



Inspectie Justitie en Veiligheid
Ministerie van Justitie en Veiligheid

Risicovol algoritmegebruik

Onderzoek naar gebruik van algoritmes
bij de reclassering

Rapport

Inhoudsopgave

Voorwoord	3
1. Samenvatting	4
2. Inleiding.....	9
2.1 Aanleiding	9
2.2 Over de reclassering	10
2.3 Onderzoek naar de reclassering	12
2.4 Leeswijzer	13
3. Inventarisatie algoritmes	14
MATCHTOOL	14
IFZO	14
SCIL	14
OXREC	15
STATIC-99R, STABLE-2007 en ACUTE-2007.....	15
4. Bevindingen nader onderzoek risicotaxatie-algoritmes	16
4.2 Technische robuustheid.....	17
OXREC	18
STATIC-99R, STABLE-2007, ACUTE-2007	23
4.3 Privacy en datagovernance	25
4.4 Transparantie	26
4.5 AI-verordening.....	27
4.6 Diversiteit, non-discriminatie en rechtvaardigheid.....	28
4.7 Menselijke controle.....	29
4.8 Governance	32
5. Conclusies	33
5.1 Deelconclusies	34
5.2 Hoofdconclusies	36
5.3 Opvolging van conclusies.....	36
6. Aanbevelingen.....	37
Bijlage A: begrippen	38
Bijlage B: opzet van het onderzoek	39
Computersimulaties OXREC.....	40
Bijlage C: de OXREC	41
Wat is de OXREC?	41
Ontwikkeling en ingebruikname van de OXREC.....	42

Hoe werkt de OXREC?	42
Bijlage D: de STATIC-99R, STABLE-2007 en ACUTE-2007.....	44
Wat zijn de STATIC-99R, STABLE-2007 en ACUTE-2007?.....	44
Ontwikkeling van de STATIC-99R, STABLE-2007 en ACUTE-2007.....	46
Bijlage E: onderbouwing dat de OXREC een AI-systeem is conform artikel 3(1) AI-verordening.....	47
Beschrijving van de OXREC	47
Definitie van AI-systeem.....	47
Toepassing van de definitie van AI-systeem op de OXREC.....	48
1. Machine gebaseerd	48
2. Autonomie	48
3. Vertonen van aanpassingsvermogen na inzet	49
4. Doelstellingen	49
5. Inferentievermogen	49
6. Output	50
7. Invloed op de omgeving	50
Conclusie	50

Voorwoord

De Inspectie Justitie en Veiligheid is dit jaar gestart met toezicht op het gebruik van algoritmes binnen de domeinen Justitie en Veiligheid en Asiel en Migratie. Dit doen wij omdat de impact van algoritmes groot kan zijn, ook in negatieve zin. Dit blijkt bijvoorbeeld uit geruchtmakende gevallen als de toeslagenaffaire en het discriminerende algoritme van DUO. Daarnaast bevinden veel personen die te maken hebben met Justitie en Veiligheid en Asiel en Migratie zich vaak in een uiterst kwetsbare positie. Denk dan bijvoorbeeld aan verdachten, kwetsbare jongeren, gedetineerden en asielzoekers, maar ook aan de maatschappij die beschermd moet worden tegen criminelen. Wanneer er algoritmes in deze situaties gebruikt worden, is het des te belangrijker dat deze op een verantwoorde manier ingezet worden. Alleen zo kunnen zowel individuen als de gehele maatschappij goed beschermd worden.

Dit rapport beschrijft het onderzoek van de Inspectie Justitie en Veiligheid naar het gebruik van algoritmes bij de drie reclasseringsorganisaties in Nederland. Dit is het eerste onderzoek dat de Inspectie Justitie en Veiligheid heeft uitgevoerd naar algoritmes.

De reclassering heeft als één van haar hoofdtaken om adviezen uit te brengen over verdachten en gedetineerden. Hierin wordt gebruik gemaakt van algoritmes die inschatten of een persoon opnieuw de fout in zal gaan. De reclassering gebruikt deze algoritmes met als doel om adviezen over verdachten en veroordeelden zo objectief mogelijk te maken. Desondanks constateert de Inspectie Justitie en Veiligheid dat er sprake is geweest van een gebrek aan sturing om verantwoord om te gaan met deze algoritmes. Dit heeft ertoe geleid dat tekortkomingen bij het gebruik van deze algoritmes jarenlang onopgemerkt zijn gebleven, die een negatieve impact hebben gehad op de adviezen die de reclassering heeft uitgebracht.

De onderzochte algoritmes van de reclassering voldoen niet aan de wettelijke vereisten en richtlijnen, zij het in verschillende mate. Het is daarom van het grootste belang dat de reclassering passende maatregelen treft. De reclassering heeft hiervoor de eerste verbetermaatregelen in gang gezet.

Dit onderzoek onderstreept het belang van een juiste, verantwoorde en rechtmatige toepassing van algoritmes waardoor risico's in de taakuitvoering worden gemitigeerd.

Liesbeth Huijzer

Waarnemend inspecteur-generaal Inspectie Justitie en Veiligheid

1. Samenvatting

1.1 Inleiding

De Inspectie Justitie en Veiligheid (hierna: Inspectie JenV) heeft onderzoek gedaan naar het gebruik van algoritmes bij de drie reclasseringsorganisaties in Nederland (hierna: de reclassering): Reclassering Nederland, het Leger des Heils Jeugdbescherming & Reclassering en Stichting Verslavingsreclassering GGZ.

De reclassering heeft als één van haar kerntaken om adviezen uit te brengen over verdachten en veroordeelde personen. De reclassering adviseert met welke bijzondere voorwaarden (zoals een locatieverbod of behandeling) het risico op recidive kan worden beperkt. De reclassering stelt het advies op aan de hand van relevante factoren als de aard van het gepleegde delict, het delictverleden, de persoonlijke omstandigheden van de verdachte/veroordeelde en een inschatting over het recidiverisico. Voor de inschatting van één van deze elementen, het risico op recidive, wordt gebruik gemaakt van risicotaxatie-algoritmes. Dit zijn algoritmes die aan de hand van enkele vragen over een verdachte of veroordeelde de kans op het opnieuw plegen van een strafbaar feit voorspellen. De uitkomsten van deze algoritmes worden meegewogen in het uiteindelijke advies van de reclasseringsmedewerker. Deze combinatie van het menselijk oordeel en het gebruik van één of meerdere algoritmes is een bewuste keuze van de reclassering om tot een gestructureerd oordeel te komen, en daarbij (onbewuste) bias van reclasseringsmedewerkers te voorkomen.

Uiteindelijk beslist de rechter of een strafmaatregel wordt opgelegd. Als de rechter beslist een straf op te leggen, dan zoekt de rechter naar een passende straf en houdt daarbij rekening met de omstandigheden. De straf die de rechter oplegt hangt af van: de ernst van het misdrijf, de omstandigheden van het misdrijf en leeftijd en voorgeschiedenis van de dader, en de relevante wettelijke kaders. De reclassering adviseert de rechter en het OM over de tweede factor: de persoon c.q. de persoonlijke omstandigheden van de verdachte/veroordeelde.

De centrale onderzoeksvraag luidt: "Worden algoritmes bij de reclassering op een verantwoorde manier gebruikt?" De Inspectie JenV heeft de volgende deelvragen geformuleerd om deze vraag te kunnen beantwoorden:

- Welke algoritmes gebruikt de reclassering?
- Voldoen de algoritmes aan de normen die gesteld zijn voor overheidsalgoritmes?
- Welke gevolgen kunnen de algoritmes hebben op mensen in een kwetsbare positie?
- In hoeverre is de reclassering voorbereid op de nieuwe wettelijke vereisten van de Artificiële Intelligentie verordening (AI-verordening)?

De Inspectie JenV heeft dit onderzoek uitgevoerd in de periode van januari tot en met oktober 2025. Voor dit onderzoek heeft de Inspectie JenV bij de reclassering documentatie geanalyseerd over de algoritmes die zij gebruikt. Dit betreffen onder

andere beleidsdocumenten, wetenschappelijke studies, documentatie over de werking van de algoritmes en Data Protection Impact Assessments (DPIA's). Ook heeft de Inspectie JenV de sturing op het gebruik van algoritmes onderzocht en beoordeeld hoe de algoritmes in de praktijk gebruikt worden door middel van data-analyses en interviews met betrokken medewerkers. De algoritmes zijn beoordeeld op grond van het toetsingskader algoritmes van de Inspectie JenV.

Na een inventarisatie van de algoritmes die de reclassering gebruikt, is er nader onderzoek gedaan naar de vier risicotaxatie-algoritmes die de reclassering gebruikt: de OXREC, STATIC-99R, STABLE-2007 en ACUTE-2007. Dit zijn algoritmes die een risico-inschatting berekenen voor de kans op recidive aan de hand van persoonsgegevens zoals leeftijd, geslacht en delictverleden. De Inspectie JenV heeft besloten nader onderzoek te doen naar deze algoritmes omdat ze een belangrijk onderdeel vormen van het primaire proces van de reclassering. Dit komt omdat het gebruik van algoritmes verplicht is bij reclasseringsadviezen en medewerkers de uitkomsten van de algoritmes expliciet meenemen in hun risico-inschattingen. De uitkomsten belanden op deze wijze ook in rechtbank. Als er tekortkomingen zouden blijken in de toepassing van deze algoritmes, dan zou dit nadelige gevolgen kunnen hebben voor verdachten, veroordeelden en de maatschappij.

1.2 Bevindingen

Technische robuustheid

OXREC

De Inspectie JenV heeft in haar onderzoek tekortkomingen geconstateerd op methodologisch en praktisch vlak met betrekking tot het gebruik van de OXREC door de reclassering. De Inspectie JenV heeft in de software-implementatie van de OXREC drie gebreken aangetroffen. Deze bevindingen maken dat sinds de ingebruikname van het algoritme in 2018 het algoritme geen correcte voorspellingen heeft gedaan. Deze gebreken zijn dusdanig groot dat de Inspectie JenV met behulp van computersimulaties schat dat in ongeveer 21% van de gevallen de OXREC tot een andere risicocategorie zou zijn gekomen als het algoritme op de bedoelde wijze zou zijn toegepast. Dit zou dan meestal een hogere risico-inschatting zijn geweest.

