de RISC-onderdelen samen een groot deel van de variantie van de StatRec-score kunnen verklaren, of met andere woorden, dat de StatRec-score goed kan worden voorspeld op basis van de RISC. Daarnaast is exploratief onderzocht óf er schalen zijn die, gegeven de invloed van alle andere RISC-schalen, een eigen bijdrage leveren aan de verklaring van de StatRec-voorspelling van de recidivekans. Bij de interpretatie van de resultaten moet rekening gehouden worden met het feit dat het hier gaat om de validering van de RISC met behulp van een *voorspelling van recidive*. In 2008 zal een onderzoek worden uitgevoerd naar de samenhang tussen de RISC en daadwerkelijke recidive. Pas dan zal blijken hoe goed de voorspelling op basis van de RISC echt is en dus of de RISC over een goede predictieve validiteit beschikt.

2.4.3 Congruente validiteit: de leefgebieden die door de RISC-schalen in kaart worden gebracht

Voor de beantwoording van de vraag naar de congruente validiteit van de afzonderlijke RISC-schalen dient bij voorkeur voor alle schalen te worden vastgesteld of ze daadwerkelijk op een valide wijze het te meten leefgebied in kaart brengen. Tijdens de voorbereiding van dit onderzoek bleek dit evenwel op praktische en inhoudelijke bezwaren te stuiten. De praktische bezwaren hadden vooral betrekking op de inspanningen die van de drie reclasseringsorganisaties werden verwacht. Het valideren van alle RISC-schalen bleek een te grote belasting te zijn voor de reclasseringsorganisaties. Daarnaast bestonden er inhoudelijke problemen met betrekking tot het valideren van de meer feitelijke RISC-schalen zoals huisvesting, opleiding /

werk en financiën. Het lag het meest voor de hand om die schalen te valideren met behulp van dossierinformatie. Echter, omdat de reclasseringswerker bij het afnemen en invullen van de RISc gebruik maakt van dezelfde dossierinformatie, was dat geen mogelijkheid. Uiteindelijk is in overleg met de drie reclasseringsorganisaties besloten de drie minst feitelijke en meest subjectieve schalen uit het instrument te valideren. Het gaat daarbij om de schalen 10 *Emotioneel welzijn*, 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding*. Dit zijn bovendien schalen die een belangrijke rol spelen in de beoordeling van de vraag of een cliënt een gedragsinterventie zou moeten volgen. Op basis van de schaal *Emotioneel welzijn* wordt in de RISc ingeschat hoe de cliënt in het leven staat, welk zelfbeeld hij heeft en of er mogelijk sprake is van psychische problemen. Een hoge score wijst op een laag zelfbeeld en de aanwezigheid van psychische problemen. De schaal *Denkpatronen, gedrag en vaardigheden* beoogt in kaart te brengen hoe de delinquent omgaat met zaken die hij in het dagelijks leven tegenkomt. De items van de schaal hebben onder meer betrekking op impulsiviteit, probleembesef en probleemhanteling. Een hoge score wijst op de aanwezigheid van verschillende cognitieve tekorten. Op basis van de RISc-schaal *Houding* wordt beoogd de houding van de delinquent te meten tegenover het delict, de maatschappij en criminaliteit in het algemeen. Een hoge score op deze schaal duidt op een pro-criminele attitude, een gebrek aan schuldgevoel, een afwijzen van de wetten en regels van de samenleving en een negatieve houding ten opzichte van de (op te leggen of opgelegde) sanctie.

Net als voor het onderzoek naar de congruente validiteit van de RISc in termen van recidivevoorspelling is voor het onderzoek naar de congruente validiteit van de schalen 10, 11 en 12 op zoek gegaan naar gevalideerde instrumenten die hetzelfde of ongeveer hetzelfde construct beogen meten. In dit soort validiteitsonderzoek is het belangrijk dat mogelijke bronnen van 'ruis' worden vermeden en de validering van een beoordelingsinstrument zou dan ook bij voorkeur moeten worden gedaan met behulp van andere beoordelingsinstrumenten. De RISc is in Nederland echter uniek in zijn soort en ondanks pogingen om geschikte (valide) instrumenten te vinden die door reclasseringswerkers ingevuld konden worden, is dat niet gelukt. De keuze viel daarom op het gebruik van gevalideerde vragenlijsten die door reclasseringscliënten moesten worden ingevuld. Het gevolg van deze keuze is dat het *beoordelingsinstrument* RISc in deze studie wordt gevalideerd met behulp van het *eigen oordeel* van reclasseringscliënten. Op de nadelen die hieraan verbonden zijn, wordt zowel bij de bespreking van de resultaten als in het conclusiehoofdstuk teruggekomen.

Volgens de handleiding van de RISc meten de schalen 10, 11 en 12 aspecten van de persoonlijkheid. De aandacht richtte zich bij de selectie van mogelijke instrumenten voor de validering van deze schalen dan ook op persoonlijkheidsvragenlijsten. Uiteraard was een vereiste dat de betrouwbaarheid en validiteit van de te selecteren instrumenten minstens voldoende was. Voor een oordeel hierover is gebruik gemaakt van de *Documentatie van tests en testresearch in Nederland* (Evers, Van Vliet-Mulder & Groot, 2000). Bij de selectie van instrumenten om de validiteit van de RISc-schalen 10, 11 en 12 te bestuderen, is er voorts naar gestreefd uitsluitend instrumenten te selecteren met een beperkte invulduur omdat de verwachting was dat de doelgroep moeite zou hebben gedurende een langere periode geconcentreerd te werken aan het invullen van een vragenlijst. Het streven was hierbij om de invulduur te beperken tot een half uur. Als een instrument in eerder onderzoek

(onderzoek door andere auteurs) is gebruikt in een gedetineerden- of delinquentenpopulatie, is dat als een voordeel beschouwd.

Op basis van de hiervoor genoemde overwegingen zijn de Nederlandse Persoonlijkheidsvragenlijst (NPV), Utrechtse Copinglijst (UCL) en de Buss-Durkee Hostility Inventory — Dutch (BDHI-D) geselecteerd³.

De Nederlandse Persoonlijkheidsvragenlijst (NPV) is een persoonlijkheidsvragenlijst die beoogt een beperkt aantal persoonlijkheidskenmerken te meten (Luteijn, Starren & Van Dijk, 2000). De NPV is een vaak gebruikt instrument en wordt bij zeer uiteenlopende doelgroepen gebruikt. Afname duurt 20 tot 30 minuten en de betrouwbaarheid en de begripsvaliditeit van het instrument zijn beide voldoende (Evers, Van Vliet-Mulder & Groot, 2000, p. 420). Het instrument bestaat uit 133 items met drie antwoordmogelijkheden: juist, ? en onjuist. De items vormen zeven schalen⁴:

- 1 *Inadequatie*. In de items uit deze schaal komen vage angsten, vage lichamelijke klachten, een gedrukte stemming en insufficiëntiegevoelens naar voren.
- 2 *Sociale inadequatie*. De items uit deze schaal verwijzen naar het vermijden van of het zich ongelukkig voelen in sociale contacten.
- 3 *Rigiditeit*. De items uit deze schaal hebben te maken met het volgens plan willen laten verlopen van gebeurtenissen en met vaste gewoonten en principes. In sommige items komt ook een zekere intellectuele starheid naar voren.
- 4 *Verongelijkheid*. In de items uit deze schaal staat kritiek op en wantrouwen van andere mensen centraal
- 5 *Zelfgenoegzaamheid*. In de items uit deze schaal komt een sterk gevoel van tevredenheid met zichzelf naar voren en tegelijkertijd een desinteresse voor een ander en diens problemen.
- 6 *Dominantie*. De items uit deze schaal hebben te maken met initiatief willen nemen en leiding willen geven aan anderen en zelfvertrouwen in een groep.
- 7 *Zelfwaardering*. De items uit deze schaal verwijzen naar een positieve houding ten opzichte van werken, zichzelf en het leven, en goed aangepast en actief zijn.

De Utrechtse Coping Lijst (UCL) is een instrument dat beoogt te meten hoe mensen omgaan met problemen en stressvolle gebeurtenissen (Scheurs, Van de Willige, Brosschot, Tellegen & Graus, 1993). De UCL is, net als de NPV, een vaak gebruikt instrument bij verschillende onderzoeksdoelgroepen. Afname van de UCL duurt ongeveer 10 minuten en de betrouwbaarheid en de begripsvaliditeit van het instrument zijn voldoende (Evers, Van Vliet-Mulder & Groot, 2000, p. 56). De vragenlijst bestaat uit 47 items die met 'zelden of niet', 'soms', 'vaak' en 'zeer vaak' beantwoord kunnen worden. Samen vormen deze items zeven schalen⁵:

- 1 *Actief aanpakken*. Deze schaal heeft betrekking op de situatie rustig van alle kanten bekijken, de zaken op een rijtje zetten; doelgericht en met vertrouwen te werk gaan om het probleem op te lossen.
- 2 *Palliatieve reactie*. Deze schaal heeft te maken met afleiding zoeken en zich met andere dingen bezighouden om niet aan het probleem te hoeven denken;

³ Voor de validering van de schaal *Emotioneel welzijn* is nog overwogen om een klachtenlijst, zoals de *Symptom Checklist* (SCL-90) of de *General Health Questionnaire* (GHQ), te gebruiken, maar omdat de NPV geschikt bleek voor de validering van elk van de drie RiSc-schalen kreeg dit instrument de voorkeur.

⁴ De beschrijvingen van de inhoud van de schalen zijn overgenomen uit Luteijn, Starren & Van Dijk (2000).

⁵ De beschrijvingen van de inhoud van de schalen zijn overgenomen uit Schreurs et al. (1993).

proberen zich wat prettiger te voelen door te roken, te drinken of zich wat te ontspannen.

- 3 *Vermijden, afwachten.* Deze schaal betreft de zaak op zijn beloop laten, de situatie uit de weg gaan of afwachten wat er gaat gebeuren.
- 4 *Sociale steun zoeken.* Deze schaal heeft betrekking op het zoeken van troost en begrip bij anderen; zorgen aan iemand vertellen of hulp vragen.
- 5 *Passief reactiepatroon.* Deze schaal betreft zich volledig door de problemen en de situatie in beslag laten nemen, de zaak somber inzien, zich piekerend in zichzelf terugtrekken, niet in staat om iets aan de situatie te doen; piekeren over het verleden.
- 6 *Expressie van emoties.* Deze schaal heeft te maken met het laten blijken van ergernis of kwaadheid; spanningen afreageren.
- 7 *Geruststellende en troostende gedachten hanteren.* Deze schaal heeft betrekking op zichzelf geruststellen met de gedachte dat na regen zonneshijn komt, dat anderen het ook wel eens moeilijk hebben of dat er nog wel ergere dingen gebeuren; jezelf moed inspreken.

De Buss-Durkee Hostility Inventory — Dutch (BDHI-D) is een vragenlijst die vijandigheid en neiging tot agressief gedrag beoogt te meten (Lange, Hoogendoorn, Wiederspahn & De Beurs, 2005). Het instrument is genormeerd voor gedetineerden waaruit blijkt dat het instrument geschikt is om te worden ingevuld door delinquenten. Afname kost 10 tot 15 minuten en de psychometrische kenmerken van de twee agressieschalen zijn goed (COTAN-beoordeling 1997, p.607). De betrouwbaarheid van de Sociale Wenselijkheidsschaal is onvoldoende, maar deze wordt niet gebruikt in het onderzoek naar de validiteit van de RISc. De BDHI-D bestaat uit 66 vragen die met waar of onwaar kunnen worden beantwoord. De items vormen drie schalen⁶:

- 1 *Indirecte agressie.* Een hoge score op deze schaal kan een aanwijzing zijn voor veel ingehouden agressie en voor een hoge mate van psychopathologie.
- 2 *Directe agressie.* Een hoge score op deze schaal wijst op veel agressie in de vorm van fysieke of verbale uitingen,
- 3 *Sociale wenselijkheid.* Deze schaal is bedoeld voor het meten van de geneigdheid van de respondent om antwoorden te geven die overeenkomen met de gangbare normen en om antwoorden te vermijden die daar niet mee overeenkomen. Een hoge score op de Sociale Wenselijkheidsschaal zou kunnen betekenen dat de kans bestaat dat de antwoorden op de Indirecte en Directe agressieschaal te laag zijn en dat de werkelijke scores iets hoger zouden kunnen liggen.

De drie reclasseringsorganisaties participeerden aan het onderzoek naar de congruente validiteit van de RISc-schalen 10, 11 en 12 en de dataverzameling hiervoor vond plaats tussen april 2006 en half september 2006. Net als in het onderzoek naar de interbeoordelaarsbetrouwbaarheid is dit met de hulp van reclasseringswerkers gedaan. Van elke reclasseringsorganisatie werkte drie regio's mee: de RN participeerde in de regio's Noord-Nederland, Utrecht en Breda, de SVG participeerde in de regio's Noord-Nederland, Zwolle en Utrecht, het LJ&R werkte mee in de regio's Groningen, Utrecht en Arnhem. Reclasseringswerkers vroegen hun cliënten na afloop van de afname van de RISc om medewerking aan het onderzoek naar

⁶ De beschrijvingen van de inhoud van de schalen zijn overgenomen uit Lange et al. (2000).

de congruente validiteit. Iedere cliënt vulde slechts één van de drie geselecteerde vragenlijsten in. Het werd een te grote belasting voor de cliënten geacht om ze te vragen elk van de drie vragenlijsten in te vullen. Bovendien was de verwachting dat slechts zeer weinig cliënten bereid zouden zijn dit te doen. Reclaseringswerkers beschikten over een pakketje met een aantal exemplaren van elk van de drie vragenlijsten en gaven hun cliënt, als hij of zij mee wilde doen, een van de vragenlijsten uit deze voorraad. Tijdens het invullen van de vragenlijst was de reclaseringswerker aanwezig voor eventuele uitleg. Als de cliënt klaar was met invullen, controleerde de werker of alle vragen waren ingevuld en of geen bladzijden waren overgeslagen. Voor hun medewerking aan het onderzoek ontvingen cliënten €10,-. Omdat het belangrijk is dat de onderzoeksgroep een goede afspiegeling van de reclaseringspopulatie vormt, is de reclaseringswerkers gevraagd al hun cliënten bij wie zij een RISC afnamen te vragen of ze mee wilden werken aan het onderzoek. Er zijn drie uitzonderingen gemaakt: cliënten die het Nederlands onvoldoende beheersen om de vragenlijst zelf te lezen, cliënten die op zwakbegaafd niveau functioneren en cliënten voor wie de werker een van de te valideren RISC-schalen om wat voor reden dan ook niet kan invullen, zijn uitgesloten van deelname aan het onderzoek. De werkers mochten uiteraard wel enige uitleg geven bij (items uit) de vragenlijsten, maar de cliënt moest de lijst zelfstandig kunnen lezen. Net als in het onderzoek naar de interbeoordelaarsbetrouwbaarheid is de reclaseringswerkers gevraagd een overzicht bij te houden van cliënten die vanwege een van de hiervoor genoemde uitsluitingsgronden niet mee konden doen of die weigerden deel te nemen aan het onderzoek. Het doel hiervan was om na te gaan of deze groep verschilt van de groep die wel participeerde. Op basis van cliëntnummers zijn de RISC-gegevens van zowel deelnemende als weigerende cliënten door de RN aan de onderzoekers verstrekt.

Voor het goede verloop van de dataverzameling was ook in dit onderzoek namens de drie reclaseringsorganisaties een landelijke onderzoekskoördinator beschikbaar. Zij werd ondersteund door zeven regiocoördinatoren: RN en SVG stelden voor elke deelnemende regio een contactpersoon, LJ&R had één coördinator voor haar drie deelnemende regio's. De door reclaseringscliënten in te vullen vragenlijsten zijn naar rato van omvang tussen RN, SVG en LJ&R verdeeld en zijn door de regiocoördinatoren uitgedeeld aan de reclaseringswerkers die zij voor dit onderzoek bereid hadden gevonden mee te werken. Voorafgaand aan het onderzoek is een streefaantal van 70 ingevulde vragenlijsten per geselecteerd instrument vastgesteld. De periode van dataverzameling zou oorspronkelijk drie maanden beslaan, maar is uitgebreid tot bijna zes maanden. Het totaalaantal van 210 ingevulde vragenlijsten is vrijwel gehaald: tijdens de dataverzamelingsperiode zijn 204 vragenlijsten ingevuld. Omdat in een aantal gevallen de RISC-gegevens niet vóór de start van de data-analyses beschikbaar waren, konden 185 van deze vragenlijsten gebruikt worden ter validering van de drie RISC-schalen. In hoofdstuk 5, het hoofdstuk waarin de resultaten van het onderzoek naar de congruente validiteit van de RISC worden beschreven, wordt nader ingegaan op deze onderzoeksgroep van 185 reclaseringscliënten.

3 Interbeoordelaarsbetrouwbaarheid

3.1 Inleiding

In dit hoofdstuk worden de resultaten van het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISC beschreven. De dataverzameling voor dit deelonderzoek heeft plaatsgevonden in de periode van november 2005 tot half mei 2006. Zoals in hoofdstuk 2 reeds is beschreven, namen 19 koppels reclasseringswerkers de RISC onafhankelijk van elkaar twee maal af bij hun reclasseringscliënten, die door henzelf om hun medewerking waren gevraagd. Volgens de instructie die zij vooraf hadden gekregen, dienden de werkers de twee RISC-afnames met een tussenpoos van ongeveer twee tot drie weken te plannen. Deze periode was nodig om te ondervangen dat cliënten de onderzoeksresultaten bewust zouden beïnvloeden door met opzet tijdens de tweede afname andere informatie te geven dan tijdens de eerste afname. Door de genoemde periode te laten verstrijken tussen beide afnames, was de kans kleiner dat de cliënt zich nog exact zou herinneren wat hij of zij tijdens het eerste gesprek heeft gezegd. De kans werd klein geacht dat zich in die periode belangrijke of grote veranderingen voordeden in de situatie van de cliënt, maar dit kon niet worden uitgesloten. Bij de interpretatie van de resultaten dient hier zo nodig rekening mee te worden gehouden.

3.2 Beschrijving van de onderzoeksgroep

In totaal hebben 90 reclasseringscliënten toegezegd mee te werken aan het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISC. In vier gevallen is een van beide RISC's of zijn beide RISC's nooit afgerond. Bij zeven cliënten is om uiteenlopende redenen geen tweede RISC afgenomen. In drie gevallen lijkt een te oude RISC te zijn opgevoerd als eerste RISC; deze is (ruim) voor de start van het onderzoek afgenomen.

Uiteindelijk is de RISC bij 75 reclasseringscliënten daadwerkelijk twee keer afgenomen. Gemiddeld lagen er 29 dagen tussen afname van de eerste en de tweede RISC. De onderzoeksgroep bestaat uit 65 mannen en 10 vrouwen (zie tabel 1). De gemiddelde leeftijd in de onderzoeksgroep was op het moment van invoer van de eerste, reguliere RISC 35 jaar en 11 maanden (35,89 jaar) met een standaarddeviatie van 13 jaar 1 maand (13,07 jaar). De jongste cliënt was 18 jaar, de oudste 73. Het merendeel van de cliënten is in Nederland geboren. Zestig procent van de onderzoeksgroep is autochtoon, 36% heeft een niet-westerse achtergrond en 2,7% heeft een westerse achtergrond. Van één persoon viel niet te achterhalen waar hij of zij geboren was, noch waar zijn of haar ouders geboren zijn. De grootste groep bestaat uit cliënten van de RN, gevolgd door de SVG en het LJ&R. Hoewel de RN inderdaad de grootste organisatie is en het LJ&R de kleinste, komt de verhouding tussen de drie reclasseringsorganisaties in deze onderzoeksgroep niet overeen met de daadwerkelijke verhouding tussen de organisaties. In werkelijkheid is de verhouding tussen de drie organisaties ongeveer 67 - 25 - 8. De afwijking van deze verdeling die in de onderzoeksgroep is gevonden, is niet significant: $\chi^2(2) = 5,66$; $p = .06$.

Tabel 1 Achtergrondkenmerken onderzoeksgroep interbeoordelaarsbetrouwbaarheid (N = 75)

Kenmerk		Aantal	%
Geslacht	Man	65	86,7
	Vrouw	10	13,3
Organisatie	RN	57	76,0
	LJ&R	8	10,7
	SVG	10	13,3
Leeftijd	Jonger dan 20 jaar	8	10,7
	20 – 30 jaar	20	26,7
	30 – 40 jaar	20	26,7
	40 – 50 jaar	16	21,3
	50 – 60 jaar	7	9,3
	60 jaar of ouder	4	5,3
Geboorteland	Nederland	54	72,0
	Marokko	1	1,3
	Nederlandse Antillen / Aruba	8	10,7
	Suriname	7	9,3
	Turkije	2	2,7
	Overig niet-westers	2	2,7
	Onbekend	1	1,3
Herkomstgroep	Autochtoon	45	60,0
	Niet-westers allochtoon	27	36,0
	Westers allochtoon	2	2,7
	Onbekend	1	1,3

De onderzoeksopzet voorzag erin dat elke deelnemende reclaseringswerker iedere cliënt bij wie hij of zij een RISC afnam, vroeg om mee te werken aan het onderzoek. Om na te kunnen gaan of cliënten die weigerden, verschillen van cliënten die meewerkten werd de werkers gevraagd een overzicht bij te houden van de cliënten die niet wilden participeren aan het onderzoek. In de praktijk bleek het in een aantal gevallen echter tot planningsproblemen te leiden wanneer alle cliënten werden gevraagd om mee te doen: te veel mensen waren hiertoe bereid waardoor de tweede reclaseringswerker te veel afspraken voor een tweede afname moest inpassen binnen zijn of haar reguliere werk. Bovendien verliep de dataverzameling in eerste instantie moeizaam. Het onderzoek interfereerde met het reguliere reclaseringswerk waardoor cliënten niet gevraagd werden voor het onderzoek. Om deze reden is ervan afgezien om de deelnemers aan het onderzoek te vergelijken met de weigeraars. In plaats daarvan is de onderzoeksgroep vergeleken met de reguliere reclaseringspopulatie. Hiervoor is gebruik gemaakt van de in hoofdstuk 2 beschreven grote RISC-database. De onderzoeksgroep uit het interbeoordelaarsbetrouwbaarheidsonderzoek is vergeleken met 11.590 afgenomen RISC's uit de genoemde database⁷. 90,6% van deze RISC's is afgenomen bij een man, 9,4% is afgenomen bij een vrouw. De verdeling in de onderhavige onderzoeksgroep blijkt niet significant af te wijken van de verdeling tussen mannen en vrouwen in de grote database: $\chi^2(1) = 1,44$; $p = .23$. De gemiddelde leeftijd op het moment van afname van de RISC

⁷ Ten opzichte van de beschrijving in paragraaf 2.3 zijn alle RISC's van de deelnemers aan het interbeoordelaarsbetrouwbaarheidsonderzoek tijdelijk verwijderd. Dit is gedaan om de vergelijking tussen onderzoeksgroep en reclaseringspopulatie zuiver te houden en te voorkomen dat de onderzoeksgroep deels met zichzelf werd vergeleken.

ligt in de database (N = 11.560; van 30 daders mist de geboortedatum) op 34 jaar en 4 maanden (m = 34,30; sd = 12,01). Een t-toets van de gemiddelde leeftijd in de onderzoeksgroep tegen de gemiddelde leeftijd in de grote database wijst uit dat de gemiddelde leeftijd van de onderzoeksgroep niet significant verschilt van de leeftijd in de database: $t = 1,06$; $df = 74$; $p = .29$ (tweezijdig). Ook wat betreft herkomst — ingedeeld in de categorieën autochtoon, niet-westers allochtoon, westers allochtoon en onbekend — bestaan er geen significante verschillen tussen de onderzoeksgroep en de groep uit de grote database: $\chi^2(3) = 0,83$; $p = .84$.

Naast een vergelijking tussen de onderzoeksgroep en de reguliere reclasseringspopulatie op eerdergenoemde achtergrondkenmerken is tevens nagegaan in hoeverre de onderzoeksgroep op hun RISC-scores afwijkt van de reguliere reclasseringspopulatie. Voor deze analyses is gebruik gemaakt van de RISC-scores van de onderzoeksgroep op de RISC die in het kader van het reguliere reclasseringswerk is afgenomen. Tabel 2 vermeldt de gemiddelde scores van onderzoeksgroep en vergelijkingsgroep en de resultaten van de uitgevoerde t-tests. Het verschil tussen beide groepen blijkt op twee schalen significant te zijn. De onderzoeksgroep heeft een hogere gemiddelde score op de schalen *Huisvesting* en *Relaties met partner, gezins- en familieleden*. Dit betekent dat zij meer problemen hebben op deze gebieden. Wanneer evenwel de effectgroottes worden bestudeerd, blijkt dat er sprake is van slechts kleine verschillen ($d=.35$ voor *Huisvesting*, $d=.32$ voor *Relaties met partner, gezins- en familieleden*).

Concluderend kan gesteld worden dat de in dit onderzoek betrokken onderzoeksgroep zowel wat hun achtergrondkenmerken betreft als wat betreft hun RISC-scores in grote mate lijkt op de reguliere reclasseringspopulatie.

Tabel 2 Vergelijking RISC-scores tussen onderzoeksgroep en reguliere reclasseringspopulatie

RISC-onderdeel	Gemiddelde score		t
	Onderzoeksgroep	Vergelijkingsgroep	
Schaal 1&2 Delictgegevens	10,05	10,44	- 0,32
Schaal 3 Huisvesting en wonen	3,08	1,78	2,47*
Schaal 4 Opleiding, werk en leren	6,89	7,34	- 0,55
Schaal 5 Inkomen en omgaan met geld	2,93	2,64	0,79
Schaal 6 Relaties met partner, gezin en familie	2,99	2,47	2,77**
Schaal 7 Relaties met vrienden en kennissen	3,81	3,49	0,68
Schaal 8 Druggebruik	3,73	3,07	1,08
Schaal 9 Alcoholgebruik	1,31	1,31	- 0,02
Schaal 10 Emotioneel welzijn	2,08	2,03	0,25
Schaal 11 Denkpatronen, gedrag en vaardigheden	5,92	6,29	- 0,84
Schaal 12 Houding	4,53	5,21	- 1,30
Totaalscore	45,89	46,53	- 0,16

* $p < .05$

** $p < .01$

3.3 Kwaliteit van de data

Met de analyse van de interbeoordelaarsbetrouwbaarheid van de RISC wordt nagegaan of verschillende beoordelaars tot een gelijklopend oordeel komen over de cliënt bij wie zij de RISC afnemen. Omdat missende waarden niet in dergelijke analyses worden betrokken, is het van belang om vooraf na te gaan of de eerste, regulier afgenomen RISC en de tweede, speciaal voor dit onderzoek afgenomen RISC op dit punt van elkaar verschillen. Vanzelfsprekend wisten de reclasseringswerkers of zij een reguliere of een tweede RISC afnamen. Omdat een tweede RISC niet van directe relevantie was voor het dagelijks werk van de reclasseringswerker is het voorstelbaar dat eventueel moeilijk te beantwoorden vragen op de tweede RISC makkelijker zijn overgeslagen. Om dit na te gaan is voor alle schalen apart en voor het totaal aantal op de RISC te scoren items met behulp van t-toetsen onderzocht of er systematische verschillen bestonden in het aantal missende waarden tussen de afnames van de reguliere (eerste) en de tweede RISC. De resultaten zijn weergegeven in tabel 3 en tonen aan dat er geen significante verschillen bestaan tussen beide RISC-afnames. Er is evenmin sprake van een duidelijke trend.

Tabel 3 Vergelijking van het aantal missende waarden tussen RISC-afnames^a

RISC-onderdeel	Gemiddeld aantal missende waarden		t
	RISC1	RISC2	
Schaal 1 Delictgeschiedenis	0,00	0,03	- 1,42 ^{n.s.}
Schaal 2 Analyse huidig delict en delictpatroon	0,52	0,51	0,15 ^{n.s.}
Schaal 3 Huisvesting en wonen	1,01	1,00	0,10 ^{n.s.}
Schaal 4 Opleiding, werk en leren	0,01	0,03	- 0,58 ^{n.s.}
Schaal 5 Inkomen en omgaan met geld	0,00	0,01	- 1,00 ^{n.s.}
Schaal 6 Relaties met partner, gezin en familie	0,15	0,09	0,82 ^{n.s.}
Schaal 7 Relaties met vrienden en kennissen	0,11	0,05	0,82 ^{n.s.}
Schaal 8 Druggebruik	3,29	3,11	0,78 ^{n.s.}
Schaal 9 Alcoholgebruik	0,59	0,57	0,28 ^{n.s.}
Schaal 10 Emotioneel welzijn	0,00	0,01	- 1,00 ^{n.s.}
Schaal 11 Denkpatronen, gedrag en vaardigheden	0,00	0,00	--
Schaal 12 Houding	0,09	0,00	1,21 ^{n.s.}
Totaal aantal te scoren items	5,77	5,41	1,15 ^{n.s.}

^a Er is tweezijdig getoetst.
^{n.s.} niet significant

Nu vastgesteld is dat er geen significante verschillen bestaan in het gemiddeld aantal missende waarden op de eerste en de tweede RISC, is een tweede controle van de kwaliteit van de data nodig. Zijn er schalen waar in vergelijking met de totale reclasseringspopulatie gemiddeld genomen (berekend over beide RISC-afnames) significant meer items missen? Per schaal is het gemiddeld aantal missende waarden over beide RISC-afnames samen berekend. Vervolgens is met behulp van een t-toets nagegaan of dit gemiddelde significant afwijkt van het gemiddeld aantal missende waarden in de totale reclasseringspopulatie⁸ (N = 11.590). Uit deze vergelij-

⁸ Net als in paragraaf 3.2 zijn alle RISC's van de deelnemers aan het interbeoordelaarsbetrouwbaarheidsonderzoek tijdelijk verwijderd. Dit is gedaan om de vergelijking tussen onderzoeksgroep en reclasseringspopulatie zuiver te houden en te voorkomen dat de onderzoeksgroep deels met zichzelf werd vergeleken.

kingen blijkt dat het gemiddeld aantal missende waarden in de onderzoeksgroep op slechts drie schalen afwijkt van het gemiddelde in de totale reclasseringspopulatie (zie tabel 4). In elk van deze gevallen is het gemiddeld aantal missende waarden in de onderzoeksgroep kleiner dan in de totale reclasseringspopulatie. In combinatie met de in tabel 3 beschreven analyses kan worden geconcludeerd dat de kwaliteit van de data voldoende is om de interbeoordelaarsbetrouwbaarheid van de RISc mee te bestuderen.

Tabel 4 Vergelijking gemiddeld aantal missende waarden per schaal met totale populatie^a

RISc-onderdeel	Gemiddeld aantal missende waarden		t
	Onderzoeksgroep	Vergelijkingsgroep	
Schaal 1 Delictgegevens	0,01	0,53	- 55,18***
Schaal 2 Analyse huidig delict en delictpatroon	0,51	0,52	- 0,06
Schaal 3 Huisvesting en wonen	1,01	0,76	1,63
Schaal 4 Opleiding, werk en leren	0,02	0,04	- 1,76
Schaal 5 Inkomen en omgaan met geld	0,01	0,02	- 2,00*
Schaal 6 Relaties met partner, gezin en familie	0,12	0,12	0,00
Schaal 7 Relaties met vrienden en kennissen	0,08	0,06	0,48
Schaal 8 Druggebruik	3,20	3,38	- 0,57
Schaal 9 Alcoholgebruik	0,58	0,63	- 0,96
Schaal 10 Emotioneel welzijn	0,01	0,02	- 2,00*
Schaal 11 Denkpatronen, gedrag en vaardigheden			
Schaal 12 Houding	0,05	0,06	- 0,37

^a Er is tweezijdig getoetst.

* p < .05

*** p < .001

3.4 Het analyseren van de interbeoordelaarsbetrouwbaarheid

Zoals in hoofdstuk 1 is beschreven, bestaat de RISc uit kwantitatieve en kwalitatieve items. Het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISc heeft zich met name gericht op de kwantitatieve, te scoren items omdat dit nu eenmaal de items zijn op basis waarvan schaal- en totaalscores worden berekend. Dit betekent dat de items van schaal 13 niet bij de analyses van de interbeoordelaarsbetrouwbaarheid zijn betrokken. Deze schaal bestaat immers in zijn geheel uit kwalitatieve items en er wordt voor deze schaal geen schaalscore berekend. Van de andere schalen zijn, naast de te scoren items, ook de vragen naar de relatie tussen enerzijds leefgebied en gevaar en anderzijds leefgebied en delictgedrag bij het onderzoek betrokken. Deze vragen, die met ja of nee moeten worden beantwoord, leveren weliswaar geen bijdrage aan de berekening van de schaalscores, maar spelen een belangrijke rol in het advies dat de reclasseringswerker uitbrengt met betrekking tot het reclasseringsaanbod. Daarnaast is een aantal vragen naar feitelijkheden, bijvoorbeeld naar detentie en naar druggebruik, bij het onderzoek betrokken.

De gebruikte analysetechnieken om de interbeoordelaarsbetrouwbaarheid van de RISc-items en —schalen vast te stellen, zijn afgestemd op het meetniveau van items en schalen. Een aantal vragen van de RISc is nominaal, zoals de vraag of de cliënt gedetineerd is en de vraag of de cliënt drugs gebruikt. Deze items worden

met ja of nee beantwoord en tellen niet mee in de score van de schaal. Om de mate van overeenstemming tussen de reclasseringswerkers op dergelijke items te onderzoeken, is coëfficiënt κ berekend (Cohen, 1960).

De items op basis waarvan de schaalscores van de RISC worden berekend, zijn ordinaal. Ook de items die aan het eind van bijna iedere schaal nagaan of het begrip dat de schaal beoogt te meten, verband houdt met gevaar en met delictgedrag, zijn als ordinale items geanalyseerd. Een bevestigend antwoord op deze vragen duidt immers op ernstiger problematiek dan een ontkennend antwoord. De schaalscores van de RISC worden, zoals in hoofdstuk 1 is beschreven, in drie categorieën ingedeeld: 1) geen criminogeen probleem, 2) criminogene factor is aanwezig, en 3) criminogene factor is in ernstige mate aanwezig. Ook de totaalscore van de RISC wordt in drie categorieën ingedeeld: 1) laag risico, 2) midden risico, en 3) hoog risico). In deze vorm zijn de RISC-schaal- en totaalscores van een ordinaal meetniveau⁹. Voor ordinale items, de schaalscores en de totaalscore is de volgende analysestrategie gehanteerd (cf. Born, 1995, p. 130-132):

- 1 Het berekenen van de proportie overeenstemming. Deze maat is de meest gebruikte en eenvoudigste maat, maar hiermee kon niet worden volstaan. Het ontbreekt namelijk zowel aan een controle voor kansovereenstemming als aan een formele toetsing van de mate van overeenstemming.
- 2 Het berekenen van de χ^2 -maat van Lawlis en Lu (1972). Deze maat geeft aan of de overeenstemming significant groter is dan op basis van toeval te verwachten is (kansovereenstemming).
- 3 T-maat van Tinsley en Weiss (1975) berekenen. Deze maat is een afgeleide van de χ^2 -maat van Lawlis en Lu en geeft een indicatie van de grootte van de overeenstemming (0=de overeenstemming is niet groter dan op basis van toeval is te verwachten, 1=perfecte overeenstemming).

De berekening van de χ^2 -maat van Lawlis en Lu bestaat onder meer uit de bepaling van kansovereenstemming. De kansovereenstemming hangt af van het aantal schaalpunten discrepantie tussen beoordelingen dat men nog toestaat als weergave van overeenstemming tussen beoordelaars (Born, 1995). Als dat toeneemt, zal ook de kansovereenstemming toenemen. Omdat de antwoordschaal van de RISC-items doorgaans uit drie en in een paar gevallen uit twee antwoordcategorieën bestaat, is een discrepantie van één of twee schaalpunten gedefinieerd als 'geen overeenstemming'.

Born (1995) hanteert in navolging van Angleitner, John en Löhr (1986, aangehaald in Born) voor de P -maat een waarde van minimaal 50% om van overeenkomst te kunnen spreken. Deze grenswaarde wordt in dit onderzoek overgenomen. Bij de interpretatie van Cohen's κ en de T-maat van Tinsley en Weiss wordt de indeling van Landis en Koch (1977) gehanteerd:

- een κ of T die kleiner is dan 0,21 duidt op een zwakke overeenstemming,
- een κ of T tussen 0,21 en 0,40 duidt op een matige overeenstemming,
- een κ of T tussen 0,41 en 0,60 duidt op een redelijke overeenstemming,

⁹ Aan de indeling in categorieën ligt een intervalschaal ten grondslag en op basis hiervan had de intraklassecorrelatie kunnen worden berekend. Dit is echter een relatieve maat, die aangeeft of beoordelaars tot eenzelfde rangordening komen. Voor de RISC is daarentegen van belang dat beoordelaars niet alleen tot eenzelfde rangordening komen, maar ook dezelfde waarden hanteren. Immers, de gecategoriseerde schaal- en totaalscores geven aan of er sprake is van een criminogeen probleem respectievelijk een risico op recidive. Om na te gaan of beoordelaars tot exact hetzelfde oordeel komen, moet een absolute maat voor interbeoordelaarsbetrouwbaarheid worden gehanteerd en de analysemethode voor ordinale items voldoet hieraan.

- een κ of T tussen 0,61 en 0,80 duidt op een sterke overeenstemming, en
- een κ of T tussen 0,81 en 1,00 duidt op een zeer sterke overeenstemming.

Analyse van de interbeoordelaarsbetrouwbaarheid is in principe alleen uitgevoerd indien minstens een kwart van de 75 cases een geldige waarde had. Op inhoudelijke gronden is in evenwel een uitzondering gemaakt in de analyse van de interbeoordelaarsbetrouwbaarheid van de schalen 3 *Huisvesting* en 8 *Druggebruik*. Omdat een ontbrekende score op deze schalen veroorzaakt kan worden doordat de schaal niet van toepassing is (wanneer de cliënt gedetineerd is respectievelijk geen drugs gebruikt), is voor deze schalen nagegaan hoeveel ontbrekende scores veroorzaakt worden door het ten onrechte niet invullen van de schaal. Indien dat aantal beperkt was, zijn toch analyses uitgevoerd waarbij de ontbrekende scores zijn gehercodeerd tot een geldige waarde, te weten ‘niet van toepassing’.