De Inspectie JenV heeft niet onderzocht in welke mate de uitkomst van de OXREC van invloed is geweest op de daadwerkelijke adviezen die de reclassering heeft gegeven, en hoe (bijvoorbeeld) OM en rechtbank die adviezen hebben gewogen.

STATIC-99R, STABLE-2007, ACUTE-2007

Deze drie algoritmes worden wereldwijd veel gebruikt om het recidiverisico te voorspellen bij mannelijke zedendelinquenten. De Inspectie JenV heeft geen gebreken aangetroffen in de software-implementatie van deze algoritmes.

Van deze drie algoritmes is alleen de STATIC-99R gevalideerd in een Nederlandse populatie met mannelijke zedendelinquenten, zij het met verouderde data uit de

periode 1995-2002. De STABLE-2007 en ACUTE-2007 zijn niet gevalideerd in een Nederlandse populatie. Hierdoor is het onduidelijk of de algoritmes ook goed functioneren bij Nederlandse mannelijke zedendelinquenten en/of -verdachten. De Inspectie JenV heeft niet onderzocht wat de gevolgen hiervan zijn geweest in de praktijk.

Privacy en datagovernance

De vier risicotaxatie-algoritmes zijn profilerende algoritmes waar een Data Protection Impact Assessment (DPIA) voor verplicht is volgens de Algemene Verordening Gegevensbescherming (AVG). Deze DPIA's ontbreken voor de vier risicotaxatie-algoritmes, waardoor er niet aan de wettelijke vereisten van de AVG wordt voldaan.

Transparantie

De vier risicotaxatie-algoritmes zijn openbaar gepubliceerd en de werking is relatief eenvoudig en inzichtelijk. De algoritmes zijn niet gepubliceerd in het algoritmeregister, maar de reclassering is voornemens dit wel te doen voor het einde van 2025.

AI-verordening

Naar inzicht van de Inspectie JenV is de OXREC een hoog-risico AI-systeem waar de Artificiële Intelligentie verordening (AI-verordening) op van toepassing is en waarvoor strenge wettelijke vereisten gelden. De reclassering heeft nog geen maatregelen getroffen om aan deze wettelijke eisen te voldoen.

Diversiteit, non-discriminatie en rechtvaardigheid

In 2020 publiceerde hoogleraar Gijs van Dijck een kritisch artikel over de OXREC. Hierin stelde hij dat gebruik van de OXREC mogelijk kan leiden tot etnisch profileren. De reclassering heeft de inschatting gemaakt dat er geen sprake was van discriminatie. Hierdoor zijn er geen stappen ondernomen om discriminatie door de algoritmes te voorkomen, te meten of te verminderen. De Inspectie JenV constateert dat de OXREC twee variabelen ('buurtscore' en 'hoogte van het inkomen') gebruikt die mogelijk kunnen leiden tot discriminatie van personen met een migratieachtergrond. Het gebruik van dit type variabelen is volgens het College voor de Rechten van de Mens verboden op basis van de grondwet en internationale verdragen als er geen aanvullende maatregelen getroffen zijn. De Inspectie JenV constateert dat deze maatregelen niet zijn genomen door de reclassering.

Menselijke controle

De risicotaxatie-algoritmes zijn belangrijke factoren in het eindoordeel van de medewerkers. Hoewel de reclassering in trainingsmateriaal aangeeft dat de risicotaxatie-algoritmes uitsluitend ter ondersteuning gebruikt mogen worden en niet het oordeel van de medewerkers vervangen, ziet de Inspectie JenV risico's ten aanzien van betekenisvolle menselijke tussenkomst. Zo wordt in sommige situaties

gestimuleerd om de uitkomsten van het algoritme over te nemen en wordt in een interne podcast aangegeven dat het oordeel van de medewerker zonder de OXREC niet beter is dan het opgooien van een muntje. Medewerkers worden niet voldoende uitgelegd hoe de OXREC werkt om in te kunnen schatten wanneer de uitkomst minder betrouwbaar is.

Governance

De Inspectie JenV constateert dat duidelijke sturing door de reclassering op het gebruik, periodieke validaties en onderhoud van de algoritmes ontbreekt.

1.3 Conclusies

De Inspectie JenV komt in haar onderzoek tot de volgende conclusies:

- De reclassering maakt gebruik van risicotaxatie-algoritmes die niet goed aansluiten op de situaties waarin ze worden toegepast. De algoritmes zijn gebaseerd op verouderde gegevens die in een andere context en doelgroep verzameld zijn.
- Er is geen eenduidig beleid om de risicotaxatie-algoritmes te valideren en te evalueren.
- De reclassering is transparant over het gebruik van risicotaxatie-algoritmes.
- De reclassering gebruikt sinds de ingebruikname de OXREC op een verkeerde manier door een onjuiste software-implementatie.
- De reclassering heeft signalen ontvangen over mogelijke discriminatie door de OXREC. Hier zijn geen concrete maatregelen op getroffen. Het algoritme bevat verboden variabelen die mogelijk kunnen leiden tot discriminatie.
- Naar inzicht van de Inspectie Justitie en Veiligheid is de OXREC een hoog-risico AI-systeem waar de AI-verordening vanaf 2030 op van toepassing is.
- De reclassering voldoet niet aan de wettelijke vereisten en richtlijnen die gesteld worden voor het gebruik van risicotaxatie-algoritmes.
- De Inspectie JenV ziet risico's ten aanzien van betekenisvolle menselijke tussenkomst bij het gebruik van risicotaxatie-algoritmes.
- De reclassering heeft geen adequate sturing georganiseerd voor het gebruik van risicotaxatie-algoritmes.

De Inspectie JenV trekt de volgende hoofdconclusies:

- De Inspectie JenV ziet risico's in het algoritmegebruik van de reclassering die kunnen leiden tot negatieve gevolgen voor zowel de maatschappij als individuele verdachten en veroordeelde personen.
- De reclassering gebruiken de STATIC-99R, STABLE-2007 en ACUTE-2007 niet op een aantoonbaar verantwoorde wijze.
- De reclassering gebruikt de OXREC niet op een verantwoorde wijze.

De Inspectie JenV heeft in juli 2025 het bestuur van de drie reclasseringsorganisaties op de hoogte gebracht van haar eerste bevindingen. De Inspectie JenV heeft daarbij aandacht gevraagd voor de risico's van het gebruik van de OXREC. Het bestuur van de 3RO heeft op basis van de melding van de Inspectie JenV de eerste verbetermaatregelen in gang gezet, zoals het opstellen van

verplichte DPIA's, het aanpassen van de software-implementatie en het opzetten van een programma om de algoritme-governance te verbeteren. Hiervoor heeft de reclassering ook een claim ingediend bij het ministerie van Justitie en Veiligheid. Deze verbetermaatregelen zijn geen onderwerp van het onderzoek geweest.

1.4 Aanbevelingen

In het licht van haar bevindingen acht de Inspectie het logisch dat alle implementatiegebreken van de OXREC zo snel mogelijk worden verholpen, en dat - wanneer dit niet op korte termijn mogelijk blijkt - de OXREC (tijdelijk) wordt stilgelegd. De Inspectie heeft echter geen onderzoek gedaan wat de impact hiervan zal zijn voor de taakuitvoering van de reclassering, en begrijpt van de reclassering dat een dergelijke aanbeveling zeer nadelige operationele gevolgen zou hebben.

De Inspectie beveelt daarnaast het volgende aan:

1. Richt een alomvattende governance-structuur in voor de ontwikkeling, toepassing en het onderhoud van algoritmes, en voer deze ook uit;
2. Valideer en evalueer periodiek de risicotaxatie-algoritmes op een wijze die aansluit op het gebruik van deze algoritmes door de reclassering. Hertrain hierbij AI-systemen zoals de OXREC periodiek;
3. Verifieer de correctheid van de software-implementatie van de algoritmes, ook na aanpassingen aan de algoritmes;
4. Voldoe bij het gebruik van risicotaxatie-algoritmes in ieder geval aan de geldende wettelijke normen en richtlijnen die gesteld zijn in het algoritmekader voor de overheid, de AVG, de richtlijn betekenisvolle menselijke tussenkomst, het toetsingskader risicoprofilering en de AI-verordening;
5. Rapporteer periodiek, iedere zes maanden, over de opvolging van de aanbevelingen.

2. Inleiding

De Inspectie Justitie en Veiligheid (hierna: de Inspectie JenV) heeft onderzoek gedaan naar de ontwikkeling en het gebruik van algoritmes bij de drie organisaties die in Nederland het reclasseringswerk uitvoeren: Reclassering Nederland, Stichting Verslavingsreclassering GGZ en Leger des Heils Jeugdbescherming & Reclassering (hierna gezamenlijk: de reclassering).

2.1 Aanleiding

De Inspectie JenV houdt toezicht op het domein van Justitie en Veiligheid (JenV) en Asiel en Migratie (AenM). In het meerjarenperspectief 2025-2028¹ vestigt de Inspectie JenV in het bijzonder aandacht op het thema digitalisering en heeft de ze expliciet oog voor mensen in een kwetsbare positie. De Inspectie JenV ziet dat overheidsorganisaties steeds vaker² gebruik maken van data en algoritmes in hun taakuitvoering en dat het gebruik van algoritmes mogelijk gevolgen heeft voor mensen in een kwetsbare positie.