3.5 Resultaten van de analyses van de interbeoordelaarsbetrouwbaarheid

3.5.1 Schaal 1&2 Delictgegevens

De schalen 1 en 2 bieden informatie over de delictgeschiedenis van de cliënt en over de aard en achtergrond van het delict waarvoor de cliënt op het moment van afname van de RISC bij de reclassering komt. Voor de schalen 1 en 2 worden geen aparte schaalscores berekend, maar wordt volstaan met een combinatiescore die inzicht biedt in de delictgegevens van de cliënt. Schaal 1 bestaat uit vijf items die meetellen bij de berekening van de schaalscore, schaal 2 uit drie items.

Voor het beantwoorden van de items van schaal 2 is het van belang of een cliënt ontkent of niet. Indien een cliënt ontkent en hij of zij is nog niet veroordeeld, wordt deze schaal niet ingevuld. Dit heeft tot direct gevolg dat de schaalscore op de combinatieschaal 1 & 2 niet kan worden berekend en dat er geen totaalscore kan worden berekend. De vragen naar ontkenning en veroordeling zijn nominale variabelen, wat betekent dat voor de bepaling van de interbeoordelaarsbetrouwbaarheid Cohen's κ is berekend. Voor de vraag of de cliënt is veroordeeld zijn, indien de cliënt bekent, missende waarden gehercodeerd tot de geldige waarde ‘niet van toepassing’.

Tabel 5 toont dat de overeenstemming tussen twee beoordelaars op de vraag of de cliënt ontkent en op de vraag of de cliënt veroordeeld is, sterk is ($0,60 < \kappa < 0,81$). Hoewel het niet ondenkbaar is dat een cliënt tegen de ene reclasseringswerker bekennt en tegen de ander ontkent, wekt het, ondanks de goede κ , toch wat bevreemding dat men het niet geheel eens is over de vraag of de cliënt is veroordeeld. Het is mogelijk dat dit wordt verklaard door de tijd die verstreek tussen beide afnames: gemiddeld lagen er 29 dagen tussen afname van de eerste en de tweede RISC. Het kan zijn dat een rechtbank in een aantal gevallen uitspraak heeft gedaan in de periode tussen de twee afnames. Een alternatieve verklaring kan echter ook liggen in een verschil in interpretatie door de reclasseringswerker wanneer hoger beroep is aangetekend; de ene werker kan van mening zijn dat een cliënt is veroordeeld — de rechter heeft immers uitspraak gedaan — terwijl de tweede werker de mening is toegedaan dat de cliënt nog niet is veroordeeld zolang het hoger beroep nog loopt. Tot slot is het ook mogelijk dat het bekenen dan wel ontkennen van een cliënt door de reclasseringswerkers niet altijd eenduidig wordt geïnterpreteerd. Een cliënt

kan het delict gedeeltelijk bekennen wat voor de ene reclasseringswerker aanleiding is om te beslissen dat hij of zij te maken heeft met een bekende cliënt terwijl de andere reclasseringswerker van oordeel is te maken te hebben met een ontkennende cliënt.

Tabel 5 Interbeoordelaarsbetrouwbaarheid: ontkenning en veroordeling

RISc-item	N	κ
Ontkennende delinquent	75	0,76
Is veroordeeld	75	0,77

Voor het berekenen van de schaal score op de combinatieschaal 1&2 wordt gebruik gemaakt van acht items. In tabel 6 worden de resultaten van de analyses gegeven voor de zes items van schaal 1. Omdat schaal 2 niet wordt ingevuld als de cliënt ontkent en nog niet is veroordeeld, ontbreken voor een aantal daders de scores op de items van deze schaal. Voor de analyse van de interbeoordelaarsbetrouwbaarheid van schaal 2 zijn daarom de missende waarden, in het geval van ontkennende, niet veroordeelde verdachten, gehercodeerd tot geldige waarden. Vervolgens konden analyses worden uitgevoerd. Omdat de items door de hercodering van missende waarden niet meer ordinaal, maar nominaal van karakter waren, is voor elk item Cohen's κ berekend. De resultaten hiervan worden gepresenteerd in tabel 7. De nummering van de items komt overeen met die in de RISc-handleiding. De resultaten tonen aan dat er met betrekking tot de items van schaal 1 overeenstemming bestaat tussen de beoordelaars. De *P*-waarden liggen ruim boven 50% en de significante χ^2 -waarden wijzen eveneens op overeenstemming. Over het algemeen is de mate van overeenstemming sterk: de T-maat loopt uiteen van 0,44 tot 0,76. Eén item valt op: de vraag of er sprake is van diverse delicten en/of een zwaar delict in de justitiële voorgeschiedenis van de cliënt (item 1.9). De weliswaar redelijke, maar toch wat lage overeenstemming tussen beoordelaars op deze vraag wekt enige verbazing. Deze vraag kan immers beantwoord worden op basis van de door de werker op te vragen justitiële documentatie van de cliënt. Hierin staat de volledige *justitiële* delictgeschiedenis van de cliënt vermeld en kan door de werker worden nagegaan in hoeverre er diverse of zware delicten voorkomen. Wat onder een zwaar delict moet worden verstaan wordt in de RISc uitgewerkt en ook de toelichting op de antwoordcategorieën laat weinig ruimte voor eigen interpretatie. De interbeoordelaarsbetrouwbaarheid van de items van schaal 2 is redelijk tot sterk. Desondanks valt item 2,10 op. Het gaat om de vraag of de cliënt zijn of haar verantwoordelijkheid neemt voor het delict. De overeenstemming op dit item is aanzienlijk lager dan op de overige items uit deze schaal. Dit kan erop wijzen dat reclasseringswerkers het moeilijk vinden om in te schatten of de cliënt zijn verantwoordelijkheid neemt. Uiteraard is het mogelijk dat cliënten, afhankelijk van hun stemming, op het ene moment meer en op het andere moment minder bereid zijn de eigen verantwoordelijkheid te accepteren voor hun delictgedrag. Wellicht wordt de matige overeenstemming echter ook in de hand gewerkt door het feit dat de werkers uit slechts twee antwoordcategorieën kunnen kiezen en daarmee te weinig nuance aan kunnen brengen in hun oordeel. De ene werker laat dan andere overwegingen zwaarder wegen bij de keuze voor een antwoord dan de andere.

Tabel 6 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 1

RISC-item	N	P	χ^2	T
1.5 Aantal keer veroordeeld vóór 18 jaar	75	80%	72,82***	0,70
1.6 Aantal veroordelingen vanaf 18 jaar	74	77%	62,93***	0,66
1.7 Niet nakomen reclasseringsafspraken	74	88%	42,39***	0,76
1.8 Huidige/eerdere tenlastelegging voor inbraak	75	88%	43,33***	0,76
1.9 Diverse delicten in delictgeschiedenis	75	63%	28,62***	0,44

** p < .01

*** p < .001

Tabel 7 Interbeoordelaarsbetrouwbaarheid van de gehercodeerde items van RISC-schaal 2

RISC-item	N	κ
2.10 Verantwoordelijkheid nemen	71	0,47
2.11a Delict zijn onderdeel van een patroon	69	0,71
2.11b Delicten worden ernstiger	71	0,70

3.5.2 Schaal 3 Huisvesting en wonen

Op de schaal *Huisvesting en wonen* wordt nagegaan wat de woongeschiedenis en de huidige huisvestingssituatie van de reclasseringscliënt is. In de handleiding van de RISC wordt opgemerkt dat de huisvesting van een cliënt in sterke mate samenhangt met de kans op recidive en tegelijkertijd een van de meest dynamische en te beïnvloeden leefgebieden is.

Indien de cliënt gedetineerd is, dient de reclasseringswerker aan te geven of hij of zij op afzienbare termijn vrijkomt (binnen drie tot zes maanden). Indien een cliënt gedetineerd is en niet binnenkort vrijkomt, wordt uitsluitend het eerste item van de schaal beantwoord. In alle andere gevallen wordt de gehele schaal ingevuld. De vraag of een cliënt gedetineerd is en de vraag of hij op afzienbare tijd vrijkomt, zijn nominaal van aard. Voor deze vragen is Cohen's κ berekend. Om te voorkomen dat een te groot aantal missende waarden de analyse onmogelijk zou maken, zijn ten behoeve van de analyses van de vraag of de detentie binnenkort afloopt missende waarden tot een geldige waarde gehercodeerd indien de cliënt niet gedetineerd was (er is een extra antwoordcategorie 'gedetineerd' aangemaakt). De resultaten worden vermeld in tabel 8 en laten zien dat de overeenstemming tussen de beoordelaars sterk tot zeer sterk is. Dat κ voor de vraag of de cliënt gedetineerd is niet gelijk is aan 1,00 heeft heel waarschijnlijk te maken met de tijd die verstreek tussen de beide afnames. Een cliënt kan tussen beide RISC-afnames onder toezicht van de reclassering zijn gekomen waarbij zijn of haar detentie wordt opgeheven (schorsing voorlopige hechtenis). Tegelijkertijd is het ook mogelijk dat een cliënt zich niet aan de voorwaarden van het toezicht houdt en tussen de 1^e en de 2^e RISC weer vast komt te zitten (persoonlijke communicatie E. de Ruijter (RN), 3-8-2006).

Tabel 8 Interbeoordelaarsbetrouwbaarheid: detentie en einde detentie

RISC-item	N	κ
Gedetineerd?	75	0,87
Loopt detentie binnenkort af?	75	0,76

In item 3.1 wordt gevraagd of de cliënt ooit dakloos is geweest. Deze vraag kan (en moet) worden ingevuld, ongeacht de vraag of iemand ten tijde van de RISC-afname gedetineerd is. Ook de vragen naar de samenhang tussen huisvesting en gevaar respectievelijk het delictgedrag kunnen worden beantwoord wanneer een cliënt gedetineerd is.

Dat er op de drie vragen sprake is van overeenstemming blijkt uit de *P*-waarden, die ruim boven de 50% liggen, en de significante χ^2 -waarden (zie tabel 9). De *T*-waarden geven aan dat de mate van overeenstemming tussen de beoordelaars sterk is met betrekking tot de huisvestingsgeschiedenis van de cliënten en met betrekking tot de vraag of huisvesting verband houdt met het gevaar dat cliënten vormen voor zichzelf of voor anderen. De mate van overeenstemming op de vraag of er een relatie is tussen huisvesting en het delictgedrag is minder sterk, maar toch redelijk te noemen.

Tabel 9 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 3

RISC-item	N	<i>P</i>	χ^2	<i>T</i>
3.1 Huisvestingsgeschiedenis ^a	75	87%	40,35***	0,73
Relatie huisvesting en delictgedrag	75	76%	20,29***	0,52
Relatie huisvesting en gevaar	75	88%	43,33***	0,76

*** $p < .001$

^a Omdat item 3.1 te scheef verdeeld was voor een normale analyse, is de oorspronkelijke antwoordschaal gehercodeerd tot een tweepuntsschaal.

Voor de analyse van de interbeoordelaarsbetrouwbaarheid van de drie overige items van de schaal dienden de missende waarden, indien deze het gevolg waren van detentie, te worden gehercodeerd tot geldige waarden. Vervolgens konden analyses worden uitgevoerd. Omdat de items door de hercodering van missende waarden niet meer ordinaal, maar nominaal van karakter waren, is voor elk item Cohen's κ berekend. De mate van overeenstemming tussen de reclasseringswerkers op deze items is in alle gevallen redelijk (zie tabel 10). Dat lijkt voor een schaal die vrij feitelijk van aard is, wellicht aan de magere kant, maar hierbij moet aangetekend worden dat in deze resultaten ook de overeenstemming 'ingebakken' zit die beoordelaars bereiken op de vraag of er sprake is van detentie en vrijlating op afzienbare termijn.

Tabel 10 Interbeoordelaarsbetrouwbaarheid van de gehercodeerde items van RISC-schaal 3

RISC-item	N	κ
3.2 Huidige woonsituatie	75	0,60
3.3 Geschiktheid van de woning	75	0,60
3.4 Woonomgeving	74	0,57

3.5.3 Schaal 4 Opleiding, werk en leren

De schaal *Opleiding, werk en leren* beoogt met behulp van zeven te scoren items het opleidingsniveau en de werkgerelateerde vaardigheden van de cliënt in kaart te brengen. Er wordt gekeken naar het verleden, de huidige situatie en de houding van de cliënt ten opzichte van opleiding, werk en leren. De handleiding van de RISC vermeldt dat opleiding, werk en leren enerzijds zeer dynamische factoren zijn

die samenhangen met delictgedrag, en anderzijds factoren zijn die goed te beïnvloeden zijn door middel van gedragsinterventies.

De vraag of het opleidingsniveau van de cliënt verband houdt met het gevaar dat hij voor zichzelf of voor anderen vormt, was te scheef verdeeld om te kunnen worden geanalyseerd. Omdat dit soort vragen uitsluitend met 'ja' en 'nee' beantwoord kunnen worden, kon geen hercodering plaatsvinden waardoor het niet mogelijk is uitspraken te doen over de mate van overeenstemming tussen de reclasseringswerkers met betrekking tot deze vraag.

In tabel 11 worden de resultaten van de interbeoordelaarsbetrouwbaarheidsanalyses van de overige items van schaal 4 weergegeven. In alle gevallen is er sprake van overeenstemming. De *P*-waarden zijn hoger dan 50% en de χ^2 -waarden zijn significant. De mate van overeenstemming tussen de beoordelaars op de items die gebruikt worden voor de berekening van de schaa score (items 4.1 t/m 4.7) is blijkens de T-maten redelijk tot sterk. Toch had wellicht verwacht mogen worden dat de overeenstemming op vragen naar feitelijke informatie zoals opleidingsniveau (afgeronde schoolopleidingen) en schoolbezoek sterker zou zijn geweest.

Met betrekking tot de inschatting of opleiding en werk samenhangen met het delictgedrag is de overeenstemming tussen de beoordelaars slechts matig. Kennelijk is het moeilijk om te beoordelen of er een relatie is tussen opleiding, werk en leren enerzijds en het delictgedrag anderzijds en komen werkers op basis van verschillende overwegingen tot hun oordeel. Het item is echter van groot belang binnen de RISC omdat de aan- of afwezigheid van een dergelijke relatie in de reclasseringspraktijk gevolgen heeft voor de vraag of vanuit justitie in de opleiding en arbeidstoeleiding van een cliënt zal worden geïnvesteerd. Indien opleiding en werk geen verband houden met delictgedrag, zullen op dit terrein, ongeacht het opleidingsniveau of de werksituatie van een cliënt, door justitieorganisaties geen interventies worden aangeboden. Een goede interbeoordelaarsbetrouwbaarheid is daarom van belang. In hoofdstuk 7 worden suggesties gedaan om een betere interbeoordelaarsbetrouwbaarheid te bereiken.

Tabel 11 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 4

RISC-item	N	<i>P</i>	χ^2	T
4.1 Opleidingsniveau	75	72%	49,90***	0,58
4.2 Schoolbezoek	73	70%	43,30***	0,55
4.3 Belemmeringen voor school en werk	75	76%	60,82***	0,64
4.4 Werkervaring en werkverleden	75	88%	57,06***	0,62
4.5 Huidige werksituatie	74	74%	55,35***	0,61
4.6 Vaardigheden tav opleiding, werk en leren	75	67%	37,02***	0,50
4.7 Houding t.o.v. opleiding, werk en leren	75	69%	43,22***	0,54
Relatie opleiding / werk en delictgedrag	75	67%	8,35**	0,33

** $p < .01$

*** $p < .001$

3.5.4 Schaal 5 Inkomen en omgaan met geld

De schaal *Inkomen en omgaan met geld* bestaat uit vier items en beoogt in kaart te brengen welke bronnen van inkomsten de cliënt heeft, of hij voor zijn financiën afhankelijk is van anderen of van criminele activiteiten en of zijn inkomsten en

uitgaven in balans zijn. De reclasseringswerker gaat aan de hand van deze schaal tevens na of er schulden zijn en of er sprake is van problematisch gokgedrag. Een slechte financiële situatie is volgens de handleiding van de RISC een risicofactor voor het herhalen van delictgedrag.

Het item 5.4 (Gokverslaving) bleek te scheef verdeeld te zijn om analyses uit te kunnen voeren. Omdat de oorspronkelijke antwoordschaal uit slechts twee antwoordcategorieën bestaat, was het niet mogelijk het item te hercoderen en zo doende kunnen geen uitspraken worden gedaan over de interbeoordelaarsbetrouwbaarheid van dit item. De interbeoordelaarsbetrouwbaarheid van de overige items wordt vermeld in tabel 12. In alle gevallen is sprake van overeenstemming tussen de beoordelaars: de *P*-waarden zijn hoger dan 50% en de χ^2 -waarden zijn significant. Uit de *T*-waarden blijkt dat de mate van overeenstemming voor alle items redelijk tot sterk is.

Omdat de inkomensschaal, net als de huisvestings- en opleidingsschaal, vrij feitelijk van aard is, zou de vraag gesteld kunnen worden of de overeenstemming op de items die meetellen bij de berekening van de schaalscore (items 5.1, 5.2 en 5.3) niet sterker had moeten zijn.

Tabel 12 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 5

RISC-item	N	<i>P</i>	χ^2	<i>T</i>
5.1 Hoofdbron inkomsten	75	76%	60,82***	0,64
5.2 Huidige financiële situatie	75	76%	60,82***	0,64
5.3 Ernstige beperkingen in budget	75	71%	46,50***	0,56
Relatie inkomen en delictgedrag	73	78%	23,04***	0,56
Relatie inkomen en gevaar	74	82%	31,15***	0,65

*** $p < .001$

3.5.5 Schaal 6 Relaties met partner, gezin en familie

De schaal *Relaties met partner, gezin- en familie* bestaat uit vijf items en meet de kwaliteit van de hechte relaties die een delinquent is zijn leven is aangegaan. De schaal beperkt zich tot de directe omgeving van de delinquent. Er wordt gekeken naar de aanwezigheid van hechte banden een niet-criminele partner en niet-criminele gezins- of familieleden. Met name de stabiliteit van en tevredenheid met bestaande relaties wordt beoordeeld. Dit geldt ook voor alleenstaanden: in hoeverre is men tevreden met de huidige situatie.

De interbeoordelaarsbetrouwbaarheid van de bij schaal 6 horende items wordt weergegeven in tabel 13. Ook bij deze schaal blijkt dat er op alle items sprake is van overeenstemming (de *P*-maten zijn groter dan 50% en de χ^2 -waarden zijn significant), maar de mate van overeenstemming lijkt ten opzichte van de in de voorgaande paragrafen besproken schalen, iets lager te zijn. Deze schaal is van een veel minder feitelijk karakter dan de voorgaande schalen en het is dan ook niet onvoorstelbaar dat het relatief moeilijk is om de kwaliteit van de hechte relaties van cliënten te beoordelen. Desondanks is de mate van overeenstemming op de meeste items die meetellen bij de berekening van de schaalscore (items 6.1 t/m 6.5), redelijk. Een van de items — item 6.5 Huiselijk geweld — valt echter op vanwege de matige overeenstemming tussen de beoordelaars. De resultaten wijzen erop dat de reclasseringswerkers het relatief vaak oneens zijn over de vraag of er in de thuissi-

tuatie van de cliënt sprake zou kunnen zijn van huiselijk geweld. Dit is een item met een tweepuntsantwoordschaal die beoordelaars niet de mogelijkheid biedt om enige twijfel uit te spreken. Beoordelaars kunnen kiezen tussen ‘er is geen enkele aanleiding om te denken dat...’ en ‘er is objectief bewijs’. Wellicht dat deze twee categorieën te weinig ruimte bieden aan beoordelaars om hun inschatting goed kwijt te kunnen.

De vragen naar het verband tussen familierelaties en het delictgedrag cq. gevaar worden door de reclasseringswerkers in redelijke tot sterke mate van overeenstemming beantwoord.

Tabel 13 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 6

RISC-item	N	P	χ^2	T
6.1 Jeugdervaringen	73	68%	40,10***	0,53
6.2 Ervaring met hechte relaties vanaf (jong) vol-wassenheid	71	63%	28,41***	0,45
6.3 Huidige relatie met partner, gezin en familie	74	62%	27,26***	0,43
6.4 Familie- of gezinslid heeft justitieel dossier	69	75%	17,77***	0,51
6.5 Huiselijk geweld	71	69%	10,28**	0,38
Relatie familierelaties en delictgedrag	75	76%	20,29***	0,52
Relatie familierelaties en gevaar	75	84%	34,69***	0,68

** p < .01

*** p < .001

3.5.6 *Schaal 7 Relaties met vrienden en kennissen*

De schaal *Relaties met vrienden en kennissen* beoogt met behulp van vier items in kaart te brengen hoe een cliënt sociaal functioneert en hoe het sociale netwerk van de cliënt eruit ziet buiten zijn familie. De manier waarop een delinquent sociaal functioneert en de aard van zijn of haar sociale netwerk zijn, volgens de RISC-handleiding, beide belangrijk om in te kunnen schatten wat het risico op recidive is. Daarnaast wordt nagegaan in hoeverre de cliënt grote risico's neemt, buitensporig sensatie zoekt en meeloopt met anderen.

In tabel 14 worden de resultaten gepresenteerd van de interbeoordelaarsbetrouwbaarheidsanalyses van de items van schaal 7. Bestudering van de P-maten en de χ^2 -waarden leert dat er in alle gevallen sprake is van overeenstemming. Net als schaal 6 is deze schaal minder feitelijk van inhoud dan de eerste vijf RISC-schalen. Dienovereenkomstig zijn de T-maten wat lager dan bij de feitelijkere schalen, maar de overeenstemming is op elk item en elke vraag redelijk.

Tabel 14 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 7

RISC-item	N	P	χ^2	T
7.1 Vrienden en kennissenkring	72	67%	35,52***	0,50
7.2 Laat zich negatief beïnvloeden	72	71%	45,02***	0,56
7.3 Manipuleert en gebruikt vrienden / kennissen	71	65%	31,16***	0,47
7.4 Zoekt sensatie, opwindning, risico's	74	73%	51,74***	0,59
Relatie vrienden en delictgedrag	75	75%	18,27***	0,49
Relatie vrienden en gevaar	75	77%	22,43***	0,55

*** p < .001

3.5.7 Schaal 8 Druggebruik

Aan de hand van zes te scoren items beoogt de schaal *Druggebruik* een eerste indicatie te geven van de mate waarin drugs een rol spelen bij het delictgedrag van de cliënt. Omdat er vele goede instrumenten zijn om druggebruik en de achterliggende motieven nauwkeurig in kaart te brengen, is de RISC-schaal die betrekking heeft op druggebruik beknopt gehouden. Indien er sprake lijkt te zijn van problematisch druggebruik dient door middel van verdiegingsdiagnostiek nader onderzoek te worden gedaan. Bij de vrijwel alle schalen van de RISC zijn steeds twee, met ja of nee te beantwoorden vragen opgenomen om na te gaan of er een relatie is tussen enerzijds het leefgebied en anderzijds het delict en het risico op gevaar. Bij deze schaal en bij de alcoholschaal zijn deze vragen opgenomen als kwantitatieve items in de schaal en tellen mee bij de berekening van de schaalscore.

Voordat de schaal wordt ingevuld, beantwoordt de reclasseringswerker de vraag of de cliënt drugs gebruikt. Indien dit niet het geval is, krijgt de cliënt een score 0 op de schaal en worden de items van de schaal niet ingevuld. De vraag of een cliënt drugs gebruikt is een nominale variabele. De beoordelaars blijken met betrekking tot deze variabele in sterke mate met elkaar overeen te stemmen (zie tabel 15).

Tabel 15 Interbeoordelaarsbetrouwbaarheid: druggebruik

RISC-item	N	κ
Druggebruik	75	0,71

Wanneer een cliënt geen drugs gebruikt, hoeven de items van schaal 8 niet te worden ingevuld en krijgt de cliënt de laagste schaalscore (0). Hierdoor heeft een groot aantal cliënten missende waarden op de items. Voor de berekening van de interbeoordelaarsbetrouwbaarheid zijn deze missende waarden, net als voor schaal 3, gehercodeerd tot geldige scores ('gebruikt geen drugs'). Vervolgens konden de analyses worden uitgevoerd. Omdat de items door de hercodering van missende waarden niet meer ordinaal, maar nominaal van karakter waren, is voor elk item Cohen's κ berekend. Tabel 16 toont aan dat de overeenstemming tussen twee beoordelaars op de vragen van schaal 8 redelijk is. Het laatste item van de schaal (de vraag of een cliënt gemotiveerd is om iets aan zijn druggebruik te doen) hoeft slechts beantwoord te worden wanneer op de eerste vier items samen een score van één of meer wordt gehaald. Om deze reden is het aantal personen in de analyse van dit item beduidend lager dan bij de andere items.

Tabel 16 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 8

RISC-item	N	κ
8.1a Soort drug	75	0,57
8.1b Frequentie van gebruik	75	0,57
8.2 Drugs staan centraal	74	0,56
8.3 Relatie druggebruik en delictgedrag	75	0,60
8.4 Relatie druggebruik en gevaar	75	0,54
8.5 Motivatie om druggebruik aan te pakken	61	0,60

3.5.8 *Schaal 9 Alcoholgebruik*

De schaal *Alcoholgebruik* bestaat uit vijf items waarmee beoogd wordt recent alcoholgebruik vast te stellen en de relatie daarvan met het delict. Het doel van de RISC-schaal is niet om na te gaan of er vanuit gezondheidsoogpunt sprake is van problematisch alcoholgebruik, maar uitsluitend of het alcoholgebruik verband houdt met het delictgedrag. Net als op de schaal *Druggebruik* zijn de vragen naar de relatie tussen het leefgebied en delictgedrag en gevaar opgenomen als te scoren items in de schaal.

Het laatste item van de schaal (de vraag of een cliënt gemotiveerd is om iets aan zijn alcoholgebruik te doen) hoeft slechts beantwoord te worden wanneer de score op de eerste vier items samen één of hoger is. Omdat dit slechts in weinig gevallen voorkwam, was de N van item 9.5 (motivatie) te laag om analyses op uit te voeren (N = 25). In tabel 17 worden de resultaten van de interbeoordelaarsbetrouwbaarheidsanalyses van de vier overige items weergegeven. In alle gevallen is sprake van overeenstemming (de *P*-maten zijn groter dan 50% en de χ^2 -waarden zijn significant). De mate van overeenstemming is voor alle items sterk.

Tabel 17 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 9

RISC-item	N	<i>P</i>	χ^2	T
9.1 Overmatig alcoholgebruik in het verleden	75	75%	57,06***	0,62
9.2 Huidige gebruik is een probleem ^a	75	84%	34,69***	0,68
9.3 Relatie alcoholgebruik en delictgedrag	75	80%	90,50***	0,78
9.4 Relatie alcoholgebruik en gevaar ^a	75	87%	40,35***	0,73

^a Omdat de items 9.2 en 9.4 te scheef verdeeld waren voor een normale analyse, is de oorspronkelijke antwoordschaal gehercodeerd tot een tweepuntsschaal.

*** $p < .001$

3.5.9 *Schaal 10 Emotioneel welzijn*

Op basis van vijf items wordt met de schaal *Emotioneel welzijn* beoogd in te schatten hoe de cliënt in het leven staat, welk zelfbeeld hij heeft en of er mogelijk sprake is van psychische problemen. De schaal heeft volgens de handleiding van de RISC betrekking op die emotionele en psychologische factoren die samenhangen met de kans op recidive. Psychische problemen zijn aan de orde als de cliënt in zijn of haar dagelijks functioneren wordt belemmerd door een psychische of psychiatrische conditie. De handleiding waarschuwt dat emotioneel welzijn en psychische problemen recidive op zichzelf niet voorspellen, maar wel bijdragen aan het in stand houden van gedragspatronen van bepaalde groepen daders.

Tabel 18 toont de resultaten van de interbeoordelaarsbetrouwbaarheidsanalyses van de items van schaal 10. Eén van de items (10.4 Zelfdestructief gedrag) was te scheef verdeeld om te kunnen analyseren (daders scoorden voornamelijk nul). Hercodering van de oorspronkelijke antwoordschaal was niet mogelijk, omdat die voor dit item uit slechts twee categorieën bestond. Analyses konden daarom niet op dit item worden uitgevoerd en er kunnen dan ook geen uitspraken worden gedaan over de interbeoordelaarsbetrouwbaarheid van dit item. Op de overige items is sprake van overeenstemming: de *P*-maten zijn groter dan 50% en de χ^2 -waarden zijn significant. De mate van overeenstemming is in vrijwel alle gevallen redelijk.

Gezien de complexiteit van de schaal — het inschatten van emotionele en psychische problemen is geen eenvoudige taak — is dit een bevredigend resultaat te noemen. Bovendien kan de beoordelaar voor de scoring van deze schaal in mindere mate terugvallen op dossierinformatie en moet hij of zij de benodigde informatie grotendeels uit het gesprek met de cliënt halen. De overeenstemming over de relatie tussen emotioneel welzijn en gevaar blijkt evenwel matig te zijn.

Tabel 18 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 10

RISC-item	N	P	χ^2	T
10.1 Moeite zich staande te houden	75	65%	34,10***	0,48
10.2 Psychische problemen	75	71%	46,50***	0,56
10.3 Zelfbeeld	75	61%	26,06***	0,42
10.5 Bijzondere omstandigheden in verleden	74	72%	48,26***	0,57
Relatie emotioneel welzijn en delictgedrag	75	72%	14,53***	0,44
Relatie emotioneel welzijn en gevaar	75	69%	11,23***	0,39

*** p < .001

Wanneer een cliënt gedetineerd is, wordt door de beoordelaar ingeschat of de score op schaal 10 toe te schrijven kan zijn aan de effecten van de detentie. Deze vraag dient met ja of nee te worden beantwoord. Voor de analyse van de interbeoordelaarsbetrouwbaarheid van deze vraag dienden de missende waarden, indien deze het gevolg waren van detentie, te worden gehercodeerd tot geldige waarden, namelijk 'niet van toepassing'. De analyse die vervolgens uitgevoerd kon worden, toont aan dat de mate van overeenstemming tussen de beoordelaars redelijk is (zie tabel 19).

Tabel 19 Interbeoordelaarsbetrouwbaarheid: scores op schaal 10 toe te schrijven aan detentie?

RISC-item	N	K
Score toe te schrijven aan detentie?	75	0,54

3.5.10 Schaal 11 Denkpatronen, gedrag en vaardigheden

De schaal *Denkpatronen, gedrag en vaardigheden* dient op basis van acht items in kaart te brengen hoe de delinquent omgaat met zaken die hij in het dagelijks leven tegenkomt. De items van de schaal hebben betrekking op uiteenlopende kenmerken van delinquenten. Voorbeelden zijn sociale en interpersoonlijke vaardigheden, impulsiviteit, dominant gedrag, zelfbeheersing, probleembesef en probleemhanteling. Tekorten of problemen op deze gebieden blijken volgens de RISC-handleiding samen te hangen met de kans op recidive. Tegelijkertijd blijkt dat cognitief-gedragsmatige interventies, die zich richten op dit soort problemen, tot de effectiefste gedragsinterventies ter voorkoming van recidive horen (zie onder meer AOs, Miller & Drake, 2006; Farrington & Welsh, 2005; McGuire, 2002; Tong & Farrington, 2006; Wilson, Bouffard & MacKenzie, 2005). De RISC-handleiding stelt dan ook dat de score op deze schaal in hoge mate bepaalt of een delinquent geschikt is voor het volgen van een cognitief trainingsprogramma. Een betrouwbare (en valide) inschatting van cognitieve problemen bij delinquenten is daarom van groot belang.

In tabel 20 worden de resultaten van de analyses weergegeven voor de acht kwantitatieve items van schaal 11 en de vragen naar de samenhang tussen denkpatroon en gevaar cq. delictgedrag. In alle gevallen is sprake van overeenstemming: de *P*-maten zijn groter dan 50% en de χ^2 -waarden zijn significant. Wat echter opvalt is dat de *P*-maten voor zes van de acht te scoren items niet veel groter zijn dan 50% en voor alle te scoren items duiden de *T*-maten op een matige overeenstemming tussen de beoordelaars. De mate van overeenstemming op de vragen of er een relatie is tussen denkpatroon en delictgedrag en denkpatroon en gevaar is redelijk.

Schaal 11 hoort, vanwege zijn relatief subjectieve karakter, met de schalen 10 en 12 tot de moeilijkst te scoren schalen. De reclasseringswerker kan bij de afname van de RISC voor deze drie schalen ook nauwelijks terugvallen op dossierinformatie en dient de schalen voornamelijk te scoren op basis van informatie uit het gesprek met de cliënt. Gezien het belang dat aan schaal 11 blijkt de handleiding wordt gehecht, is de matige overeenstemming op de items echter wel degelijk problematisch.

Tabel 20 Interbeoordelaarsbetrouwbaarheid van de items van RISC-schaal 11

RISC-item	N	<i>P</i>	χ^2	<i>T</i>
11.1 Sociale en interpersoonlijke vaardigheden	75	68%	40,06***	0,52
11.2 Impulsiviteit	75	56%	17,02***	0,34
11.3 Dominant gedrag	75	55%	15,06***	0,32
11.4 Zelfbeheersing	75	55%	15,06***	0,32
11.5 Probleembesef	75	57%	19,10***	0,36
11.6 Probleemhantering	75	60%	23,62***	0,40
11.7 Doelgerichtheid, toekomstperspectief	75	57%	19,10***	0,36
11.8 Denktrant en leerbaarheid	75	59%	21,30***	0,38
Relatie denkpatronen en delictgedrag	75	71%	12,83***	0,41
Relatie denkpatronen en gevaar	75	71%	12,83***	0,41

*** $p < .001$

Wanneer een cliënt gedetineerd is, wordt door de beoordelaar ingeschat of de score op schaal 11 toe te schrijven kan zijn aan de effecten van de detentie. Deze vraag dient met ja of nee te worden beantwoord. De analyse is op dezelfde wijze uitgevoerd als voor de identieke vraag die bij schaal 10 wordt gesteld (hercodering van missende waarden, indien deze het gevolg waren van detentie) en laat een sterke mate van overeenstemming tussen de beoordelaars zien (zie tabel 21).

Tabel 21 Interbeoordelaarsbetrouwbaarheid: scores op schaal 11 toe te schrijven aan detentie?

RISC-item	N	κ
Score toe te schrijven aan detentie?	75	0,65

3.5.11 Schaal 12 Houding

Op basis van vijf items beoogt de schaal *Houding* de houding van de delinquent te meten tegenover het delict, de maatschappij, criminaliteit in het algemeen evenals de houding tegenover (gedrags)interventies en beoogde (gedrags)verandering. De houding van de dader zou volgens de RISC-handleiding één van de belangrijkste

factoren zijn die bijdraagt aan recidive en zou daarom altijd inzet moeten zijn van op recidivevermindering gerichte interventie door justitie.

Op de items die gebruikt worden in de berekening van de schaalscore blijkt sprake te zijn van overeenstemming tussen de beoordelaars (de *P*-maten zijn groter dan 50% en de χ^2 -waarden zijn significant; zie tabel 22). De mate van overeenstemming is op drie van deze vijf items is echter matig. Onder deze drie bevindt zich ook de vraag of de cliënt gemotiveerd is om te veranderen en bereid is om mogelijkheden tot verandering te gebruiken (item 12.5). Dit is een belangrijk item binnen de RISc omdat het iets zegt over de responsiviteit van de cliënt — de (on)mogelijkheden van de dader om te veranderen. Een inschatting van de responsiviteit moet door de reclaseringswerker worden betrokken in zijn of haar oordeel over de vraag of het zinnig is om gedragsinterventies aan te bieden aan de cliënt. Ook de beoordelingen van de manier waarop de cliënt het delict en zijn eigen aandeel nu beleeft (item 12.1 Pro-criminele houding) stemmen slechts matig overeen tussen de beoordelaars. Een pro-criminele houding zou volgens de handleiding echter een adequate voorspeller zijn van recidive en het breken van voorwaarden. De inschatting van deze houding blijkt echter niet op een heel betrouwbare manier te plaats te vinden. Verontrustend is ook dat er géén overeenstemming bestaat tussen de beoordelaars met betrekking tot de vraag of de houding van de cliënt samenhangt met het risico op recidive (de *P*-maat is weliswaar groter dan 50%, maar de χ^2 -waarde is niet significant). Zoals eerder is opgemerkt, wordt de houding van de dader in de RISc-handleiding beschouwd als een van de belangrijkste factoren in de voorspelling van recidive en juist over deze samenhang zijn de beoordelaars het niet eens.

Schaal 12 beoogt een uitzonderlijk moeilijk te meten construct in kaart te brengen. In de handleiding van de RISc wordt in de inleiding bij deze schaal al gesteld dat attitudes, of houdingen lastig te meten zijn. Toch wordt de houding van de cliënt als een bijzonder belangrijke factor gezien in de voorspelling van recidive én de mogelijkheden tot interventie. De resultaten van de interbeoordelaarsbetrouwbaarheidsanalyses stemmen daarom niet geheel positief.

Tabel 22 Interbeoordelaarsbetrouwbaarheid van de items van RISc-schaal 12

RISc-item	N	<i>P</i>	χ^2	T
12.1 Pro-criminele houding	73	59%	21,11***	0,38
12.2 Houding tegenover de sanctie ^a	74	72%	13,85***	0,43
12.3 Houding t.a.v. de samenleving	74	70%	44,89***	0,55
12.4 Inzicht en houding to. zichzelf en delictgedrag	73	53%	12,97***	0,30
12.5 Veranderingsgezindheid	74	58%	20,08***	0,37
Relatie houding en delictgedrag	75	60%	3,01 ^{n.s.}	0,20
Relatie houding en gevaar	75	71%	12,83***	0,41

^a Omdat item 12.2 te scheef verdeeld was voor een normale analyse, is de oorspronkelijke antwoordschaal gehercodeerd tot een tweepuntsschaal.

^{n.s.} niet significant

*** $p < .001$

Wanneer een cliënt gedetineerd is, wordt door de beoordelaar ingeschat of de score op schaal 12 toe te schrijven kan zijn aan de effecten van de detentie. Deze vraag dient met ja of nee te worden beantwoord. De analyse is op dezelfde wijze uitgevoerd als voor de identieke vraag die bij de schalen 10 en 11 worden gesteld

(hercodering van missende waarden, indien deze het gevolg waren van detentie) en laat een sterke mate van overeenstemming tussen de beoordelaars zien (zie tabel 23).

Tabel 23 Interbeoordelaarsbetrouwbaarheid: scores op schaal 12 toe te schrijven aan detentie?

RISc-item	N	κ
Score toe te schrijven aan detentie?	75	0,67

3.5.12 Schaal- en totaalscores

De resultaten van de analyses van de interbeoordelaarsbetrouwbaarheid van de schaalscores en de totaalscore worden in tabel 24 getoond. Bij alle schaalscores en bij de totaalscore is sprake van overeenstemming: de P -maten zijn groter dan 50% en de χ^2 -waarden zijn significant. De T -maten wijzen in alle gevallen op een redelijke tot sterke mate van overeenstemming. Ook de schalen waarvan de items tot meer problemen leiden, vertonen een redelijke tot goede interbeoordelaarsbetrouwbaarheid. De bepaling van de schaal- en totaalscores van het instrument lijkt dus relatief vrij te zijn van beoordelaarsinvloeden. Anders geformuleerd komen reclasseringswerkers die onafhankelijk van elkaar de RISc afnemen bij dezelfde cliënt, tot redelijk goed overeenstemmende oordelen over de aan- dan wel afwezigheid van criminogene problemen en over het recidiverisico van de cliënt.

Omdat de analyses van de schaalscores 1&2 *Delictgegevens* en 3 *Huisvesting* uitsluitend op een selectieve groep konden worden uitgevoerd — ontkennende, niet veroordeelde verdachten en respectievelijk gedetineerde daders — zijn de analyses herhaald waarbij, wanneer dat aan de orde was, de ontbrekende scores werden gehercodeerd tot de geldige waarde ‘ontkennend en niet veroordeeld’ respectievelijk ‘gedetineerd’. Omdat de schaalscores door deze hercodering niet meer ordinaal, maar nominaal van karakter waren, is voor deze schalen Cohen’s κ berekend. Ook op basis van deze analyses is de interbeoordelaarsbetrouwbaarheid van beide schalen redelijk: $\kappa_{1\&2} = 0,48$ en $\kappa_3 = 0,54$.

Tabel 24 Interbeoordelaarsbetrouwbaarheid van de RISc-schalen en de RISc-totaalscore

RISc-onderdeel	N	P	χ^2	T
Schaal 1&2 Delictgegevens	60	72%	39,12***	0,58
Schaal 3 Huisvesting en wonen	45	67%	22,03***	0,50
Schaal 4 Opleiding, werk en leren	75	72%	49,90***	0,58
Schaal 5 Inkomen en omgaan met geld	75	80%	72,82***	0,70
Schaal 6 Relaties met partner, gezin en familie	75	65%	34,10***	0,48
Schaal 7 Relaties met vrienden en kennissen	73	78%	65,13***	0,67
Schaal 8 Druggebruik	75	79%	68,70***	0,68
Schaal 9 Alcoholgebruik	75	79%	68,70***	0,68
Schaal 10 Emotioneel welzijn	75	85%	90,50***	0,78
Schaal 11 Denkpatronen, gedrag en vaardigheden	75	72%	49,90***	0,58
Schaal 12 Houding	74	62%	27,26***	0,43
Totaalscore	60	78%	54,02***	0,68

*** $p < .001$

4 Schaalstructuur en interne consistentie van de RISC

4.1 Het beoordelen van structuur en de interne consistentie van de RISC

Zoals is hoofdstuk 2 is toegelicht, is het onderzoek naar de schaalstructuur en de interne consistentie van de RISC uitgevoerd met behulp van een van de Reclasse-ring Nederland (RN) verkregen database. Hierin zijn alle RISC's zijn opgenomen die tussen november 2004 en mei 2006 door de drie reclasseringsorganisaties zijn ge-start. Het aantal RISC's waarop analyses konden worden uitgevoerd, bedroeg na de noodzakelijke datacleaning 11.666 RISC's. De onderzoeksgroep bestaat uit 10.572 mannen en 1.094 vrouwen (zie tabel 25). Van één persoon was het geslacht niet bekend. De gemiddelde leeftijd in de onderzoeksgroep was 34 jaar en 4 maanden (34,30 jaar) met een standaarddeviatie van 12 jaar (12,01 jaar). Van 34 personen was de leeftijd onbekend. De jongste cliënt was 17 jaar en 1 maand (17,11 jaar) en de oudste 86 jaar en 6 maanden (86,48 jaar). Het grootste deel van de onderzoeksgroep is in Nederland geboren (66,6%). Ongeveer 60% van de onderzoeksgroep is autochtoon, 33,1% is niet-westers allochtoon en 3,9 % heeft een westerse allochto-ne achtergrond. Van 297 personen was onbekend waar hij of zij geboren was en van 300 cliënten kon niet worden achterhaald wat zijn of haar herkomst is.

Tabel 25 Achtergrondkenmerken onderzoeksgroep interne consistentie (N = 11.666)

Kenmerk		Aantal	%
Geslacht	Man	10.572	90,6
	Vrouw	1.094	9,4
Leeftijd	Jonger dan 20 jaar	1.167	10,0
	20 – 30 jaar	3.751	32,2
	30 – 40 jaar	3.137	26,9
	40 – 50 jaar	2.307	19,8
	50 – 60 jaar	917	7,9
	60 jaar of ouder	353	3,0
	Onbekend	34	0,3
Geboorteland	Nederland	7.999	68,6
	Marokko	582	5,0
	Nederlandse Antillen / Aruba	590	5,1
	Suriname	754	6,5
	Turkije	352	3,0
	Overig niet-westers	677	5,8
	Westers	415	3,6
	Onbekend	297	2,5
Herkomstgroep	Autochtoon	7.051	60,4
	Niet-westers allochtoon	3.856	33,1
	Westers allochtoon	459	3,9
	Onbekend	300	2,6

Om de schaalstructuur van het instrument te onderzoeken, is steeds voor elke a priori vastgestelde schaal en voor de totaalscore een principale componentenana-

lyse (pca) uitgevoerd op de bijbehorende items respectievelijk schalen (met iteraties en varimaxrotatie). Iedere keer is nagegaan in hoeverre de pca één factor opleverde en zijn de factorladingen bestudeerd. Daarnaast is voor elke schaal en voor de totaalscore een betrouwbaarheidsanalyse uitgevoerd. De interne consistentie van elke schaal en van de totaalscore is berekend met behulp van Cronbach's coëfficiënt alfa. De Commissie Testaangelegenheden Nederland (COTAN; Evers, Van Vliet-Mulder & Groot, 2000) verbindt de volgende kwalificaties aan alfacoëfficiënten:

- Bij tests voor belangrijke beslissingen op individueel niveau: $\alpha < 0,80$ = onvoldoende; $0,80 \leq \alpha < 0,90$ = voldoende; $\alpha \geq 0,90$ = goed.
- Bij tests voor minder belangrijke beslissingen op individueel niveau: $\alpha < 0,70$ = onvoldoende; $0,70 \leq \alpha < 0,80$ = voldoende; $\alpha \geq 0,80$ = goed.
- Bij tests voor onderzoek op groepsniveau: $\alpha < 0,60$ = onvoldoende; $0,60 \leq \alpha < 0,70$ = voldoende; $\alpha \geq 0,70$ = goed.

Belangrijke beslissingen zijn “die beslissingen die op basis van de testscore worden genomen, die in principe, of op korte termijn, onomkeerbaar zijn en die voor een belangrijk deel buiten de betrokkenen om worden genomen” (Evers, Van Vliet-Mulder & Groot, 2000, p. 1411). Het risico op recidive, zoals bepaald met de RISc-totaalscore, weegt zwaar in het advies dat een reclasseringswerker uitbrengt in zijn of haar voorlichtingsrapportage voor de Officier van Justitie, in de advisering aan Gevangeniswezen (adviesrapport reïntegratieplan) en in het formuleren van benodigd toezicht en interventies in een reclasseringstraject. Bij de beoordeling van de interne consistentie van de totaalscore is daarom uitgegaan van de richtlijnen voor belangrijke beslissingen op individueel niveau. De schalen van de RISc zijn daarentegen niet bedoeld voor het nemen van dergelijke belangrijke beslissingen. De aanwezigheid van problemen op de leefgebieden die door de schalen in kaart worden gebracht, speelt weliswaar een rol bij de inhoudelijke invulling van het reclasseringsaanbod, maar dergelijke beslissingen betreffen minder belangrijke beslissingen op individueel niveau. Bij de beoordeling van de interne consistentie is dan ook gekozen voor de richtlijn die voor dergelijke beslissingen geldt. Tijdens de ontwikkeling van de RISc is reeds aandacht besteed aan de interne consistentie van de schalen van het instrument. Deze bleek in alle gevallen bevredigend ($> 0,70$). In het onderhavige onderzoek wordt getracht deze bevindingen te repliceren in een grote onderzoeksgroep.

Naast de berekening van alfa zijn in de betrouwbaarheidsanalyses ook de gemiddelde interitemcorrelatie, de itemrestcorrelaties per item en de waarde van alfa als een item wordt verwijderd, berekend. De gemiddelde interitemcorrelatie geeft informatie over de samenhang tussen de items die een schaal vormen. Deze correlatie mag zeker niet negatief zijn, maar vooral ook niet te hoog omdat in dat geval de items te veel hetzelfde meten. Interitemcorrelaties tussen de 0,20 en 0,70 zijn gewenst (Stouthard, 1998). De itemrestcorrelatie is de correlatie tussen de score op een item en de score op de andere items die bij dezelfde schaal horen. De itemrestcorrelatie geeft aan in hoeverre een item onderscheid kan maken tussen personen die hoog scoren op de schaal en personen die laag scoren (discriminatie). Deze correlatie mag niet te laag zijn en doorgaans wordt een ondergrens van 0,20 gehanteerd (Van der Knaap, 2003). Als coëfficiënt alfa stijgt wanneer een item uit de schaal wordt verwijderd, betekent dat grofweg dat het item niet bijdraagt aan de schaal.

4.2 Resultaten: de schaalstructuur en interne consistentie van de RISC

In eerste instantie zijn voor elke RISC-schaal en voor de totaalscore principale componentenanalyses uitgevoerd op de bijbehorende items respectievelijk schalen en werden factoren met een eigenwaarde groter dan één geëxtraheerd. Bij de meeste schalen leverde de pca op deze manier een éénfactoroplossing op. De analyses op de schalen 1&2 *Delictgegevens*, 6 *Relatie met partner, gezin en familie*, 8 *Druggebruik* en 9 *Alcoholgebruik* en de analyse op de totaalscore leverden meerdere factoren een eigenwaarde groter dan één op. Om na te gaan in hoeverre met minder factoren kon worden volstaan, zijn de scree plots bestudeerd. Omdat in alle gevallen bleek dat een éénfactoroplossing afdoende was, zijn voor elke schaal en voor de totaalscore de bijbehorende items respectievelijk schalen nogmaals ingevoerd in een pca en gedwongen tot een éénfactoroplossing¹⁰.

Over het algemeen laden de items / schalen goed op de bijbehorende factor. In de meeste gevallen is de interne consistentie voldoende tot goed en ook de gemiddelde interitemcorrelaties zijn vrijwel allemaal goed te noemen (zie tabel 26). De afzonderlijke correlaties tussen de verschillende RISC-schalen worden in tabel 27 gegeven. De itemrestcorrelaties, die niet een tabel zijn opgenomen maar die bij de onderzoekers zijn op te vragen, zijn in de meeste gevallen voldoende¹¹. Wat uit de bestudering van tabel 27 opgemaakt kan worden is dat de samenhang tussen de meeste RISC-schalen varieert tussen de 0,20 en 0,70. Voor een intern consistente schaal (in casu de totaalscore) is dit gunstig omdat het aangeeft dat ieder onderdeel een eigen bijdrage levert aan de totaalscore. De sterke samenhang tussen schaal 11 *Denkpatronen, gedrag en vaardigheden* en schaal 12 *Houding* ($r=0,65$) valt op omdat geen van de andere correlaties zo sterk is. Daarnaast valt op dat schaal 9 *Alcoholgebruik* zwak samenhangt met alle andere schalen. Ook schaal 10 *Emotioneel welzijn* vertoont met veel andere schalen een zwakke samenhang en hangt met slechts drie schalen middelsterk samen (opleiding, relaties met partner en denkpatronen).

Tabel 26 Coëfficiënt alfa en gemiddelde interitemcorrelatie per RISC-schaal en RISC-totaalscore

RISC-onderdeel	N	Aantal items / schalen	alfa	interitemcorrelatie
Schaal 1&2 Delictgegevens	4.526	8	0,80	0,34
Schaal 3 Huisvesting en wonen	8.647	4	0,78	0,47
Schaal 4 Opleiding, werk en leren	11.364	7	0,88	0,53
Schaal 5 Inkomen en omgaan met geld	11.480	4	0,71	0,38
Schaal 6 Relaties met partner, gezin en familie	10.556	5	0,61	0,25
Schaal 7 Relaties met vrienden en kennissen	11.256	4	0,76	0,45
Schaal 8 Druggebruik	4.230	6	0,70	0,27
Schaal 9 Alcoholgebruik	4.581	5	0,65	0,27
Schaal 10 Emotioneel welzijn	11.554	5	0,79	0,43
Schaal 11 Denkpatronen, gedrag en vaardigheden	11.554	8	0,87	0,45
Schaal 12 Houding	11.405	5	0,84	0,52
Totaalscore	5.900	11	0,79	0,32

¹⁰ Voor een overzicht van de resultaten van de pca's waarin één factor werd geëxtraheerd, wordt naar bijlage 2 verwezen.

¹¹ Een overzicht met de volledige resultaten van de betrouwbaarheidsanalyses, waaronder ook de itemrestcorrelaties, is op te vragen bij de onderzoekers.

De meeste schalen van de RISC kunnen als goede schalen worden beschouwd. Bij de schalen 1&2, 6, 8 en 9 en bij de totaalscore is, zoals hiervoor al is aangegeven, echter iets aan de hand. Per schaal worden de resultaten besproken.

Schaal 1&2 Delictgegevens

De schalen 1 en 2 geven informatie over de delictgeschiedenis van de cliënt en over de aard en achtergrond van het delict waarvoor de cliënt op het moment van afname van de RISC wordt vervolgd. Voor de schalen 1 en 2 worden geen aparte schaalscores berekend, maar wordt volstaan met een combinatiescore die inzicht biedt in de delictgegevens van de cliënt. De principale componentenanalyses zijn dan ook uitgevoerd op de items van deze twee schalen samen.

In de éénfactoroplossing die is geëxtraheerd nadat uit de scree plot was gebleken dat met dit aantal factoren kon worden volstaan, laadt item 2.10 'Verantwoordelijkheid nemen voor het delictgedrag' laag (zie bijlage 2 voor de factorladingen) op de factor die in de éénfactoroplossing is geëxtraheerd. De betrouwbaarheidsanalyse toont dat de interne consistentie van de schaal goed is en dat de gemiddelde interitemcorrelatie normaal is (zie tabel 26). Item 2.10 vertoont echter lage correlaties met de overige items, heeft een lage itemrestcorrelatie (0,20) en drukt de alfa (alfa stijgt als dit item wordt verwijderd).

Een herhaling van de analyses op deze schaal zonder item 2.10 levert een bevredigende éénfactoroplossing op waarbij alle items hoog op de factor laden. De interne consistentie van de schaal neemt toe naar 0,82 en de gemiddelde interitemcorrelatie stijgt naar 0,40. Item 2.10 hoort kennelijk niet bij de rest van de items: het meet een ander concept. Dat is verklaarbaar gezien de inhoud van het item. Het is immers niet per definitie zo dat iemand met een ernstiger delictgeschiedenis meer verantwoordelijkheid neemt voor zijn of haar delictgedrag. Het item is, omdat het te maken lijkt te hebben met de houding van een cliënt ten opzicht van criminaliteit, ingevoerd in een principale componentenanalyse op de items van schaal 12 *Houding*. Uit deze analyse blijkt dat het daar goed bij past. De pca levert een bevredigende éénfactoroplossing op en de lading van het item op de factor is goed. Hoewel het item niet echt bijdraagt aan de interne consistentie van de schaal (die met 0,84 goed te noemen is), beïnvloedt het deze ook niet negatief. Tevens vertoont het item een goede itemrestcorrelatie.

Blijkens de resultaten van de betrouwbaarheidsanalyse zou de interne consistentie van de schaal nog iets verbeteren als item 2.11b 'Delicten worden steeds ernstiger' zou worden verwijderd. De itemrestcorrelatie van dit item en de lading van dit item op de schaal geven echter geen aanleiding om te concluderen dat het niet bij de schaal zou horen. Het lijkt erop dat dit item een zeer specifiek kenmerk van de delictgegevens van de cliënt in kaart brengt.

Schaal 6 Relaties met partner, gezin en familie

Nadat de vijf items van schaal zes in een éénfactoroplossing zijn gedwongen, blijkt dat item 6.4 'Familie- of gezinslid heeft een justitieel dossier' een lage factorlading heeft (zie bijlage 2). Uit de betrouwbaarheidsanalyse blijkt dat het de, toch al lage, interne consistentie behoorlijk drukt en dat het een te lage itemrestcorrelatie heeft (0,17). De gemiddelde interitemcorrelatie tussen de items is laag en ligt tegen de ondergrens aan.

Een herhaling van de analyses op de schaal 6 zonder item 6.4, leverde wel een bevredigende éénfactoroplossing. De items laden hoog op de factor, de gemiddelde interitemcorrelatie is 0,34 en de itemrestcorrelaties zijn normaal. De betrouwbaar-

heid is echter nog steeds onvoldoende (0,67). Toch lijkt de conclusie gerechtvaardigd dat item 6.4 niet bij de overige items hoort en kennelijk iets anders meet. Schaal 6 bestaat uit items die betrekking hebben op de *kwaliteit* van relaties met familie- of gezinsleden. Een mogelijke verklaring voor de bevinding dat item 6.4 niet bij deze schaal hoort, is dat de kwaliteit van dergelijke relaties niet per definitie negatief wordt beïnvloed wanneer een familie- of gezinslid een justitieel dossier heeft. Met andere woorden, met een familie- of gezinslid met een justitieel dossier kan een cliënt een heel goede en warme relatie hebben. Het persoonlijke karakter van deze vraag kan eventueel ook een verklaring zijn. Als de informatie niet uit het dossier van de cliënt valt te halen, dient de reclasseringswerker het in het gesprek met de cliënt te achterhalen. Niet alle cliënten zullen in gelijke mate geneigd zijn dergelijke informatie met de werker te bespreken.

Schaal 8 Druggebruik

De items van schaal 8 laden overwegend hoog op de factor die op basis van de gedwongen éénfactoroplossing is geëxtraheerd. Dit geldt echter niet voor item 8.5 'Motivatie om druggebruik aan te pakken'. Het item laadt laag (zie bijlage 2) en correleert negatief met een aantal andere items waardoor de gemiddelde interitemcorrelatie ook laag is (0,27). De itemrestcorrelatie van item 8.5 is erg laag (0,04) en uit de betrouwbaarheidsanalyse blijkt tevens dat de interne consistentie van de schaal, die in deze samenstelling voldoende is, goed zou zijn als item 8.5 zou worden verwijderd.

Een herhaling van de analyses op schaal 8, zonder item 8.5, levert een bevredigende éénfactoroplossing op. Alle items laden hoog op de factor, de interne consistentie stijgt naar 0,83, de gemiddelde interitemcorrelatie is 0,50 en alle items dragen bij aan de schaal. Het is duidelijk dat item 8.5 iets anders meet dan de overige items en niet bijdraagt aan een schaal die de ernst van de druggebruik beoogt te meten. Bovendien hoeft item 8.5 slechts beantwoord te worden wanneer op de overige items van de schaal een score van 1 of meer wordt gehaald. Hierdoor wordt dit item alleen ingevuld voor een bepaalde subgroep, namelijk de groep met minstens enige problemen op het gebied van druggebruik. Daarnaast is het goed voorstelbaar dat de motivatie om iets aan dergelijke problemen te doen geen lineaire samenhang vertoont met de ernst van de problemen.

Schaal 9 Alcoholgebruik

In een gedwongen éénfactoroplossing laden vier van de vijf items van schaal 9 goed op de factor. Net als bij schaal 8 wordt de uitzondering gevormd door het item dat de motivatie van de cliënt om het gebruik aan te pakken, beoogt te meten (item 9.5). De interne consistentie van schaal 9 is onvoldoende (0,65), maar zou aanzienlijk verbeteren als item 9.5 zou worden verwijderd. Dit item blijkt ook een lage itemrestcorrelatie te hebben (0,10). De overige items hebben een bevredigende itemrestcorrelatie.

Wanneer de analyses worden herhaald zonder item 9.5 levert dit een bevredigende éénfactoroplossing op. De interne consistentie van de schaal is goed (0,89), de itemrestcorrelaties zijn goed en de interitemcorrelatie is ook goed (0,67). Het is duidelijk dat item 9.5 iets anders meet dan de overige items en niet bijdraagt aan een schaal die de ernst van het alcoholgebruik beoogt te meten. Net als bij schaal 8 geldt dat het item slechts beantwoord hoeft te worden wanneer op de overige items van de schaal een score van 1 of meer wordt gehaald. Hierdoor wordt dit item alleen ingevuld voor de subgroep daders met minstens enige problemen op

het gebied van alcoholgebruik. Het is goed voorstelbaar dat de motivatie om iets aan dergelijke problemen te doen geen lineaire samenhang vertoont met de ernst van de problemen.

Totaalscore

De principale componentenanalyse op de totaalscore leverde een oplossing op met drie componenten met eigenwaarden groter dan één. De scree plot toonde evenwel aan dat met één factor volstaan kon worden. De analyse is vervolgens gedwongen tot een éénfactoroplossing waaruit blijkt dat de schaalscores overwegend hoog laden op de factor. Schaal 9 *Alcoholgebruik* vormt hierop de uitzondering (zie bijlage 2) en ook de itemrestcorrelatie van schaal 9 is eveneens aan de lage kant: 0,23. De interne consistentie van de totaalscore kan, ook volgens de strenge normen die gelden voor schalen op basis waarvan belangrijke beslissingen op individueel niveau worden genomen, voldoende worden genoemd (0,79). Wanneer schaal 9 zou worden verwijderd, zou de interne consistentie evenwel iets beter zijn. De gemiddelde interitemcorrelatie van de schaal is voldoende, maar wat opvalt bij de afzonderlijke interitemcorrelaties (zie tabel 27) is dat die van schaal 9 met de overige schalen zonder uitzondering laag is (in alle gevallen lager dan 0,25). Omdat schaal 9 een leefgebied in kaart beoogt te brengen dat een belangrijke rol kan spelen bij herhaling van crimineel gedrag, ligt het niet erg voor de hand om deze schaal uit de totaalscore te verwijderen. Bovendien is de invloed van de schaal op de interne consistentie beperkt. De analyses op de totaalscore zijn daarom niet herhaald zonder schaal 9.

Het algemene beeld van structuur en interne consistentie van de RISC

Het algemene beeld op basis van de principale componentenanalyses en de betrouwbaarheidsanalyses van de RISC-schalen en RISC-totaalscore is dat de kwaliteit van de meeste schalen en van de totaalscore goed is. Er zijn echter een paar uitzonderingen waarop in het slothoofdstuk van dit rapport zal worden teruggekomen. Op die plek zal bovendien een aantal suggesties worden gedaan voor aanpassing van het instrument.

Tabel 27 Correlaties tussen de elf RISC-schalen^a

		1&2	3	4	5	6	7	8	9	10	11	12
Schaal 1&2 Delictgegevens	r	1,00										
	N	(9.612)										
Schaal 3 Huisvesting en wonen	r	0,38	1,00									
	N	(7.257)	(8.733)									
Schaal 4 Opleiding, werk en leren	r	0,47	0,38	1,00								
	N	(9.605)	(8.725)	(11.656)								
Schaal 5 Inkomen en omgaan met geld	r	0,42	0,50	0,47	1,00							
	N	(9.604)	(8.724)	(11.641)	(11.651)							
Schaal 6 Relaties met partner, gezin en familie	r	0,26	0,30	0,34	0,25	1,00						
	N	(9.607)	(8.721)	(11.641)	(11.638)	(11.650)						
Schaal 7 Relaties met vrienden en kennissen	r	0,53	0,38	0,49	0,45	0,27	1,00					
	N	(9.573)	(8.691)	(11.601)	(11.597)	(11.598)	(11.608)					
Schaal 8 Druggebruik	r	0,43	0,39	0,38	0,46	0,22	0,42	1,00				
	N	(8.103)	(7.140)	(9.790)	(9.785)	(9.787)	(9.748)	(9.799)				
Schaal 9 Alcoholgebruik	r	0,19	0,15	0,15	0,19	0,19	0,17	0,22	1,00			
	N	(9.588)	(8.709)	(11.628)	(11.622)	(11.621)	(11.581)	(9.772)	(11.636)			
Schaal 10 Emotioneel welzijn	r	0,20	0,25	0,36	0,24	0,41	0,25	0,24	0,19	1,00		
	N	(9.603)	(8.725)	(11.641)	(11.637)	(11.636)	(11.597)	(9.784)	(11.621)	(11.650)		
Schaal 11 Denkpatronen, gedrag en vaardigheden	r	0,45	0,31	0,51	0,35	0,45	0,49	0,29	0,22	0,47	1,00	
	N	(9.595)	(8.717)	(11.629)	(11.626)	(11.625)	(11.588)	(9.774)	(11.610)	(11.631)	(11.638)	
Schaal 12 Houding	r	0,47	0,26	0,40	0,26	0,28	0,45	0,23	0,13	0,22	0,65	1,00
	N	(9.593)	(8.647)	(11.550)	(11.545)	(11.544)	(11.508)	(9.708)	(11.528)	(11.546)	(11.546)	(11.558)

^a Alle correlaties zijn significant ($p < .001$, tweezijdig)

5 Congruente validiteit

5.1 Inleiding

Dit hoofdstuk beschrijft de resultaten van het onderzoek naar de congruente validiteit van de RISC. Op basis van deze resultaten kan de vraag worden beantwoord of de RISC meet wat het veronderstelt te meten, namelijk recidiverisico en criminogene factoren die daaraan ten grondslag liggen. Zoals in hoofdstuk 2 al is besproken, is de congruente validiteit van de RISC op twee manieren bestudeerd. Op de eerste plaats zijn RISC-scores gerelateerd aan de voorspelling van de recidivekans op basis van de door het WODC ontwikkelde StatRec-schaal (Wartna & Tollenaar, 2006). Op de tweede plaats is de samenhang bestudeerd tussen scores op de schalen 10 *Emotioneel welzijn*, 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding* en de scores op gevalideerde persoonlijkheidsvragenlijsten die (vrijwel) dezelfde concepten beogen te meten: de NPV, de UCL en de BDHI-D. In dit hoofdstuk wordt eerst de samenhang tussen de RISC en StatRec beschreven. Vervolgens wordt ingegaan op de validering van de RISC-schalen 10, 11 en 12 aan de hand van de door reclasseringscliënten ingevulde persoonlijkheidsvragenlijsten.

5.2 Resultaten van de validering met behulp van een recidivevoorspellingsmodel

5.2.1 Selectie van de onderzoeksgroep

In het onderzoek naar de congruente validiteit van de RISC in termen van de voorspelling van recidive, is gebruik gemaakt van de eerder in dit rapport beschreven RISC-database (zie paragraaf 2.3). Deze database, bestaande uit 11.666 RISC's, is voor het validiteitsonderzoek andermaal aan een ronde datacleaning onderworpen. Op basis hiervan is besloten dat, wanneer naar aanleiding van *dezelfde* strafzaak *meerdere* RISC's zijn afgenomen, uitsluitend de oudste RISC in het onderzoek is betrokken¹². De StatRec-score wordt namelijk berekend op basis van justitiële achtergrondinformatie die *ten tijde van de uitgangszaak* geldt. Het verdiende in het onderhavige validiteitsonderzoek de voorkeur dat de datum van de RISC-afname zo dicht mogelijk bij de 'inschrijfdatum' van de strafzaak ligt. Wanneer meerdere RISC's naar aanleiding van dezelfde strafzaak zijn afgenomen, komt de afnamedatum van de oudste RISC het meest overeen met de inschrijfdatum van de strafzaak. Het is overigens interessant om in het onderzoek naar de predictieve validiteit van de RISC, dat in 2008 wordt uitgevoerd, na te gaan of de RISC's die op verschillende momenten, maar wel naar aanleiding van dezelfde strafzaak bij dezelfde cliënt worden afgenomen een verschillende samenhang vertonen met daadwerkelijke recidive.

In totaal waren voor het StatRec-onderzoek naar de congruente validiteit van de RISC 11.528 RISC's beschikbaar. Om StatRec-scores te berekenen, dienden de gege-

¹² Een simpel voorbeeld waarbij zich dit kan voordoen, is de situatie waarin aan de start van het strafproces een voorlichtingsrapportage is uitgebracht op basis van de RISC en het instrument in een latere fase nogmaals wordt afgenomen ten behoeve van detentiefasering. De oorspronkelijke RISC kan inmiddels 'verlopen' zijn.

vens over de justitiële voorgeschiedenis van deze groep uit de Onderzoeksdatabase Justitiële Documentatie (OBJD) te worden geëxtraheerd. Deze informatie is op basis van parketnummers en persoonsgegevens¹³ verkregen volgens de standaardmethoden van de WODC-Recidivemonitor. In het kort komt deze procedure erop neer dat een bestand met parketnummers en persoonsgegevens wordt verstuurd naar JustId in Almelo. Daar vindt een versleuteling plaats waarna op geanonimiseerde wijze de benodigde justitiële achtergrondgegevens uit de OBJD kunnen worden geëxtraheerd (zie voor een uitgebreidere beschrijving Wartna, Blom & Tollenaar, 2004).

Niet bij elke RISC was het parketnummer bekend van de zaak naar aanleiding waarvan de RISC werd afgenomen. Zonder parketnummer kon geen justitiële voorgeschiedenis worden geëxtraheerd en kon derhalve geen StatRec-score worden berekend. Om deze reden kon de RISC niet worden betrokken in het validiteitsonderzoek. In totaal zijn 10.638 combinaties van parketnummer met persoonsgegevens bij JustId aangeboden. In 10.174 (95,6%) gevallen kon de combinatie tussen persoon en parketnummer door JustId worden teruggevonden en versleuteld. De overige combinaties van persoonsgegevens en parketnummer waren niet te traceren en in deze gevallen kon daarom geen Statrec-score worden berekend. Na het verkrijgen van de versleutelde gegevens is informatie over de justitiële voorgeschiedenis van elke persoon in de onderzoeksgroep geëxtraheerd uit de OBJD. Vervolgens zijn deze data bewerkt volgens de standaardmethoden van de WODC-Recidivemonitor. Tijdens de data-extractie en de databewerking moest een aantal records om uiteenlopende redenen uit het onderzoeksbestand worden verwijderd zodat in totaal in 9.985 gevallen een StatRec-score kon worden berekend. Dit betekent een uitval van 6,1% van de 10.638 oorspronkelijke combinaties van parketnummer en persoonsgegevens. Het grootste deel van de uitval wordt veroorzaakt doordat uitsluitend een hofnummer beschikbaar was (het nummer waaronder een zaak in hoger beroep wordt ingeschreven bij de gerechtshoven). Zonder het parketnummer van de strafzaak die voorafgaand aan het hoger beroep bij een rechtbank heeft gediend, kunnen de medewerkers van JustId deze hofnummers niet terugvinden in hun databases. Het gaat om 405 van de 653 uitgevallen cases ofwel om 62,0% van de totale uitval.

De 9.985 RISC's waarbij een StatRec-score kon worden berekend, zijn afgenomen bij 9.930 unieke personen. Van 55 daders zijn meerdere RISC's beschikbaar, die naar aanleiding van verschillende strafzaken en op andere tijdstippen zijn afgenomen. Van deze 9.930 unieke personen is 90,2% mannelijk en 9,8% vrouwelijk (zie tabel 28). De meeste daders zijn in Nederland geboren, gevolgd door de groep daders die in Suriname zijn geboren. Ingedeeld naar herkomst — dus rekening houdend met het geboorteland van de ouders — blijkt ruim drievijfde van de onderzoeksgroep een Nederlandse achtergrond te hebben. Bijna eenderde van de onderzoeksgroep is niet-westers allochtoon, en een kleine groep bestaat uit westerse allochtonen. Van 0,9% van de onderzoeksgroep is de herkomst onbekend. De gemiddelde leeftijd in de onderzoeksgroep ten tijde van de afname van de RISC was 34 jaar en 2 maanden ($N = 9.985$; $m = 34,16$; $sd = 11,99$; $range = 17,11 - 78,37$)¹⁴.

¹³ Persoonsgegevens zijn na het doorlopen van een aparte toestemmingsprocedure aangevraagd bij en verkregen via de RN.

¹⁴ Tabel 28 geeft, in tegenstelling tot soortgelijke tabellen elders in dit rapport, geen overzicht van de indeling in leeftijdsgroepen. Hier is voor gekozen omdat de N waarover leeftijd wordt berekend afwijkt van de N in de tabel.

Tabel 28 Congruente validiteit in termen van recidivekans: achtergrondkenmerken van de onderzoeksgroep (N = 9.930)

Kenmerk		Aantal	%
Geslacht	Man	8.961	90,2
	Vrouw	969	9,8
Geboorteland	Nederland	7.269	73,2
	Marokko	435	4,4
	Nederlandse Antillen / Aruba	463	4,7
	Suriname	603	6,1
	Turkije	282	2,8
	Overig niet-westers	528	5,3
	Overig westers	263	2,6
	Onbekend	87	0,9
Herkomstgroep	Autochtoon	6.285	63,3
	Niet-westers allochtoon	3.171	31,9
	Westers allochtoon	387	3,9
	Onbekend	87	0,9

5.2.2 De relatie tussen de RISC-scores en StatRec: vergelijking van vier risicogroepen

De relatie tussen de RISC-scores en de StatRec-scores is op een aantal manieren onderzocht. Als eerste stap is de onderzoeksgroep op basis van de gewogen RISC-totaalscores¹⁵ in vier risicogroepen verdeeld: daders met een lage RISC-totaalscore, daders met een middellage totaalscore, daders met een middelhoge totaalscore en daders met een hoge RISC-totaalscore. De grenswaarden van deze groepen zijn vastgesteld in de totale database (met 11.666 RISC's) en vormen in die onderzoeksgroep vier gelijke groepen van 25% van de totale groep. De groep van 9.985 RISC's waarbij een StatRec-score kon worden berekend bestaat voor 22,2% uit RISC's met een lage totaalscore, voor 20,3% uit RISC's met een middellage totaalscore, voor 19,8% uit RISC's met een middelhoge totaalscore en voor 18,5% uit RISC's met een hoge score. Voor 19,1% van de RISC's kon geen totaalscore worden berekend. In de totale groep van 9.985 cases ligt de gemiddelde StatRec-score op 0,54 met een standaarddeviatie van 0,32. In tabel 29 worden de gemiddelde StatRec-scores voor de vier categorieën van de RISC-totaalscore weergegeven. De gemiddelde RISC-scores vertonen een oplopende trend: des te hoger de RISC-totaalscore, des te hoger de StatRec-score. Tabel 29 geeft tevens de resultaten weer van de ANOVA die is uitgevoerd om na te gaan of de StatRec-scores van de vier groepen significant van elkaar afwijken. Dit blijkt het geval te zijn en de maat voor de effectgrootte (η^2) geeft aan dat er sprake is van een groot verschil (Cohen, 1988). Uit de uitgevoerde post-hoc tests (Tukey HSD) blijkt ten slotte dat elke groep significant van de drie andere groepen verschilt.

¹⁵ De gewogen RISC-totaalscore wordt berekend op basis van de gewogen RISC-schaalscores. Vervolgens wordt de gewogen RISC-totaalscore op basis van grenswaarden in drie categorieën verdeeld: laag recidiverisico, midden recidiverisico, hoog recidiverisico. Zie de toelichting op de scoring van de RISC in hoofdstuk 1.

Tabel 29 Gemiddelde StatRec-score voor vier categorieën van de RISC-totaalscore en de vergelijking tussen deze vier groepen (ANOVA)

	Aantal	Gemiddelde StatRec-score (standaarddeviatie)	F	p	η^2
Lage totaalscore	2.215	0,31 (0,21)			
Middellage totaalscore	2.028	0,44 (0,26)			
Middelhoge totaalscore	1.982	0,59 (0,29)	1.667,90	.000	0,38
Hoge totaalscore	1.851	0,84 (0,22)			

5.2.3 *Samenhang tussen de RISC-scores en StatRec: correlaties tussen RISC en StatRec*

Als tweede stap in de analyse van de samenhang tussen RISC en StatRec zijn correlaties berekend. In de tabellen 30, 31 en 32 worden voor verschillende groepen correlaties gegeven tussen enerzijds de RISC-schalen en de RISC-totaalscore en anderzijds de StatRec-score. De N die in de tabellen staat vermeld bij de groepen (in de bovenste rij) geeft voor elke groep het maximum aantal RISC's aan waarvoor een StatRec-score kon worden berekend. De aantallen kunnen daarom afwijken van de aantallen die elders in dit rapport zijn genoemd. Cohen (1988) geeft de volgende grenzen voor de interpretatie van correlaties: bij correlaties tot 0,30 is sprake van een klein effect, correlaties vanaf 0,30 tot 0,50 duiden op een middelgroot effect en bij correlaties vanaf 0,50 is sprake van een groot effect. Voor de beschrijving van de correlaties in dit hoofdstuk wordt bij correlaties tot 0,30 gesproken van een zwakke samenhang, bij correlaties van 0,30 tot 0,50 van een middelsterke samenhang, en bij correlaties vanaf 0,50 van een sterke samenhang. Omdat het doel van de RISC is om de kans op recidive te voorspellen, wordt een sterke, positieve samenhang verwacht tussen de RISC-totaalscore en StatRec. Ook de schalen van de RISC worden verwacht positief met de StatRec-score samen te hangen, maar omdat elke schaal slechts een onderdeel vormt van de RISC-voorspelling van de recidivekans worden middelsterke samenhangen verwacht. Dit geldt uiteraard niet voor schaal 1&2 *Delictgegevens*. De score op deze schaal wordt berekend op basis van veel van dezelfde informatie als in de StatRec-schaal wordt gebruikt en daarom wordt dan ook een sterke, positieve correlatie met StatRec verwacht.