Algoritmes bieden kansen om processen te automatiseren en efficiënter in te richten en geven de mogelijkheid om meer en betere inzichten te verwerven uit beschikbare data. Tegelijkertijd zijn er steeds meer zorgen over het gebruik van algoritmes door de overheid³ en hebben algoritmes bijvoorbeeld bijgedragen aan de toeslagenaffaire⁴ en aan het indirect discrimineren door DUO⁵. Algoritmes vormen steeds vaker een belangrijk onderdeel van de taakuitvoering van organisaties. Wanneer ze niet goed functioneren kan de impact groot zijn, op zowel mensen, de maatschappij als op de organisatie zelf.

Zo geven algoritmes niet altijd de juiste voorspellingen, lijden ze soms aan een gebrek aan transparantie, worden ze niet altijd correct toegepast en sluit het gebruik in de praktijk niet altijd aan bij de wijze waarop de algoritmes zijn ontwikkeld. Ook kan de afhankelijkheid van algoritmes groot zijn, wat gevolgen kan hebben voor de taakuitvoering bij bijvoorbeeld een storing of een ransomware-aanval. Daarnaast zijn er de laatste jaren steeds meer zorgen over privacy-overtredingen en discriminatie door algoritmes. De impact van een slecht functionerend algoritme op mensen, de maatschappij en organisaties is hierbij potentieel groot, aangezien algoritmes vaak gebruikt worden om grote aantallen beslissingen te nemen of voorspellingen te doen.

De Inspectie JenV heeft onderzoek gedaan naar het gebruik van algoritmes bij de reclassering omdat algoritmes een belangrijk onderdeel vormen van het primaire proces en worden gebruikt om risico's in te schatten over verdachten en

¹ Inspectie Justitie en Veiligheid. Meerjarenperspectief 2025-2028 (<https://www.inspectie-jenv.nl/Publicaties/jaarplannen/2024/10/17/meerjarenperspectief-2025---2028-inspectie-jenv>).

² Algemene rekenkamer. Focus op AI bij de Rijksoverheid (2024).

³ Autoriteit Persoonsgegevens. Rapportage AI- & Algoritmerisico's Nederland (RAN) (2024).

⁴ <https://www.amnesty.nl/wat-we-doen/tech-en-mensenrechten/algoritmes-big-data-overheid>

⁵ <https://www.rijksoverheid.nl/actueel/nieuws/2024/03/01/kabinet-maakt-excuses-voor-indirecte-discriminatie-bij-controles-op-de-uitwonendenbeurs>

veroordeelde personen. Het gebruik van algoritmes verplicht is bij reclasseringsadviezen en de uitkomsten belanden op deze wijze ook in de rechtbank. Vanwege de kwetsbare positie van deze personen is het belangrijk dat deze algoritmes op een verantwoorde manier worden ontwikkeld en toegepast.

2.2 Over de reclassering

De reclassering is een belangrijke schakel in het Nederlandse strafrechtelijk systeem. Zo begeleidt ze gedetineerde personen bij de re-integratie in de maatschappij, en organiseert ze werkstraffen voor veroordeelden. Daarnaast heeft de reclassering als één van haar kerntaken om onderzoek te doen naar en advies te geven over personen die worden verdacht van of zijn veroordeeld voor een strafbaar feit.⁶ De reclassering stelt het advies op aan de hand van relevante factoren als de aard van het gepleegde delict, het delictverleden, de persoonlijke omstandigheden van de verdachte/veroordeelde en het recidiverisico.

De reclassering adviseert met welke bijzondere voorwaarden (zoals een locatieverbod of behandeling) het risico op recidive kan worden beperkt. De adviezen die de reclassering opstelt worden in de regel gegeven aan het Openbaar Ministerie (OM) of de rechtbank. Zo kan het advies van invloed zijn op de strafeis van het OM, en kan de rechter het advies van de reclassering meewegen om verdachten een gevangenisstraf op te leggen of een gedetineerde juist vervroegd vrij te laten.⁷

Uiteindelijk beslist de rechter of een straf wordt opgelegd. Als de rechter beslist een stafmaatregel op te leggen, dan zoekt de rechter naar een passende straf en houdt daarbij rekening met de omstandigheden (artikel 351 WvSv). De straf die de rechter oplegt hangt af van: de ernst van het misdrijf, de omstandigheden van het misdrijf en leeftijd en voorgeschiedenis van de dader, en de relevante wettelijke kaders. De reclassering adviseert de rechter en het OM over de tweede factor: de persoon c.q. de persoonlijke omstandigheden van de verdachte/veroordeelde.

Om tot gestructureerde adviezen te komen heeft de reclassering de RISC⁸ ontwikkeld. Dit is een gestructureerd risicotaxatie- en adviesinstrument dat door reclasseringsmedewerkers wordt gebruikt om in te schatten wat de kans is dat een verdachte of veroordeelde opnieuw de fout in zal gaan, oftewel zal recidiveren. In situaties waarin de reclassering advies geeft over personen, wordt de RISC ingezet. Met de RISC beoogt de reclassering factoren in kaart te brengen die van invloed kunnen zijn op het toekomstige gedrag van een verdachte of veroordeelde. Deze factoren worden ook wel *leefgebieden* genoemd. Ze omvatten negen verschillende factoren: delictverleden, huisvesting, dagbesteding, financiën, relatie (met partner, gezin en familie), sociaal netwerk, middelengebruik en verslaving, psychosociaal functioneren en houding. De reclassering geeft aan te hechten aan de wetenschappelijke onderbouwing van haar adviesproces. Zo zijn alle leefgebieden

⁶ Artikel 8(1a) Reclasseringsregeling 1995.

⁷ Maas, M., E. Legters, and S. Fazal. "Professional en risicotaxatieinstrument hand in hand: Hoe de reclassering risico's inschat." *Nederlands Juristenblad* 28 (2020): 2055-2058. Pagina 1-2.

⁸ Van den Berg, Chantal, et al. "Validatiestudie risico-en beschermende factoren van de RISC: Een evaluatieonderzoek naar de leefgebieden van het risicotaxatie en adviesinstrument van de 3RO." (2021). Pagina 19.

zorgvuldig geselecteerd op basis van onderzoeken die de relatie tussen die leefgebieden en de kans op recidiveren aantonen.⁹

De reclassering beoogt met de RISC te komen tot gestructureerde en zorgvuldige adviezen. Opgevolgde adviezen kunnen immers grote impact hebben. Zo kunnen de maatschappelijke gevolgen ernstig zijn als een dader vervroegd vrijkomt en opnieuw de fout ingaat. Maar ook bij te streng ingrijpen zijn er negatieve gevolgen zoals onnodige vrijheidsbeperking van een veroordeelde en hoge kosten voor het uitvoeren van straffen en maatregelen.

In de RISC worden naast de leefgebieden ook algoritmes gebruikt om de kans op recidive in te schatten. De OXREC¹⁰ is zo'n algoritme, en is een vast onderdeel van de RISC. Daarnaast gebruikt de reclassering nog andere algoritmes om de kans op recidive in te schatten bij mannelijke zedendelinquenten. Dit zijn de STATIC-99R, STABLE-2007 en ACUTE-2007. Net als bij de leefgebieden beantwoordt de medewerker vragen over de verdachte of veroordeelde die als invoer dienen voor de algoritmes. De algoritmes berekenen een risico-inschatting voor de kans op recidive. De reclasseringsmedewerker gebruikt deze risico-inschatting bij het opstellen van het advies.

Deze combinatie van het klinisch oordeel en het gebruik van één of meerdere algoritmes in de RISC is een bewuste keuze van de reclassering om tot een gestructureerd oordeel te komen. De reclassering verwijst hiervoor naar wetenschappelijk onderzoek waaruit naar voren komt dat de combinatie van deze verschillende aanpakken kan leiden tot betere adviezen dan wanneer slechts gebruik wordt gemaakt van uitsluitend de expertise van de reclasseringsmedewerker.¹¹ De reclassering beoogt hiermee tot een goede oordeelsvorming te komen waarbij voorkomen wordt dat medewerkers te veel op hun 'onderbuikgevoel' af gaan. De uitkomsten van de algoritmes worden in principe ingezet om het oordeel van de individuele medewerker te spiegelen. Daarnaast wordt het gebruik van deze algoritmes gezien als middel om reclasseringsmedewerkers feedback te geven. Reclasseringsmedewerkers weten vaak niet of verdachten of veroordeelden recidiveren. Door een algoritme te gebruiken om recidive te voorspellen is het doel van de reclassering om objectieve informatie over soortgelijke personen mee te laten wegen in het eindoordeel van de reclasseringsmedewerker. Daarnaast geeft de reclassering aan dat het gebruik van risicotaxatie-algoritmes ook voortkomt uit diverse adviezen van bijvoorbeeld de Raad voor Strafrechttoepassing en Jeugdbescherming¹², en de Nationaal Rapporteur Mensenhandel en Seksueel Geweld tegen Kinderen¹³. Het gebruik van

⁹ Van den Berg, Chantal, et al. "Validatiestudie risico-en beschermende factoren van de RISC: Een evaluatieonderzoek naar de leefgebieden van het risicotaxatie en adviesinstrument van de 3RO." (2021). Pagina 21-27.

¹⁰ Fazel, Seena, et al. "Prediction of violent reoffending in prisoners and individuals on probation: A Dutch validation study (OxRec)." *Scientific Reports* 9.1 (2019): 841.