De correlaties tussen RISC en StatRec in de totale onderzoeksgroep zijn in tabel 30 weergegeven. De RISC-totaalscore en de StatRec-score vertonen in deze groep de verwachte sterke samenhang. Van de afzonderlijke RISC-schalen vertoont schaal 1&2 *Delictgegevens* conform de verwachting een sterke samenhang met de StatRec-score. De meeste dynamische criminogene factoren, zoals die door de RISC in kaart worden gebracht, hangen zoals verwacht middelsterk samen met de voorspelling van de recidivekans op basis van StatRec. De dynamische factoren met de sterkste correlaties met StatRec zijn de schalen 4 *Opleiding, werk en leren*, 5 *Inkomen en omgaan met geld*, 7 *Relaties met vrienden en kennissen* en 8 *Druggebruik* (correlaties groter dan 0,40). De samenhang tussen de schalen 3 *Huisvesting en wonen*, 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding* en StatRec zijn minder sterk

maar ook voor deze schalen is sprake van een middelsterke samenhang met de voorspelde recidivekans (correlaties groter of gelijk aan 0,30). Wat opvalt, is dat de schalen 6 *Relaties met partner, gezin en familie*, 9 *Alcoholgebruik* en 10 *Emotioneel welzijn* lage correlaties vertonen met de StatRec-score. Tegen de verwachting in blijkt er in de totale onderzoeksgroep een zwakke samenhang te zijn tussen deze schalen en de recidivekans zoals die op basis van de StatRec-schaal is berekend.

Op basis van achtergrondkenmerken is een aantal subgroepen gevormd waarvoor de correlaties tussen enerzijds de afzonderlijke RISC-schalen en de RISC-totaalscore en anderzijds de StatRec-score zijn berekend. In tabel 30 staan de correlaties voor mannen, vrouwen en drie leeftijdsgroepen¹⁶. De correlatie tussen de RISC-totaalscore en de StatRec-score is voor deze vijf groepen sterk. Opvallend is echter dat de correlaties in de groep vrouwen aanzienlijk lager zijn dan voor mannen. Dit heeft onder andere te maken met het feit dat er beduidend meer vrouwelijke first offenders zijn (geen geregistreerde justitiecontacten in het verleden). Gezien de statische kenmerken op basis waarvan de StatRec-score wordt berekend, is het voorspellen van de recidivekans voor first offenders moeilijker dan voor daders met eerdere justitiecontacten. Het onderzoek naar de predictieve validiteit van de RISC, dat in 2008 wordt uitgevoerd, zal uitsluitsel moeten geven over de vraag of de RISC de recidivekans van vrouwen goed voorspelt.

Wanneer correlaties tussen de RISC-schaalscores en de StatRec-score in de verschillende groepen (mannen, vrouwen en de drie leeftijdsgroepen) uit tabel 30 worden bestudeerd, valt direct op dat in elke groep de schaal 1&2 *Delictgegevens*, de sterkste samenhang vertoont met de StatRec-score. Daarnaast valt op dat, na schaal 1&2, de schalen 8 *Druggebruik*, 4 *Opleiding, werk en leren* en 7 *Relaties met vrienden en kennissen* binnen vrijwel elke groep de drie belangrijkste samenhangen met de StatRec vertonen. In elke groep is de samenhang tussen deze schalen en StatRec middelsterk of zelfs sterk. In de jongste leeftijdsgroep is de samenhang tussen druggebruik en StatRec daarentegen laag. Wanneer de correlaties in aflopende volgorde worden geordend, blijkt ook dat de samenhang tussen druggebruik en StatRec binnen deze groep een lagere plaats inneemt dan binnen de anderen groepen. Deze bevindingen lijken erop te wijzen dat het belang van druggebruik bij het (blijven) plegen van delicten toeneemt naarmate daders ouder worden. Hierbij moet echter in het oog worden gehouden dat het ten eerste gaat om correlaties en niet om causale verbanden en ten tweede om correlaties met een *voorspelling van recidive* en niet om correlaties met daadwerkelijke recidive. Naast de leeftijdsverschillen voor druggebruik valt op dat schaal 11 *Denkpatronen, gedrag en vaardigheden* binnen de jongste groep een hogere plek in de ordening van correlaties inneemt dan bij de andere leeftijdsgroepen.

Net als in de totale onderzoeksgroep vertonen de schalen 6 *Relaties met partner, gezin en familie*, 9 *Alcoholgebruik* en 10 *Emotioneel welzijn* in alle subgroepen die in tabel 30 zijn opgenomen, zwakke samenhangen met de recidivevoorspelling door StatRec. De enige uitzondering wordt gevormd door de middelgrote correlatie tussen alcoholgebruik en StatRec binnen de middelste leeftijdsgroep. Wat verder opvalt is dat bij vrouwen alleen de schalen *Druggebruik*, *Relaties met vrienden en kennissen* en *Opleiding, werk en leren* de verwachte middelsterke samenhang ver-

¹⁶ De complete groep is op basis van leeftijd in drie gelijke groepen verdeeld. De jongste groep bestaat uit daders tot en met 26,48 jaar ($m = 21,72$; $sd = 2,40$). De middelste groep bestaat uit daders van 26,49 jaar tot en met 39,10 jaar ($m = 32,98$; $sd = 3,64$). De oudste groep bestaat uit daders vanaf 39,11 jaar ($m = 48,15$; $sd = 7,67$). De oudste dader in de onderhavige onderzoeksgroep was 78 jaar en vier maanden..

tonen met de StatRec-voorspelling van de recidivekans. Met name de zwakke samenhang van de schalen 3 *Huisvesting en wonen* en 12 *Houding* met StatRec is opmerkelijk, omdat deze correlaties significant afwijken van de grens voor een middelsterke samenhang (95%CI is respectievelijk 0,14 - 0,26 en 0,16 - 0,28). De lage correlaties voor inkomen en denkpatronen wijken niet significant af van 0,30. De betrouwbaarheidsintervallen zijn in de groep vrouwen immers wat ruimer dan in andere, grotere onderzoeksgroepen.

Tabel 31 geeft de correlaties tussen RISC en StatRec voor verschillende herkomstgroepen¹⁷. Net als in de totale onderzoeksgroep is de samenhang tussen de RISC-totaalscore en de StatRec-score sterk positief. Ook de samenhang tussen RISC-schaal 1&2 en StatRec is binnen alle herkomstgroepen sterk positief. Voorts vertonen de meeste RISC-schalen middelsterke, positieve samenhangen met de StatRec-score. Net als voor de totale groep geldt dit niet voor de schalen 6 *Relaties met partner, gezin en familie*, 9 *Alcoholgebruik* en 10 *Emotioneel welzijn*. Daarnaast valt een aantal zaken op. Binnen de groep Surinamers vertoont een drietal dynamische factoren een sterke samenhang met de voorspelde recidivekans. Het gaat om druggebruik, opleiding / werk en relaties met vrienden. Deze laatste factor blijkt ook bij Turken een sterke samenhang te vertonen met StatRec. Opvallend is verder de zwakke samenhang tussen schaal 12 *Houding* en StatRec binnen de groep Marokkanen. Ook in andere herkomstgroepen vertoont deze schaal een zwakke of nét middelsterke samenhang met StatRec, maar binnen die groepen wijkt de correlatie, in tegenstelling tot de correlatie van de groep Marokkanen, niet significant af van 0,30 (95%CI bij Marokkanen is 0,15 - 0,29).

Wanneer binnen elke groep de correlaties op aflopende volgorde worden gesorteerd, blijkt dat binnen de meeste groepen, na schaal 1&2 *Delictgegevens*, de schalen 4 *Opleiding, werk en leren*, 7 *Relaties met vrienden en kennissen* en 8 *Druggebruik* tot de belangrijkste samenhangen met de StatRec-score horen. Schaal 11 *Denkpatronen, gedrag en vaardigheden* blijkt bij Antillianen / Arubanen en bij Turken een hogere plek in te nemen in de ordening van samenhangen dan bij de andere groepen.

Op basis van de strafzaken naar aanleiding waarvan de RISC is ingevuld, is een indeling gemaakt in delictcategorieën. Hierbij is uitsluitend gekeken naar het delict met de hoogste strafdreiging. Dat wil zeggen dat de RISC van een dader die werd vervolgd wegens eenvoudige diefstal én verkrachting in de zedencategorie is ingedeeld. De RISC van een dader die daarentegen wegens verkrachting én moord werd vervolgd is in de agressie categorie ingedeeld omdat moord de hoogste strafdreiging heeft¹⁸. De correlaties tussen RISC en StatRec voor de verschillende delictcategorieën worden in tabel 32 weergegeven.

In overeenstemming met de verwachtingen en met de bevindingen voor de totale onderzoeksgroep vertoont de RISC-totaalscore in elke delictcategorie een sterke, positieve samenhang met de StatRec-score. Ook de samenhang tussen RISC-schaal 1&2 *Delictgegevens* en StatRec is in alle delictgroepen sterk positief. Het grootste deel van de dynamische schalen vertoont binnen de meeste groepen de verwachte middelsterke samenhang met StatRec. Dit geldt net als bij mannen en vrouwen, de leeftijdsgroepen en de herkomstgroepen opnieuw niet voor de schalen 6 *Relaties*

¹⁷ Indeling in deze groepen is gedaan op basis van geboorteland van de cliënt en geboorteland van zijn/haar ouders.

¹⁸ Een overzicht van de delictindeling is bij de auteurs op te vragen.

met partner, gezin en familie en 10 *Emotioneel welzijn*; binnen geen van de delictcategorieën worden de verwachte middelsterke correlaties gevonden. De andere schaal die in de hiervoor beschreven subgroepen laag correleert met StatRec — schaal 9 *Alcoholgebruik* — vertoont bij één van de delictcategorieën echter wel een middelsterke samenhang met de StatRec-voorspelling van recidive. Alcoholgebruik hangt bij plegers van vermogensdelicten zonder geweld middelsterk samen met de StatRec-score. Hierbij dient echter opgemerkt te worden dat het 95%-betrouwbaarheidsinterval ook lage correlaties omvat.

In de groep de groep plegers van vermogensdelicten zonder geweld valt op dat relatief veel correlaties tegen de grens van een sterke samenhang zitten. Er lijken binnen deze groep méér correlaties dan in de andere delictcategorieën redelijk sterk te zijn. In de groep plegers van agressiedelicten valt daarentegen op dat veel RISC-schalen een zwakke samenhang vertonen met de StatRec-voorspelling van het recidiverisico. In het geval van huisvesting, relaties met partner en gezin, alcoholgebruik, emotioneel welzijn en houding wijken deze correlaties significant af van de grens van 0,30. Uitsluitend de leefgebieden opleiding en werk, inkomen en omgaan met geld, relaties met vrienden en kennissen en druggebruik vertonen in deze groep de verwachte middelsterke samenhang met de StatRec-score. Binnen de meeste groepen horen, na schaal 1&2 *Delictgegevens*, de schalen 4 *Opleiding, werk en leren*, 7 *Relaties met vrienden en kennissen* en 8 *Druggebruik* opnieuw tot de belangrijkste samenhangen met de StatRec-score. Opvallend is echter dat schaal 5 *Inkomen en omgaan met geld* binnen de groep zedendelinquenten, na schaal 1&2, een tweede plaats inneemt en dat de samenhang tussen inkomen en StatRec binnen deze groep sterker is dan de correlatie van opleiding, druggebruik en relaties met vrienden en de StatRec-score. Binnen de groep plegers van opiumdelicten neemt schaal 5 daarentegen een relatief lage positie in. Bij de plegers van verkeersdelicten valt op dat schaal 7 *Relaties met vrienden en kennissen* een veel lagere plek in de rangorde inneemt dan bij de meeste andere groepen het geval is.

Tabel 30 Correlaties tussen RISC-scores en StatRec voor de totale onderzoeksgroep en voor verschillende deelgroepen^a

		Totale groep (N=9.985)	Mannen (N=9.013)	Vrouwen (N=972)	Jongste leeftijds- groep (N=3.403)	Middelste leef- tijdsgroep (N=3.278)	Oudste leeftijds- groep (N=3.304)
Schaal 1&2 Delictgegevens	r	0,67	0,67	0,58	0,65	0,70	0,64
	N	(8.122)	(7.294)	(828)	(2.842)	(2.658)	(2.622)
Schaal 3 Huisvesting en wonen	r	0,35	0,37	0,20	0,29	0,38	0,38
	N	(7.718)	(6.866)	(852)	(2.521)	(2.462)	(2.735)
Schaal 4 Opleiding, werk en leren	r	0,46	0,49	0,35	0,39	0,50	0,46
	N	(9.976)	(9.005)	(971)	(3.402)	(3.274)	(3.300)
Schaal 5 Inkomen en omgaan met geld	r	0,43	0,45	0,28	0,35	0,47	0,45
	N	(9.973)	(9.003)	(970)	(3.396)	(3.276)	(3.301)
Schaal 6 Relaties met partner, gezin en familie	r	0,17	0,19	0,18	0,17	0,19	0,18
	N	(9.970)	(9.000)	(970)	(3.400)	(3.274)	(3.296)
Schaal 7 Relaties met vrienden en kennissen	r	0,49	0,50	0,36	0,42	0,49	0,45
	N	(9.934)	(8.965)	(969)	(3.392)	(3.255)	(3.287)
Schaal 8 Druggebruik	r	0,46	0,46	0,42	0,29	0,54	0,57
	N	(8.327)	(7.577)	(750)	(2.930)	(2.811)	(2.586)
Schaal 9 Alcoholgebruik	r	0,22	0,21	0,11	0,18	0,30	0,28
	N	(9.961)	(8.994)	(967)	(3.394)	(3.270)	(3.297)
Schaal 10 Emotioneel welzijn	r	0,12	0,15	0,15	0,12	0,16	0,15
	N	(9.971)	(9.000)	(971)	(3.398)	(3.273)	(3.300)
Schaal 11 Denkpatronen, gedrag en vaardig- heden	r	0,35	0,35	0,27	0,36	0,36	0,33
	N	(9.960)	(8.988)	(972)	(3.393)	(3.270)	(3.297)
Schaal 12 Houding	r	0,30	0,30	0,22	0,34	0,32	0,30
	N	(9.889)	(8.921)	(968)	(3.373)	(3.249)	(3.267)
Totaalscore	r	0,64	0,65	0,50	0,59	0,68	0,65
	N	(8.076)	(7.251)	(825)	(2.826)	(2.644)	(2.606)

^a Alle correlaties zijn significant ($p < .001$, enkelzijdig).

Tabel 31 Correlaties tussen RISC-scores en StatRec voor verschillende herkomstgroepen^a

		Nederlands (N=6.289)	Marokkaans (N=759)	Antilliaans / Arubaans (N=531)	Surinaams (N=844)	Turks (N=445)	Overig niet- westers (N=597)	Westers (N=387)
Schaal 1&2 Delictgegevens	r	0,68	0,65	0,56	0,67	0,70	0,66	0,68
	N	(5.348)	(509)	(418)	(669)	(339)	(413)	(320)
Schaal 3 Huisvesting en wonen	r	0,37	0,26	0,33	0,35	0,25	0,33	0,26
	N	(5.183)	(486)	(326)	(579)	(340)	(435)	(304)
Schaal 4 Opleiding, werk en leren	r	0,48	0,35	0,39	0,51	0,33	0,31	0,45
	N	(6.284)	(759)	(531)	(844)	(443)	(595)	(387)
Schaal 5 Inkomen en omgaan met geld	r	0,44	0,34	0,29	0,39	0,31	0,37	0,42
	N	(6.281)	(759)	(529)	(843)	(444)	(597)	(387)
Schaal 6 Relaties met partner, gezin en familie	r	0,22	0,08*	0,11**	0,17	0,09*	0,06 ^{n.s.}	0,11*
	N	(6.282)	(757)	(530)	(844)	(445)	(592)	(387)
Schaal 7 Relaties met vrienden en kennissen	r	0,49	0,45	0,37	0,51	0,55	0,47	0,49
	N	(6.269)	(752)	(524)	(840)	(445)	(590)	(383)
Schaal 8 Druggebruik	r	0,48	0,38	0,36	0,51	0,41	0,37	0,45
	N	(5.202)	(645)	(481)	(734)	(348)	(477)	(315)
Schaal 9 Alcoholgebruik	r	0,27	0,26	0,14	0,19	0,28	0,25	0,24
	N	(6.273)	(758)	(530)	(839)	(445)	(596)	(387)
Schaal 10 Emotioneel welzijn	r	0,15	0,10**	0,17	0,24	0,10*	0,16	0,15**
	N	(6.281)	(757)	(530)	(842)	(445)	(596)	(387)
Schaal 11 Denkpatronen, gedrag en vaardigheden	r	0,35	0,29	0,37	0,42	0,34	0,29	0,26
	N	(6.275)	(758)	(528)	(842)	(444)	(593)	(387)
Schaal 12 Houding	r	0,30	0,22	0,28	0,30	0,28	0,26	0,28
	N	(6.242)	(751)	(524)	(835)	(440)	(582)	(383)
Totaalscore	r	0,65	0,60	0,56	0,68	0,61	0,59	0,60
	N	(5.323)	(507)	(415)	(666)	(338)	(405)	(316)

^a Tenzij anders vermeld zijn alle correlaties significant ($p < .001$, enkelzijdig).

^{n.s.} niet significant

* $p < .05$

** $p < .01$ (in alle gevallen is enkelzijdig getoetst)

Tabel 32 Correlaties tussen RISC-scores en StatRec voor verschillende delictcategorieën

		Vermogen geen geweld (N=1.990)	Vermogen met geweld (N=965)	Agressie (N=4.917)	Zeden ^b (N=835)	Opium (N=650)	Verkeer (N=159)	Overig (N=469)
Schaal 1&2 Delictgegevens	r	0,69	0,62	0,63	0,58	0,62	0,64	0,68
	N	(1.671)	(785)	(4.027)	(581)	(547)	(149)	(362)
Schaal 3 Huisvesting en wonen	r	0,40	0,31	0,23	0,29	0,26	0,25	0,47
	N	(1.431)	(493)	(4.212)	(687)	(369)	(153)	(373)
Schaal 4 Opleiding, werk en leren	r	0,48	0,43	0,37	0,40	0,34	0,40	0,53
	N	(1.988)	(965)	(4.912)	(833)	(650)	(159)	(469)
Schaal 5 Inkomen en omgaan met geld	r	0,37	0,31	0,31	0,44	0,29	0,36	0,43
	N	(1.988)	(964)	(4.909)	(835)	(650)	(158)	(469)
Schaal 6 Relaties met partner, gezin en familie	r	0,29	0,25	0,11	0,28	0,12	0,14*	0,26
	N	(1.988)	(965)	(4.905)	(834)	(650)	(159)	(469)
Schaal 7 Relaties met vrienden en kennissen	r	0,48	0,34	0,42	0,35	0,34	0,30	0,50
	N	(1.985)	(960)	(4.882)	(832)	(648)	(159)	(468)
Schaal 8 Druggebruik	r	0,48	0,34	0,38	0,38	0,40	0,38	0,43
	N	(1.769)	(859)	(3.983)	(622)	(570)	(131)	(393)
Schaal 9 Alcoholgebruik	r	0,30	0,23	0,25	0,27	0,21	0,10 ^{n.s.}	0,22
	N	(1.984)	(963)	(4.904)	(833)	(649)	(159)	(469)
Schaal 10 Emotioneel welzijn	r	0,14	0,16	0,09	0,15	0,03 ^{n.s.}	-0,05 ^{n.s.}	0,06 ^{n.s.}
	N	(1.986)	(965)	(4.910)	(833)	(650)	(159)	(468)
Schaal 11 Denkpatronen, gedrag en vaardigheden	r	0,46	0,39	0,29	0,30	0,34	0,35	0,46
	N	(1.985)	(963)	(4.904)	(832)	(649)	(159)	(468)
Schaal 12 Houding	r	0,41	0,30	0,24	0,32	0,30	0,36	0,43
	N	(1.973)	(957)	(4.872)	(820)	(642)	(159)	(466)
Totaalscore	r	0,69	0,59	0,55	0,55	0,58	0,59	0,69
	N	(1.661)	(779)	(4.004)	(580)	(543)	(149)	(360)

^a Tenzij anders vermeld zijn alle correlaties significant ($p < .001$, enkelzijdig).

^b De groep zedendelicten bestaat in deze analyse uit de groep RISC's die naar aanleiding van een strafzaak werden aangenomen waarin een zedendelict het hoogst bedreigde delict was. Als in deze strafzaak een ander feit een hogere strafdreiging had, is de RISC in de bijbehorende groep ingedeeld (bijvoorbeeld bij agressie indien het hoger bedreigde feit moord betreft).

^{n.s.} niet significant

* $p < .05$

** $p < .01$ (in alle gevallen is enkelzijdig getoetst)

5.2.4 *Multivariate analyses van de samenhang tussen de RISC-scores en StatRec*

Om na te gaan in hoeverre de RISC de scores op de StatRec verklaart, zijn regressieanalyses uitgevoerd. In eerste instantie kon dit slechts op een beperkte selectie van de onderzoeksgroep gebeuren omdat het aantal ontbrekende schaalscores, met name op de schalen 1&2 *Delictgegevens* en 3 *Huisvesting*, aanzienlijk was. Omdat dit betekende dat uitsluitend een zeer specifiek deel van de onderzoeksgroep in de analyses kon worden betrokken — de daders die bekennen dan wel veroordeeld zijn én die niet gedetineerd zijn of op afzienbare termijn vrij komen — is ervoor gekozen ontbrekende waarden te imputeren. Dat wil zeggen dat ontbrekende waarden zijn vervangen door een geldige score. De imputatie van ontbrekende waarden is gedaan op basis van *person mean substitution* (Huisman, 2000). Omdat de schaal van de gewogen scores per RISC-schaal verschilt — gewogen schaalscores op schaal 1&2 kunnen bijvoorbeeld variëren van 0 tot 50, terwijl gewogen schaalscores op schaal 10 kunnen variëren van 0 tot 6 — zijn eerst de gewogen schaalscores van de RISC getransformeerd naar z-scores. Vervolgens is voor elk individu het gemiddelde van alle getransformeerde schaalscores uitgerekend. De ontbrekende z-scores per schaal van een dader zijn daarna vervangen door de gemiddelde z-score van die dader. RISC's waarbij meer dan drie z-schaalscores ontbraken, zijn hiervan uitgezonderd en niet in de regressieanalyses betrokken (het gaat om 0,1% van het totaal).

In de regressieanalyse is de StatRec-score als afhankelijke variabele ingevoerd en zijn de RISC-schaalscores (z-waarden) als onafhankelijke variabelen ingevoerd. Omdat het uitgangspunt van de RISC is dat de schalen van het instrument criminogene factoren in kaart brengen die elk herhaling van het delictgedrag voorspellen, zijn alle schalen tegelijk in de analyse ingevoerd. In hoofdstuk 2 is al stilgestaan bij het feit dat het niet reëel is om te verwachten dat elke RISC-schaal, rekening houdend met de invloed van alle andere schalen, een unieke bijdrage levert aan de verklaring van de StatRec-voorspelling van recidive. Door de tweeledige doelstelling van de RISC — voorspellen en indiceren — is er in het instrument sprake van redundantie: schalen hangen met elkaar samen en hebben elk betrekking op een deel van hetzelfde achterliggende construct. Ze vormen immers een totaalscore met een voldoende interne consistentie. Bovendien is de score op de meeste RISC-schalen goed te voorspellen op basis van de andere schalen (zie tabel 33). Met name de score op schaal 11 *Denkpatronen, gedrag en vaardigheden* is goed te voorspellen op basis van de andere schalen, maar ook vrijwel alle andere schalen zijn goed te voorspellen door de score op de andere RISC-schalen. De enige uitzondering wordt gevormd door schaal 9 *Alcoholgebruik*.

Wanneer alle schalen als onafhankelijke variabelen in een regressieanalyse worden ingevoerd, is het door de gebleken samenhang tussen de schalen mogelijk dat regressiecoëfficiënten erg klein worden en zelfs negatief kunnen worden. Dit betekent echter niet dat de congruente validiteit van het instrument niet in orde is. De meeste bivariate correlaties tussen de afzonderlijke RISC-schalen en de StatRec-score zijn immers middelsterk en ze zijn alle positief. De resultaten van de regressieanalyses zijn voor het onderzoek naar de congruente validiteit van de RISC met name van belang omdat zij aangeven in hoeverre de StatRec-voorspelling door de RISC kan worden verklaard. De RISC wordt verwacht een aanzienlijk deel van de variantie van de StatRec-recidivekans te kunnen verklaren. Aan de interpretatie van de regressiecoëfficiënten per RISC-schaal zijn enkele beperkingen verbonden. Aan kleine regressiecoëfficiënten (tussen -0,10 en 0,10) en aan negatieve regressiecoëffi-

ciënten wordt geen aandacht besteed¹⁹. Grotere, positieve regressiecoëfficiënten kunnen daarentegen wel worden geïnterpreteerd. Deze interpretatie van de regressiecoëfficiënten moet als een exploratieve analyse worden gezien: zijn er schalen die ondanks de samenhang tussen de RISC-schalen een eigen bijdrage leveren aan de verklaring van de StatRec-score?

Tabel 33 Percentage verklaarde variantie per schaal door de andere RISC-schalen (op basis van regressieanalyses)

	Verklaarde variantie
Schaal 1&2 Delictgegevens	45%
Schaal 3 Huisvesting en wonen	45%
Schaal 4 Opleiding, werk en leren	43%
Schaal 5 Inkomen en omgaan met geld	41%
Schaal 6 Relaties met partner, gezin en familie	28%
Schaal 7 Relaties met vrienden en kennissen	44%
Schaal 8 Druggebruik	28%
Schaal 9 Alcoholgebruik	9%
Schaal 10 Emotioneel welzijn	31%
Schaal 11 Denkpatronen, gedrag en vaardigheden	60%
Schaal 12 Houding	49%

Tabel 34 geeft de resultaten van twee regressieanalyses voor de totale onderzoeksgroep: in het eerste model is schaal 1&2 *Delictgegevens* wel betrokken, in het tweede model is deze schaal buiten de analyses gehouden. Hier is voor gekozen omdat schaal 1&2 en StatRec op veel van dezelfde informatie zijn gebaseerd. Echter, bij de interpretatie van model 2 dient in het oog te worden gehouden dat dit niet de manier is waarop de RISC wordt gebruikt. In de berekening van de RISC-totaalscore telt schaal 1&2 immers gewoon mee.

Het model met schaal 1&2, model 1, verklaart 48% van de variantie in de afhankelijke variabele, namelijk de recidivekans zoals berekend met StatRec. Het model zonder schaal 1&2, model 2, verklaart 38% van de variantie van de StatRec-score. Dit is in beide gevallen een aanzienlijk deel van de variantie: een R^2 van 0,48 komt overeen met een correlatie van 0,69 (een sterke samenhang) en een R^2 van 0,38 komt overeen met een correlatie van 0,62 (eveneens een sterke samenhang).

Desalniettemin blijft 52% respectievelijk 62% van de variantie onverklaard. Hoewel het onderhavige rapport niet de aangewezen plek om hier uitgebreid op in te gaan, is een korte toelichting op zijn plaats. Er zijn verschillende mogelijke redenen waarom een groot deel van de variantie van de recidivevoorspelling (de StatRec-score) niet wordt verklaard door de RISC-schalen. Ten eerste zijn meetfouten natuurlijk niet uit te sluiten en ten tweede kan het zijn dat bepaalde persoonskenmerken, die niet in de RISC zitten, een deel van de resterende variantie verklaren. Benadrukt moet echter worden dat de gevonden percentages verklaarde variantie hoog zijn.

In model 1 levert schaal 1&2 *Delictgegevens* zoals verwacht mocht worden de belangrijkste bijdrage aan de verklaring van de StatRec-score: hoe hoger de score op schaal 1&2, des te hoger de door StatRec voorspelde recidivekans. Naast schaal 1&2 leveren ook twee dynamische factoren een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-voorspelling van recidive. Schaal 4 *Opleiding, werk en*

¹⁹ Regressiecoëfficiënten tussen -0,10 en 0,10 worden niet in de tabellen gerapporteerd.

leren levert de grootste bijdrage, gevolgd door de schaal 7 *Relaties met vrienden en kennissen*. Voor de overige schalen geldt dat door de onderlinge samenhang met alle andere schalen zeer lage of zelfs negatieve regressiecoëfficiënten werden gevonden. Nogmaals wordt benadrukt dat dit niet betekent dat de congruente validiteit van de RISC onvoldoende is, maar dat dit inherent is aan de dubbele doelstelling van het instrument: voorspellen en indiceren. Het feit dat opleiding en werk en relaties met vrienden en kennissen ondanks de samenhang met de andere RISC-schalen een unieke bijdrage leveren aan de verklaring van de StatRec-voorspelling van de recidivekans biedt mogelijkwijs aanknopingspunten voor het indiceren van gedragsinterventies. Voor deze schalen geldt dat problemen op deze gebieden, onafhankelijk van problemen op andere leefgebieden, samengaan met een grotere kans op recidive. Wellicht bieden interventies op deze twee leefgebieden de beste kansen om het recidiverisico direct te beïnvloeden. In dit onderzoek gaat het echter om de relatie tussen RISC en de *voorspelling* van de recidivekans en niet om de relatie met daadwerkelijke recidive. Het onderzoek naar de predictieve validiteit van de RISC, dat in 2008 wordt uitgevoerd, zal moeten uitwijzen in hoeverre deze RISC-schalen een unieke bijdrage leveren aan de voorspelling van daadwerkelijke recidive.

Wanneer model 2 wordt bestudeerd, wordt duidelijk wat de bijdrage van de verschillende dynamische RISC-schalen aan de verklaring van de StatRec-score is als in die verklaring geen rekening wordt gehouden met de justitiële voorgeschiedenis van de dader. De schalen die ondanks de samenhang met alle andere schalen een eigen bijdrage leveren aan de verklaring van het StatRec-recidiverisico zijn de schalen die betrekking hebben op opleiding en werk, relaties met vrienden en huisvesting. Daarnaast leveren druggebruik, alcoholgebruik en inkomen een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-score. Opnieuw zal uit het onderzoek naar de predictieve validiteit van de RISC moeten blijken welke dynamische RISC-schalen een unieke bijdrage leveren aan de voorspelling van daadwerkelijke recidive wanneer geen rekening wordt gehouden met de justitiële voorgeschiedenis van daders.

Tabel 34 Regressieanalyses van RISC op StatRec voor de totale groep^a

	Model 1	Model 2
N	9.973	9.973
R ²	0,48	0,38
β (gestandaardiseerd)		
Schaal 1&2 Delictgegevens	0,43***	--
Schaal 3 Huisvesting en wonen		0,14***
Schaal 4 Opleiding, werk en leren	0,16***	0,22***
Schaal 5 Inkomen en omgaan met geld		0,11***
Schaal 6 Relaties met partner, gezin en familie		
Schaal 7 Relaties met vrienden en kennissen	0,14***	0,22***
Schaal 8 Druggebruik		0,12***
Schaal 9 Alcoholgebruik		0,10***
Schaal 10 Emotioneel welzijn	-0,12***	-0,15***
Schaal 11 Denkpatronen, gedrag en vaardigheden		
Schaal 12 Houding	-0,12***	

^a Schaalscores zijn getransformeerd naar z-scores

*** p < .001

In de correlatiematrices (tabellen 30, 31 en 32) kon een aantal verschillen tussen diverse subgroepen worden vastgesteld. Om na te gaan in hoeverre er tussen deze subgroepen verschillen bestaan in het percentage variantie van de StatRec- recidivevoorspelling dat door de RISC kan worden verklaard, zijn regressie-analyses uitgevoerd voor de afzonderlijke groepen. Daarnaast is wederom exploratief nagegaan of er dynamische RISC-schalen zijn die, ondanks de samenhang tussen de schalen, een eigen bijdrage leveren aan de verklaring van de StatRec-score. Kleine en negatieve regressiecoëfficiënten worden net als in de totale onderzoeksgroep buiten beschouwing gelaten. Specifieke aandacht is uitgegaan naar opvallende verschillen tussen subgroepen. Verschillen tussen groepen zijn in voorkomende gevallen getoetst met behulp van een door Paternoster et al. (1998) opgestelde formule om na te gaan of de effecten van de RISC-schalen voor bepaalde groepen sterker of minder sterk zijn dan voor andere groepen. In deze formule worden de ongestandaardiseerde coëfficiënten tegen elkaar getoetst (z-toets). In de regressieanalyses per subgroep zijn steeds alle schalen van de RISC opgenomen, dus ook schaal 1&2. Hier is voor gekozen omdat dit het meeste recht doet aan de manier waarop de RISC in de praktijk wordt gebruikt: het bepalen van het risico op recidive op basis van statische én dynamische factoren.

In tabel 35 zijn de resultaten van de regressieanalyses bij mannen en vrouwen weergegeven. De RISC-schalen verklaren bij mannen veel meer variantie van de StatRec-score dan bij vrouwen: 49% respectievelijk 28%. Hoewel 28% verklaarde variantie nog steeds niet weinig is, is het wel aanzienlijk lager dan de percentages verklaarde variantie in de totale onderzoeksgroep, de groep mannen en in andere onderscheiden subgroepen (zie verderop). Hieruit kan echter niet worden geconcludeerd dat de congruente validiteit van de RISC in termen van de voorspelling van de recidivekans minder goed is voor vrouwen dan voor mannen. Zoals eerder in dit hoofdstuk al is opgemerkt, is het aantal first offenders binnen de groep vrouwen relatief groot: 39,9% van de vrouwen met een StatRec-score heeft geen eerdere justitiecontacten. Het voorspellen van de recidivekans met behulp van de StatRec is voor first offenders moeilijker dan voor daders die wel eerdere justitiecontacten hebben. Daarnaast is de StatRec-score voor vrouwen erg rechts-scheef verdeeld (veel vrouwen hebben een laag StatRec-recidiverisico). Mogelijkerwijs vormen deze twee gegevens een verklaring voor de relatief lagere verklaringskracht van de RISC voor de StatRec bij vrouwen.

In beide groepen levert schaal 1&2, zoals verwacht mocht worden, de belangrijkste bijdrage aan de verklaring van de StatRec-score. Binnen beide groepen blijken ook opleiding en werk en relaties met vrienden en kennissen een noemenswaardige, eigen bijdrage te leveren aan de verklaring van het recidiverisico zoals dat met StatRec wordt voorspeld. Uit de vergelijking van de effecten van de RISC-schalen tussen mannen en vrouwen blijkt dat het effect van schaal 4 *Opleiding, werk en leren* voor mannen significant sterker is dan voor vrouwen ($z=2,26$). Met betrekking tot relaties met vrienden en kennissen bestaan er geen significante verschillen tussen mannen en vrouwen in de sterkte van het effect op de verklaring van de recidivekans zoals die is berekend op basis van StatRec.

Tabel 35 Regressieanalyses van RISc op StatRec voor mannen en vrouwen^a

	Mannen	Vrouwen
N	9.002	971
R ²	0,49	0,28
β (gestandaardiseerd)		
Schaal 1&2 Delictgegevens	0,42***	0,31***
Schaal 3 Huisvesting en wonen		
Schaal 4 Opleiding, werk en leren	0,18***	0,14***
Schaal 5 Inkomen en omgaan met geld		
Schaal 6 Relaties met partner, gezin en familie		
Schaal 7 Relaties met vrienden en kennissen	0,15***	0,15***
Schaal 8 Druggebruik		
Schaal 9 Alcoholgebruik		
Schaal 10 Emotioneel welzijn	-0,11***	
Schaal 11 Denkpatronen, gedrag en vaardigheden		
Schaal 12 Houding	-0,12***	-0,12**

^a Schaalscores zijn getransformeerd naar z-scores

n.s. niet significant

* p < .05

** p < .01

*** p < .001

In tabel 36 worden de resultaten gepresenteerd van regressieanalyses per leeftijdsgroep. De RISc-schalen verklaren in de middelste groep de grootste proportie variantie van de StatRec-score, maar ook in beide andere groepen is het aandeel verklaarde variantie zeer aanzienlijk. In elke leeftijdsgroep leveren delictgegevens logischerwijs de grootste bijdrage aan de verklaring van het recidiverisico zoals ingeschat met StatRec. Daarnaast heeft in elke groep de dynamische RISc-schaal 4 *Opleiding, werk en leren* een unieke invloed in de verklaring van de StatRec-score. Voor de jongste groep is dit samen met schaal 7 *Relaties met vrienden en kennissen* de enige dynamische criminogene factor met een unieke invloed op de verklaring van de StatRec-score. Binnen de middelste en de oudste leeftijdsgroepen leveren ook de schalen 8 *Druggebruik* en 9 *Alcoholgebruik* een noemenswaardige, unieke bijdrage aan de verklaring van het recidiverisico. In de oudste groep, ten slotte, hebben ook huisvesting en inkomen nog een eigen invloed, ondanks de samenhang met alle andere schalen.

Uit de vergelijking van de verschillende leeftijdsgroepen blijkt dat het effect van schaal 1&2 *Delictgegevens* voor de jongste en de middelste leeftijdsgroep sterker is dan voor de oudste leeftijdsgroep (z is respectievelijk 2,29 en 2,58). Tussen de jongste en de middelste leeftijdsgroep bestaan geen significante verschillen met betrekking tot de sterkte van het effect van delictgeschiedenis op de verklaring van de recidivekans zoals die is berekend op basis van StatRec (z = -0,43). De effecten van de schalen 4 *Opleiding, werk en leren*, 8 *Druggebruik* en 9 *Alcoholgebruik* zijn voor de middelste leeftijdsgroep wel significant sterker dan voor de jongste groep (z is respectievelijk -2,80; -3,43; en -2,26). De jongste groep verschilt ook van de oudste groep: de effecten van schaal 3 *Huisvesting en wonen*, 5 *Inkomen en omgaan met geld*, 8 *Druggebruik* en 9 *Alcoholgebruik* zijn sterker voor de oudste groep dan voor de jongste groep (z is respectievelijk -2,94; -3,09; -6,57; en -2,70). De middelste leeftijdsgroep verschilt ten slotte ook weer van de oudste leeftijdsgroep: het effect van de schalen 3 *Huisvesting en wonen*, 5 *Inkomen en omgaan met geld* en 8 *Druggebruik* is sterker voor de oudste groep dan voor de middelste groep (z is respectievelijk

lijk -2,03; -2,19; en -3,23). Deze resultaten lijken erop te wijzen dat dergelijke factoren — middelengebruik, huisvesting, inkomen, opleiding en werk — bij oudere daders een grotere rol spelen dan bij jongere daders. Veel delictplegers stoppen na een bepaalde tijd met het plegen van delicten; naarmate daders ouder worden, stoppen er meer met het plegen van delicten. Wanneer mensen ondanks het ouder worden door blijven gaan met het plegen van delicten zou dat wellicht te maken kunnen hebben met dit soort kenmerken. Op basis van deze analyses kunnen daarover geen harde conclusies worden getrokken, maar het is wel interessant om in de studie naar de predictieve validiteit verder in te gaan op dit soort verschillen tussen jongere en oudere daders.

Tabel 36 Regressieanalyses leeftijdsgroepen^a

	Jongste groep	Middelste groep	Oudste groep
N	3.399	3.276	3.298
R ²	0,43	0,53	0,46
β (gestandaardiseerd)			
Schaal 1&2 Delictgegevens	0,50***	0,44***	0,34***
Schaal 3 Huisvesting en wonen			0,11***
Schaal 4 Opleiding, werk en leren	0,12***	0,17***	0,16***
Schaal 5 Inkomen en omgaan met geld			0,12***
Schaal 6 Relaties met partner, gezin en familie			
Schaal 7 Relaties met vrienden en kennissen	0,10***		
Schaal 8 Druggebruik		0,10***	0,16***
Schaal 9 Alcoholgebruik		0,11***	0,12***
Schaal 10 Emotioneel welzijn	-0,11***	-0,11***	
Schaal 11 Denkpatronen, gedrag en vaardigheden			
Schaal 12 Houding			

^a Schaalscores zijn getransformeerd naar z-scores

n.s. niet significant

* p < .05

** p < .01

*** p < .001

Tabel 37 geeft de resultaten van de regressieanalyses voor verschillende herkomstgroepen. Alleen de resultaten van daders met een Nederlandse achtergrond en daders uit de vier grootste migrantengroepen in Nederland (Marokkanen, Antillianen / Arubanen, Surinamers en Turken) zijn weergegeven²⁰. Wat in tabel 37 opvalt, is dat de percentages verklaarde variantie in alle groepen ongeveer even hoog liggen, behalve bij de groep Antillianen / Arubanen. In deze groep verklaren de RISC-schalen 32% van de variantie in de StatRec-score, wat overeenkomt met een sterke correlatie van 0,57. In de andere groepen verklaart de RISC gemiddeld 47% van de variantie in de StatRec-voorspelling van de recidivekans, overeenkomend met een correlatie van 0,69.

In alle groepen levert schaal 1&2 *Delictgegevens* logischerwijs de grootste bijdrage aan de verklaring van de StatRec-score. Wat opvalt is dat het effect van deze schaal groter is voor Nederlanders dan voor daders uit andere herkomstgroepen ($z = 3,07$). Voor Antillianen / Arubanen en voor Surinamers is de invloed van deze schaal op de verklaring van de StatRec-voorspelling van de recidivekans in vergelijking met

²⁰ Resultaten voor de groepen met een anderszins niet westerse of westerse achtergrond kunnen bij de onderzoekers worden opgevraagd.

dadere uit andere herkomstgroepen juist kleiner (z is respectievelijk $-3,82$ en $-3,11$). Uit de exploratieve bestudering van de regressiecoëfficiënten van de dynamische criminogene factoren komt een aantal opmerkelijke resultaten. Zo blijken opleiding en werk voor Marokkanen en Turken geen unieke bijdrage te leveren aan de verklaring van de StatRec-recidivekans. Dit zijn de enige twee subgroepen die in het kader van het onderzoek naar de congruente validiteit van de RISC zijn bestudeerd, waarvoor deze schaal geen unieke invloed heeft op de verklaring van de recidivekans. Het verschil met dadere uit andere herkomstgroepen blijkt voor zowel Marokkanen ($z = -2,90$) als Turken ($z = -2,52$) significant te zijn; ten opzichte van dadere uit de andere herkomstgroepen is het effect van opleiding en werk op de verklaring van de StatRec-voorspelling van recidive voor Marokkanen en voor Turken minder sterk. Opleiding en werk hebben in de groep Nederlanders daarentegen een sterker effect op de verklaring van de StatRec-voorspelling van de recidivekans dan bij dadere uit andere herkomstgroepen ($z = 3,53$). Relaties met vrienden en kennissen leveren in alle groepen, behalve in de groep Antillianen / Arubanen een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-score. Het verschil met andere herkomstgroepen is voor Nederlanders en Turken significant (z is respectievelijk $-3,16$ en $4,02$); in vergelijking met dadere uit de andere herkomstgroepen levert de relatie met vrienden en kennissen voor Nederlanders een minder sterke bijdrage en voor Turken een significant grotere bijdrage aan de verklaring van de StatRec-score. Wanneer Antillianen en Arubanen met betrekking tot de invloed van de relaties met vrienden en kennissen worden vergeleken met dadere uit andere herkomstgroepen is dit verschil significant: relaties met vrienden verklaren in de groep Antillianen / Arubanen een kleiner deel van de variantie van de StatRec-score dan in de groep dadere uit andere herkomstgroepen ($z = -2,96$). In de groepen van Marokkaanse, Antilliaanse en Surinaamse afkomst levert druggebruik een unieke bijdrage aan de verklaring van de recidivevoorspelling op basis van StatRec. Wanneer elk van deze groepen wordt vergeleken met alle dadere uit andere herkomstgroepen zijn er geen significante verschillen met betrekking tot de sterkte van dit effect. Voor Nederlanders en Turken levert deze schaal, gegeven de samenhang met alle andere schalen geen eigen bijdrage, maar is daarentegen alcoholgebruik van belang voor de verklaring van de StatRec-recidivekans. Dit effect is voor Nederlanders significant sterker dan voor dadere uit andere herkomstgroepen ($z = 2,79$). Binnen de groep Antillianen / Arubanen en de groep Surinamers levert, tot slot, schaal 11 *Denkpatronen, gedrag en vaardigheden* een unieke bijdrage aan de verklaring van de op basis van StatRec voorspelde recidivekans. Het verschil met dadere uit andere herkomstgroepen is echter niet significant (z is respectievelijk $1,79$ en $1,41$).

Tabel 37 Regressieanalyses herkomstgroepen^a

	Nederlands	Marok- kaans	Antilliaans Arubaans	Surinaams	Turks
N	6.284	759	529	841	445
R ²	0,50	0,41	0,32	0,49	0,48
β (gestandaardiseerd)					
Schaal 1&2 Delictgegevens	0,44***	0,43***	0,34***	0,38***	0,42***
Schaal 3 Huisvesting en wonen					
Schaal 4 Opleiding, werk en leren	0,18***		0,15**	0,15***	
Schaal 5 Inkomen en omgaan met geld					
Schaal 6 Relaties met partner, gezin en familie		-0,11**	-0,10*		
Schaal 7 Relaties met vrienden en kennissen	0,12***	0,22***		0,13***	0,32***
Schaal 8 Druggebruik		0,11***	0,13**	0,15***	
Schaal 9 Alcoholgebruik	0,11***				0,10**
Schaal 10 Emotioneel welzijn	-0,11***				
Schaal 11 Denkpatronen, gedrag en vaardigheden			0,17**	0,13**	
Schaal 12 Houding	-0,13***	-0,15***	-0,11*	-0,12***	

^a Schaalscores zijn getransformeerd naar z-scores

n.s. niet significant

* p < .05

** p < .01

*** p < .001

In tabel 38 worden de resultaten gepresenteerd van de regressieanalyses van RISC op StatRec binnen verschillende delictgroepen. Het percentage verklaarde variantie loopt uiteen van 39% voor zeden- en agressiedelinquenten tot 54% voor plegers van verkeersdelicten en plegers van overige delicten. In alle gevallen gaat het om een aanzienlijk deel van de variantie.

Zoals verwacht mocht worden, draagt schaal 1&2 *Delictgegevens* in alle groepen het sterkst bij aan de verklaring van de StatRec-score. Daarnaast levert schaal 4 *Opleiding, werk en leren* in alle groepen een noemenswaardige, unieke bijdrage aan de verklaring van de recidivekans zoals voorspeld op basis van StatRec. Wat opvalt is dat het effect van deze schaal in de groep verkeersdelinquenten sterker is dan voor plegers die wegens andere delicten werden vervolgd ($z = 2,81$). In de groep plegers van vermogensdelicten zonder geweld leveren naast opleiding en werk ook huisvesting, relaties met vrienden, druggebruik en denkpatronen een eigen bijdrage aan de verklaring van de StatRec-recidivekans. In de groep plegers van vermogensdelicten met geweld heeft naast opleiding en werk alleen schaal 11 *Denkpatronen, gedrag en vaardigheden* een noemenswaardig, uniek effect op de StatRec-score. Binnen de groep plegers van agressiedelicten zijn, naast uiteraard opleiding en werk, relaties met vrienden en kennissen en alcoholgebruik van belang voor de verklaring van de StatRec-recidivekans. In feite is dit niet verwonderlijk aangezien veel agressiedelicten onder invloed van alcohol worden gepleegd (Greenfeld, 1998; Roizen, 1997). Het verschil met plegers van andere delicten is evenwel niet significant ($z = 1,60$).

Voor zedendelinquenten leveren drie van de tien dynamische factoren een unieke bijdrage aan de verklaring van de op basis van StatRec voorspelde recidivekans. Het gaat hierbij om opleiding en werk, inkomen en alcoholgebruik. Voor plegers

van opiumdelicten levert, naast opleiding en werk alleen druggebruik een eigen bijdrage aan de verklaring van de StatRec-score. In de groep verkeersdelinquenten is naast opleiding en werk het druggebruik van belang voor de verklaring van het StatRec-recidiverisico. Hoewel de unieke bijdrage van houding niet significant is, valt hij op omdat verkeersdelinquenten de enige subgroep vormen waarin deze schaal een positieve bijdrage levert gegeven de samenhang met alle andere schalen. In een grotere onderzoeksgroep zou een dergelijk effect wel significantie bereiken en dit verdient zeker aandacht in toekomstig onderzoek. In de groep plegers van 'overige delicten' zijn, tot slot, huisvesting, opleiding en denkpatronen de leefgebieden met een uniek effect op de verklaring van de StatRec-score, ondanks de samenhang tussen deze schalen en de andere RISC-schalen.

Op een aantal verschillen tussen subgroepen is hiervoor al ingegaan. De verschillen tussen delictgroepen zijn onderzocht door steeds een delictgroep af te zetten tegen daders die wegens andere delicten werden vervolgd. Plegers van vermogensdelicten zonder geweld zijn dus vergeleken met plegers van andere delicten. Hierbij moet de groep plegers van 'andere delicten' worden onderscheiden van de *delictcategorie* 'overige delicten'. De groep plegers van andere delicten omvat in elke analyse alle daders die wegens een ander delict werden vervolgd dan het delict waarvoor de te beschrijven groep werd vervolgd. Daders uit de delictcategorie 'overige delicten' werden vervolgd voor een delict dat niet was in te delen in een van de andere delictcategorieën (vermogen zonder / met geweld, agressie, zeden, opium of verkeer).

Plegers van vermogensdelicten zonder geweld verschillen uitsluitend met betrekking tot relaties met vrienden en kennissen van plegers van overige delicten: het effect van dit leefgebied op de verklaring van de StatRec-recidivekans is minder sterk voor plegers van vermogensdelicten zonder geweld dan voor plegers van andere delicten ($z = -2,32$). Wat opvalt is dat het effect van schaal 1&2 *Delictgegevens* in de groep plegers van vermogensdelicten met geweld minder sterk is dan in de groep plegers van andere delicten ($z = -3,64$). In de groep plegers van agressiedelicten leveren opleiding en werk, net als in vrijwel alle andere onderzochte subgroepen, leveren opleiding en werk een unieke bijdrage aan de verklaring van het recidiverisico zoals berekend met StatRec. In vergelijking met plegers van andere delicten is het effect van deze schaal in de groep agressiedaders evenwel significant minder sterk ($z = -2,42$).

Zedendelinquenten wijken in een aantal opzichten af van plegers van andere delicten. Sowieso viel binnen de groep al het belang op van schaal 5 *Inkomen en omgaan met geld*. Uit de vergelijkingen met de groep plegers van andersoortige delicten dan zedendelicten blijkt dat het effect van inkomen voor de zedendelinquenten significant sterker is dan voor de plegers van andere delicten ($z = 3,74$). De bijdrage die schaal 1&2 *Delictgegevens* in de groep zedendelinquenten levert aan de verklaring van de StatRec-score is significant kleiner dan voor plegers van andere delicten ($z = -2,23$). Het effect van druggebruik op de verklaring van de StatRecscore is voor daders van opiumdelicten sterker dan voor plegers van andere delicten ($z = 2,14$).

Er blijkt een aantal opmerkelijke verschillen te bestaan tussen plegers van verkeersdelicten en plegers van andere delicten. Het effect van druggebruik en houding op de verklaring van de StatRec-score is voor plegers van verkeersdelicten sterker dan voor plegers van andere delicten (z is respectievelijk 3,33 en 3,37). Zoals hiervoor al is opgemerkt, vormen verkeersdelinquenten de enige bestudeerde sub-

groep voor wie houding op de verklaring van de recidivekans, gegeven alle andere RISC-schalen, een positieve bijdrage levert aan de verklaring van de StatRec-score. Ook het effect van opleiding en werk is voor verkeersdelinquenten sterker dan voor plegers van andere delicten ($z = 2,81$). Bij plegers van overige delicten is, tot slot, de bijdrage die huisvesting en wonen leveren aan de verklaring van de StatRec-recidivekans groter dan bij plegers van andere delicten ($z = 2,65$).

Tabel 38 Regressieanalyse van RISC-schalen op StatRec voor verschillende delictcategorieën^a

	Vermogen geen geweld	Vermogen met geweld	Agressie	Zeden ^b	Opium	Verkeer	Overig
N	1.990	965	4.909	833	648	159	469
R ²	0,53	0,42	0,39	0,39	0,43	0,54	0,54
β (gestandaardiseerd)							
Schaal 1&2 Delictgegevens	0,45***	0,49***	0,44***	0,36***	0,46***	0,49***	0,39***
Schaal 3 Huisvesting en wonen	0,11***						0,19***
Schaal 4 Opleiding, werk en leren	0,14***	0,17***	0,15***	0,14***	0,10**	0,36***	0,20***
Schaal 5 Inkomen en omgaan met geld				0,18***			
Schaal 6 Relaties met partner, gezin en familie							
Schaal 7 Relaties met vrienden en kennissen	0,10***		0,14***				
Schaal 8 Druggebruik	0,11***				0,16***	0,26***	
Schaal 9 Alcoholgebruik				0,13***			
Schaal 10 Emotioneel welzijn	-0,11***	-0,13***	-0,13***		-0,17***	-0,23***	-0,17***
Schaal 11 Denkpatronen, gedrag en vaardigheden	0,10***	0,12**					0,16**
Schaal 12 Houding		-0,13***	-0,13***			0,15 ^{n.s.}	-0,10 ^{n.s.}

^a Schaalscores zijn getransformeerd naar z-scores

^b De groep zedendelicten bestaat in deze analyse uit de groep RISC's die naar aanleiding van een strafzaak werden afgenomen waarin een zedendelict het hoogst bedreigde delict was. Als in deze strafzaak een ander feit een hogere strafdreiging had, is de RISC in de bijbehorende groep ingedeeld (bijvoorbeeld bij agressie indien het hoger bedreigde feit moord betreft).

^{n.s.} niet significant

* p < .05

** p < .01

*** p < .001

5.3 Congruente validiteit van de RISC-schalen 10, 11 en 12: schaalstructuur en interne consistentie van de voor validering gebruikte instrumenten

5.3.1 Onderzoeksgroep

In totaal hebben 204 reclasseringscliënten tussen april en half september 2006 in het kader van het onderzoek naar de congruente validiteit van de RISC een van de drie in het onderzoek gebruikte vragenlijsten — NPV, UCL en BDHI-D — ingevuld. Dit zijn allemaal andere cliënten dan degenen die meewerkten aan het onderzoek naar de interbeoordelaarsbetrouwbaarheid. De RISC's van zes cliënten waren onbruikbaar voor de analyses omdat de invoerdata van deze RISC's aangaven dat ze ruim voor de start van het onderzoek waren afgenomen. Drie cliënten werden vanwege fouten in hun cliëntnummer niet teruggevonden in de door de RN beheerde RISC-database en van elf cliënten was de RISC bij de start van de data-analyses in oktober 2006 (nog) niet beschikbaar. Om deze reden zijn deze personen uit de onderzoeksgroep verwijderd. Het ging om vier mensen die een NPV hadden ingevuld, acht mensen die een UCL hadden ingevuld en zeven mensen die een BDHI-D hadden ingevuld. In de totale validiteitsonderzoeksgroep resteren 185 personen.

In tabel 39 wordt, gegroepeerd naar de ingevulde vragenlijst, een aantal achtergrondkenmerken weergegeven²¹. Elke groep bestaat voor ongeveer 85% uit mannen. De verschillen in de verdeling mannen en vrouwen tussen de drie groepen zijn niet significant: $\chi^2(2) = 0,24$; $p = .89$. De gemiddelde leeftijd van de NPV-groep is 32 jaar en 8 maanden (sd = 11,21; range = 19,40 - 65,62). De UCL-groep is gemiddeld 35 jaar en 1 maand (sd = 13,01; range = 18,64 - 74,45), en de BDHI-D-groep is gemiddeld 33 jaar en 1 maand (sd = 12,66; range = 18,82 - 70,48). De verschillen tussen de drie groepen zijn niet significant: $F = 0,66$; $df = 2$; $p = .52$.

De verdeling over de drie reclasseringsorganisaties blijkt voor de drie groepen grotendeels overeen te komen: er bestaan geen significante verschillen tussen de drie groepen: $\chi^2(4) = 1,02$; $p = .91$. De indeling in drie herkomstgroepen (autochtoon, niet-westers allochtoon en westers allochtoon) die elders in dit rapport is gehanteerd, heeft in de onderhavige onderzoeksgroepen tot gevolg dat de subgroepen te klein worden om nog betekenisvolle analyses uit te kunnen voeren²². Om deze reden is ervoor gekozen de twee allochtone subgroepen samen te nemen. In elk van de drie groepen is het merendeel van de daders van autochtone herkomst. Dit percentage varieert van groep tot groep, maar er bestaan geen significante verschillen tussen de drie groepen met betrekking tot de verdeling in autochtonen en allochtonen: $\chi^2(2) = 2,63$; $p = .27$.

²¹ Deze tabel verschilt van de tabel met achtergrondkenmerken van de onderzoeksgroep uit het interbeoordelaarsbetrouwbaarheidsonderzoek omdat bij de variabelen leeftijdsklasse en geboorteland het aantal personen in bepaalde categorieën erg klein was. Daarom is ervoor gekozen de groepen niet uit te splitsen naar leeftijdsklasse en geboorteland.

²² Met speciale analysetechnieken is het wel mogelijk om na te gaan of er significante verschillen bestaan tussen de drie onderzoeksgroepen (NPV, UCL en BDHI-D) in de verdeling in de drie herkomstgroepen. Een parametrische bootstrap-analyse wees uit dat dergelijke verschillen niet aanwezig waren: $p(B=10.000)=.24$.

Tabel 39 Congruente validiteit van de schalen 10, 11 en 12: achtergrondkenmerken van de onderzoeksgroepen per vragenlijst

Kenmerk	Ingevulde vragenlijst			
	NPV	UCL	BDHI-D	
Omvang groep	61	60	64	
Geslacht	Man	51 (83,6%)	52 (86,7%)	54 (84,4%)
	Vrouw	10 (16,4%)	8 (13,3%)	10 (15,6%)
Organisatie	RN	37 (60,7%)	34 (56,7%)	39 (60,9%)
	LJ&R	9 (14,8%)	10 (16,7%)	7 (10,9%)
	SVG	15 (24,6%)	16 (26,7%)	18 (28,1%)
Gemiddelde leeftijd (in jaren)	32,67	35,08	33,11	
Herkomstgroep	Autochtoon	51 (83,6%)	47 (78,3%)	57 (89,1%)
	Allochtoon	10 (16,4%)	13 (21,7%)	7 (10,9%)

Omdat de drie onderzoeksgroepen niet significant van elkaar verschillen met betrekking tot hun achtergrondkenmerken, zijn de groepen samengevoegd tot één onderzoeksgroep (N = 185). In tabel 40 worden de achtergrondkenmerken van deze groep gegeven. Net als in het onderzoek naar de interbeoordelaarsbetrouwbaarheid was het oorspronkelijk de bedoeling deze onderzoeksgroep te vergelijken met daders die wel gevraagd waren deel te nemen aan het onderzoek naar de congruente validiteit, maar die niet wilden meewerken. Hiertoe was de reclasseringswerkers in dit onderzoek gevraagd iedere cliënt te benaderen voor deelname aan het onderzoek en een administratie bij te houden met de cliëntnummers van de daders die weigerden mee te doen. De omvang van de geregistreerde groep weigeraars was echter dusdanig klein dat, gezien het feit dat de periode van dataverzameling verlengd diende te worden om voldoende vragenlijsten te verzamelen, twijfels bestaan over de vraag of de groep geregistreerde weigeraars een representatieve afspiegeling vormt van de non-respons. Net als in het onderzoek naar de interbeoordelaarsbetrouwbaarheid is ervoor gekozen de onderzoeksgroep niet te vergelijken met de weigeraars, maar met de reguliere reclasseringspopulatie. Hiervoor is gebruik gemaakt van de RISC-database die van de RN is verkregen en die in paragraaf 2.3 is beschreven. De personen die deelnamen aan het onderzoek naar de congruente validiteit zijn ten behoeve van deze vergelijking tijdelijk uit de grote database verwijderd opdat zij niet 'met zichzelf' zouden worden vergeleken. Zo doende kon de onderzoeksgroep worden vergeleken met 11.638 RISC's uit de RISC-database van de drie reclasseringsorganisaties.

90,6% van de 11.638 RISC's uit de RISC-database is afgenomen bij een man, 9,4% is afgenomen bij een vrouw. De verdeling in de onderhavige onderzoeksgroep blijkt significant af te wijken van de verdeling tussen mannen en vrouwen in de grote database: $\chi^2(1) = 7,15$; $p = .01$. Het percentage vrouwen in de onderzoeksgroep is met 15,1% groter dan mocht worden verwacht op basis van de verdeling in de reguliere reclasseringspopulatie. Een verklaring hiervoor is moeilijk te geven. Wellicht zijn vrouwen sneller bereid een vragenlijst in te vullen dan mannen. De gemiddelde leeftijd op het moment van afname van de RISC ligt in de database (N = 11.604; van 34 daders mist de geboortedatum) op 34 jaar en 4 maanden ($m = 34,30$; $sd = 12,01$; range = 17,11 - 86,48). Een t-toets om de gemiddelde leeftijd in de onderzoeksgroep ($m = 33,60$; $sd = 12,31$; range = 18,64 - 74,45) te vergelijken met de ge-

middelste leeftijd in de grote database wijst uit dat er geen significant verschil bestaat tussen beide groepen: $t = -0,77$; $df = 184$; $p = .44$ (tweezijdig). Met betrekking tot herkomst — ingedeeld in de categorieën autochtoon, niet-westers allochtoon en westers allochtoon — wijkt de onderzoeksgroep wel significant af van de reguliere reclasseringspopulatie: $\chi^2(2) = 38,44$; $p = .00$. Het aandeel autochtonen is in de onderzoeksgroep groter dan in de reguliere populatie (83,8% tegen 60,4%), terwijl het aandeel niet-westerse allochtonen in de onderzoeksgroep kleiner dan in de reguliere populatie (13,0% tegen 33,1%). De percentages westerse allochtonen in beide groepen ontlopen elkaar nauwelijks (3,2% tegen 3,9%). Hoe dit verschil in herkomstgroepen verklaard kan worden, is onduidelijk. Het is mogelijk dat de regio's waarin het onderzoek is uitgevoerd van invloed zijn geweest op het percentage deelnemers met een niet-westerse achtergrond: in Noord-Nederland komen wellicht minder niet-westerse allochtonen in aanraking met politie en justitie dan elders in Nederland (de Randstad). Aan de andere kant deden SVG, LJ&R en RN ook in Utrecht mee aan het onderzoek. Het is mogelijk dat niet-westerse allochtonen minder bereid waren een vragenlijst in te vullen, maar het is ook mogelijk dat reclasseringswerkers minder vaak aan cliënten met een niet-westerse achtergrond hebben gevraagd of ze mee zouden willen werken aan het onderzoek. Wellicht omdat zij de Nederlandse taal minder goed (zouden) beheersen, iets wat een voorwaarde was voor deelname aan het onderzoek.

De grootste groep bestaat uit cliënten van de Reclassering Nederland (RN), gevolgd door de verslavingsreclassering en het Leger des Heils. Hoewel de RN inderdaad de grootste organisatie is en het Leger des Heils de kleinste, komt de verhouding tussen de drie reclasseringsorganisaties in deze onderzoeksgroep niet overeen met de daadwerkelijke verhouding tussen de organisaties. In werkelijkheid is de verhouding tussen de drie organisaties ongeveer 67 - 25 - 8 (zie ook hoofdstuk 3). De afwijking van deze verdeling die in de onderzoeksgroep is gevonden, is significant: $\chi^2(2) = 10,21$; $p = .01$.

Tabel 40 Congruente validiteit van de schalen 10, 11 en 12: achtergrondkenmerken van de onderzoeksgroep (N = 185)

Kenmerk		Aantal	%
Geslacht	Man	157	84,9
	Vrouw	28	15,1
Organisatie	RN	110	59,5
	LJ&R	26	14,1
	SVG	49	26,5
Leeftijd	Jonger dan 20 jaar	16	8,6
	20 – 30 jaar	68	36,8
	30 – 40 jaar	54	29,2
	40 – 50 jaar	23	12,4
	50 – 60 jaar	17	9,2
	60 jaar of ouder	7	3,8
Geboorteland	Nederland	165	89,2
	Marokko	2	1,1
	Nederlandse Antillen / Aruba	5	2,7
	Suriname	4	2,2
	Turkije	1	0,5
	Overig niet-westers	3	1,6
	Overig westers	5	2,7
Herkomstgroep	Autochtoon	155	83,8
	Niet-westers allochtoon	24	13,0
	Westers allochtoon	6	3,2

De onderzoeksgroep blijkt op een aantal opvallende achtergrondkenmerken te verschillen van de reguliere reclasseringspopulatie (N = 11.638). Om na te gaan in hoeverre ze wat hun RISC-scores betreft afwijken van de reguliere populatie zijn t-toetsen uitgevoerd (tweezijdig getoetst). De resultaten, die in tabel 41 worden gepresenteerd, laten zien dat het verschil tussen de onderzoeksgroep en de reguliere reclasseringspopulatie in vier gevallen significant is. De onderzoeksgroep heeft minder problemen op het gebied van opleiding en werk, meer problemen in hun relaties met partner en gezin, meer problemen met betrekking tot alcoholgebruik en *minder* problemen op het gebied van houding. In alle gevallen blijkt echter uit de bestudering van de effectgroottes dat het om kleine verschillen gaat. De effectgroottes (d) zijn respectievelijk 0,15, 0,16, 0,16 en 0,25.

Tabel 41 **Vergelijking RISC-scores tussen onderzoeksgroep en reguliere reclasseringspopulatie**

RISC-onderdeel	Gemiddelde score		t
	Onderzoeksgroep	Vergelijken- groep	
Schaal 1&2 Delictgegevens	10,43	10,44	-0,12
Schaal 3 Huisvesting en wonen	2,21	1,79	1,69
Schaal 4 Opleiding, werk en leren	6,34	7,34	-1,99*
Schaal 5 Inkomen en omgaan met geld	2,86	2,65	0,82
Schaal 6 Relaties met partner, gezin en familie	2,76	2,47	2,15*
Schaal 7 Relaties met vrienden en kennissen	3,36	3,49	-0,42
Schaal 8 Druggebruik	3,05	3,08	1,10
Schaal 9 Alcoholgebruik	1,61	1,31	2,11*
Schaal 10 Emotioneel welzijn	2,24	2,03	1,61
Schaal 11 Denkpatronen, gedrag en vaardigheden	5,86	6,29	-1,63
Schaal 12 Houding	4,16	5,21	-3,33**
Totaalscore	44,80	46,51	-0,67

* $p < .05$ ** $p < .01$

De verschillen tussen de onderzoeksgroep en de reguliere reclasseringspopulatie met betrekking tot geslacht, herkomst en verdeling over de drie reclasseringsorganisaties zijn opmerkelijk. Met betrekking tot de RISC blijkt de onderzoeksgroep evenwel niet bijzonder veel te verschillen van de reguliere reclasseringspopulatie. Op een aantal schalen wijkt hun score weliswaar significant af van die in de reguliere populatie, maar het gaat nooit om grote verschillen. Desondanks zal bij de interpretatie van de resultaten van het onderzoek naar de congruente validiteit de nodige voorzichtigheid moeten worden betracht.

5.3.2 *Controle van de schaalstructuur van NPV, UCL en BDHI-D*

Voorafgaand aan het beantwoorden van de onderzoeksvragen naar de congruente validiteit van de RISC, is nagegaan in hoeverre de schaalstructuur van de hiervoor gebruikte vragenlijsten — NPV, UCL en BDHI-D — in de onderhavige onderzoeksgroep overeenkomt met de resultaten die in de handleidingen van de verschillende instrumenten worden gerapporteerd (Lange, Hoogendoorn, Wiederspahn & De Beurs, 2005; Luteijn, Starren & Van Dijk, 2000; Schreurs, Van de Willige, Brosschot, Tellegen & Graus, 1993). Ten behoeve van dit replicatieonderzoek is voor iedere vragenlijst per schaal een principale componentenanalyse (pca) uitgevoerd (met iteraties en varimaxrotatie). In eerste instantie zijn factoren met een eigenwaarde groter dan één geëxtraheerd. Vervolgens werden de items horend bij een schaal nogmaals ingevoerd in een pca en gedwongen tot een éénfactoroplossing en werd Cronbach's alfa berekend. Tevens zijn per vragenlijst de correlaties tussen de schalen berekend en is nagegaan hoe de scores van de RISC-onderzoeksgroep verdeeld zijn over de klassen van de normeringsschalen.

Een verslag met een volledige beschrijving van de resultaten is bij de onderzoekers op te vragen. Hier wordt volstaan met een beschrijving van de belangrijkste bevindingen. Bij de interpretatie van deze bevindingen wordt benadrukt dat de onderzoeksgroep per vragenlijst klein is. De NPV is door 61 personen ingevuld, de UCL

door 60 personen en de BDHI-D door 64 personen. Deze aantallen hebben tot gevolg dat de resultaten niet erg stabiel zijn. Zelfs op het oog aanzienlijke afwijkingen ten opzichte van de resultaten uit de handleidingen van de instrumenten, kunnen daarom nog vallen binnen de te verwachten steekproeffluctuaties.

Over het algemeen wijzen de resultaten erop dat de gebruikte vragenlijsten in de onderhavige onderzoeksgroep dezelfde karakteristieken vertonen als in de handleidingen is vermeld. Gegeven de kleine aantallen personen die de vragenlijsten hebben ingevuld, wijken de resultaten van de *pca*'s, de interne consistentie van de schalen en de intercorrelaties tussen de schalen doorgaans niet significant af van de resultaten uit de handleidingen. Voor de meeste schalen van de drie vragenlijsten wijkt de verdeling van de normscores weliswaar significant af van de verdelingen die in de handleidingen worden gegeven, maar vrijwel nergens is dit zodanig dat voor plafond- of vloereffecten moet worden gevreesd²³. Opvallend zijn echter de lage interne consistentie van de NPV-schaal *Zelfgenoegzaamheid* ($\alpha = 0,47$) en de gepiekte verdeling van normscores op de NPV-schaal *Zelfwaardering* en de UCL-schaal *Geruststellende gedachten*. Uitgaande van 95% betrouwbaarheidsintervallen lijkt de samenhang tussen de UCL-schaal *Steun zoeken* en een aantal andere UCL-schalen, zelfs gezien het kleine aantal personen in de onderzoeksgroep, sterk af te wijken van de resultaten die in de UCL-handleiding worden gerapporteerd. De in de handleiding gerapporteerde correlaties vallen buiten de in de RISC-groep gevonden 95% betrouwbaarheidsintervallen.

5.4 Resultaten van de validering van RISC met behulp van NPV, UCL en BDHI-D

In dit onderzoek is er, bij gebrek aan andere mogelijkheden, voor gekozen de RISC te valideren aan de hand van instrumenten die door reclasseringscliënten worden ingevuld. Het beoordelingsinstrument RISC wordt dus gevalideerd met behulp van het eigen oordeel van delinquenten. Dat dit een nadeel met zich meebrengt, is in hoofdstuk 2 al besproken. Uit onderzoek naar probleemgedrag bij kinderen en jongeren is immers bekend dat de overeenstemming tussen verschillende informanten doorgaans niet bijzonder groot is (Achenbach, McConaughy & Howell, 1987; Ten Brink, 1998; Kroes, Veerman & De Bruyn, 2000). Het ligt dan ook voor de hand dat dit in het geval van delinquenten niet anders zal zijn en daarom wordt geen al te hoge overeenstemming tussen delinquenten en reclasseringswerkers verwacht. In het onderzoek naar de congruente validiteit van de RISC-schalen 10, 11 en 12 wordt daarom uitgegaan van een correlatie (Pearson's *r*) van minstens 0,30 (de ondergrens voor een middelgroot effect, zie Cohen (1988)).

5.4.1 Verwachte samenhangen tussen RISC-schalen en schalen van NPV, UCL en BDHI-D

Op basis van de schaal *Emotioneel welzijn* (schaal 10) wordt in de RISC ingeschat hoe de cliënt in het leven staat, welk zelfbeeld hij heeft en of er mogelijk sprake is

²³ Plafond- of vloereffecten verwijzen naar een beperkte range aan scores op een bepaalde variabele (restriction of range). Een dergelijk beperkt scorebereik heeft tot gevolg dat correlaties tussen die variabele en andere variabelen worden verlaagd.

van psychische problemen. Deze schaal wordt daarom verwacht positief samen te hangen met de NPV-schalen *Inadequatie* en *Sociale inadequatie*. Deze schalen meten immers vage angsten en insufficiëntiegevoelens, respectievelijk het vermijden of het zich ongelukkig voelen in sociale contacten. Tevens wordt een negatieve samenhang verwacht met *Zelfwaardering* (NPV), een schaal die verwijst naar een positieve houding ten opzichte van werken, zichzelf en het leven. Omdat de schaal *Passief reactiepatroon* van de UCL betrekking heeft op de zaak somber inzien, zich piekerend in zichzelf terugtrekken en niet in staat zijn om iets aan de situatie te doen, wordt met deze schaal eveneens een positieve samenhang verwacht.

Schaal 11 *Denkpatronen, gedrag en vaardigheden* beoogt in kaart te brengen hoe de delinquent omgaat met zaken die hij in het dagelijks leven tegenkomt. De items van de schaal hebben onder meer betrekking op probleembesef en probleemhanteling. De verwachting is daarom dat deze schaal positief samenhangt met de ‘inadequater’ copingstrategieën *Palliatieve reactie*, *Vermijden* en *Passief reactiepatroon* (elk een UCL-schaal). Deze schalen hebben betrekking op copingstrategieën zoals het zoeken van afleiding en zich met andere dingen bezighouden om niet aan het probleem te hoeven denken, zaken op zijn beloop laten of afwachten wat er gaat gebeuren, het zich door een probleem in beslag laten nemen, piekeren en de zaak somber inzien. Tegelijkertijd wordt een negatief verband verwacht met de adequatere copingstrategie *Actief aanpakken* (UCL), een schaal die beoogt te meten in hoeverre iemand de situatie rustig van alle kanten bekijkt en doelgericht te werk gaat om problemen op te lossen. Omdat schaal 11 ook betrekking heeft op impulsiviteit en zelfbeheersing werd tevens een positieve correlatie met de BDHI-D-schaal *Directe agressie* (fysieke en verbale agressie) verwacht.

Op basis van de RISC-schaal *Houding* (schaal 12) wordt beoogd de houding van de delinquent te meten tegenover het delict, de maatschappij en criminaliteit in het algemeen. Een hoge score op deze schaal duidt op een pro-criminele attitude, een gebrek aan schuldgevoel, een afwijzen van de wetten en regels van de samenleving en een negatieve houding ten opzichte van de (op te leggen of opgelegde) sanctie. Op basis hiervan wordt verwacht dat deze schaal positief samenhangt met de NPV-schalen *Verongelijkheid* en *Zelfgenoegzaamheid* en met de schaal *Directe agressie* van de BDHI-D. De twee NPV-schalen meten kritiek hebben op en wantrouwen van andere mensen, respectievelijk een gevoel van tevredenheid met zichzelf en een desinteresse voor een ander en diens problemen. De BDHI-D-schaal *Directe agressie* heeft betrekking op agressie in de vorm van fysieke of verbale uitingen.

5.4.2 Resultaten congruente validiteit schalen 10, 11 en 12

In de tabellen 42, 43 en 44 worden de resultaten van de correlatieanalyses weergegeven die zijn uitgevoerd om de verwachtingen met betrekking tot respectievelijk schaal 10, 11 en 12 te toetsen (er is eenzijdig getoetst). Tevens wordt in elke tabel (onder de stippellijn) de samenhang getoond tussen de schalen van de gebruikte vragenlijsten. Omdat elke reclasseringscliënt slechts één vragenlijst invulde, kunnen deze correlaties alleen tussen schalen van hetzelfde instrument worden berekend. De reden om deze correlaties in de tabellen weer te geven, is dat tussen schalen die worden gebruikt om een andere schaal te valideren ook aanzienlijke samenhangen worden verwacht; ze beogen immers (delen van) hetzelfde concept te meten. Over het algemeen vertonen de verschillende NPV- en UCL-schalen de

verwachte middelsterke tot sterke samenhangen. De uitzonderingen worden gevormd door een aantal van de UCL-schalen die zijn gebruikt bij de validering van schaal 11 van de RISC. Met name de lage correlaties tussen enerzijds *Actief aanpakken* en anderzijds *Palliatieve reactie* en *Vermijden* vallen op. De gevonden correlaties onderbouwen in grote lijnen de vooraf geformuleerde verwachtingen. Schaal 10 *Emotioneel welzijn* vertoont met alle schalen die een soortgelijk begrip meten, de verwachte, middelsterke samenhang (tabel 42). Met de inadequatieschaal van de NPV vertoont schaal 10 zelfs een grote samenhang (0,50 is de ondergrens voor een groot effect (Cohen, 1988)). Ook schaal 11 *Denkpatronen, gedrag en vaardigheden* hangt in grote lijnen op de verwachte manier samen met de schalen die (delen van) hetzelfde begrip beogen te meten (tabel 43). De correlaties met de UCL-schalen *Palliatieve reactie* en *Vermijden* raken vrijwel de grenswaarde om van een middelsterke samenhang te kunnen spreken en de overige correlaties komen eveneens overeen met de vooraf opgestelde verwachtingen. Een aantal verwachtingen wordt echter niet ondersteund. Schaal 12 *Houding* vertoont niet de verwachte, middelsterke samenhang met de NPV-schalen *Verongelijkheid* en *Zelfgenoegzaamheid* (tabel 44). Integendeel, de samenhang tussen schaal 12 en de twee NPV-schalen is vrijwel nul. Er dient echter opgemerkt te worden dat de NPV-schaal *Zelfgenoegzaamheid* een zeer lage interne consistentie vertoont in de RISC-onderzoeksgroep. Het is daarom de vraag of deze schaal in deze onderzoeksgroep wel hetzelfde meet als in de onderzoeksgroepen waarover in de handleiding van het instrument wordt gerapporteerd. Bovendien vertoont schaal 12 wél de verwachte, middelsterke samenhang met de schaal *Directe agressie* van de BDHI-D. Desondanks kan naar aanleiding van deze resultaten de vraag gesteld worden in hoeverre schaal 12 daadwerkelijk de houding van reclasseringscliënten ten opzichte van andere mensen meet. Dat de schaal dit wel beoogt te doen, mag blijken uit de specifieke aandacht die volgens de handleiding bij het invullen van deze schaal moet worden besteed aan vijandigheid en het niet willen erkennen van de persoonlijke grenzen van anderen (bijvoorbeeld van slachtoffers).