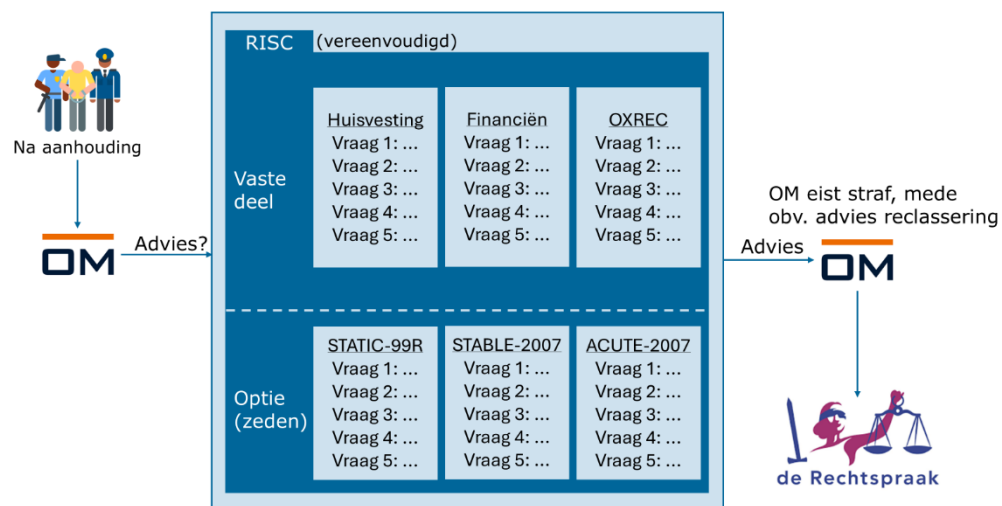
¹¹ Maas, M., E. Legters, and S. Fazel. "Professional en risicotaxatieinstrument hand in hand: Hoe de reclassering risico's inschat." *Nederlands Juristenblad* 28 (2020): 2055-2058. Pagina 2.

¹² Raad voor Strafrechttoepassing en Jeugdbescherming. "Advies Risicotaxatie in de strafrechtstoepassing" (2021).

¹³ Nationaal Rapporteur Mensenhandel en Seksueel Geweld tegen Kinderen. "Gewogen risico. Deel 2: Behandeling opleggen aan zedendelinquenten". (2017).

risicotaxatie-algoritmes staat ook beschreven in een algemene maatregel van bestuur.¹⁴

Het proces waarin de RISC ingezet wordt, inclusief de algoritmes die daarin gebruikt worden, is schematisch weergegeven in figuur 1. Dit voorbeeld betreft een casus waarin een verdachte wordt aangehouden en het OM de reclassering om advies vraagt. De reclassering gebruikt hiervoor de RISC waarin ook de OXREC gebruikt wordt. Bij mannelijke zedendelicten kunnen de STATIC-99R, STABLE-2007 en ACUTE-2007 ook worden ingezet.



Figuur 1. Gebruik van de RISC bij aanhouding van een verdachte. In dit voorbeeld wordt het gebruik van de RISC vereenvoudigd toegelicht. Het OM kan de reclassering om advies vragen over deze verdachte. Hierbij maakt de reclassering gebruik van de RISC die ook algoritmes bevat om het risico op recidive te voorspellen. Het OM besluit mede op basis van het advies van de reclassering om een bepaalde straf te eisen bij de rechter, waarna de rechter een uitspraak doet.

2.3 Onderzoek naar de reclassering

In een vooronderzoek van de Inspectie JenV zijn risico's naar voren gekomen over de algoritmes die gebruikt worden door de reclassering. Zo zijn er in het verleden zorgen geuit over het gebruik van mogelijk discriminerende variabelen in de OXREC.¹⁵ Daarnaast hebben de uitkomsten van deze algoritmes potentieel grote consequenties voor mensen. De algoritmes nemen een belangrijke plek in het adviseringsproces van de reclassering in en bij gerechtelijke uitspraken wordt verwezen naar de uitkomsten van die algoritmes.

¹⁴ Besluit van 12 juni 2008, houdende regels inzake de beslissing over de voorwaarden en het uit te oefenen toezicht bij voorwaardelijke invrijheidstelling (Uitvoeringsbesluit voorwaardelijke invrijheidstelling)
¹⁵ van Dijck, Gijs. "Algoritmische risicotaxatie van recidive. Over de Oxford Risk of Recidivism tool (OXREC), ongelijke behandeling en discriminatie in strafzaken." *Nederlands Juristenblad* 95.25 (2020): 1784-1790.

In het licht van de geconstateerde risico's en de mogelijke impact op mensen, heeft de Inspectie JenV besloten om onderzoek te verrichten naar de algoritmes die de reclassering gebruikt. Het onderzoek heeft zich daarbij gericht op algoritmes die van invloed zijn op de kwaliteit van de taakuitvoering en/of die mensen in een kwetsbare positie kunnen raken.

2.4 Leeswijzer

De Inspectie JenV heeft voor dit onderzoek geïnventariseerd welke algoritmes gebruikt worden door de reclassering. Deze inventarisatie staat beschreven in Hoofdstuk 3. Vervolgens is er nader onderzoek gedaan naar de meest risicovolle algoritmes. Hierin zijn de algoritmes getoetst op:

- Menselijke controle;
- Technische robuustheid;
- Privacy en datagovernance;
- Transparantie;
- Diversiteit, non-discriminatie en rechtvaardigheid;
- Governance.

Deze bevindingen zijn beschreven in Hoofdstuk 4. De wijze waarop de Inspectie JenV het onderzoek heeft uitgevoerd is beschreven in bijlage B.

3. Inventarisatie algoritmes

In dit onderzoek is geïnventariseerd welke algoritmes de reclassering gebruikt. Naar aanleiding van deze uitvraag geeft de reclassering aan de volgende zeven algoritmes te gebruiken bij hun taakuitvoering.

MATCHTOOL

Veroordeelde personen die een werkstraf moeten uitvoeren, kunnen via dit zelf-ontwikkelde query-algoritme worden gekoppeld aan werkstraflocaties. Op basis van onder andere reisafstand, de beschikbaarheid van de verdachte of veroordeelde en de beschikbaarheid van de werkstrafprojecten wordt een aantal werkstrafprojecten getoond. Hierbij wordt ook problematiek van de persoon meegenomen, omdat niet iedereen geschikt is voor werkstraffen waar bijvoorbeeld ook kinderen aanwezig kunnen zijn. De reclassering geeft aan dat de reclasseringsmedewerker en niet het algoritme uiteindelijk bepaalt waar een persoon geplaatst wordt.

IFZO

Het IFZO (*Informatiesysteem Forensische ZOrg*) is vergelijkbaar met de hierboven genoemde MATCHTOOL, en wordt gebruikt om personen te koppelen aan ambulante of klinische forensische zorgaanbieders. Dit gebeurt op basis van onder andere geslacht, leeftijd, geografisch gebied en klinische problematiek. Het algoritme is ontwikkeld door de Dienst Justitiële Inrichtingen (DJI), die het algoritme gepubliceerd¹⁶ heeft in het algoritmeregister van de Nederlandse overheid.

SCIL

De SCIL (*SCreener voor Intelligentie en Licht-verstandelijke beperking*)¹⁷ is een algoritme dat de reclassering gebruikt om in te schatten of een persoon een licht-verstandelijke beperking heeft, oftewel een IQ tussen de 50 en 85. Het algoritme is extern ontwikkeld en gevalideerd in een Nederlandse populatie. De SCIL bestaat uit veertien vragen die de reclasseringsmedewerker afneemt. Voor elke vraag kan een aantal punten gescoord worden. Wanneer de persoon minder dan 20 punten behaalt, dan is er een vermoeden dat deze persoon een licht verstandelijke beperking heeft. De reclassering geeft aan dat deze indicatie medewerkers helpt om hun gesprekstechnieken aan te passen. Dit gebeurt bijvoorbeeld door vragen korter en concreter te maken in gesprekken met personen met een mogelijke lichtverstandelijke beperking.

¹⁶ <https://algoritmes.overheid.nl/nl/algoritme/oorg10114/85258929/matchen-justitiabele-aan-forensische-zorglocatie-ifzo-algoritme> (geraadpleegd op 25 juli 2025)

¹⁷ Kaal, H., Henk LI Nijman, and X. M. H. Moonen. "SCIL." *Voor volwassenen (SCIL 18+) en jongeren van 14* (2015): 14-17.

OXREC

De OXREC (*OXford risk of REcidivism tool*) is een risicotaxatie-algoritme dat op basis van dertien persoonskenmerken (zoals leeftijd, geslacht en middelenmisbruik) inschat wat de kans op recidive is, oftewel wat de kans is dat de persoon opnieuw de fout ingaat. Ook wordt een risicocategorie (laag, gemiddeld of hoog) bepaald aan de hand van deze kans. Het algoritme is extern ontwikkeld door Oxford University en is ook gevalideerd in een Nederlandse populatie. De OXREC wordt in elk adviestraject en in toezichttrajecten toegepast. Medewerkers worden geïnstrueerd om hun eigen risico-inschatting naast die van de OXREC te leggen. Medewerkers kunnen de inschatting van de OXREC overnemen of hier beargumenteerd van afwijken (zie hoofdstuk 4). Een uitgebreide omschrijving van dit algoritme is te vinden in Bijlage C.

STATIC-99R, STABLE-2007 en ACUTE-2007

De STATIC-99R, STABLE-2007 en ACUTE-2007 zijn drie risicotaxatie-algoritmes die gebruikt worden om het risico op recidive in te schatten bij mannen die verdacht zijn van of veroordeeld zijn voor zedenmisdrijven zoals kindermisbruik, aanranding en verkrachting. Na het invullen van de algoritmes volgt een risicocategorie (laag, matig-laag, matig-hoog of hoog). De algoritmes zijn extern ontwikkeld door een Noord-Amerikaanse onderzoeksgroep. De drie algoritmes, of een deel daarvan, worden toegepast bij alle adviestrajecten en toezichttrajecten met mannen die verdacht of veroordeeld zijn van zedenmisdrijven. Net als de OXREC geldt dat medewerkers de inschatting van deze drie algoritmes kunnen overnemen of hier beargumenteerd van kunnen afwijken. Een uitgebreide omschrijving van deze drie algoritmes is te vinden in bijlage D.