Tabel 42 Correlaties tussen RISC-schaal 10 en instrumenten die hetzelfde begrip beogen te meten (NPV en UCL)

		Inadequatie	Sociale inadequatie	Zelfwaardering	Passief reactiepatroon
Schaal 10 Emotioneel welzijn	r	0,53***	0,37**	-0,45***	0,45***
	(CI) [†]	(0,32 – 0,69)	(0,13 – 0,57)	(-0,63 – -0,22)	(0,22 – 0,63)
	N	(61)	(61)	(61)	(60)
<hr/>					
Inadequatie	r	1,00			--
	(CI)				
	N	(61)			
Sociale inadequatie	r	0,58***	1,00		--
	(CI)	(0,39 – 0,73)			
	N	(61)	(61)		
Zelfwaardering	r	-0,54***	-0,50***	1,00	--
	(CI)	(-0,70 – -0,33)	(-0,67 – -0,28)		
	N	(61)	(61)	(61)	

[†] CI = 95% betrouwbaarheidsinterval

** p < .01

*** p < .001

Tabel 43 Correlaties tussen RISC-schaal 11 en instrumenten die hetzelfde begrip beogen te meten (NPV, UCL, BDHI-D)

		Palliatieve reactie	Vermijden	Passief reac- tiepatroon	Actief aan- pakken	Directe agressie
Schaal 11	r	0,28*	0,28*	0,45***	-0,31**	0,33**
	(CI) [†]	(0,03 – 0,50)	(0,03 – 0,28)	(0,22 – 0,63)	(-0,52 – -0,06)	(0,09 – 0,53)
	N	(60)	(60)	(60)	(60)	(64)
<hr/>						
Palliatieve reactie	r	1,00				--
	(CI)					
	N	(60)				
Vermijden	r	0,46***	1,00			--
	(CI)	(0,23 – 0,64)				
	N	(60)	(60)			
Passief reactiepatroon	r	0,26*	0,22 ^{n.s.}	1,00		--
	(CI)	(0,01 – 0,48)	(-0,04 – 0,45)			
	N	(60)	(60)	(60)		
Actief aanpakken	r	-0,11 ^{n.s.}	-0,11 ^{n.s.}	-0,30*	1,00	--
	(CI)	(-0,35 – 0,15)	(-0,35 – 0,15)	(-0,51 – -0,05)		
	N	(60)	(60)	(60)	(60)	

[†] CI = 95% betrouwbaarheidsinterval

^{n.s.} niet significant

* p < .05

** p < .01

*** p < .001

Tabel 44 Correlaties tussen RISC-schaal 12 en instrumenten die hetzelfde begrip beogen te meten (NPV en BDHI-D)

		Verongelijkheid	Zelfgenoegzaamheid	Directe agressie
Schaal 12 Houding	r	0,04	-0,07	0,37**
	(CI) [†]	(-0,21 - 0,29)	(-0,32 - 0,19)	(0,13 - 0,57)
	N	(61)	(61)	(62)
<hr/>				
Verongelijkheid	r	1,00		--
	(CI)			
	N	(61)		
Zelfgenoegzaamheid	r	0,45***	1,00	--
	(CI)	(0,22 - 0,63)		
	N	(61)	(61)	

[†] CI = 95% betrouwbaarheidsinterval

** p < .01

*** p < .001

De hier gerapporteerde resultaten moeten met de nodige voorzichtigheid worden geïnterpreteerd en kunnen niet worden beschouwd als 'het bewijs' dat de schalen 10, 11 en 12 al dan niet meten wat ze beogen te meten. Zoals eerder in dit hoofdstuk al is opgemerkt, zijn de onderzoeksgroepen klein. Dit betekent dat de resultaten onderhevig kunnen zijn aan steekproeffluctuaties en dat de resultaten weinig stabiel zijn. De 95%-betrouwbaarheidsintervallen van de correlaties (CI) geven dit ook aan: ze zijn zeer ruim. Daarnaast bleek bij de controle van de structuren van de gebruikte vragenlijsten dat de NPV-schalen *Zelfgenoegzaamheid* en *Zelfwaardering* reden tot zorg gaven (zie paragraaf 5.2.2). Een ander punt waarom de correlaties tussen RISC-schalen en de andere vragenlijsten voorzichtig moeten worden geïnterpreteerd, is dat de groep reclasseringscliënten die meewerkte aan dit onderzoek, in bepaalde opzichten significant verschilt van de reguliere reclasseringspopulatie. Naar verhouding participeerden meer vrouwen en autochtonen in het onderzoek naar de congruente validiteit (zie paragraaf 5.2.1). Tot slot moet hier nogmaals genoemd worden dat de keuze voor de validering van het *beoordelingsinstrument* RISC met behulp van *zelfbeoordelingsvragenlijsten* uit noodzaak is geboren. Het had de voorkeur verdiend de RISC te valideren met behulp van instrumenten die door dezelfde informant, in casu de reclasseringswerker, konden worden ingevuld.

Ondanks al deze kanttekeningen leveren de resultaten van dit onderzoek een redelijke indicatie voor de congruente validiteit van de RISC-schalen 10, 11 en 12. De RISC-schalen 10 en 11 vertonen middelsterke, positieve samenhangen met NPV-, UCL- en BDHI-D-schalen die soortgelijke concepten beogen te meten. Schaal 11 hangt daarnaast negatief samen met de UCL-schaal *Actief aanpakken*, een schaal die een tegengesteld begrip beoogt te meten. Deze resultaten wijzen erop dat deze schalen daadwerkelijk meten wat ze beogen te meten. Daarnaast kunnen de resultaten van dit onderzoek twijfels opwekken over de vraag of de RISC-schaal 12 er voldoende in slaagt de houding van de reclasseringscliënt ten opzichte van andere mensen, de maatschappij, het delict en criminaliteit in het algemeen in kaart te brengen.

6 RISc en specifieke doelgroepen

6.1 De relatie tussen de RISc en StatRec bij specifieke doelgroepen

Op verzoek van de Directie Sanctie- & Preventiebeleid van het Ministerie van Justitie is de oorspronkelijke onderzoeksvraag naar de relatie tussen de RISc en StatRec uitgebreid met de vraag naar deze relatie bij een aantal specifieke doelgroepen. Zoals in hoofdstuk 1 is toegelicht is dit gedaan naar aanleiding van een door de drie reclasseringsorganisaties uitgevoerde knelpuntenevaluatie (Bosker, 2006). Hieruit kwam naar voren dat de RISc bij een aantal specifieke doelgroepen tot een lagere inschatting van het recidiverisico lijkt te leiden dan reclasseringswerkers in de praktijk aannemelijk achten. Als dit daadwerkelijk het geval zou zijn, betekent dit dat de validiteit van het instrument voor deze doelgroepen onvoldoende zou zijn. Aparte aandacht is daarom besteed aan de congruente validiteit van de RISc in termen van de voorspelling van de recidivekans in deze groepen. Het gaat om de volgende doelgroepen: plegers van huiselijk geweld, zedendelinquenten, veelplegers, oudere daders, oplichters en bestuurders onder invloed. In deze paragraaf wordt beschreven welke definities zijn gehanteerd voor elk van deze groepen en wordt een aantal kenmerken van deze groepen beschreven.

Plegers van huiselijk geweld zijn geïdentificeerd op basis van de RISc. Eén van de RISc-variabelen biedt de reclasseringswerker de mogelijkheid om door het aanvinken van bepaalde omschreven categorieën meer informatie te geven over het delict of de delicten waar een cliënt van wordt verdacht of voor is veroordeeld. Indien de categorie 'Fysiek geweld jegens partner/iemand in de huiselijke kring' was aangevinkt, is een persoon beschouwd als een pleger van huiselijk geweld. De groep plegers van huiselijk geweld bestaat derhalve uit daders die ten tijde van de RISc-afname verdacht werden van of veroordeeld waren voor het plegen van huiselijk geweld. In de totale onderzoeksgroep ging het om 2.308 RISc's die zijn afgenomen bij plegers van huiselijk geweld. In totaal kon voor 2.110 van deze RISc's de StatRec-score worden berekend. De meeste van deze daders zijn mannen (94%) en autochtoon (65,4%) (zie tabel 45). De gemiddelde leeftijd van de groep plegers van huiselijk geweld ten tijde van de afname van de RISc was 36 jaar en zes maanden ($m = 36,53$; $sd = 11,30$; $range = 17,34 - 76,96$) en hun gemiddelde StatRec-score is 0,41 ($sd = 0,23$). Dit komt overeen met een voorspelde recidivekans van 41%.

Om te bepalen wie tot de groep zedendelinquenten hoort, is nagegaan of in de strafzaak naar aanleiding waarvan de RISc werd afgenomen, sprake was van een zedendelict. In tegenstelling tot de analyses in hoofdstuk 5 is hierbij niet alleen gekeken naar het hoogst bedreigde feit, maar naar alle feiten die in de zaak waren opgenomen. Iedereen met een zaak waarin een van de artikels 239 tot en met 250 uit het Wetboek van Strafrecht voorkomen, is als zedendelinquent beschouwd. Deze artikels hebben betrekking op schennis van de eerbaarheid, verspreiden of aanbieden van pornografische afbeeldingen, jeugdigen gebruiken voor pornografie, verkrachting, gemeenschap met een bewusteloze of geestelijk gestoorde, gemeenschap met een persoon beneden de 12, gemeenschap met een persoon tussen de 12 en 16 jaar, aanranding, ontucht en dwingen tot seksuele handelingen met een derde tegen betaling. Informatie over de feiten waarvoor een dader werd vervolgd was beschikbaar in het bestand dat voor de berekening van de StatRec-score is aangemaakt. Voor elk van de 856 RISc's waarbij voor een zedendelict werd ver-

volgd kon dan ook een StatRec-score worden berekend. Slechts een heel klein deel van deze daders bestaat uit vrouwen (2,1%) (zie tabel 45). Bijna driekwart van de daders is autochtoon (73,8%) en de gemiddelde leeftijd ten tijde van de RISC-afname was 39 jaar en 10 maanden ($m = 39,81$; $sd = 13,45$; range = 18,19 — 77,19). De gemiddelde recidivekans van deze groep, zoals berekend met StatRec, bedraagt 28% ($m = 0,28$; $sd = 0,21$)

Veelplegers plegen met grote regelmaat strafbare feiten. In dit onderzoek is iedereen met elf of meer eerdere zaken als veelpleger beschouwd²⁴. Deze informatie is gehaald uit het StatRec-bestand dat is aangemaakt voor het berekenen van de StatRec-score. Logischerwijs kon voor elk van de 2.154 RISC's die zijn afgenomen bij daders met elf of meer eerdere zaken een StatRec-score worden berekend. De meeste veelplegers zijn mannen (95,6%) en drie vijfde is autochtoon (zie tabel 45). De gemiddelde leeftijd van de groep veelplegers was op het moment van de afname van de RISC 36 jaar en vier maanden ($m = 36,33$; $sd = 9,55$; range = 18,26 — 73,30). De gemiddelde StatRec-score van deze groep is 0,79 ($sd = 0,13$), wat overeenkomt met een gemiddelde recidivekans van 79%. Niet onverwacht is deze kans voor deze groep het hoogst van de bestudeerde specifieke doelgroepen.

Om oudere daders te identificeren is er in het onderhavige onderzoek voor gekozen om de grens te leggen bij 50 jaar en 10 maanden (50,85 jaar). Tien procent van de RISC's in de totale database is namelijk afgenomen bij daders die ouder waren dan 50 jaar en 10 maanden. Statistisch gezien zijn deze mensen dus 'oude' delinquenten. Het gaat om 1.151 RISC's uit het StatRec-analysebestand. Van deze mensen hebben 975 mensen een StatRec-score. De meesten van deze 975 daders zijn mannen (88,3%). Bijna vier vijfde van de groep oudere daders is autochtoon (78,4%) en de gemiddelde leeftijd tijdens de RISC-afname was 58 jaar ($m = 58,03$; $sd = 6,00$; range = 50,86 — 78,37). De gemiddelde StatRec-recidivekans in deze groep is 24% ($m = 0,24$; $sd = 0,19$).

Tot de groep oplichters behoort iedereen met een zaak waarin een van de artikels 326 tot en met 329 uit het Wetboek van Strafrecht voorkomen. Dat wil zeggen dat mensen onder meer zijn vervolgd voor oplichting, flessentrekkerij, bedrog, oneerlijke mededinging, steekpenningen aannemen of betalen, bedrog met handelsnaam of merk. In de onderzoeksgroep kwamen alleen artikel 326 (oplichting) en 337 (bedrog met handelsnaam of merk) voor. Het gaat om 172 RISC's uit het StatRecbestand. Omdat ook in dit geval de informatie over het delict waarvoor werd vervolgd afkomstig is uit het bestand dat voor de berekening van de StatRec-score is aangemaakt, hebben alle oplichters een StatRec-score. Net als bij oudere daders bestaat een relatief groot deel van de groep oplichters uit vrouwen (16,3%) (zie tabel 45). 65,1% bestaat uit autochtonen en de gemiddelde leeftijd van de groep oplichters was, toen de RISC bij hen werd afgenomen, 35 jaar en zes maanden ($m = 35,54$; $sd = 10,86$; range = 18,77 — 67,87). De gemiddelde StatRec-score in deze groep is 0,55 ($sd = 0,25$). Dit komt overeen met een recidivekans van 55%.

Bestuurders onder invloed, ten slotte, zijn daders die vervolgd werden voor artikel 8 uit de Wegenverkeerswet (besturen onder invloed). Deze informatie was beschikbaar in het bestand dat voor de berekening van de StatRec-score is aangemaakt. Voor alle 166 RISC's die zijn afgenomen naar aanleiding van rijden onder invloed kon dan ook een StatRec-score worden berekend. Ook in deze groep bevinden zich

²⁴ Deze operationalisering maakt, in tegenstelling tot de definitie die in de WODC-Veelplegersmonitor wordt gebruikt, geen onderscheid tussen actieve en niet-actieve veelplegers. Omdat extra gegevens bij de operationalisering betrokken hadden moeten worden om dit onderscheid wel te maken, is uit praktische overwegingen voor de huidige operationalisering gekozen.

relatief vrij veel vrouwen (15,7%) (zie tabel 45). Ongeveer vier vijfde van de bestuurders onder invloed is autochtoon (81,3%) en de gemiddelde leeftijd ten tijde van de RISC-afname was 36 jaar en zeven maanden ($m = 36,63$; $sd = 11,39$; range = 18,36 — 71,52). De gemiddelde recidivekans zoals berekend met StatRec bedraagt voor de groep bestuurders onder invloed 48% ($m = 0,48$; $sd = 0,24$).

Tabel 45 Achtergrondkenmerken van specifieke doelgroepen

Kenmerk		Plegers van huiselijk geweld	Zedendelin- quenten	Veelplegers	Oudere daders	Oplichters	Bestuurders onder in- vloed
Omvang groep		2.110	856	2.154	975	172	166
Geslacht	Man	94,0%	97,9%	95,6%	88,3%	83,7%	84,3%
	Vrouw	6,0%	2,1%	4,4%	11,7%	16,3%	15,7%
Herkomstgroep	Autochtoon	65,4%	73,8%	60,1%	78,4%	65,1%	81,3%
	Niet-westers allochtoon	29,1%	21,4%	34,5%	13,9%	27,9%	16,9%
	Westers allochtoon	4,1%	4,4%	3,2%	6,8%	5,8%	1,8%
	Onbekend	1,4%	0,4%	2,1%	0,9%	1,1%	-
Gemiddelde leeftijd (in jaren)		36,53	39,81	36,33	58,03	35,54	36,63
Gemiddelde StatRec-score		0,41	0,28	0,79	0,24	0,55	0,48

6.2 Correlaties tussen RISC en StatRec bij specifieke doelgroepen

In tabel 46 wordt voor de verschillende groepen de samenhang tussen enerzijds de RISC-schalen en de RISC-totaalscore en anderzijds de StatRec-score gepresenteerd. De N die in de tabel staat vermeld bij de groepen (in de bovenste rij) geeft de maximale N weer. De interpretatie van de resultaten is op dezelfde manier gedaan als in hoofdstuk 5. Correlaties tot 0,30 zijn als zwak beschouwd, correlaties van 0,30 tot 0,50 als middelsterk en correlaties vanaf 0,50 zijn als sterk beschouwd (Cohen, 1988). Ook de verwachtingen waren gelijk aan die in hoofdstuk 5. De RISC-totaalscore werd verwacht in alle groepen een sterke samenhang te vertonen met de StatRec-voorspelling van de recidivekans. Ook schaal 1&2 *Delictgegevens* werd geacht sterk samen te hangen met de StatRec-score. Van de overige RISC-schalen werd verwacht dat zij een middelsterke samenhang zouden vertonen met de StatRec-recidivekans. Bij de interpretatie van de resultaten dient steeds rekening te worden gehouden met het feit dat StatRec de kans voorspelt dat een dader nogmaals met justitie in aanraking komt. Op de gevolgen hiervan voor de beantwoording van de vraag of de indruk van reclasseringswerkers correct is dat de RISC voor de onderzochte doelgroepen ten onrechte een lager recidiverisico voorspelt, wordt in hoofdstuk 7 teruggekomen.

De verwachtingen met betrekking tot de samenhang tussen de RISC-totaalscore en de StatRec-score worden in alle subgroepen bevestigd behalve in de groep veelplegers. De samenhang tussen deze variabelen is in deze groep middelsterk: $r = 0,45$. Wanneer het 95% betrouwbaarheidsinterval wordt berekend, blijkt dit zelfs significant af te wijken van de grens voor een sterke correlatie: 95% CI = 0,41 — 0,49. Ook de RISC-schalen vertonen in de groep veelplegers in veel gevallen een zwakke sa-

menhang met de StatRec-score. Een verklaring hiervoor kan liggen in het feit dat de spreiding van StatRec-scores in de groep veelplegers heel beperkt is ($m=0,79$; $sd=0,13$). De meeste veelplegers hebben dus een hoog recidiverisico waardoor het hebben van veel of weinig problemen op de RISC-leefgebieden niet sterk differentiëert naar recidiverisico.

Schaal 1&2 *Delictgegevens* vertoont in alle groepen, behalve de groep veelplegers, de verwachte sterke samenhang met de op basis van StatRec voorspelde recidivekans. Wat voorts opvalt is dat schaal 10 *Emotioneel welzijn* in geen enkele subgroep voldoet aan de verwachtingen: de samenhang tussen de scores op deze schaal en de StatRec-score is in elke subgroep zwak en in de groep bestuurders onder invloed zelfs negatief (95% CI = -0,28 — 0,02). Ook de schalen 6 *Relaties met partner, gezin en familie* en 9 *Alcoholgebruik* vertonen in geen van de groepen een middelsterke samenhang, maar deze samenhangen wijken in de groepen zedendelinquenten en oplichters niet significant af van de grenswaarde van een middelsterke correlatie. De zwakke samenhang tussen alcoholgebruik en de StatRec-score in de groep bestuurders onder invloed mag op het eerste gezicht wellicht verbazen, maar in deze groep blijkt vrijwel iedereen een hoge score te hebben op schaal 9. Met andere woorden, problemen op het gebied van alcoholgebruik differentiëren in deze groep niet tussen daders met een hoog en daders met een laag recidiverisico omdat bijna elke bestuurder onder invloed problemen heeft met betrekking tot alcoholgebruik.

Wanneer per subgroep wordt gekeken valt op dat binnen de groep plegers van huiselijk geweld schaal 6 *Relaties met partner, gezin en familie* een zwakke samenhang vertoont met de StatRec-score. Deze samenhang wijkt net niet significant af van de grens voor een middelsterke correlatie: 95%CI = 0,22 — 0,30. Hoewel plegers van huiselijk geweld gemiddeld significant hoger scoren op deze schaal dan daders van andere delicten²⁵, is de verdeling van de scores op deze schaal niet significant scheef of gepiekt. Dit kan dan ook geen verklaring zijn voor de zwakke samenhang tussen schaal 6 en de StatRec-voorspelling van het recidiverisico. Naast een lage correlatie met schaal 6 vertonen in deze groep ook de schalen 3 *Huisvesting en wonen*, 9 *Alcoholgebruik*, 10 *Emotioneel welzijn* en 12 *Houding* een zwakke samenhang met de StatRec-score. De schalen 4 *Opleiding, werk en leren*, 7 *Relaties met vrienden en kennissen* en 8 *Druggebruik* hebben binnen de groep plegers van huiselijk geweld de sterkste samenhang met de op basis van StatRec voorspelde recidivekans. In de groep zedendelinquenten vertonen de meeste dynamische RISC-schalen middelsterke, positieve samenhangen met de StatRec-score. Dit geldt echter niet voor de eerdergenoemde schalen 6, 9 en 10. Wat in deze groep verder opvalt is het relatieve belang van inkomen: wanneer de correlaties naar grootte worden geordend, is dit de op-één-na-sterkste correlatie (na schaal 1&2). Ook opleiding en werk en druggebruik nemen in deze groep een prominente plaats in. Omdat bij veelplegers sprake is van een zekere mate van restriction of range — de meeste veelplegers hebben een hoog StatRec-recidiverisico — moeten de correlaties tussen RISC en StatRec voor deze groep met de nodige terughoudendheid geïnterpreteerd worden. De enige dynamische RISC-schalen die, gegeven de hoge recidivekans van de meeste daders, de verwachte middelsterke samenhang met de StatRec-score vertonen zijn de schalen die betrekking hebben op opleiding en werk, inkomen en omgaan met geld, relaties met vrienden en kennissen en drug-

²⁵ Kolmogorov-Smirnov test: $Z=15,25$; $p<.001$. Vanwege heterogeniteit van variantie en omdat de omvang van beide groepen te zeer verschilde, was het niet mogelijk een ANOVA uit te voeren.

gebruik. De overige schalen hangen zwak of zelfs vrijwel niet (alcoholgebruik en emotioneel welzijn) samen met de recidivekans zoals op basis van StatRec is bekend.

In de groep oudere daders valt op dat een relatief groot aantal schalen zwak samenhangt met de StatRec-score. Naast de schalen 6, 9 en 10 vertonen ook schaal 3 *Huisvesting en wonen*, schaal 11 *Denkpatronen, gedrag en vaardigheden* en schaal 12 *Houding* een zwakke samenhang met de StatRec-recidivekans. Na schaal 1&2 *Delictgegevens* vertonen opleiding en werk, inkomen en omgaan met geld en druggebruik binnen deze groep de sterkste correlaties met de door StatRec voorspelde recidivekans.

In de groep oplichters hangt schaal 11 *Denkpatronen, gedrag en vaardigheden* sterk samen met de StatRec-voorspelling van de recidivekans. Deze samenhang is opmerkelijk genoeg bijna net zo sterk als de samenhang tussen StatRec en schaal 1&2 *Delictgegevens*. Naast schaal 1&2 en schaal 11 vertonen binnen deze groep relaties met vrienden en kennissen, opleiding en werk en houding de sterkste samenhangen met de StatRec-voorspelling van recidive. De overige correlaties zijn middelsterk of wijken niet significant af van de grens voor een middelsterke correlatie (6 *Relaties met partner, gezin en familie* en 9 *Alcoholgebruik*). Zelfs de samenhang tussen schaal 10 *Emotioneel welzijn* en StatRec wijkt niet significant af van 0,30. De groep oplichters is echter klein waardoor er sprake is van ruime betrouwbaarheidsintervallen.

De groep bestuurders onder invloed is, net als de groep oplichters, aan de kleine kant. Desondanks wijken de correlaties van StatRec met schaal 6 *Relaties met partner, gezin en familie*, 9 *Alcoholgebruik* en 10 *Emotioneel welzijn* significant af van de grens voor een middelsterke correlatie. Met betrekking tot alcoholgebruik is hiervoor al opgemerkt dat dit veroorzaakt wordt door de scheve verdeling van de scores op deze schaal. De zwakke samenhang tussen huisvesting en recidivekans wijkt niet significant af van 0,30 (95% CI = 0,11 — 0,40). Wat verder in deze groep opvalt is het belang van de schalen 12 *Houding* en 5 *Inkomen en omgaan met geld*. Wanneer de correlaties binnen deze groep naar grootte worden geordend, zijn dit na schaal 1&2 de schalen met de sterkste samenhang met de StatRec-voorspelling van het recidiverisico.

Binnen elke bestudeerde subgroep is een aantal opmerkelijke resultaten te benoemen, maar over het algemeen worden de verwachtingen met betrekking tot de samenhang tussen enerzijds RISC-schalen en RISC-totaalscore en anderzijds de StatRec-voorspelling van de recidivekans bevestigd. De belangrijkste uitzondering is schaal 10 *Emotioneel welzijn*. De samenhang tussen deze schaal en de StatRec-score is in alle subgroepen zwak en wijkt in bijna alle gevallen significant af van de grens voor een middelsterke correlatie (95% CI bij oplichters is 0,05 — 0,34). Ook de schalen 6 *Relaties met partner, gezin en familie* en 9 *Alcoholgebruik* vertonen in alle subgroepen zwakke correlaties met de door StatRec voorspelde recidivekans. De leefgebieden met de sterkste samenhang, na schaal 1&2 *Delictgegevens* zijn over het algemeen opleiding en werk, relaties met vrienden en kennissen en druggebruik, maar ook inkomen en omgaan met geld hangt in een aantal subgroepen in belangrijke mate samen met de StatRec-score. Onderzoek naar de predictieve validiteit van de RISC zal moeten uitwijzen in hoeverre deze bevindingen kunnen worden gerepliceerd met betrekking tot de samenhang met daadwerkelijke recidive.

Tabel 46 Correlaties tussen RISC-scores en StatRec voor verschillende doelgroepen^a

		Plegers van huiselijk geweld (N=2.110)	Zedendelinquenten ^b (N=856)	Veelplegers (N=2.154)	Oudere daders (N=975)	Oplichters (N=172)	Bestuurders onder invloed (N=166)
Schaal 1&2 Delictgegevens	r	0,62	0,59	0,44	0,55	0,63	0,63
	N	(2.084)	(598)	(1.686)	(760)	(148)	(154)
Schaal 3 Huisvesting en wonen	r	0,26	0,30	0,28	0,25	0,35	0,26
	N	(1.925)	(698)	(1.293)	(846)	(119)	(154)
Schaal 4 Opleiding, werk en leren	r	0,40	0,42	0,30	0,36	0,48	0,36
	N	(2.108)	(854)	(2.150)	(975)	(172)	(166)
Schaal 5 Inkomen en omgaan met geld	r	0,34	0,46	0,35	0,40	0,39	0,38
	N	(2.108)	(856)	(2.154)	(975)	(171)	(165)
Schaal 6 Relaties met partner, gezin en familie	r	0,26	0,28	0,06**	0,12	0,27	0,07 ^{n.s.}
	N	(2.109)	(855)	(2.149)	(971)	(172)	(166)
Schaal 7 Relaties met vrienden en kennissen	r	0,40	0,36	0,33	0,31	0,51	0,35
	N	(2.099)	(853)	(2.139)	(969)	(171)	(166)
Schaal 8 Druggebruik	r	0,39	0,41	0,33	0,45	0,36	0,37
	N	(1.662)	(640)	(2.098)	(723)	(154)	(145)
Schaal 9 Alcoholgebruik	r	0,22	0,29	-0,01 ^{n.s.}	0,26	0,29	0,04 ^{n.s.}
	N	(2.106)	(854)	(2.147)	(974)	(172)	(166)
Schaal 10 Emotioneel welzijn	r	0,18	0,15	0,05**	0,12	0,20**	-0,13*
	N	(2.107)	(854)	(2.147)	(974)	(171)	(166)
Schaal 11 Denkpatronen, gedrag en vaardigheden	r	0,31	0,31	0,17	0,25	0,57	0,34
	N	(2.105)	(853)	(2.149)	(974)	(172)	(166)
Schaal 12 Houding	r	0,24	0,33	0,12	0,25	0,42	0,42
	N	(2.106)	(841)	(2.142)	(962)	(172)	(166)
Totaalscore	r	0,57	0,58	0,45	0,52	0,69	0,54
	N	(2.074)	(597)	(1.676)	(753)	(148)	(154)

^a Tenzij anders vermeld zijn alle correlaties significant ($p < .001$, enkelzijdig).

^b De groep zedendelinquenten bestaat uit alle daders die wegens een zedendelict werden vervolgd, ook als dit binnen de strafzaak niet het feit met de hoogste strafdreiging was.

^{n.s.} niet significant

* $p < .05$

** $p < .01$ (in alle gevallen is enkelzijdig getoetst)

6.3 Multivariate analyses van de samenhang tussen de RISC-scores en StatRec bij specifieke doelgroepen

In tabel 47 zijn de resultaten van de regressieanalyses voor de specifieke doelgroepen weergegeven. Net als in hoofdstuk 5 zijn de RISC-schalen ten behoeve van deze regressieanalyses getransformeerd naar z-scores, opdat met behulp van *person mean substitution* ontbrekende schaalscores konden worden geïmputeerd (zie paragraaf 5.2.4). De analyses zijn op dezelfde manier uitgevoerd als in hoofdstuk 5: de StatRec-score is als afhankelijke variabele ingevoerd en de RISC-schaalscores (z-waarden) zijn, alle tegelijk, als onafhankelijke variabelen ingevoerd. Opnieuw is het niet reëel om te verwachten dat elke RISC-schaal, rekening houdend met de invloed van alle andere schalen, een unieke bijdrage levert aan de verklaring van de StatRec-voorspelling van recidive. Net als in hoofdstuk 5 zijn de resultaten van de regressieanalyses met name van belang omdat zij aangeven in hoeverre de StatRec-voorspelling door de RISC kan worden verklaard. De RISC wordt ook in de specifieke doelgroepen die in dit hoofdstuk worden bestudeerd, verwacht een aanzienlijk deel van de variantie van de StatRec-recidivekans te kunnen verklaren. In het kader van een exploratieve analyse van de vraag of er schalen zijn die ondanks de samenhang tussen de RISC-schalen een eigen bijdrage leveren aan de verklaring van de StatRec-score zijn grotere, positieve regressiecoëfficiënten geïnterpreteerd. Aan kleine regressiecoëfficiënten (tussen -0,10 en 0,10) en aan negatieve regressiecoëfficiënten wordt net als in hoofdstuk 5 geen aandacht besteed²⁶.

De RISC-schalen verklaren bij bestuurders onder invloed en bij oplichters de meeste variantie van de StatRec-score: 53% respectievelijk 52%. Dit zijn zeer aanzienlijke percentages. Ook in de groep plegers van huiselijk geweld en de groep zedendelinquenten verklaart de RISC een aanzienlijk deel van de variantie van de StatRec-voorspelling van het recidiverisico (respectievelijk 43% en 41%). De RISC-schalen verklaren de minste variantie van de StatRec-score bij de groep veelplegers: 28% van de variantie van de StatRec-score wordt verklaard door de RISC. Hoewel dit overeenkomt met een sterke correlatie van 0,53 en ook 28% verklaarde variantie aanzienlijk is, is het beduidend lager dan in de andere groepen. De RISC lijkt hiermee voor veelplegers minder geschikt om het recidiverisico te voorspellen. Dit kan echter niet met zekerheid worden geconcludeerd. Ten eerste heeft de StatRec betrekking op de voorspelling van recidive en kon in dit onderzoek niet worden nagegaan in hoeverre de RISC daadwerkelijke recidive voorspelt. Ten tweede vertoont de StatRec-score van veelplegers vrij weinig variantie. De verwachting is op basis van StatRec echter dat 79% van de veelplegers in de eerste vier jaar na de uitgangszaak naar aanleiding waarvan de RISC is afgenomen opnieuw met justitie in aanraking komt. Mogelijkerwijs voorspelt de RISC voor bepaalde veelplegers een te laag recidiverisico. Van de veelplegers met een geldige RISC-totaalscore heeft 6,7% een laag risico, 47,5% een middelhoog recidiverisico en 45,8% een hoog risico. De veelplegers met een lage RISC-totaalscore (N = 113) hebben een gemiddelde StatRec-score van 0,66. De veelplegers met een middelhoge RISC-totaalscore (N = 796) hebben een gemiddelde StatRec-score van 0,76. Hoewel discussie mogelijk is over de vraag wat als laag of middelhoog recidiverisico beschouwd moet worden, ligt het niet erg voor de hand om een recidiverisico van 66% als laag en een recidiverisico van 76% als middelhoog te beschouwen. Dit onderbouwt het vermoeden van reclasseringswerkers dat RISC het recidiverisico bij veelplegers niet correct inschat.

²⁶ Regressiecoëfficiënten tussen -0,10 en 0,10 worden niet in de tabellen gerapporteerd.

Hierbij dient echter opgemerkt te worden dat de StatRec-recidivekans betrekking heeft op het opnieuw in aanraking komen met justitie. In hoofdstuk 7 wordt hier op teruggekomen.

In elke doelgroep levert schaal 1&2 *Delictgegevens*, zoals verwacht mocht worden, de sterkste, unieke bijdrage aan de verklaring van de StatRec-voorspelling van de recidivekans. Binnen de groep plegers van huiselijk geweld levert uitsluitend de dynamische RISC-schaal 4 *Opleiding, werk en leren* een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-score. Dit effect bestaat ondanks de onderlinge samenhang met de andere RISC-schalen. Ook in de groep zedendelinquenten levert deze schaal een unieke bijdrage aan de verklaring van de door StatRec voorspelde recidivekans. Daarnaast leveren in deze groep de schalen 5 *Inkomen en omgaan met geld* en 9 *Alcoholgebruik* een noemenswaardige, unieke bijdrage. Uit de vergelijking van de groep zedendelinquenten met alle overige daders blijkt dat het effect van inkomen voor de groep zedendelinquenten significant sterker is dan voor de overige delinquenten ($z = 4,08$). De bijdrage die schaal 1&2 *Delictgegevens* levert aan de verklaring van de StatRec-score is voor zedendelinquenten minder sterk dan voor daders van overige delicten ($z = -2,30$). Deze resultaten komen overeen met de resultaten die werden gevonden in de groep zedendelinquenten die in hoofdstuk 5 werd bestudeerd.

In de groep veelplegers levert een relatief groot aantal dynamische RISC-schalen een noemenswaardige, unieke bijdrage aan de verklaring van de op basis van StatRec ingeschatte recidivekans. Het gaat om de leefgebieden opleiding en werk, inkomen en omgaan met geld, relaties met vrienden en kennissen en druggebruik. In vergelijking met niet-veelplegers blijkt dat het effect van schaal 1&2 *Delictgegevens* voor veelplegers minder sterk is ($z = -15,64$). Ook de effecten van opleiding en relaties met vrienden zijn minder sterk voor veelplegers dan voor andere daders (z is respectievelijk $-4,50$ en $-9,05$). Nogmaals wordt benadrukt dat de spreiding in StatRec-scores bij veelplegers erg beperkt is waardoor deze resultaten met de nodige terughoudendheid moeten worden geïnterpreteerd.

In de groep oudere daders leveren drie dynamische RISC-schalen een noemenswaardige, eigen bijdrage aan de verklaring van de voorspelde StatRec-recidivekans. Het gaat om de schalen die betrekking hebben op opleiding en werk, inkomen en omgaan met geld en alcoholgebruik. De oudere daders verschillen van daders die jonger zijn dan 50 jaar en tien maanden met betrekking tot het effect van schaal 5 *Inkomen en omgaan met geld*. Het effect is voor oudere daders sterker dan voor jongere daders ($z = 3,19$).

Voor de groep oplichters is een aantal niet significante effecten in tabel 47 weergegeven. Het gaat om de schalen 4 *Opleiding, werk en leren* en 7 *Relaties met vrienden en kennissen*. Of deze effecten in een grotere groep oplichters wel significant zijn, zou moeten blijken uit een herhaling van dit onderzoek onder een grotere groep oplichters. Wat opvalt is de sterke bijdrage die schaal 11 *Denkpatronen, gedrag en vaardigheden* levert aan de verklaring van de StatRec-voorspelling van het recidiverisico. In geen enkele andere groep, ook niet in de groepen die in hoofdstuk 5 werden bestudeerd, is de bijdrage van schaal 11 zo sterk. Het verschil tussen oplichters en plegers van andere delicten is echter niet significant ($z=1,58$). Naast schaal 11 levert ook de dynamische RISC-schaal 9 *Alcoholgebruik* een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-score in de groep oplichters. Binnen de groep bestuurders onder invloed levert geen van de dynamische criminogene factoren een significante noemenswaardige bijdrage aan de verklaring van de StatRec-score. Net als bij oplichters is het effect van opleiding en werk welis-

waar in de tabel vermeld, maar niet significant. Nader onderzoek zal moeten uitwijzen of dit leefgebied in een grotere groep bestuurders onder invloed wel een significante unieke bijdrage levert aan de verklaring van de StatRec-score.

De RISc blijkt in de onderscheiden specifieke doelgroepen de StatRec-voorspelling van de recidivekans goed te kunnen verklaren. Alleen in de groep veelplegers is het percentage verklaarde variantie in vergelijking met de andere groepen aan de lage kant. Desondanks kan ook voor deze groep van een sterk effect worden gesproken. Van de dynamische criminogene factoren levert opleiding en werk in de meeste groepen een noemenswaardige unieke bijdrage. Daarnaast is er een aantal verschillen tussen de groepen. In hoofdstuk 7 wordt hierop teruggekomen.

Tabel 47 Regressieanalyse van RISC-schalen op StatRec voor specifieke doelgroepen^a

	Plegers van huiselijk ge- weld	Zedendelin- quenten ^b	Veelplegers	Oudere daders	Oplichters	Bestuurders onder invloed
N	2.109	854	2.154	974	172	166
R ²	0,43	0,41	0,28	0,34	0,53	0,52
β (gestandaardiseerd)						
Schaal 1&2 Delictgegevens	0,51***	0,35***	0,31***	0,30***	0,34***	0,50***
Schaal 3 Huisvesting en wonen						
Schaal 4 Opleiding, werk en leren	0,14***	0,15***	0,11***	0,15***	0,13 ^{n.s.}	0,11 ^{n.s.}
Schaal 5 Inkomen en omgaan met geld		0,19***	0,13***	0,18***		
Schaal 6 Relaties met partner, gezin en familie						-0,11 ^{n.s.}
Schaal 7 Relaties met vrienden en kennissen			0,12***		0,10 ^{n.s.}	
Schaal 8 Druggebruik			0,12***			
Schaal 9 Alcoholgebruik		0,12***		0,14***	0,12*	
Schaal 10 Emotioneel welzijn	-0,11***				-0,12 ^{n.s.}	-0,26***
Schaal 11 Denkpatronen, gedrag en vaardigheden					0,23*	
Schaal 12 Houding			-0,13***			

^a Schaalscores zijn getransformeerd naar z-scores

^b De groep zedendelinquenten bestaat uit alle daders die wegens een zedendelict werden vervolgd, ook als dit binnen de strafzaak niet het feit met de hoogste strafdreiging was.

^{n.s.} niet significant

* p < .05

** p < .01

*** p < .001

7 Conclusies en aanbevelingen

7.1 De interbeoordelaarsbetrouwbaarheid, interne consistentie en congruente validiteit van de RISC

In drie deelstudies heeft het WODC in de periode van september 2005 tot maart 2007 onderzoek uitgevoerd naar de interbeoordelaarsbetrouwbaarheid, structuur en congruente validiteit van de RISC. Het algemene beeld dat uit de resultaten van de drie deelstudies naar voren komt, is vrij positief. De interbeoordelaarsbetrouwbaarheid van het instrument is voldoende. Met betrekking tot de items van de RISC is de mate van overeenstemming tussen reclasseringswerkers over het algemeen redelijk tot sterk en ook op schaalniveau is de overeenstemming tussen de beoordelaars in alle gevallen redelijk tot sterk. Met betrekking tot de totaalscore van de RISC komen reclasseringswerkers in sterke mate tot hetzelfde oordeel. Wel valt op dat de interbeoordelaarsbetrouwbaarheid van items afneemt naarmate minder naar feitelijkheden wordt gevraagd en dat een groot aantal items van de schalen 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding* slechts een matige overeenstemming laat zien. Om de interbeoordelaarsbetrouwbaarheid van de RISC op punten te verbeteren kan een aantal aanbevelingen worden gedaan. Deze worden in de volgende paragraaf uiteengezet.

De RISC-schalen vormen over het algemeen goede schalen: de meeste items van elke schaal meten ieder een deel van hetzelfde achterliggende concept dat de schaal in kaart beoogt te brengen en de scores zijn in de meeste schalen voldoende repliceerbaar over de items. Dat wil zeggen dat met betrekking tot de meeste schalen sprake is van een betrouwbare meting. Dit blijkt uit de interne consistentie van de schalen, die voor de meeste schalen voldoende tot goed is. De uitzonderingen worden gevormd door de schalen 6 *Relaties met partner, gezin en familie* en 9 *Alcoholgebruik*. Deze schalen bestaan uit items die in onvoldoende mate eenzelfde achterliggend begrip in kaart brengen. Aanbevelingen om deze schalen te verbeteren worden in paragraaf 7.3 gedaan. Daarnaast worden suggesties gedaan om de schalen 1&2 *Delictgegevens* en 8 *Druggebruik* te verbeteren. Aan de interne consistentie van de totaalscore worden strengere eisen gesteld omdat op basis van deze score belangrijkere beslissingen over individuele daders worden genomen dan op basis van de afzonderlijke schaalcores. De betrouwbaarheidsanalyse van de totaalscore toont aan dat aan deze strengere eisen wordt voldaan.

Het onderzoek naar de congruente validiteit van de RISC in termen van de voorspelling van de recidivekans levert eveneens gunstige resultaten op. De samenhang tussen de RISC-totaalscore en de StatRec-score is in de totale onderzoeksgroep sterk: de correlatie is 0,64. Dit betekent dat er een sterk verband bestaat tussen de RISC-totaalscore, die een inschatting beoogt te geven van de recidivekans, en de gevalideerde voorspelling van de recidive zoals berekend met StatRec. Alle correlaties tussen de afzonderlijke RISC-schalen en de StatRec-recidivekans zijn in de totale onderzoeksgroep positief. Ze zijn echter niet allemaal zo sterk als werd verwacht op basis van het uitgangspunt van de RISC dat elke schaal verband houdt met (het risico op) delictgedrag. Dit geldt met name voor de schalen 6 *Relaties met partner, gezin en familie* en 10 *Emotioneel welzijn*, maar ook de correlatie tussen schaal 9 *Alcoholgebruik* en de StatRec-voorspelling van recidive is zwak. De samenhang tussen RISC en StatRec is ook in verschillende subgroepen bestudeerd. Over het alge-

meen werd het beeld bevestigd dat uit de analyses van de totale onderzoeksgroep naar voren kwam. Naast het berekenen van correlaties zijn regressieanalyses uitgevoerd om na te gaan in hoeverre de score op de StatRec te verklaren is op basis van de RISC-schalen. Hieruit bleek dat de RISC-schalen samen 48% van de variantie van de StatRec-recidivekans verklaren. Dit is een zeer aanzienlijk deel en deze bevinding ondersteunt dan ook de congruente validiteit van de RISC. Ook in vrijwel alle bestudeerde subgroepen — mannen, vrouwen, verschillende leeftijdsgroepen, herkomstgroepen, delictgroepen en de groepen die als specifieke doelgroepen werden bestudeerd — slaagt de RISC er goed in de StatRec-score te verklaren. Uitzonderingen hierop worden gevormd door vrouwen en veelplegers. Hoewel de congruente validiteit van de RISC voor vrouwen niet in het geding is — de RISC verklaart ook in deze groep een aanzienlijk deel van de variantie van de StatRec-score — valt op dat de mate waarin de RISC de StatRec-score in deze groep verklaart aanzienlijk minder groot is dan in de andere groepen. Met betrekking tot veelplegers zijn er aanwijzingen dat de RISC een te laag recidiverisico voorspelt. Voorts valt te concluderen dat er tussen de bestudeerde subgroepen verschillen bestaan in de aard van de dynamische factoren die, ondanks de onderlinge samenhang tussen de RISC-schalen, een unieke bijdrage leveren aan de verklaring van de StatRec-voorspelling van de recidivekans. Ook bestaan er tussen diverse groepen verschillen in de sterkte van de effecten van RISC-schalen op de verklaring van de StatRec-score. Het is duidelijk dat voor sommige groepen de invloed van bepaalde RISC-schalen sterker of minder sterk is dan voor andere groepen. Op deze conclusies wordt in paragraaf 7.4 nader ingegaan.

De congruente validiteit van de schalen 10 *Emotioneel welzijn*, 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding* is niet alleen in termen van de voorspelling van de recidivekans onderzocht, maar tevens inhoudelijk in termen van de begrippen die ze beogen te meten. Met behulp van drie persoonlijkheidsvragenlijsten — de Nederlandse Persoonlijksheidsvragenlijst (NPV), de Utrechtse Copinglijst (UCL) en de Buss-Durkee Hostility Inventory-Dutch (BDHI-D) — is nagegaan in hoeverre de drie RISC-schalen correleren met instrumenten die (delen van) hetzelfde concept beogen te meten. Hoewel er een aantal kanttekeningen is te plaatsen bij dit deel van het onderzoek — een kleine onderzoeksgroep die niet representatief is voor de gehele reclasseringspopulatie en het feit dat een beoordelingsinstrument wordt gevalideerd met behulp van zelfbeoordelingsvragenlijsten — kan toch gesteld worden dat de resultaten in grote lijnen een gunstig beeld geven van de congruente validiteit van de schalen 10 en 11. Schaal 10 *Emotioneel welzijn* vertoont met alle schalen die een soortgelijk begrip meten, de verwachte, middelsterke samenhang. Ook schaal 11 *Denkpatronen, gedrag en vaardigheden* hangt in grote lijnen op de verwachte manier samen met de schalen die (delen van) hetzelfde begrip beogen te meten. Schaal 12 *Houding* vertoont echter niet de verwachte, middelsterke samenhang met de NPV-schalen *Verongelijkheid* en *Zelfgenoegzaamheid*. Naar aanleiding hiervan zou de vraag gesteld kunnen worden in hoeverre schaal 12 daadwerkelijk de houding van reclasseringscliënten ten opzichte van andere mensen, de maatschappij, het delict en criminaliteit in het algemeen meet. Tegelijkertijd moet echter rekening worden gehouden met het feit dat er naast de genoemde kanttekeningen, twijfels zijn over de vraag of de NPV-schaal *Zelfgenoegzaamheid* in deze onderzoeksgroep wel hetzelfde meet als in de onderzoeksgroepen waarover in de handleiding van het instrument wordt gerapporteerd. Bovendien vertoont schaal 12 wel de verwachte, middelsterke samenhang met de BDHI-D-schaal *Directe agressie*.

Naar aanleiding van de deelstudies naar de interbeoordelaarsbetrouwbaarheid en de structuur van de RISC kan een aantal aanbevelingen worden gedaan om het instrument verder te ontwikkelen en te verbeteren. In de volgende twee paragrafen worden deze per deelonderzoek besproken. In paragraaf 7.4 zal iets dieper ingegaan worden op de conclusies die uit de studie naar de congruente validiteit van de RISC kunnen worden getrokken en in de slotparagraaf wordt getracht de bevindingen uit de drie deelstudies te integreren.

7.2 Aanbevelingen met betrekking tot de interbeoordelaarsbetrouwbaarheid van de RISC

De interbeoordelaarsbetrouwbaarheid van de verschillende RISC-onderdelen — items, schalen en de totaalscore — is over het algemeen redelijk tot sterk. Op basis van de resultaten van het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISC is een aantal aanbevelingen te doen die de interbeoordelaarsbetrouwbaarheid van het instrument verder kunnen verbeteren.

1 Formuleer instructies nog eenduidiger.

Ambivalente vraagstellingen of instructies, die onbedoeld ruimte laten voor eigen interpretatie door degene die de test invult, zijn onwenselijk in elk (psychologisch) instrument omdat zij de betrouwbaarheid van het instrument ondermijnen. De noodzaak van eenduidige instructies voor de RISC blijkt onder meer uit de interbeoordelaarsbetrouwbaarheid van de vraag of de cliënt is veroordeeld. Hoewel de overeenstemming tussen de beoordelaars op deze vraag sterk is, was minstens een zeer sterke overeenstemming verwacht. Er wordt immers gevraagd naar een puur feitelijk gegeven. Zoals in hoofdstuk 3 al is opgemerkt, is het mogelijk dat er sprake is van een verschil in interpretatie door de reclasseringswerker wanneer hoger beroep is aangetekend; de ene werker kan van mening zijn dat een cliënt is veroordeeld — de rechter heeft immers uitspraak gedaan — terwijl de tweede werker de mening is toegedaan dat de cliënt nog niet is veroordeeld zolang het hoger beroep nog loopt. Een eenduidiger instructie bij dit item kan een mogelijk verschil in interpretatie verhelpen en tot een hogere interbeoordelaarsbetrouwbaarheid leiden. Dat een goede interbeoordelaarsbetrouwbaarheid ook van belang is voor items die niet meetellen in de berekening van een schaalscore, moge blijken uit het feit dat voor een dader die ontkent en niet veroordeeld is, geen RISC-totaalscore wordt berekend. De vraag of een cliënt ontkent is daarmee ook een van de items die kunnen profiteren van een eenduidiger instructie.

2 Pas de antwoordschalen aan op de inhoud van het item (vermijd tweepuntsschalen bij moeilijker te beoordelen items).

Bij een aantal items in de RISC is gebruik gemaakt van een antwoordschaal met twee antwoordcategorieën. Veelal gaat het hierbij in feite om een item waar met ja of nee op kan worden geantwoord, maar in een aantal gevallen lijkt de tweepuntsschaal de reclasseringswerker te weinig ruimte voor nuancering te bieden. Dit lijkt in die gevallen een matige overeenstemming in de hand te werken. De reclasseringswerker wordt in feite gedwongen te kiezen tussen 'er zijn geen problemen' en 'er zijn ernstige problemen' waardoor bij twijfel de ene werker een andere afwijking maakt dan de andere. Een van de items waar zich dit bij voordoet is het item

6.5 *Huiselijk geweld*. Beoordelaars kunnen voor dit item kiezen tussen ‘er is geen enkele aanleiding om te denken dat...’ en ‘er is objectief bewijs’. De verwachting is dat het toevoegen van een derde antwoordcategorie tot een betere interbeoordelaarsbetrouwbaarheid leidt. Een andere mogelijkheid zou kunnen zijn om twee antwoordcategorieën te handhaven, maar de tweede antwoordcategorie uit te breiden opdat deze ook in geval van twijfel gescoord kan worden.

3 *Besteed nog meer aandacht aan opleiding en deskundigheidsbevordering voor reclasseringswerkers.*

Reclasseringswerkers zijn professionals en werkers die de RISC afnemen, worden getraind in het gebruik van het instrument. Toch wordt op basis van de bevindingen uit het onderzoek naar de interbeoordelaarsbetrouwbaarheid van de RISC aanbevolen nog meer aandacht te besteden aan opleiding en deskundigheidsbevordering. De aanleiding om deze aanbeveling te doen is tweeledig. Enerzijds werd bij een aantal heel feitelijke items een lagere interbeoordelaarsbetrouwbaarheid geconstateerd dan op basis van de inhoud van de items werd verwacht. Zo verbaasde de redelijke, maar toch onverwacht lage overeenstemming tussen beoordelaars op de vraag of er sprake is van diverse en/of zware delicten in de justitiële voorgeschiedenis van de dader (item 1.9). Deze informatie kan worden nagezocht in de door de werker op te vragen justitiële documentatie van de cliënt en de instructie van het item is heel eenduidig over de vraag wat onder een zwaar delict moet worden verstaan. Anderzijds zijn veel items echter lastig in te schatten en vragen de RISC-items een grote mate van deskundigheid bij de reclasseringswerker. Hoe moet bijvoorbeeld beoordeeld worden of problemen die op een bepaald leefgebied aanwezig zijn, verband houden met het delictgedrag? De interbeoordelaarsbetrouwbaarheid van een aantal van dergelijke vragen, zoals de vraag of er een relatie is tussen opleiding en werk en delictgedrag, is matig terwijl dit soort items van groot belang zijn voor de vraag of in het kader van reclasseringstoezicht of reïntegratie gewerkt zal gaan worden aan deze problemen. Als een probleem geen verband houdt met het risico op recidive, acht justitie het immers niet haar taak om te helpen bij het oplossen van dat probleem. Andere voorbeelden van items met een matige interbeoordelaarsbetrouwbaarheid of een interbeoordelaarsbetrouwbaarheid die lager is dan verwacht en waarbij deskundigheidsbevordering mogelijk van nut kan zijn voor het vergroten van de interbeoordelaarsbetrouwbaarheid, hebben betrekking op de vraag of iemand verantwoordelijkheid neemt voor zijn of haar delictgedrag en de vraag of een cliënt drugs gebruikt.

4 *Pak de matige interbeoordelaarsbetrouwbaarheid van de items van de schalen 11 en 12 aan.*

Hoewel de interbeoordelaarsbetrouwbaarheid van de schaalscores van de schalen 11 *Denkpatronen, gedrag en vaardigheden* en 12 *Houding* redelijk is, hebben veel items van deze schalen een matige interbeoordelaarsbetrouwbaarheid. Omdat veel belang wordt gehecht aan deze schalen is een goede interbeoordelaarsbetrouwbaarheid, ook van de items, essentieel. Zeven van de acht te scoren items van schaal 11 en drie van de vijf te scoren items van schaal 12 hebben een matige interbeoordelaarsbetrouwbaarheid. Verschillende mogelijkheden lijken te bestaan voor het verbeteren van de interbeoordelaarsbetrouwbaarheid van deze items. Op de eerst plaats kan in het kader van deskundigheidsbevordering speciale aandacht worden besteed aan het scoren van de schalen 11 en 12. Daarnaast is het laten be-

studeren van de scoringsinstructies door ervaren gedragsdeskundigen (bijvoorbeeld klinisch psychologen) wellicht een optie.

7.3 Aanbevelingen met betrekking tot de structuur van de RISc

Op basis van de resultaten van het onderzoek naar de structuur van de RISc is geconcludeerd dat de schaalstructuur en de interne consistentie van de RISc over het algemeen goed te noemen zijn. Een paar items gedragen zich echter niet zoals gewenst: ze blijken niet te passen binnen de schalen waar ze bij horen. Het gaat om de items 2.10 *Verantwoordelijkheid nemen voor het delictgedrag*, 6.4 *Familie- of gezinslid heeft een justitieel dossier*, 8.5 *Motivatie om druggebruik aan te pakken* en 9.5 *Motivatie om alcoholgebruik aan te pakken*. Elk van deze items meet een ander concept dan datgene wat de schaal waar ze bij horen beoogt te meten. Verantwoordelijkheid nemen voor delictgedrag gaat niet per definitie samen met een beperkt of weinig ernstig delictverleden. Warme relaties met partner en gezin sluiten niet uit dat een familie- of gezinslid in het verleden met justitie in aanraking is geweest en de motivatie om drug- of alcoholgebruik aan te pakken vertoont geen direct verband met de ernst van het middelengebruik. De kwaliteit van de schalen waar het om gaat, zou verbeteren als deze items niet meer in de berekening van de schaal score zouden worden betrokken — voor schaal 9 *Alcoholgebruik* zou het zelfs het verschil uitmaken tussen een onvoldoende en een voldoende interne consistentie. Dit is dan ook de aanbeveling die op basis van het onderzoek naar de structuur van de RISc wordt gedaan. Hiermee is evenwel niet gezegd dat de items in kwestie uit de RISc verwijderd zouden moeten worden.

Indien men om inhoudelijke redenen vast wil houden aan item 2.10 valt aan te raden het item bij schaal 12 *Houding* onder te brengen. Het item draagt niet echt bij aan de goede interne consistentie van deze schaal, maar beïnvloedt deze ook niet negatief en het heeft een goede itemrestcorrelatie. Het item is in deze schaal op zijn plek. Hoewel het valt af te raden de items 8.5 en 9.5 nog langer bij de berekening van de scores van hun respectievelijke schalen te betrekken, is er niets op tegen om ze in de RISc te handhaven als aparte indicatoren voor de motivatie van de cliënt om zijn of haar middelengebruik aan te pakken. Wel verdient het aanbeveling een paar extra items op te nemen die hetzelfde construct beogen te meten om zodoende een intern consistente ‘motivatieschaal’ te construeren. Het inschatten van motivatie op basis van een enkel item is namelijk een hachelijke onderneming. Met betrekking tot item 6.4, ten slotte, wordt aanbevolen het uit schaal 6 te verwijderen. Om inhoudelijke redenen zou ervoor gekozen kunnen worden het item als aparte indicator wel in de RISc te behouden. Het wordt immers geacht een samenhang te vertonen met de kans op recidive. De correlatie tussen dit item en de StatRec-voorspelling van de recidivekans is echter laag ($r = 0,26$; $N = 9.274$). Prospectief onderzoek naar de predictieve validiteit van de RISc zal uitsluitend moeten geven over de werkelijke waarde van dit item, maar met de kwaliteit van de relaties met partner, gezin en familie heeft het niets te maken.

Met betrekking tot de totaalscore worden naar aanleiding van het huidige onderzoek geen aanbevelingen gedaan. Hoewel verbeteringen wellicht mogelijk zijn, wordt voorgesteld eerst op schaalniveau aanpassingen door te voeren voordat iets aan de samenstelling van de totaalscore wordt veranderd. Vervolgens zouden opnieuw analyses op de totaalscore moeten worden uitgevoerd voordat eventueel besloten kan worden aan de berekening hiervan iets te veranderen.

7.4 Conclusies naar aanleiding van het validiteitsonderzoek van de RISC

7.4.1 *Congruente validiteit in termen van recidivevoorspelling*

Zoals in paragraaf 7.1 al is geconcludeerd hangt de RISC-totaalscore zowel in de totale onderzoeksgroep als in de bestudeerde subgroepen in sterke mate samen met de StatRec-voorspelling van de recidivekans. Uit de resultaten van de regressieanalyses blijkt dat de score op de StatRec in de totale onderzoeksgroep en in vrijwel alle onderzochte subgroepen goed te verklaren is op basis van de scores op de elf RISC-schalen tezamen. Deze resultaten vormen een overtuigende onderbouwing voor de congruente validiteit van de totaalscore van de RISC. In de groep vrouwen slaagt de RISC er weliswaar minder goed in om de StatRec-recidivekans te verklaren, maar ook in deze groep wordt de congruente validiteit van de RISC ondersteund.

Tussen de RISC-schalen die dynamische criminogene factoren in kaart beogen te brengen, en de StatRec-score werd een middelsterke samenhang verwacht. In de meeste gevallen wordt deze verwachting zowel in de totale onderzoeksgroep als in de bestudeerde subgroepen door de resultaten onderbouwd. Uitzonderingen doen zich voornamelijk voor met betrekking tot de schalen 6 *Relaties met partner, gezin en familie*, 9 *Alcoholgebruik* en 10 *Emotioneel welzijn*. Terwijl de schalen 6 en 9 in een aantal onderzochte subgroepen nog (bijna) middelsterke correlaties vertonen met StatRec, is de samenhang tussen emotioneel welzijn en het risico op recidive zoals gemeten met StatRec in alle subgroepen zwak. De conclusie is dat de congruente validiteit van de afzonderlijke RISC-schalen in termen van de relatie met de recidivekans over het algemeen duidelijk onderbouwd kon worden, maar dat dit niet geldt voor met name schaal 10 *Emotioneel welzijn*, en in iets mindere mate ook voor schaal 6 *Relaties met partner, gezin en familie*. Daarnaast vertoont in de groep vrouwen slechts een klein deel van de dynamische RISC-schalen de verwachte middelsterke samenhang met de StatRec-voorspelling van de recidivekans. Hoewel het, gezien de onderlinge samenhang tussen de afzonderlijke RISC-schalen, niet reëel was om te verwachten dat van elke afzonderlijke schaal de unieke bijdrage aan de verklaring van de StatRec-recidivekans kon worden bepaald, zijn enkele opmerkelijke resultaten gevonden. De schalen 4 *Opleiding, werk en leren* en 7 *Relaties met vrienden en kennissen* leveren in de totale onderzoeksgroep een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-score. De invloed van opleiding en vrienden en kennissen op de verklaring van de op basis van StatRec berekende recidivekans is ook in veel van de onderzochte subgroepen gevonden — zowel de groepen waarover in hoofdstuk 5 wordt gerapporteerd als die uit hoofdstuk 6. Bovendien levert druggebruik (schaal 8) in acht van de zeventien in hoofdstuk 5 bestudeerde subgroepen een noemenswaardige, unieke bijdrage aan de voorspelling van de StatRec-recidivekans. Daarnaast bestaan er verschillen tussen groepen in de aard van de dynamische factoren die, ondanks de onderlinge samenhang tussen de RISC-schalen, een unieke bijdrage leveren aan de verklaring van de StatRec-voorspelling van de recidivekans. Dergelijke verschillen zouden aanleiding kunnen geven om de indicatie van gedragsinterventies af te stemmen op de achtergrond van de cliënt. Dit moet beschouwd worden als een voorzichtige conclusie omdat deze bevindingen eerst gerepliceerd zouden moeten worden in onderzoek naar de mate waarin de RISC daadwerkelijke recidive voorspelt — de predictieve validiteit van de RISC.

7.4.2 *Congruente validiteit van de afzonderlijke RISC-schalen: meten ze wat ze beogen te meten?*

Met betrekking tot de congruente validiteit in termen van de inhoudelijke betekenis van de RISC-schalen wordt aanbevolen nader onderzoek te doen. De congruente validiteit van de schalen 10, 11 en 12 is in het onderhavige onderzoek bestudeerd en in het geval van de schalen 10 en 11 vrij gunstig gebleken. Vervolgonderzoek naar de congruente validiteit van schaal 12 is echter aan te bevelen omdat de resultaten uit het onderhavige onderzoek niet eenduidig waren. Daarnaast zal de congruente validiteit van de overige schalen van de RISC in een of meerdere vervolgstudies moeten worden onderzocht.

7.4.3 *Congruente validiteit in een aantal specifieke doelgroepen*

Omdat onder reclasseringswerkers de indruk bestond dat de RISC in het geval van een aantal specifieke doelgroepen een te laag recidiverisico inschat, is de samenhang tussen RISC en StatRec in deze groepen afzonderlijk bestudeerd. Uit de resultaten van de analyses blijkt dat de bivariate samenhang (correlatie) tussen de RISC-totaalscore en de StatRec-recidivekans in de meeste groepen sterk is. Daarnaast kan worden geconcludeerd dat de RISC ook voor de meeste onderzochte specifieke doelgroepen goed in staat is de StatRec-recidivevoorspelling te verklaren en dus een goede congruente validiteit bezit. De RISC-schalen verklaren de minste variatie van de StatRec-score bij de groep veelplegers en hiermee lijkt de RISC iets minder geschikt om het recidiverisico bij deze groep te voorspellen. De indruk dat de RISC voor deze groep ten onrechte een te laag recidiverisico voorspelt wordt verder ondersteund door de bevinding dat daders die op basis van de RISC een laag of een middelhoog recidiverisico hebben, een StatRec-recidivekans van respectievelijk 66% en 76% hebben. Onderzoek naar de samenhang tussen RISC en daadwerkelijke recidive zal definitief uitsluitsel moeten geven.

Veelplegers zijn hiermee de enige van de specifieke doelgroepen waarvoor er aanwijzingen zijn dat de inschatting van reclasseringswerkers correct is dat de RISC tot een onterechte lagere inschatting van het recidiverisico zou leiden. Toch is het wellicht moeilijk om de indruk die onder reclasseringswerkers bestaat dat de RISC bij een aantal specifieke doelgroepen tot een lagere inschatting van het recidiverisico leidt dan zijzelf op basis van hun ervaring aannemelijk achten, op basis van dit onderzoek te weerleggen. In dit onderzoek is immers de vraag naar de congruente validiteit van het instrument bij deze specifieke groepen beantwoord en nog niet de vraag naar de predictieve validiteit. In dit kader lijkt er echter ook een belangrijk verschil in definitie van recidive te bestaan. StatRec voorspelt de kans dat een dader opnieuw met justitie in aanraking komt en dit is ook het doel van de RISC. Reclasseringswerkers, daarentegen, gaan in de praktijk bij hun inschatting wellicht uit van het opnieuw plegen van een delict zonder dat er sprake van hoeft te zijn dat de dader hiervoor in aanraking komt met de politie, laat staan dat er sprake hoeft te zijn van vervolging. Hiermee raakt deze definitiekwestie aan een fundamenteel probleem: het is met behulp van bestaande registraties van politie en justitie niet mogelijk om zicht te krijgen op gepleegde delicten waar geen aangifte van wordt gedaan en / of waar geen dader voor aangewezen kan worden. Heel formeel gerekend is er vanuit het oogpunt van reclassering en justitie ook geen sprake van gedrag waar zij iets 'mee moeten doen'. Er hoeft geen strafzaak te worden gevoerd,

er hoeft geen voorlichtingsrapportage te worden uitgebracht en er hoeft ook geen tijd en energie te worden gestoken in het begeleiden van de dader.

De enige manier waarop recidive met bestaande registraties in kaart is te brengen is met behulp van politie- en justitiedata. WODC-onderzoek naar recidive gaat uit van justitiedata omdat zodoende rekening gehouden kan worden met gevallen waarin de verdachte toch onschuldig blijkt te zijn en wordt vrijgesproken. Zo is ook in dit onderzoek gewerkt met een definitie van recidive als nieuw justitiecontact en dit zal ook in het onderzoek naar de predictieve validiteit van de RISC het geval zijn. Dit heeft naar alle waarschijnlijkheid tot gevolg dat het in voorkomende gevallen niet mogelijk zal zijn op basis van de resultaten van dit onderzoek de zorgen weg te nemen van reclasseringswerkers over de validiteit van de RISC bij bepaalde doelgroepen. Wellicht ten overvloede wordt hier echter nog een keer benadrukt dat het het doel van de RISC is om de kans in te schatten dat een dader opnieuw wegens een delict wordt veroordeeld tot een sanctie (Adviesbureau Van Montfoort & Reclassering Nederland, 2004, p. 9). Er zijn, met uitzondering van de groep veelplegers, op basis van het onderhavige onderzoek geen aanwijzingen dat de RISC hier voor bepaalde doelgroepen minder goed in slaagt.

7.4.4 Enkele opmerkingen naar aanleiding van het onderzoek naar de congruente validiteit in termen van recidivevoorspelling

De RISC is gevalideerd met behulp van een *voorspelling van recidive*. De samenhang tussen RISC en StatRec, bestudeerd met behulp van correlaties en regressie-analyses, geeft dan ook geen uitsluitsel over de daadwerkelijke voorspellingskracht van de RISC. Bij de validering van RISC met behulp van StatRec kunnen kanttekeningen worden geplaatst. StatRec voorspelt de kans op recidive op basis van een aantal achtergrondkenmerken waarvan de meeste betrekking hebben op het justitieel verleden van een dader. Een kritische lezer zou zich daarom af kunnen vragen of met de validering van de RISC met behulp van de StatRec wel iets méér wordt gedaan dan het bestuderen van de samenhang tussen het justitieel verleden van daders en de problemen die zij op verschillende leefgebieden hebben. Hierbij moet echter wel bedacht worden dat de StatRec-score een *gevalideerde* voorspelling van het recidiverisico geeft. Het gebruik van StatRec bij het onderzoek naar de congruente validiteit van de RISC heeft dus wel degelijk zin, maar men moet ervoor waken de resultaten te interpreteren als indicaties van de mate waarin RISC daadwerkelijk recidive voorspelt. Dit kan slechts blijken uit het onderzoek dat in 2008 zal worden uitgevoerd naar de predictieve validiteit van de RISC. Speciale aandacht zal hierbij uit moeten gaan naar de voorspelling van recidive bij vrouwen en veelplegers. Voor vrouwen lijkt de RISC er wat minder goed in te slagen de recidivekans te voorspellen. Hoewel ook in deze groep sprake is van een sterk effect blijft het percentage verklaarde variantie duidelijk achter bij de overige subgroepen. Met betrekking tot veelplegers zijn er aanwijzingen dat de RISC er onvoldoende in slaagt de recidivekans correct in te schatten. De RISC-schalen verklaren een relatief gering deel van de variantie van de StatRec-voorspelling van de recidivekans en meer dan de helft van de groep veelplegers had volgens de RISC een lage of middelhoge kans op recidive, terwijl hun StatRec-score onmiskenbaar hoog was. De vraag of RISC inderdaad een te laag recidiverisico voorspelt, zal in het onderzoek naar de predictieve validiteit definitief beantwoord moeten worden.

Over het algemeen kan echter geconcludeerd worden dat, als het onderzoek naar de congruente validiteit van de RISC in termen van recidivevoorspelling enige ver-

wachtingen mag wekken met betrekking tot het onderzoek naar de predictieve validiteit, dit gunstige verwachtingen zijn. De RISc-totaalscore vertoont een sterke positieve samenhang met de op basis van StatRec voorspelde recidivekans. Mocht naar aanleiding hiervan de vraag gesteld worden wat de RISc dan nog toevoegt aan de, veel eenvoudiger te scoren StatRec dan zijn daar twee antwoorden op mogelijk. Ten eerste *verklaren* de kenmerken op basis waarvan de StatRec recidive voorspelt niet *waarom* een dader een hoger of een lager recidiverisico heeft. De StatRec voorspelt recidive op basis van predictoren: geslacht, bijvoorbeeld, is een goede predictor van recidive maar het geslacht van een dader biedt geen verklaring voor het recidiverisico. Het tweede antwoord op de vraag wat RISc toevoegt boven de StatRec heeft te maken met de mogelijkheden om het recidiverisico door middel van gedragsinterventies te verkleinen. StatRec voorspelt recidive op basis van kenmerken die niet kunnen worden beïnvloed met behulp van gedragsinterventies. In de voorspelling van RISc wordt rekening gehouden met dynamische factoren die wel met behulp van gedragsinterventies veranderd kunnen worden. Deze factoren moeten dan wel gerelateerd zijn aan daadwerkelijke recidive. Of dit het geval is, zal ook nu weer moeten blijken uit het onderzoek naar de predictieve validiteit van de RISc, maar het onderhavige onderzoek wekt positieve verwachtingen.

7.5 Tot slot

De resultaten van de drie uitgevoerde studies naar de interbeoordelaarsbetrouwbaarheid, de structuur en de congruente validiteit van de RISc in ogenschouw nemend, luidt de algemene conclusie dat de RISc in termen van de onderzochte kenmerken over gunstige psychometrische kwaliteiten beschikt. Om het instrument verder te verbeteren zijn op verschillende punten aanpassingen mogelijk en vervolgonderzoek, met name naar de predictieve validiteit van het instrument, is nodig. In de voorafgaande paragrafen van dit hoofdstuk werden per deelonderzoek aanbevelingen gedaan of de conclusies nader besproken. In deze slotparagraaf worden de resultaten van de drie deelonderzoeken met betrekking tot een viertal potentieel ‘problematische’ schalen geïntegreerd. Met betrekking tot de RISc in zijn geheel en de overige onderdelen van de RISc blijft de conclusie dat er zonder meer sprake is van gunstige psychometrische eigenschappen.

Ten eerste kan de waarde van schaal 6 *Relaties met partner, gezin en familie* in twijfel worden getrokken. De vraag kan gesteld worden in hoeverre deze schaal daadwerkelijk thuishoort in een instrument zoals de RISc, dat beoogt het recidiverisico in kaart te brengen en de criminogene factoren die aan dat risico ten grondslag liggen. De interbeoordelaarsbetrouwbaarheid van de schaal is weliswaar redelijk, maar een andere vorm van betrouwbaarheid — de interne consistentie — is onvoldoende. Een onvoldoende betrouwbare schaal kan nooit een valide meting opleveren van het begrip dat gemeten moet worden. Daarnaast draagt de schaal nauwelijks bij aan de betrouwbare meting van de RISc-totaalscore (opnieuw de interne consistentie) en tot slot vertoont de schaal in geen van de onderzochte subgroepen de verwachte middelsterke samenhang met het door StatRec voorspelde recidiverisico. De betrouwbaarheid en de congruente validiteit van schaal 6 *Relaties met partner, gezin en familie* konden, met andere woorden, in het onderhavige onderzoek niet onderbouwd worden. Een wijziging in de samenstelling van de schaal (in casu het verwijderen van item 6.4 *Familie- of gezinslid heeft een justitieel dossier*) kan tot verbetering van de betrouwbaarheid van de schaal leiden, maar het is de

vraag of dit ook het gewenste effect zal hebben op de congruente validiteit van de schaal. Het onderzoek naar de predictieve validiteit van de RISc kan hier inzicht in bieden.

Op de tweede plaats worden in de drie deelonderzoeken verschillende problemen met betrekking tot schaal 9 *Alcoholgebruik* gevonden. De interbeoordelaarsbetrouwbaarheid van de schaal is weliswaar sterk, maar opnieuw is de interne consistentie van de schaal onvoldoende en ook deze schaal draagt niet bij aan de betrouwbare meting van de totaalscore (interne consistentie). De schaal hangt zwak samen met de overige RISc-schalen en vertoont in de totale onderzoeksgroep een zwakke samenhang met de StatRec-voorspelling van de recidivekans. Er zijn slechts een paar subgroepen waarin deze schaal wel de verwachte middelsterke samenhang met de StatRec-score vertoont. Desondanks is de waarde van deze schaal evident. Alcoholgebruik levert namelijk in een redelijk aantal van de bestudeerde subgroepen een noemenswaardige, unieke bijdrage aan de verklaring van de StatRec-recidivekans. Dit lijkt aan te geven dat als een dader problemen heeft met betrekking tot alcoholgebruik, dit in bepaalde gevallen een belangrijke voorspeller is van de kans op recidive. Uiteraard dient deze bevinding te worden gerepliceerd in het onderzoek naar de predictieve validiteit van de RISc. Daarnaast zal ondanks deze gunstige bevindingen de lage interne consistentie van de schaal verholpen moeten worden door de samenstelling van de schaal te wijzigen (aanbevolen wordt om het item 9.5 *Motivatie om alcoholgebruik aan te pakken* uit de berekening van de schaalscore te verwijderen).

Ten derde kan de waarde van schaal 10 *Emotioneel welzijn* in twijfel worden getrokken. De congruente validiteit van de schaal in termen van *inhoud* (emotioneel welzijn) wordt weliswaar ondersteund in dit onderzoek, maar de schaal hangt zowel in de totale onderzoeksgroep als in alle onderzochte subgroepen slechts zwak samen met de voorspelling van recidive. In een enkel geval is de samenhang tussen schaal 10 en de StatRec-voorspelling van recidive zelfs negatief. De congruente validiteit van deze schaal in termen van de samenhang met de recidivekans kan in dit onderzoek dan ook niet worden onderbouwd. De conclusie dat een dergelijke schaal weinig bijdraagt aan de risicotaxatie van de RISc komt overeen met de keuze die door de ontwikkelaars van de LSI-R is gemaakt om in de doorontwikkeling van dat instrument naar de huidige LS/CMI de schaal *Emotions / Personal*, die veel overeenkomsten vertoonde met de RISc-schaal *Emotioneel welzijn*, om te werken tot een schaal *Antisocial pattern* (Andrews, Bonta & Wormith, 2004). Deze schaal beoogt psychopathie en antisociaal gedrag beoogt te meten. Een dergelijke aanpassing zou ook voor de RISc meerwaarde op kunnen leveren.