4. Bevindingen nader onderzoek risicotaxatie-algoritmes

De Inspectie JenV heeft besloten om nader onderzoek te doen naar de vier risicotaxatie-algoritmes, omdat deze algoritmes een belangrijk onderdeel van het primaire proces van de reclassering vormen doordat medewerkers de uitkomsten van de algoritmes expliciet meenemen in hun risico-inschattingen. Als zou blijken dat de algoritmes niet goed functioneren, dan zou dit nadelige gevolgen kunnen hebben voor zowel verdachten of veroordeelden als de maatschappij. Immers, het is mogelijk dat het opleggen van een gevangenisstraf door een rechter waarschijnlijker is als de reclassering inschat dat het risico op herhaling groot is. De feitelijke impact van de adviezen van de reclassering is niet onderzocht door de Inspectie JenV. Om deze risico's beter te kunnen inschatten is er nader onderzoek gedaan naar de vier risicotaxatie-algoritmes.

4.1 Leeswijzer

Hieronder volgen de bevindingen van de Inspectie JenV naar aanleiding van het nader onderzoek naar de risicotaxatie-algoritmes die de reclassering gebruikt: OXREC, STATIC-99R, STABLE-2007 en ACUTE-2007. Voor een beschrijving hoe deze algoritmes werken, op basis van welke vragen ze tot een risico-inschatting komen en hoe ze ontwikkeld zijn wordt verwezen naar Bijlage C voor de OXREC en Bijlage D voor de STATIC-99R, STABLE-2007 en ACUTE-2007.

Deze bevindingen hebben betrekking op de onderzoeksperiode van januari tot en met oktober 2025 en zijn gegroepeerd in de thema's die noodzakelijk zijn voor het verantwoord gebruik van algoritmes. De wijze waarop de Inspectie JenV te werk is gegaan evenals deze thema's zijn beschreven in Bijlage B.

4.2 Technische robuustheid

Hoe wordt de kwaliteit van algoritmes gemeten?

Algoritmes zoals de OXREC, STATIC-99R, STABLE-2007 en ACUTE-2007 maken voorspellingen over toekomstig gedrag. Om er achter te komen of deze voorspellingen goed zijn, is het nodig om data te verzamelen. Zo wordt de OXREC gebruikt om te voorspellen of een persoon binnen twee jaar gaat recidiveren. Hiervoor worden dan de gegevens van alle verdachten of veroordeelden van tenminste twee jaar geleden verzameld. Van deze personen is het immers bekend of ze gerecidiveerd hebben, omdat de gegevens nu twee jaar later verzameld worden.

Voor al deze personen wordt vervolgens de OXREC ingevuld en de kans op recidive volgens de OXREC berekend. Deze wordt vergeleken met de daadwerkelijke recidive. Hiermee worden twee belangrijke statistieken uitgerekend: de *Area Under the Curve* en de kallibratie.

Area Under the Curve (AUC)

De AUC geeft aan in hoeverre personen die recidiveren een hogere OXREC score hebben dan personen die niet recidiveren. De AUC is een getal tussen 0.5 en 1. Met een AUC van 0.5 kun je de conclusie trekken dat de OXREC score niks zegt over het wel of niet recidiveren. Het werkt dan niet beter dan het opgooien van een munt. Met een score van 1 krijgen alle personen die recidiveren een hogere score dan personen die niet recidiveren. In dit geval kan de OXREC dus perfect voorspellen. Tussen de 0.5-0.6 wordt de AUC doorgaans geïnterpreteerd als "niet-voorspellend", 0.6-0.7: "beperkt", 0.7-0.8: "redelijk", 0.8-0.9: "(zeer) goed" en een score van 0.9-1 als "perfect".¹⁸

Kallibratie

De kallibratie van een algoritme zoals de OXREC geeft aan in hoeverre de *hoogte* van de voorspelde kans op recidive klopt met de werkelijkheid. Stel dat er 100 personen zijn met een voorspelde kans op recidive van 20%. Dan verwacht je dat er binnen deze groep 20 personen recidiveren en 80 dus niet. Als blijkt dat in werkelijkheid slechts 10 personen recidiveren, dan is het risico op recidive voor deze 100 personen dus overschat. De kallibratie geeft dus aan of de voorspelde kansen op recidive overeenkomen met de werkelijkheid.

Door de AUC en de kallibratie te bepalen op basis van historische data krijg je een indicatie of de voorspellingen over het algemeen goed zijn. Dit geeft echter geen garantie dat het algoritme in de toekomst goed blijft werken, of dat het goede voorspelling geeft voor specifieke individuen. Het is hierom belangrijk dat de historische data waar het algoritme op getraind is zo nauw mogelijk aansluit bij de dagelijkse praktijk en dat de kwaliteit van de algoritmes regelmatig gemeten wordt.

¹⁸ Çorbacioğlu, Şeref Kerem, and Gökhan Aksel. "Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value." *Turkish journal of emergency medicine* 23.4 (2023): 195-198.

OXREC

De reclassering heeft ervoor gekozen om gebruik te maken van de OXREC mede omdat deze wetenschappelijk en internationaal ontwikkeld is en ook gevalideerd¹⁹ is op een Nederlandse populatie.

De Inspectie JenV heeft tekortkomingen geconstateerd bij de implementatie van de OXREC in Nederland, op zowel technisch als methodologisch vlak.

Software-implementatie

De Inspectie JenV heeft bij haar onderzoek ook de software-implementatie van de OXREC in het systeem van de reclassering onderzocht op mogelijke gebreken. Hierbij zijn drie gebreken aangetroffen:

1. De formules die gebruikt worden voor de inschatting van recidive bij gedetineerden en niet-gedetineerden zijn met elkaar verwisseld. Hierdoor wordt de kans op recidive bij niet-gedetineerden hoger ingeschat, en bij gedetineerden juist lager;
2. In de software-implementatie van de reclassering worden gedetineerden die minder lang dan 6 maanden gedetineerd zijn geweest beschouwd als niet-gedetineerden. Door de verwisseling van de formules voor gedetineerden en niet-gedetineerden (punt 1) heeft dit gebrek voor deze specifieke subgroep geen verdere consequenties;
3. De modelparameter voor drugsproblematiek is onjuist.²⁰ Hierdoor wordt hun risico op recidive te laag ingeschat.

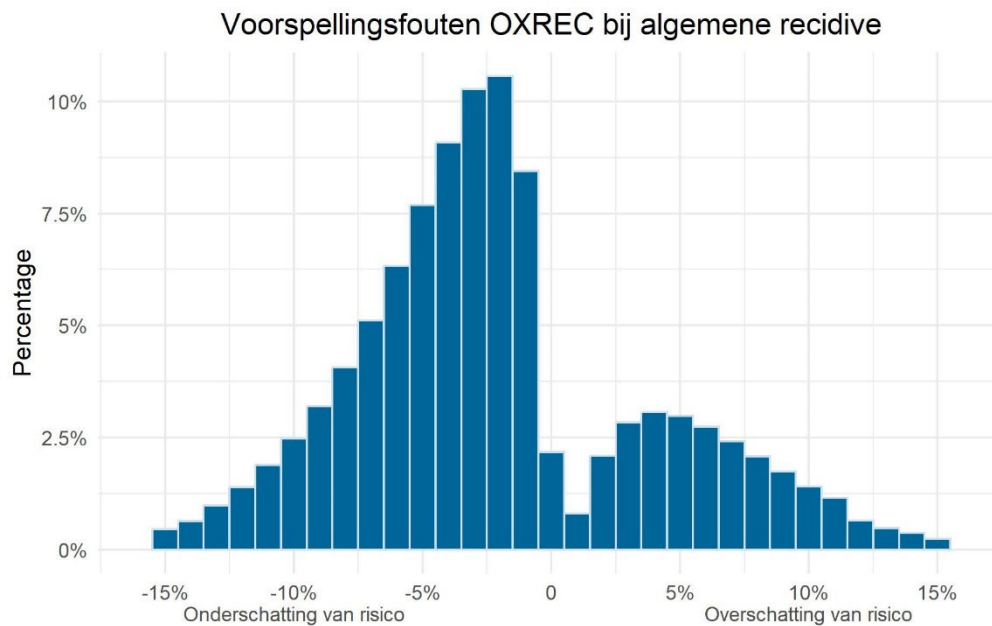
De Inspectie JenV heeft aan de hand van het versiebeheersysteem, de programmacode en het functioneel ontwerp geconstateerd dat deze gebreken aanwezig waren in de software sinds de ingebruikname van de OXREC in 2018.

¹⁹ Fazel, Seena, et al. "Prediction of violent reoffending in prisoners and individuals on probation: A Dutch validation study (OxRec)." *Scientific Reports* 9.1 (2019): 841.

²⁰ Bij de berekening van algemene recidive volgt uit de wetenschappelijke publicatie van de OXREC dat het de modelparameter voor drugsmisbruik 0,571 is. In de software-implementatie is echter het getal 0,512 aangetroffen.