Met betrekking tot schaal 12 *Houding* is ten slotte het nodige doorontwikkelwerk nodig. De interbeoordelaarsbetrouwbaarheid van deze schaal is, hoewel redelijk, de laagste van alle en de interbeoordelaarsbetrouwbaarheid van de afzonderlijke items is in veel gevallen matig. Daarnaast waren de resultaten van het onderzoek naar de congruente validiteit van de schaal in termen van inhoud, niet eenduidig. Er kunnen op basis van het onderhavige onderzoek geen conclusies worden getrokken met betrekking tot de mate waarin schaal 12 daadwerkelijk de houding van daders ten opzichte van andere mensen, de maatschappij, het delict en criminaliteit in het algemeen in kaart brengt. Nader onderzoek naar de validiteit van deze schaal is daarom nodig. Voordat dit wordt uitgevoerd is echter eerst een verbetering van de interbeoordelaarsbetrouwbaarheid aan te bevelen. Immers, betrouwbaarheid is een voorwaarde voor validiteit. Deze schaal beschikt echter ook over belangrijke gunstige psychometrische kwaliteiten: de betrouwbaarheid in termen

van interne consistentie is goed en in de totale onderzoeksgroep vertoont de schaal de verwachte middelsterke samenhang met de StatRec-voorspelling van de recidivekans. Ook in een redelijk aantal subgroepen wordt de congruente validiteit van schaal 12 *Houding* in termen van de relatie met de recidivekans onderbouwd. Alle resultaten overziend zijn er twee RISC-schalen waarbij vraagtekens kunnen worden geplaatst met betrekking tot hun waarde voor het instrument: schaal 6 *Relaties met partner, gezin en familie* en schaal 10 *Emotioneel welzijn*. Er zijn daarnaast verbeteringen van het instrument mogelijk, maar interbeoordelaarsbetrouwbaarheid, interne consistentie en congruente validiteit zijn in orde.

Summary

Psychometric qualities of the Dutch Risk Assessment Scales (RISc)

Inter-rater reliability, internal consistency and concurrent validity

1 Cause, objective and research questions

The ‘*Recidive InschattingsSchalen*’ (Risk Assessment Scales, hereinafter: RISc) is the diagnostic tool of the Dutch probation services developed in 2002-2003 by *Adviesbureau Van Montfoort*. The development of RISc was commissioned by the three Dutch probation organizations — *Reclassering Nederland* (RN), *Stichting Verslavingsreclassering GGZ Nederland* (SVG) and *Leger des Heils Jeugdzorg en Reclassering* (LJ&R) — and the instrument was developed in the context of the policy programme called ‘*Terugdringen Recidive*’ (Reducing Recidivism, or TR). For the development of RISc, the ‘What Works’ approach served as the starting point (see for example McGuire, 1995). This approach assumes that behavioural interventions aimed at reducing recidivism must be tailored to an offender’s risk of recidivism and must address the factors which put the offender at risk of reoffending in the future. In agreement with these ‘What Works’ key principles, the aim of RISc is to assess an offender’s likelihood of recidivism (defined as a new conviction) and to assess static and dynamic criminogenic factors that form the basis of this risk. The former being factors that cannot be changed, such as age, sex and prior convictions, and the latter relating to factors that are in principle changeable and influenceable. RISc comprises twelve sections which each intend to assess one of the criminogenic factors: (1) Offending history; (2) Present offence and pattern of offences; (3) Accommodation; (4) Education, work and training; (5) Financial management and income; (6) Relationships with partner, family and relatives; (7) Relationships with friends and acquaintances; (8) Drug misuse; (9) Alcohol misuse; (10) Emotional well-being; (11) Thinking and behaviour; and (12) Attitudes. Together, these sections form the overall score indicating the risk of reconviction.

RISc is based on the British Offender Assessment System (OASys; Howard, Clark & Garnham, 2003), the instrument used by the probation service and prison system in England and Wales to assess an offender’s level of likelihood of reconviction, to provide an offending-related needs profile and to allow staff to formulate supervision plans. OASys was adapted to the Dutch context and 465 offenders were assessed with it between April and September 2003. Based on the data collected in this way, the first user version of the instrument was developed (*Adviesbureau Van Montfoort & Reclassering Nederland*, 2004). During the development of RISc, the focus was on the quality of the items and attention was paid to internal consistency of the instrument’s sections. Further studies of the reliability and validity

were extremely important, in light of the nature of the instrument and the purpose for which it is used. After all, RISC is an assessment tool on the basis of which important decisions about individuals are made and whether probation officer A or B uses the instrument should not make a difference for the results. In other words: it should be possible to generalize assessments to different raters. This is the issue of the instrument's inter-rater reliability. As far-reaching decisions regarding an offender can be made on the basis of RISC, it is essential that the instrument indeed measures what it intends to measure, namely the risk of reconviction and the factors that are related to this risk. The issue thus is: what is the instrument's construct validity? In addition, a clear factor structure, which can be easily interpreted, is vital for any sound instrument. During the development of RISC, a number of a priori sections was assumed, which together form an overall score. This structure had to be tested in a large sample using principal component analyses. The same large sample can be used to study the internal consistency of the RISC overall score and the RISC sections. This type of reliability measures the consistency of results across items within a single section.

In 2005, the WODC (Research and Documentation Centre) started a study of inter-rater reliability, factor structure, internal consistency and construct validity of RISC. The purpose of the study was to gain better understanding of these psychometric characteristics of RISC, but primarily to make recommendations to psychometrically optimize the instrument. The following questions were essential in the study:

- 1 What is the inter-rater reliability of RISC?
- 2 What is the construct validity of RISC like?
In this study, the question of the instrument's construct validity concentrates on the concurrent validity: the extent to which a test test correlates well with a validated measure for the same or a related construct.
- 3 What recommendations can be made to improve the instrument on the basis of the answers to questions 1 and 2?
- 4 What is RISC's factor structure and what is the internal consistency of the RISC sections and the RISC overall score? What recommendations can be made to improve RISC's structure and internal consistency?

Based on the experience in the first years of using RISC, the three probation services have found that a fair number of probation officers have the impression that RISC, used for a few specific groups of offenders, results in the assessment of a lower risk of reconviction than they deem plausible in practice. If this really were the case, it means that the validity of the instrument is insufficient for these groups. This was the reason to ask the WODC to include the following question in the study of concurrent validity:

- 5 What is the relationship between the RISC overall score and the score of the StatRec reconviction prediction model for the following sub-groups:
 - a. Domestic violence offenders
 - b. Sexual offenders
 - c. Prolific offenders
 - d. Swindlers
 - e. Older offenders
 - f. Drunk drivers / DUI offenders

After a discussion of the designs used to answer our research questions and a presentation of the results per sub-study, the final part of this summary will provide an overview of the main conclusions and recommendations.

2 Design

The study of inter-rater reliability, factor structure, internal consistency and concurrent validity of RISC has been carried out in three sub-studies.

RISC's inter-rater reliability

In order to study RISC's inter-rater reliability, seventy-five clients of the probation services were assessed twice between November 2005 and the middle of May 2006 by two different probation officers. In addition to the standard (first) assessment, a second RISC was completed by a different probation officer, who did so independently from the first one. The probation officers worked in pairs, but had no contact with each other about the clients they assessed with RISC. The probation officers taking part in the study, who were randomly selected by the researchers, asked the clients whom they assessed with RISC whether they were willing to co-operate in the study. In total, there were nineteen pairs of probation officers involved in the study: eleven pairs from RN, five from SVG and three from LJ&R. When a client indicated that he or she was willing to co-operate in the study, this was communicated to the colleague probation officer, so that he or she could make an appointment with the client for a second assessment. This second appointment had to be scheduled about two to three weeks after the first assessment. Clients were paid €25 for taking part in the study.

The factor structure and internal consistency

The study of RISC's factor structure and internal consistency has been carried out on the basis of a database provided by *Reclassering Nederland* (RN) containing all RISC assessments initiated between November 2004 and May 2006 by the three probation organizations. After the necessary data cleaning, 11,666 RISC assessments could be analyzed.

RISC's concurrent validity

The study of RISC's concurrent validity consists of two components. The main purpose of RISC is to assess the risk of reconviction. The question as to whether RISC correlates sufficiently with an instrument that also intends to assess the risk of reconviction forms the first part of the study of concurrent validity. For this part of the study, the same database was used as the one used for the analyses of RISC's factor structure and internal consistency. For 9,985 of the 11,666 RISC assessments from this database, it was possible to calculate a score on the StatRec scale. This validated instrument predicts the risk of reconviction on the basis of a number of static offender characteristics. By studying the correlation between both instruments, it was possible to determine the concurrent validity of RISC's assessment of the level of likelihood of reconviction. The same design was used for the study of RISC's concurrent validity in the specific offender groups.

The second part of the study of RISC's concurrent validity relates to the content of the RISC sections. These sections assess specific social and personal factors and the presence or absence of criminogenic problems in these areas is used to determine

what steps should be taken by the probation services, for example: is an offender eligible for behavioural interventions. The question as to whether these sections really measure the constructs they intend to measure has been examined in the second part of the validation study. Validating all RISC sections proved to be too much of a burden for the probation services and, what was more, there were difficulties regarding validating the more factual RISC sections, such as accommodation, education/work and finances. After all, it would have been most obvious to validate such sections on the basis of probation records, but, because the probation officers use the same records when completing RISC assessments, this was not possible. In the end, in consultation with the three probation organizations, it was decided to validate the three least factual and most subjective sections of the instrument. These are sections 10 *Emotional well-being*, 11 *Thinking and behaviour* and 12 *Attitudes*. These are, moreover, sections that play an important role in answering the question as to whether a client is eligible for specific behavioural interventions. To study the concurrent validity of these sections, reliable and validated instruments that measure the same or related concepts were selected. Three questionnaires — the *'Nederlandse Persoonlijkheidsvragenlijst'* (Dutch Personality Questionnaire, NPV), the *'Utrechtse Copinglijst'* (Utrecht Coping List, UCL) and the Buss-Durkee Hostility Inventory - Dutch (BDHI-D) — were selected for the purpose of the study. Probation officers asked their clients after the RISC assessment to co-operate in the concurrent validity study. Every client completed only one of the three selected questionnaires. During the completion of the questionnaire the probation officer was present to give an explanation, where needed. Once the client had finished the questionnaire, the officer made sure that all questions had been answered and no pages had been skipped. For their co-operation in the study, the clients received €10. Two hundred and four clients of the probation services completed questionnaires between April and the middle of September 2006. In a number of cases, the client's RISC could not be retrieved, could not be used or had not been completed on the deadline of the data gathering period. Because of this, the sample for this part of the validation study consists of 185 clients.

3 Results of the study of RISC's inter-rater reliability

The nineteen pairs of probation officers independently from one another completed double RISC assessments for seventy-five clients. There was an average period of 29 days between the first and the second RISC assessment. Hardly any significant differences on background characteristics and RISC scores existed between the study sample and the population of probation service clients. The quality of the data in terms of missing values was examined and turned out to be good. RISC's inter-rater reliability was studied at both item level, section level and for the overall score. To study the degree of agreement between the probation officers on *nominal* items, coefficient κ was calculated (Cohen, 1960). For the *ordinal* items, the section scores and the overall score, the following strategy was used (cf. Born, 1995, pp. 130-132).

- 1 Calculating the proportion of agreement. This parameter is used most frequently and is the easiest to use. However, this parameter does not suffice, for both control of chance agreement and a formal test of the degree of agreement is lacking.

- 2 Calculating Lawlis and Lu's χ^2 (1972). This parameter shows whether or not agreement is significantly greater than could be expected on the basis of chance.
- 3 Calculating Tinsley and Weiss' value T (1975). This index is a derivative of Lawlis and Lu's χ^2 and indicates the degree of agreement (0=the agreement is not greater than could be expected on the basis of chance, 1=perfect agreement).

For the interpretation of both kappa and T Landis and Koch's guidelines were used, rating the strength of agreement as poor, slight, fair, moderate, substantial or almost perfect. With regard to the RISC items, the degree of agreement between probation officers was in general moderate to substantial; at section level, the agreement between the assessors was in all cases moderate to substantial. Agreement between probation officers with regard to the RISC overall score was substantial. Based on these results, RISC's inter-rater reliability can be judged as good. However, inter-rater reliability of items is poorer as the questions become less factual; a fair level of agreement is shown in respect of a substantial number of items in sections 11 and 12.

4 Results of the study of RISC's scale structure and internal consistency

To examine RISC's factor structure, principal components analyses were used to analyse each section and the overall score. For each section and for the overall score, the number of factors was checked and factor loadings were studied. This way, it was examined whether or not the items and sections respectively are related to the same underlying construct. In addition, for each section and for the overall score reliability analyses were conducted. Cronbach's coefficient alpha was calculated to assess the internal consistency of each section and of the overall score.

In general, the RISC sections form adequate scales: most items of the RISC sections each measure part of the same underlying construct that the section intends to assess. The internal consistency of the sections is adequate to good for most sections and measurements regarding most sections can thus be considered reliable. The exceptions are sections 6 *Relationships with partner, family and relatives* and 9 *Alcohol misuse*. These sections are made up of items that insufficiently relate to the same underlying construct and that show insufficient internal consistency. The analysis of the RISC overall score also revealed that the RISC sections together form a good scale: in general, sections load well on the extracted factor and therefore each section measures a part of the same underlying construct. The only section that contributes little to the overall score is section 9 *Alcohol misuse*. Stricter criteria for interpreting the overall score's internal consistency were used than for separate sections, as the more important decisions about individual offenders are taken on the basis of this score rather than on the basis of the separate sections. The reliability analysis of the overall score shows that these stricter criteria are met.

5 Results of the study of RISC's concurrent validity

To study the concurrent validity of RISC in terms of predicting the risk of reconviction, the correlation between RISC and the StatRec scale was studied. This was done in several ways. As a first step, the sample was divided into four risk groups on the basis of their RISC overall scores and the extent to which the StatRec scores of the four groups differed significantly from each other was studied. In the next stage, correlations between RISC and StatRec were computed, and finally, regression analyses were carried out.

The results of the study of RISC's concurrent validity in terms of predicting the risk of reconviction are favourable. As expected, there was a strong correlation between the RISC overall score and the validated prediction of the risk of reconviction as calculated with StatRec. On the basis of the premise of RISC — each section is related to criminal behaviour or the risk thereof — all correlations between the individual RISC sections and the StatRec risk of reconviction were expected to be moderately positive. In most cases, this expectation was corroborated in the total study sample. The exceptions are formed by sections 6 *Relationship with partner, family and relatives*, 9 *Alcohol misuse* and 10 *Emotional well-being*. The correlation between these sections and the StatRec prediction of reconviction is weak. Based on background characteristics (gender, age, ethnicity and offence type), the total sample was divided into a number of sub-samples for which the correlation between RISC and StatRec was studied. The results from these analyses are comparable to those in the total sample.

Regression analyses were carried out to verify the extent to which the StatRec score can be explained on the basis of the RISC sections. The results show that the RISC sections together explain 48% of the variance of the StatRec risk of reconviction. This is quite a substantial percentage and this finding corroborates the concurrent validity of RISC. These favourable results were replicated in almost all of the studied sub-samples. The group of female offenders is the exception. The extent to which the StatRec score of this group is explained by their RISC scores may still be substantial, but it is considerably smaller than is the case with male offenders. Although given the correlations between the individual RISC sections, it was not realistic to expect that each separate section would uniquely contribute to the explanation of the StatRec risk of reconviction, some significant results can be reported. Apart from section 1&2 *Information on offences*, sections 4 *Education, work and training*, and 7 *Relationships with friends and acquaintances* contribute substantially contribution to the explanation of the StatRec score in the total sample. This means that these factors, despite the correlation they show with the other sections, have their own impact on the explanation of the StatRec score. The impact of education and relationships with friends and acquaintances as an explanation of the risk of reconviction was also found in many of the sub-samples studied. In a number of sub-samples, also section 8 *Drug misuse* contributes to the explanation of the StatRec risk of reconviction. In addition, there are differences between the sub-samples studied in terms of the nature of dynamic factors which, despite the correlation between the RISC sections, contribute substantially to the explanation of the StatRec prediction of the risk of reconviction. A distinction between the various groups can also be made in terms of the strength of the effects of RISC sections on the explanation of the StatRec score.

The second part of the study of RISC's concurrent validity studied the correlation between RISC scores on sections 10 *Emotional well-being*, 11 *Thinking and behaviour* and 12 *Attitudes* and scores on three questionnaires (NPV, UCL, and BDHI-D) measuring similar or related constructs. Prior to the analyses, a number of expectations were drawn up with regard to the correlation between the RISC sections and specific, selected scales of the NPV, UCL and BDHI-D. The results of this study are favourable with regard to the concurrent validity of sections 10 *Emotional well-being*, and 11 *Thinking and behaviour*. The anticipated, moderate correlations between section 10 *Emotional well-being* and all scales that measure a similar construct were reported. The same goes for section 11 *Thinking and behaviour*. This constitutes a substantiation for the assumption that these sections map out the constructs they intend to measure. Section 12 *Attitudes*, on the other hand, does not show moderate correlations with two of the three sections examined. Despite some serious difficulties with regard to this sub-study and despite the fact that one of the questionnaire scales did show a moderate correlation with section 12, the results of this study may raise doubts as to whether RISC section 12 does in fact assess the attitude of the probation service's client towards other people, society, the offence and crime in general.

6 Results of the study of the correlation between RISC and StatRec in a number of specific target groups

As several probation officers had the impression that RISC did not properly assess the risk of reconviction for a number of specific offender groups, the correlation between RISC and StatRec in these groups was separately studied. Both RISC and StatRec intend to assess the risk of an offender being reconvicted. Thus, it was expected that the RISC overall score and the StatRec prediction of the risk of reconviction would show a strong correlation in the specific offender groups. This is confirmed in all groups, except for the group of prolific offenders (with eleven or more previous criminal cases). In this group, the correlation between RISC and StatRec is moderate. The separate RISC sections that intend assess dynamic criminogenic factors were expected to show moderate correlations with StatRec. These expectations are in general corroborated. As in the total sample studied, sections 6 *Relationships with partner, family and relatives*, 9 *Alcohol misuse* and 10 *Emotional well-being* are the most important exceptions. These sections show a weak correlation with StatRec-scores in all sub-groups.

Regression analyses show that RISC sections explain the largest amount of StatRec score variance in the group of drunk drivers and the group of swindlers. In the group of domestic violence offenders and the group of sexual offenders, RISC also explains a substantial part of the variance of the StatRec prediction of the risk of reconviction. The RISC sections explain the least variance of the StatRec score among the group of prolific offenders. Although this percentage does correspond with a strong correlation, RISC is significantly less successful in explaining the StatRec prediction of the risk of reconviction than it is in the other groups.

In most groups, section 4 *Education, work and training* of the dynamic RISC sections contributes substantially to explaining the StatRec score, despite the correlation with all other RISC sections. Sections 5 *Financial management and income* and 9 *Alcohol misuse* each explain a substantial, unique amount of variance in three of the six sub-groups. Between the specific sub-groups, there were few differences in

the nature of dynamic factors that make a unique contribution to the explanation of StatRec's prediction of the risk of reconviction.

In general, the analyses do not indicate that RISC gives too low a risk of reconviction for the offender groups studied. However, this is not true for the group of prolific offenders. In light of the risk of reconviction for this group estimated with the help of StatRec, the probation officers seem to have a point when they claim that RISC sometimes assesses too low a risk of reconviction for some prolific offenders. After all, prolific offenders who have a low or a medium risk of reconviction according to RISC have a respective StatRec risk of reconviction of 66% and 76%.

7 Conclusions and recommendations

Taking stock of the results of the three studies of RISC's inter-rater reliability, structure and concurrent validity, the general conclusion is that RISC, in terms of the characteristics studied, has favourable psychometric qualities. In order to further improve the instrument, adjustments are possible and more research, in particular of the instrument's predictive validity, is required. For the purpose of increasing RISC's inter-rater reliability, the following recommendations are made:

- Rephrase the instructions of RISC in such a way that they are as unambiguous as possible and avoid potential ambivalence and differences in interpretation by probation officers.
- Consider adjusting the response scales to item content. The response scale of items that are hard to assess should give room for subtle distinctions and response scales with just two response categories should be avoided for those items.
- More emphasis should be given to training probation officers in using RISC and allowing them to increase their expertise.
- The mediocre inter-rater reliability of the items in sections 11 and 12 must be dealt with. First by allowing the probation officers to increase their expertise, and perhaps by subjecting the scoring instructions to a study by a behavioural scientist.

The quality of RISC's structure can be improved in a number of ways. By disregarding items 2.10 (*Taking responsibility for the offence committed*), 6.4 (*Family member or relative has a criminal record*), 8.5 (*Motivation to tackle drug use*) and 9.5 (*Motivation to alcohol use*) in the calculation of the score on the sections to which they currently belong, the quality of these sections can be improved. This is not to say that these items should be removed from RISC, but rather to encourage finding other, more useful ways to incorporate the information gathered via these items. Item 2.10 can be added to section 12 *Attitudes*. Items 8.5 and 9.5 could be retained as separate indicators for motivation and could be included in a new 'motivation section'. Research on the correlation between item 6.4 and actual reconvictions will have to give a definitive answer for the actual value of this item. No recommendations are made in respect of the structure of the overall score. Although there is room for improvement, it is wise to first make adjustments at section level and then see what impact this has on the psychometric characteristics of the overall score.

The concurrent validity of RISC in terms of predicting reconviction is good. The RISC overall score, which intends to predict whether an offender will be reconvicted, correlates strongly with the StatRec prediction of risk of reconviction, both in the total sample and in almost all sub-groups studied. The results of the regression analyses show that a large amount of variance of the StatRec score in the total sample and in almost all sub-groups studied can be explained on the basis of the scores of the RISC sections together. Further, based on the results of this study, there are no indications that RISC underassesses the risk of reconviction in specific groups, apart from the group of prolific offenders. These results form a convincing substantiation for the concurrent validity of the RISC overall score. However, it should be noted that RISC was validated using a *prediction the risk of reconviction*. The correlation between RISC and StatRec therefore is not conclusive for RISC's actual predictive power and a study of the extent to which RISC correlates with actual reconvictions is therefore required. Special attention will have to be given to the prediction of reconviction by female offenders and prolific offenders. As for female offenders, although having good validity, RISC seems to be somewhat less successful in explaining the risk of reconviction. For prolific offenders it seems RISC is making incorrect, low risk assessments.

The question as to whether sections 10, 11 and 12 measure what they intend to measure could be answered fairly positively for sections 10 *Emotional well-being* and 11 *Thinking and behaviour*. However, further research on the concurrent validity of section 12 *Attitudes* is necessary because the results of this study were not unambiguous. In addition, the concurrent validity of the other RISC sections will have to be examined in one or several follow-up studies.

Overseeing the results of the three sub-studies, the conclusion can be drawn that there are two RISC sections that raise questions as to their value for the instrument: section 6 *Relationships with partner, family and relatives* and section 10 *Emotional well-being*. This study did neither corroborate the reliability (internal consistency) of section 6 nor the concurrent validity. The concurrent validity of section 10 in terms of substance *Emotional well-being* is corroborated, but the section shows a weak correlation with the prediction of the risk of reconviction in both the total sample and the sub-groups studied. The concurrent validity of this section in terms of correlation with the risk of reconviction therefore cannot be substantiated by this study.

Literatuur

- Abracen, J., D.L. Mailloux, R.C. Serin, C. Cousineau, P.B. Malcolm, J. Looman**
A model for the assessment of static and dynamic factors in sexual offenders
Journal of Sex Research, jrg. 41, nr. 4, 2004, pp. 321-328
- Achenbach, T. M., S.H. McConaughy, C.T. Howell**
Child/adolescent behavioral and emotional problems: Implications of cross-informant correlations for situational specificity
Psychological Bulletin, jrg. 101, nr. 2, 1987, pp. 213-232
- Adviesbureau Van Montfoort, Reclassering Nederland**
RISc versie 1.0. Recidive Inschattings Schalen. Handleiding
Harderwijk, Flevodruk, 2004
- Andrews, D.A., J.L. Bonta**
LISI-R Level of Service Inventory — Revised
Toronto, Multi-Health Systems, 1996
- Andrews, D.A., J.L. Bonta, J.S. Wormith**
Level of Service/Case Management Inventory (LS/CMI): An offender assessment system
Toronto, Multi-Health Systems, 2004
- Angleitner, A., O.P. John, F.-J. Löhr**
It's what you ask and how you ask it: An itemmetric analysis of personality questionnaires. In: A. Angleitner, J.S. Wiggins (red.) *Personality assessment via questionnaires*
Berlijn, Springer-Verlag, 1986, pp. 61-108
- Aos, S., M. Miller, E. Drake**
Evidence-based adult corrections programs: What works and what does not
Olympia, Washington State Institute for Public Policy, 2006
- Born, M.Ph.**
Het meten van prestatiegerichtheid, een situatie-response vragenlijst
Amsterdam, Vrije Universiteit, 1995 (proefschrift)
- Bosker, J.**
Knelpuntenevaluatie toepassing RISc. Inventarisatie van knelpunten bij de toepassing van RISc en aanbevelingen voor de doorontwikkeling van RISc
RN, SVG, LJ&R, Niet gepubliceerd intern document, 2006
- Brink, L.T. ten**
De ontwikkeling van kinderen tijdens een periode van klinische jeugdzorg: Status, beloop en prognose
Amsterdam, Vrije Universiteit, 1998 (proefschrift)
- Cohen, J.**
Statistical power analysis for the behavioral sciences
Hillsdale, NJ, Lawrence Erlbaum Associates, 1988, 2nd edition.
- Drenth, P.J.D., K. Sijtsma**
Testtheorie: inleiding in de theorie van psychologische tests en zijn toepassingen
Houten, Bohn Stafleu & Van Loghum, 1990
- Evers, A., J.C. van Vliet-Mulder, C.J. Groot**
Documentatie van tests en testresearch in Nederland. Deel I. Testbeschrijvingen
Assen, Van Gorcum, 2000

Farrington, D.P., B.C. Welsh

Randomized experiments in criminology: What have we learned in the last two decades?

Journal of Experimental Criminology, jrg. 1, 2005, pp. 9-38

Gendreau, P., T. Little, C. Goggin

A meta-analysis of the predictors of adult offender recidivism: What works

Criminology, jrg. 34, nr. 4, 1996, pp. 575-607

Gibbs, A.

The assessment, case management and evaluation system

Probation Journal, jrg. 46, nr. 3, 1999, pp. 182-186

Greenfeld, L.A.

Alcohol and crime. An analysis of national data on the prevalence of alcohol involvement in crime

Washington, D.C., U.S. Department of Justice, Bureau of Justice Statistics, 1998, NCJ-168632

Hanson, R.K., A.J.R. Harris

Where should we intervene? Dynamic predictors of sexual offense recidivism

Criminal Justice and Behavior, jrg. 27, nr. 1, 2000, pp. 6-35

Hanson, R.K., K.E. Morton-Bourgon

The characteristics of persistent sexual offenders: A meta-analysis of recidivism studies

Journal of Consulting and Clinical Psychology, jrg. 73, nr. 6, 2005, pp. 1154-1163

Howard, P., D. Clark, N. Garnham

Evaluation and validation of the Offender Assessment System (OASys)

OASys Central Research Unit. Report to HM Prison Service and National Probations Service, 2003

Huisman, M.

Imputation of missing item responses: Some simple techniques

Quality & Quantity, jrg. 34, 2000, pp. 331-351

Knaap, L.M. van der

Competentiegerichte assessment voor jongeren in de jeugdzorg. Ontwikkeling van een instrument

Amsterdam / Duivendrecht, PI Research, 2003 (proefschrift)

Kroes, G., J.W. Veerman, E.E.J. de Bruyn

Realiteit en vertekening bij het beoordelen van probleemgedrag van kinderen.

In: J. D. Bosch, H. A. Bosma, R. J. van der Gaag, A. J. J. M. Ruijsenaars, A. Vyt (red.) *Jaarboek ontwikkelingspsychologie, orthopedagogiek en kinderpsychiatrie 4 (2000-2001)*

Houten, Bohn Stafleu/Van Loghum, 2000, pp. 87-118

Landis, J.R., G.G. Koch

The measurement of observer agreement for categorical data

Biometrics, jrg. 33, nr. 1, 1977, pp. 159-174

Lange, A., M. Hoogendoorn, A. Wiederspahn, E. Beurs de

Buss-Durkee Hostility Inventory - Dutch, BDHI-D. Handleiding, verantwoording en normering van de Nederlandse Buss-Durkee-agressievragenlijst

Houten, Bohn Stafleu Van Loghum, 2005

Lawlis, G.F., E. Lu

Judgment of counseling process: reliability, agreement, and error

Psychological Bulletin, jrg. 78, nr. 1, 1972, pp. 17-20.

- Luteijn, F., J. Starren, H. Dijk van**
Nederlandse Persoonlijheids Vragenlijst, Handleiding
 Amsterdam, Harcourt Assessment B.V., 2000
- McGuire, J.**
What Works: Reducing Re-offending: Guidelines from Research and Practice
 Chichester, John Wiley & Sons, 1995
- McGuire, J.**
 Integrating findings from research reviews. In: J. McGuire (red.) *Offender rehabilitation and treatment. Effective programmes and policies to reduce re-offending*
 Chichester, John Wiley and Sons, 2002, pp. 3-38
- Paternoster, R., R. Brame, P. Mazerolle, A. Piquero**
 Using The Correct Statistical test for the equality of regression coefficients
Criminology, jrg. 36, nr. 4, 1998, pp. 859-866
- Philipse, M.W.G., M.W.J. Koeter, C.P.F. van der Staak, W. van den Brink**
 Static and dynamic patient characteristics as predictors of criminal recidivism: A prospective study in a Dutch forensic psychiatric sample
Law and Human Behavior, jrg. 30, nr. 3, 2006, pp. 309-327
- Roizen, J.**
 Epidemiological issues in alcohol-related violence. In: M. Galanter (Red.) *Recent developments in alcoholism. Volume 13. Alcohol and violence*
 New York, Plenum Press, 1997, pp. 7-40
- Schreurs, P.J.G., G. Willige van de, J.F. Brosschot, B. Tellegen, G.M.H. Graus**
De Utrechtse Coping Lijst: UCL. Omgaan met problemen en gebeurtenissen. Herziene uitgave
 Amsterdam, Harcourt Assessment B.V., 1993
- Stouthard, M.E.A.**
 Validiteit. In: W.P. van den Brink, G. J. Mellenbergh (red.) *Testleer en testconstructie*
 Amsterdam, Boom, 1998, pp. 269-301
- Tinsley, H.E.A., D.J. Weiss**
 Interrater Reliability and Agreement of Subjective Judgements
Journal of Counseling Psychology, jrg. 22, nr. 4, 1975, pp. 358-374
- Tong, L.S.J., D.P. Farrington**
 How effective is the "Reasoning and Rehabilitation" programme in reducing reoffending? A meta-analysis of evaluations in four countries
Psychology, Crime and Law, jrg. 12, nr. 1, 2006, pp. 3-24
- Vinke, A., B.O. Vogelvang, L. Erfemeijer, E. Veltkamp, M. Bruggeman**
Recidive Inschattings Schalen. Concept Handleiding Gebruikersversie 1.0
 Woerden, Adviesbureau van Montfoort, december 2003
- Wartna, B.S.J., M. Blom, N. Tollenaar**
De WODC-Recidivemonitor
 Den Haag, WODC, 2004
- Wartna, B.S.J., N. Tollenaar**
De StatRec-schaal. Inschatting van het recidivegevaar op basis van justitiële documentatie
 Den Haag, WODC, 2006

Wilson, D.B., L.A. Bouffard, D.L. MacKenzie

A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders

Criminal Justice and Behavior, jrg. 32, nr. 2, 2005, pp. 172-204

Bijlage 1 Klankbordgroep

Voorzitter

Mw. prof. dr. M. Ph. Born

Bijzonder hoogleraar Personeelspsychologie, Erasmus Universiteit Rotterdam

Leden

Dhr. dr. P. Oosterveld

Adjunct-uitgever, Boom test uitgevers, Amsterdam

Zelfstandig methodologisch adviseur, Methodologie, Amsterdam

Mw. drs. E. de Ruijter (lid tot 01-08-2006)

Beleidsmedewerker, Reclassering Nederland

Dhr. J. H. Stegeman (lid vanaf 01-08-2006)

Beleidsmedewerker / Projectleider doorontwikkeling RISC, Reclassering Nederland

Mw. drs. P. L. M. Steinmann

Senior beleidsmedewerker, Directie Sanctie & Preventiebeleid, Ministerie van Justitie

Bijlage 2 Factorladingen RIsc-schalen

Tabel A1 Factorladingen (gedwongen tot een éénfactoroplossing) van de items horend bij schaal 1 en 2 (N=4.526)

Item	Factorlading
1.5 Aantal keer veroordeeld tot 18 jaar	0,77
1.6 Aantal veroordelingen vanaf 18 jaar	0,78
1.7 Niet nakomen afspraken	0,67
1.8 Huidige tenlastelegging	0,71
1.9 Diverse delicten en delictgeschiedenis	0,83
2.10 Verantwoordelijkheid nemen	0,27
2.11a Delictenpatroon	0,67
2.11b Delicten worden ernstiger	0,42

Tabel A2 Factorladingen van de items horend bij schaal 1 en 2 (zonder item 2.10) (N=4.535)

Item	Factorlading
1.5 Aantal keer veroordeeld tot 18 jaar	0,77
1.6 Aantal veroordelingen vanaf 18 jaar	0,79
1.7 Niet nakomen afspraken	0,67
1.8 Huidige tenlastelegging	0,72
1.9 Diverse delicten en delictgeschiedenis	0,83
2.11a Delictenpatroon	0,67
2.11b Delicten worden ernstiger	0,42

Tabel A3 Factorladingen van de items horend bij schaal 3 (N=8.647)

Item	Factorlading
3.1 Huisvestingsgeschiedenis	0,73
3.2 Huidige woonsituatie	0,88
3.3 Geschiktheid en duurzaamheid van de woning	0,87
3.4 Woonomgeving	0,62

Tabel A4 Factorladingen van de items horend bij schaal 4 (N=11.364)

Item	Factorlading
4.1 Opleidingsniveau	0,72
4.2 Schoolbezoek	0,75
4.3 Belemmering voor scholing en werk	0,82
4.4 Werkervaring en werkverleden	0,81
4.5 Huidige werksituatie	0,65
4.6 Vaardigheden ten aanzien van opleiding, werk en leren	0,86
4.7 Houding ten opzichte van opleiding, werk en leren	0,81

Tabel A5 Factorladingen van de items horend bij schaal 5 (N=11.480)

Item	Factorlading
5.1 Hoofdbron inkomsten	0,63
5.2 Huidige financiële situatie	0,89
5.3 Ernstige beperkingen in het budget	0,80
5.4 Gokverslaving of ander verslavingsgedrag	0,59

Tabel A6 Factorladingen (gedwongen tot een éénfactoroplossing) van de items horend bij schaal 6 (N=10.556)

Item	Factorlading
6.1 Jeugdervaringen	0,70
6.2 Ervaring met hechte (partner)relaties vanaf (jong)volwassenheid	0,74
6.3 Huidige partner, familie en gezinsrelaties	0,75
6.4 Familie- of gezinslid heeft een justitieel dossier	0,32
6.5 Huiselijk geweld	0,61

Tabel A7 Factorladingen van de items horend bij schaal 6 (zonder item 6.4) (N=11.217)

Item	Factorlading
6.1 Jeugdervaringen	0,69
6.2 Ervaring met hechte (partner)relaties vanaf (jong)volwassenheid	0,75
6.3 Huidige partner, familie en gezinsrelaties	0,77
6.5 Huiselijk geweld	0,64

Tabel A8 Factorladingen van de items horend bij schaal 7 (N=11.256)

Item	Factorlading
7.1 Vrienden	0,80
7.2 Negatief beïnvloeden	0,75
7.3 Manipuleert	0,75
7.4 Zoekt sensatie	0,76

Tabel A9 Factorladingen (gedwongen tot een éénfactoroplossing) van de items horend bij schaal 8 (N=4.230)

Item	Factorlading
8.1a Type drug	0,60
8.1b Frequentie	0,56
8.2 Drugs staan centraal	0,83
8.3 Relatie drug en delict	0,84
8.4 Relatie drug en gevaar	0,74
8.5 Motivatie om druggebruik aan te pakken	0,06

Tabel A10 Factorladingen van de items horend bij schaal 8 (zonder item 8.5) (N=5.157)

Item	Factorlading
8.1a Type drug	0,73
8.1b Frequentie	0,70
8.2 Drugs staan centraal	0,84
8.3 Relatie delict drug	0,85
8.4 Relatie drug en gevaar	0,75

Tabel A11 Factorladingen (gedwongen tot een éénfactoroplossing) van de items horend bij schaal 9 (N=4.581)

Item	Factorlading
9.1 Alcoholgebruik verleden	0,50
9.2 Huidig gebruik is een probleem	0,76
9.3 Relatie alcohol en delict	0,82
9.4 Relatie alcohol en gevaar	0,84
9.5 Motivatie om alcoholgebruik aan te pakken	0,20

Tabel A12 Factorladingen van de items horend bij schaal 9 (zonder item 9.5) (N=11.493)

Item	Factorlading
9.1 Alcoholgebruik verleden	0,83
9.2 Huidig gebruik is een probleem	0,86
9.3 Relatie alcohol en delict	0,90
9.4 Relatie alcohol en gevaar	0,89

Tabel A13 Factorladingen van de items horend bij schaal 10 (N=11.554)

Item	Factorlading
10.1 Moeite zich staande te houden	0,76
10.2 Psychische problemen	0,84
10.3 Zelfbeeld	0,73
10.4 Zelfdestructief gedrag	0,62
10.5 Bijzondere omstandigheden	0,74

Tabel A14 Factorladingen van de items horend bij schaal 11 (N=11.554)

Item	Factorlading
11.1 Sociale vaardigheden	0,77
11.2 Impulsiviteit	0,69
11.3 Dominant gedrag	0,62
11.4 Zelfbeheersing	0,62
11.5 Probleembesef	0,76
11.6 Probleemhantering	0,80
11.7 Doelgericht zijn	0,72

Tabel A15 Factorladingen van de items horend bij schaal 12 (N=11.405)

Item	Factorlading
12.1 Pro-criminele houding	0,81
12.2 Houding tegenover sanctie	0,77
12.3 Houding ten aanzien van samenleving	0,72
12.4 Inzicht tegenover zichzelf en delictgedrag	0,80
12.5 Veranderingsgezindheid	0,82

Tabel A16 Factorladingen (gedwongen tot een éénfactoroplossing) van de schalen die de totaalscore vormen (N=5.900)

Schaal	Factorlading
Schaal 1&2 Delictgegevens	0,70
Schaal 3 Huisvesting en wonen	0,62
Schaal 4 Opleiding, werk en leren	0,72
Schaal 5 Inkomen en omgaan met geld	0,64
Schaal 6 Relaties met partner, gezin en familie	0,55
Schaal 7 Relaties met vrienden en kennissen	0,71
Schaal 8 Druggebruik	0,57
Schaal 9 Alcoholgebruik	0,30
Schaal 10 Emotioneel welzijn ^a	0,53
Schaal 11 Denkpatronen, gedrag en vaardigheden	0,75
Schaal 12 Houding	0,66