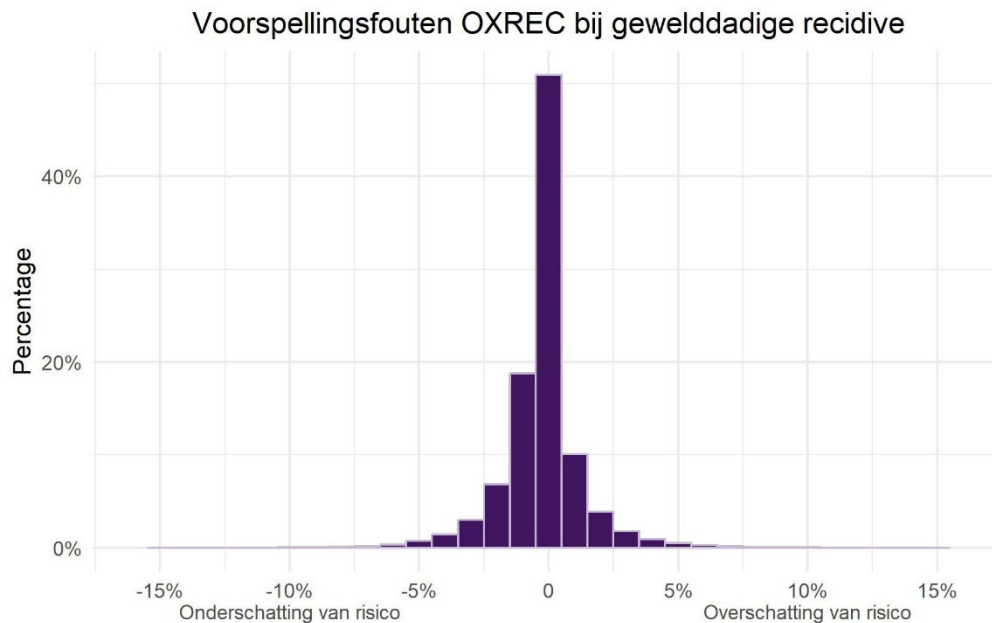
Effect van methodologische en software-implementatiegebreken

De Inspectie JenV heeft met behulp van computersimulaties het effect berekend van de aangetroffen methodologische en software-implementatiegebreken. Hierin zijn alle mogelijkheden van inputvariabelen genomen en is berekend wat de voorspellingen met en zonder de gebreken zijn. De wijze waarop de Inspectie JenV dit gedaan heeft is beschreven in Bijlage B. In figuur 2 (algemene recidive) en figuur 3 (gewelddadige recidive) zijn de voorspellingsfouten weergegeven.



Figuur 2. Voorspellingsfouten door methodologische en software-implementatiegebreken in het gebruik van de OXREC door de reclassering bij algemene recidive. In deze grafiek is weergegeven in hoeverre de gebreken hebben geleid tot een onderschatting of juist een overschatting van het risico op algemene recidive. Een onderschatting betekent dat het algoritme bij de door de ontwikkelaar beoogde toepassing geleid zou hebben tot een hogere inschatting (bijvoorbeeld 39% in plaats van 34% kans op recidive levert een onderschatting van 5% op). Een overschatting betekent juist dat de inschatting lager had moeten zijn. Hoewel er altijd een voorspellingsfout aanwezig is, zijn er gevallen waarin deze afgerond 0 zijn, waardoor deze wel zichtbaar zijn de in bovenstaande grafieken.

Uit dit figuur is op te maken dat door de voorspellingsfouten het risico op *algemene recidive* in de meeste gevallen (75%) werd onderschat, maar dat het ook voorkwam dat de risico's juist werden overschat. De gemiddelde onderschatting van het risico was 4.9% terwijl de gemiddelde overschatting 6.2% was. In beide gevallen waren er bij de voorspellingsfouten uitschieters tot meer dan 15%.



Figuur 3. Voorspellingsfouten door methodologische en software-implementatiegebreken in het gebruik van de OXREC door de reclassering bij gewelddadige recidive. In deze grafiek is weergegeven in hoeverre de gebreken hebben geleid tot een onderschatting of juist een overschatting van het risico op gewelddadige recidive. Een onderschatting betekent dat het algoritme bij de door de ontwikkelaar beoogde toepassing geleid zou hebben tot een hogere inschatting (bijvoorbeeld 39% in plaats van 34% kans op recidive levert een onderschatting van 5% op). Een overschatting betekent juist dat de inschatting lager had moeten zijn. Hoewel er altijd een voorspellingsfout aanwezig is, zijn er gevallen waarin deze afgerond 0 zijn, waardoor deze wel zichtbaar zijn de in bovenstaande grafieken.

Bij de voorspelling van de kans op *gewelddadige recidive* was in iets meer dan de helft (58%) van de gevallen sprake van een onderschatting van het risico. De voorspellingsfouten waren kleiner dan bij algemene recidive. De reden hiervoor is dat er in de OXREC er minder verschil is in gewelddadige recidive tussen gedetineerden en niet-gedetineerden. De verwisseling van de formules in de software-implementatie van de reclassering heeft hierdoor kleinere gevolgen. De gemiddelde onderschatting betrof 1% en de gemiddelde overschatting was 0.8%, met uitschieters tot ongeveer 5%.

Hoewel de voorspelde kans op recidive bij zowel algemene als gewelddadige recidive nooit klopte, hoeft dit niet altijd te leiden tot een verkeerde risicocategorie. Dit gebeurt bijvoorbeeld wanneer de voorspelde kans op algemene recidive 35% is, terwijl dit met een juiste toepassing van de berekeningen 37% had moeten zijn. In dit geval wijken de voorspellingen van elkaar af, maar vallen beide voorspellingen alsnog in de categorie 'gemiddeld' die een bandbreedte heeft tussen 30%-50%.

De Inspectie JenV heeft hierom ook onderzocht in hoeverre de aangetroffen gebreken hebben geleid tot een verandering in de risicocategorie die de OXREC oplevert (laag, gemiddeld of hoog risico). De resultaten hiervan zijn weergegeven in Tabel 1.

Afwijkingen in risicocategorie bij algemene recidive	
Correcte ²¹ risicocategorie	78,8%
Te lage risicocategorie	14,4%
Te hoge risicocategorie	6,7%
Afwijkingen in risicocategorie bij gewelddadige recidive	
Correcte risicocategorie	94,2%
Te lage risicocategorie	4,2%
Te hoge risicocategorie	1,7%

Tabel 1. Afwijkingen in risicocategorieën van de OXREC door methodologische en implementatiegebreken.

Te zien is dat bij algemene recidive de voorspellingen in ruim 21% van de gevallen leidden tot een afwijkende risicocategorie. In ruim 14% van de gevallen was de inschatting te laag. Dit gebeurde bijvoorbeeld wanneer het label 'gemiddelde kans op recidive' werd gegeven terwijl dit eigenlijk 'hoge kans op recidive' had moeten zijn. In bijna 7% van de gevallen was de inschatting juist te hoog, bijvoorbeeld wanneer de OXREC-implementatie van de reclassering op de categorie 'hoog' uitkwam, terwijl dit 'gemiddeld' had moeten zijn.

Bij gewelddadige recidive was de impact lager, met ongeveer in 6% van de gevallen een afwijkende risico-inschatting.

Methodologische bevindingen

Het algoritme inclusief modelparameters is gebaseerd op een Zweedse dataset waarin uitsluitend de gegevens van gedetineerde personen uit de periode 2001-2009 zijn meegenomen. Deze data zijn inmiddels 16-24 jaar oud. De reclassering gebruikt het algoritme echter op een Nederlandse reclasseringspopulatie anno 2025 op zowel gedetineerde personen, niet-gedeteneerde veroordeelde personen en verdachte personen. Na een validatie in 2018 door de ontwikkelaar, in opdracht van de reclassering, is er vastgesteld dat het algoritme te hoge risico-inschattingen opleverde. Hierom is het model gerekalibreerd. Dit betekent dat alle modelparameters gebaseerd op de oorspronkelijke Zweedse data nog aanwezig zijn in het algoritme, maar dat de uitkomst voor iedereen op gelijke wijze verlaagd zijn (zie Bijlage C).

Daarnaast constateert de Inspectie JenV dat gebruikte definities van variabelen afwijken van hoe het algoritme oorspronkelijk getraind is. Zo is bij de ontwikkeling van het algoritme gedefinieerd dat iemand werkt als deze persoon minimaal vier uur per maand een betaald inkomen heeft. Echter bij de toepassing in de Nederlandse reclasseringspopulatie telt vrijwilligerswerk ook mee, evenals voor de kinderen zorgen. Bij elf van de dertien variabelen wijken de definities af.²² Omdat de modelparameters van de OXREC afhankelijk zijn van de gebruikte definities, kunnen andere variabelendefinities leiden tot afwijkende risico-inschattingen.

²¹ Bij een correcte risicocategorie komt de berekende risicocategorie overeen met categorie zoals berekend zonder methodologische fouten en met de juiste modelparameters. Zie bijlage B voor meer informatie.

²² Een totaaloverzicht van de afwijkende definities is te vinden in Tabel 1 van https://static-content.springer.com/esm/art%3A10.1038%2Fs41598-018-37539-x/MediaObjects/41598_2018_37539_MOESM1_ESM.pdf

Wat zijn modelparameters?

Modelparameters zijn belangrijke instellingen in algoritmes. Het zijn de getallen die nodig zijn om de algoritmeberekening uit te voeren. Het vertelt hoe zwaar elk antwoord meetelt in de uiteindelijke berekening. De OXREC heeft voor elke antwoordoptie een eigen modelparameter. Dit is bijvoorbeeld 0,4117 voor personen die drugs misbruiken, maar het lagere getal 0 wordt toegepast voor personen die geen drugs misbruiken. Op deze wijze krijgen drugsmisbruikers een hogere voorspelde risico op recidive door het algoritme.

Bij de validatie op de Nederlandse reclasseringspopulatie is ook gebleken dat de OXREC minder goed functioneert dan bij de ontwikkeling ervan. Het vermogen van het algoritme om in te schatten of een persoon wel of niet recidiveert is 'beperkt', terwijl dit 'redelijk' was in de Zweedse populatie.²³ De onderzoekers van de validatiestudie geven aan dat het wijzigen van de definities mogelijk heeft bijgedragen aan een verslechtering van de voorspellingen.²⁴

Ook zijn er in de validatiestudie verschillen gevonden tussen de Zweedse gevangenispopulatie en de Nederlandse reclasseringspopulatie. Zo was er in de Nederlandse reclasseringspopulatie sprake van meer drugsgebruik en meer gewelddadige delicten, maar werd er juist minder gerecidiveerd.²⁵ Dit heeft ertoe geleid dat de ontwikkelaar het algoritme gekalibreerd heeft waardoor de risicoscores van alle personen op dezelfde wijze verlaagd zijn. Echter, omdat het model niet opnieuw getraind is op Nederlandse reclasseringsdata, zijn de modelparameters van de OXREC nog altijd gebaseerd op Zweedse gedetineerden die significant afwijken van de hedendaagse Nederlandse reclasseringspopulatie. Wanneer de reclassering de OXREC dus toepast op een Nederlandse reclasseringspopulatie, dan kan dit leiden tot minder accurate voorspellingen.

Tot slot constateert de Inspectie JenV dat er twee variabelen ontbreken in de toepassing van de OXREC, namelijk migratieachtergrond en de aanwezigheid van ernstige psychische problematiek. In de toepassing van de OXREC door de reclassering krijgt elke persoon, ongeacht zijn of haar afkomst, 30,75% van de modelparameter behorende bij een persoon met een migratieachtergrond.²⁶

²³ Zie eerder kader. De AUC in de originele Zweedse studie was 0.75 (tussen 0.7-0.8 wordt geclassificeerd als 'redelijk'), maar in de Nederlandse validatiestudie was deze 0.68 (tussen 0.6-0.7 wordt deze geclassificeerd als 'beperkt').

²⁴ Fazel, Seena, et al. "Prediction of violent reoffending in prisoners and individuals on probation: A Dutch validation study (OxRec)." *Scientific Reports* 9.1 (2019): 841. "Variations in the characteristics of the samples (Table 1) and these definitions may have contributed to the different performance of the Dutch validation compared to the Swedish one."

²⁵ Fazel, Seena, et al. "Prediction of violent reoffending in prisoners and individuals on probation: A Dutch validation study (OxRec)." *Scientific Reports* 9.1 (2019): 841. Pagina 2.

²⁶ De oorspronkelijke OXREC bevatte de variabele migratieachtergrond als input. De reclassering heeft er voor gekozen deze variabele uit de Nederlandse toepassing te laten door voor elke persoon, ongeacht afkomst, 30,75% van de modelparameter migratieachtergrond toe te kennen. Hier heeft de ontwikkelaar het getal 30,75% gebruikt omdat in de oorspronkelijke Zweedse populatie in de periode van 2001-2009 30,75% van de gedetineerde personen een migratieachtergrond had.

Daarnaast wordt de OXREC toegepast alsof geen enkele persoon een ernstige psychische aandoening (zoals psychoses) heeft, ook als ze die wel hebben. Deze wijze van variabelen weglaten uit het model wordt normaal gesproken toegepast als in een gehele populatie risico's berekend worden (zoals bij een model validatie). Maar bij de toepassing op individuen, zoals bij het gebruik van de OXREC door de reclassering, leidt dit tot voorspellingsfouten op individueel niveau. Zo is de OXREC ontwikkeld om individuele personen met bijvoorbeeld een ernstige psychische aandoening een hogere risicoscore toe te kennen, maar gebeurt dit niet in de toepassing door de reclassering. Reclasseringsmedewerkers kunnen in de toepassing van het algoritme niet aangeven of een persoon een ernstige psychische aandoening heeft. Hierdoor krijgen personen met een ernstige psychische aandoening een te lage risico-inschatting.

Hoewel er goede redenen kunnen zijn om variabelen weg te laten uit een algoritme, is de manier waarop van belang om tot goede voorspellingen te komen. Het is hierbij belangrijk om op te merken dat alle modelparameters in samenhang vastgesteld worden en gecorrigeerd worden voor elkaar. Het wijzigen van de definities van een voorspeller, het toevoegen of juist het verwijderen van één of meer variabelen heeft in de praktijk invloed op alle modelparameters. Het verwijderen van variabelen uit het algoritme wordt gedaan door alle modelparameters opnieuw te bepalen middels het hertrainen van de OXREC. Dit had bijvoorbeeld gedaan kunnen worden bij de validatie van de OXREC op de Nederlandse reclasseringspopulatie in 2017. Ook had de reclassering de gegevens die verzameld zijn sinds de ingebruikname van het algoritme in 2018 kunnen gebruiken om de OXREC te hertrainen. Dit is echter niet gebeurd.

STATIC-99R, STABLE-2007, ACUTE-2007

De STATIC-99R, STABLE-2007, ACUTE-2007 zijn drie algoritmes die door de reclassering gebruikt worden om de kans op recidive in te schatten bij mannelijke zedendelinquenten en/of -verdachten. Dit betreffen misdrijven zoals verkrachting en aanranding. In tegenstelling tot de OXREC, heeft de reclassering deze algoritmes niet zelf gevalideerd of laten valideren.

Software-implementatie

De Inspectie JenV heeft geen gebreken aangetroffen in de software-implementatie van deze drie algoritmes, maar plaatst wel een kanttekening wanneer er gebruik gemaakt wordt van meer dan één algoritme. De handleidingen van de STATIC-99R, STABLE-2007 en ACUTE-2007 stellen dat er een combinatiescore berekend moet worden wanneer er meerdere algoritmes tegelijkertijd gebruikt worden (zie Bijlage D). Dit is echter niet geïmplementeerd in de software van de reclassering. Hierdoor ontstaat er het risico dat medewerkers de scores op onjuiste wijze samenvoegen in hun risico-inschattingen.

Methodologische bevindingen

De STATIC-99R²⁷ is gepubliceerd in 2012 met enkele aanpassingen op het oorspronkelijke STATIC-99²⁸ algoritme uit 2000. Het algoritme is ontwikkeld aan de hand van een voornamelijk Noord-Amerikaanse populatie waar data voor verzameld is afkomstig uit de periode 1957-2007. De STABLE-2007 en ACUTE-2007 zijn allebei ontwikkeld in 2007²⁹ aan de hand van een Noord-Amerikaanse populatie. De wetenschappelijk studie maakt niet duidelijk wanneer de data precies zijn verzameld.

De drie algoritmes worden wereldwijd veel gebruikt om het recidiverisico te voorspellen bij mannelijke zedendelinquenten. Uit documentatie van de reclassering blijkt dat alleen de STATIC-99R gevalideerd is in een Nederlandse populatie met mannelijke zedendelinquenten, maar niet in opdracht van de reclassering. Deze validatie³⁰ omvatte data uit 1995-2002 van Nederlandse gedetineerden en personen die opgenomen waren in een zorginstelling. Daarmee verschilt deze validatie van de reclasseringspopulatie die ook personen bevat die (nog) niet veroordeeld zijn of niet gedetineerd zijn. Ook zijn de gebruikte data inmiddels tussen de 23 en 30 jaar oud. Het vermogen van het algoritme om in te schatten of een persoon wel of niet recidiveert is 'redelijk'³¹ in deze populatie. De kallibratie van de STATIC-99R is echter niet onderzocht. Dit betekent dat het onbekend is of het terecht is dat personen die bijvoorbeeld een score van 6 krijgen geclassificeerd worden als 'hoog risico' of dat bijvoorbeeld de categorie 'matig-hoog' passender zou zijn.

Uit de documentatie van de reclassering blijkt niet dat de STABLE-2007 en ACUTE-2007 gevalideerd zijn in een Nederlandse populatie, hierdoor is het onduidelijk of de algoritmes goed functioneren bij Nederlandse mannelijke zedendelinquenten en/of -verdachten.

Wanneer er een combinatie van de STATIC-99R, STABLE-2007 en/of ACUTE-2007 wordt afgenomen, dan volgt een samengestelde risicocategorie (zie Bijlage D). Ook de mate waarin deze samengestelde score samenhangt met het wel/niet recidiveren is niet onderzocht in een Nederlandse populatie.

Tot slot constateert de Inspectie JenV dat de software-implementatie van de drie zedenrecidive-algoritmes, in tegenstelling tot de OXREC, enkel risicocategorieën opleveren. Na het invullen van deze algoritmes volgt de uitkomst 'laag' 'laag-matig', 'matig-hoog' of 'hoog' zonder dat daar een (numerieke) kans aan gekoppeld wordt. Uit documentatie van de ontwikkelaar van de STATIC-99R blijkt dat de kwalificatie 'laag' overeenkomt met 0%-7.8% kans op recidive, 'matig-laag' met

²⁷ Helmus, Leslie, et al. "Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights." *Sexual Abuse* 24.1 (2012): 64-101.

²⁸ Hanson, R. Karl, and David Thornton. "Improving risk assessments for sex offenders: A comparison of three actuarial scales." *Law and Human Behavior* 24.1 (2000): 119-136.

²⁹ Hanson, R. Karl, et al. *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*. Vol. 5. No. 6. Ottawa, Ontario: Public Safety Canada, 2007.

³⁰ Smid, Wineke J., et al. "A comparison of the predictive properties of nine sex offender risk assessment instruments." *Psychological assessment* 26.3 (2014): 691.

³¹ Zie kader hierboven. De AUC in de Nederlandse validatiestudie was 0.74 (tussen 0.7-0.8 wordt deze geclassificeerd als 'redelijk').

7.8%-9.7%, 'matig-hoog' met 9.7%-19.6% en 'hoog' vanaf 19.6%.³² Dit verschilt met de risicocategorieën van de OXREC waar bijvoorbeeld gesproken wordt over een hoog risico op recidive vanaf 30% (gewelddadige recidive) en 50% (algemene recidive). Omdat de risicocategorieën per algoritme verschillen en de reclasseringsmedewerker geen numerieke kans op recidive ziet bij toepassing van de STATIC-99R, STABLE-2007 en ACUTE-2007, ziet de Inspectie JenV een risico op het onjuist interpreteren van de risicocategorieën.

4.3 Privacy en datagovernance

De Algemene Verordening Gegevensbescherming (AVG) schrijft de regels voor bij het verwerken van persoonsgegevens. Eén van de verplichtingen is het uitvoeren van een zogeheten *Data Protection Impact Assessment* (DPIA) bij gegevensverwerkingen die een hoog privacy risico opleveren. In een DPIA worden privacyrisico's in kaart gebracht, geanalyseerd en worden maatregelen beschreven om deze risico's zoveel mogelijk te verkleinen.³³ Dit is van belang om de rechten en vrijheden van personen zoveel mogelijk te beschermen. Omdat de vier risicotaxatie-algoritmes op basis van persoonskenmerken toekomstig gedrag (recidive) beogen te voorspellen, zijn het profilerende algoritmes. De AVG schrijft voor dat in dit geval een DPIA verplicht is.³⁴

Bij het onderzoek heeft de reclassering aangegeven dat er geen DPIA opgesteld is voor het gebruik van de risicotaxatie-algoritmes, maar dat er wel een DPIA is uitgevoerd voor het complete adviestraject. De Inspectie JenV constateert dat deze DPIA echter geen informatie of verwijzingen bevat naar het gebruik van (profilerende) algoritmes en dat er geen risico's of maatregelen zijn beschreven ten aanzien van het gebruik van algoritmes.

Van de drie reclasseringsorganisaties vermeldt de Reclassering Nederland in haar privacyverklaring op haar website³⁵ dat de gegevens van verdachten of veroordeelden gebruikt kunnen worden voor de risico-inschattingen (RISC) die zij verricht, en daarmee ook impliciet voor de risicotaxatie-algoritmes. In de privacyverklaringen van het Leger des Heils Jeugdbescherming en Reclassering, en Stichting Verslavingsreclassering GGZ (inclusief de negen onderliggende instellingen) is het gebruik van de RISC en de risicotaxatie-algoritmes niet opgenomen.

³² Helmus, L. Maaïke, et al. "Static-99R & Static-2002R Evaluators' Workbook" (2021).

³³ <https://www.autoriteitpersoonsgegevens.nl/themas/basis-avg/praktisch-avg/data-protection-impact-assessment-dpia> onder "Uitvoering DPIA".

³⁴ Besluit lijst verwerkingen persoonsgegevens waarvoor een gegevensbeschermingseffectbeoordeling (DPIA) verplicht is(15).

³⁵ <https://www.reclassering.nl/privacy/privacyverklaring/>

4.4 Transparantie

De OXREC, inclusief modelparameters en formules, is openbaar gepubliceerd.³⁶ Doordat de OXREC een lineair regressiemodel is, is de interpretatie van de modelparameters gemakkelijker dan veel andere type algoritmes. Variabelen zijn ofwel risicoverhogend ofwel risicoverlagend, en een uiteindelijke voorspelling kan herleid worden tot deze risicofactoren. Medewerkers krijgen echter bij het gebruik van de OXREC in het softwaresysteem van de reclassering geen inzicht in deze modelparameters. Hiermee is het voor een medewerker niet na te gaan welke factoren het meest hebben bijgedragen aan bijvoorbeeld een hoogrisico-inschatting. Dit is belangrijk om de risico-inschatting in context te plaatsen en om het resultaat van de OXREC beter te begrijpen.

De STATIC-99R, STABLE-2007 en ACUTE-2007 zijn ook alle drie openbaar gepubliceerd.³⁷ De werking van deze algoritmes is erg transparant: elke vraag levert een geheel aantal punten op en de optelsom leidt tot een risicocategorie. Medewerkers krijgen in tegenstelling tot de OXREC wel inzicht in deze punten waardoor ze achteraf na kunnen gaan welke factoren het meest hebben bijgedragen aan de uiteindelijke risico-inschatting.

Op alle adviesrapportages staat duidelijk vermeld dat de OXREC is toegepast en wat de voorspelde numerieke kans op recidive is, inclusief de aanduiding laag/gemiddeld/hog. Indien de zedenrecidive-algoritmes zijn toegepast, dan wordt dit ook vermeld op rapportages. Op deze wijze belanden de uitkomsten van de risicotaxatie-algoritmes ook in de rechtbank.

Een subset van gerechtelijke uitspraken wordt online gepseudonimiseerd openbaar gemaakt. De Inspectie JenV heeft in 97 uitspraken³⁸ verwijzingen gevonden naar inschattingen van de OXREC. Dit ging in alle gevallen over de voorspelling van algemene recidive door de OXREC. Bij deze vermeldingen werden de risicocategorieën beschreven en in enkele gevallen ook de bijbehorende numerieke kansen als onderbouwing van het reclasseringsadvies. Van de STATIC-99R zijn 65 verwijzingen³⁹ aangetroffen in openbare gerechtelijke uitspraken, en van de STABLE-2007 30 vermeldingen⁴⁰ als onderbouwing van het reclasseringsadvies. De Inspectie JenV heeft geen vermeldingen van de ACUTE-2007 aangetroffen.

Van de drie reclasseringsorganisaties vermeldt enkel de Reclassering Nederland op haar website dat de risicotaxatie-algoritmes toegepast worden.⁴¹ Vooralsnog heeft

³⁶ Fazel, Seena, et al. "Prediction of violent reoffending in prisoners and individuals on probation: A Dutch validation study (OxRec)." *Scientific Reports* 9.1 (2019): 841

³⁷ Helmus, Leslie, et al. "Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights." *Sexual Abuse* 24.1 (2012): 64-101 en Hanson, R. Karl, et al. *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*. Vol. 5. No. 6. Ottawa, Ontario: Public Safety Canada, 2007.

³⁸ <https://uitspraken.rechtspraak.nl/resultaat?zoekterm=OXREC&inhoudsindicatie=zt0&publicatiestatus=ps1> (bezoekt op 25-07-2025)

³⁹ <https://uitspraken.rechtspraak.nl/resultaat?zoekterm=STATIC-99R%20reclassering&inhoudsindicatie=zt0&publicatiestatus=ps1&sort=Relevance> (bezoekt op 25-07-2025)

⁴⁰ <https://uitspraken.rechtspraak.nl/resultaat?zoekterm=%22stable-2007%22%20%22reclassering%22&inhoudsindicatie=zt0&publicatiestatus=ps1&sort=Relevance> (bezoekt op 25-07-2025)

⁴¹ <https://www.reclassering.nl/naar-de-reclassering/advies/risc/>

de reclassering de risicotaxatie-algoritmes niet gepubliceerd in het algoritmeregister van de overheid, maar uit de reclassering het voornemen dit wel te doen voor het einde van 2025.

4.5 AI-verordening

Naar inzicht van de Inspectie JenV is de OXREC een AI-systeem⁴² waar de AI-verordening op van toepassing is. De AI-verordening is nieuwe wetgeving die in 2024 in werking is getreden en wettelijke vereisten stelt voor het gebruik van AI-systemen. De reclassering heeft tijdens het onderzoek aangegeven dat de OXREC geen AI-systeem is, omdat de OXREC niet *continu* zelflerend is en zichzelf dus niet voortdurend aanpast. Dit is echter geen verplichte eigenschap van een AI-systeem volgens de AI-verordening. Een volledige onderbouwing waarom de Inspectie JenV de OXREC als AI-systeem beschouwt, is opgenomen in Bijlage E.

Omdat de reclassering oordeelt dat de OXREC geen AI-systeem is, zijn er vanuit de reclassering geen stappen ondernomen om aan de wettelijke verplichtingen voortkomend uit de AI-verordening te gaan voldoen.

De OXREC voldoet op het eerste gezicht aan de definitie van een verboden AI-systeem: het is immers een algoritme dat beoogt te voorspellen of personen strafbare feiten zullen plegen op basis van profilering.⁴³ De uiteindelijke risico-inschatting die de medewerker maakt vindt echter niet *uitsluitend* plaats door middel van profilering.⁴⁴ Hiermee voldoet de OXREC aan de uitzonderingclausule van verboden AI-systemen en beschouwt de Inspectie JenV de OXREC als een hoog-risico⁴⁵ AI-systeem.

Hoog-risico AI-systemen moeten aan strenge eisen voldoen om gebruikt te mogen worden. Deze eisen staan beschreven in artikelen 8 tot en met 15 van de AI-verordening. De systemen moeten voldoen aan vereisten ten aanzien van een systeem voor risicobeheer, datagovernance, technische documentatie van het AI-systeem, logging, transparantie en informatieverstrekking, menselijk toezicht, nauwkeurigheid, robuustheid, grondrechtenbeoordeling en cyberbeveiliging. Voor de reclassering geldt dat er vanaf 2 augustus 2030 aan de meeste eisen voldaan moet worden, omdat de OXREC een bestaand AI-systeem is dat al langer in gebruik is. Aan de eis van AI-geletterdheid, het opleiden van reclasseringsmedewerkers om verantwoord met AI om te gaan, moet wel nu al voldaan worden.

Omdat de rekenregels in de STATIC-99R, STABLE-2007 en ACUTE-2007 door de mens bedacht zijn, vertonen deze algoritmes geen inferentievermogen en beschouwt de Inspectie JenV deze algoritmes derhalve niet als AI-systemen. Hiermee vallen deze algoritmes niet onder de AI-verordening en hoeft de

⁴² Artikel 3(1) AI-verordening.

⁴³ Artikel 5(1d) AI-verordening.

⁴⁴ Overweging 201 van "Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)".

⁴⁵ Overweging 204 van "Commission Guidelines on prohibited artificial intelligence practices established by Regulation (EU) 2024/1689 (AI Act)" en Bijlage III 6(d) AI-verordening.

reclassering niet te voldoen aan de regels die in deze verordening beschreven staan.

4.6 Diversiteit, non-discriminatie en rechtvaardigheid

Voordat de OXREC in gebruik is genomen bij de reclassering, is er voor gekozen om de variabele "migratiestatus" te verwijderen uit de software-implementatie van de reclassering. Dit is gedaan omdat het algoritme anders direct onderscheid zou maken op grond van iemands etniciteit dat in strijd is met artikel 1 van de grondwet.

In 2020 publiceerde hoogleraar Gijs van Dijck een kritisch artikel⁴⁶ over de OXREC. Hij wees erop dat het gebruik van de buurtscore als variabele in het algoritme mogelijk leidt tot etnisch profileren. Van Dijck stelde dat doordat er wijken zijn waarin relatief veel mensen wonen met een migratieachtergrond, iemands buurtscore indirect kan wijzen op iemands etniciteit.⁴⁷ Hiermee zou de buurtscore fungeren als een zogeheten *proxy*-variabele voor iemands etniciteit en daarmee bij kunnen dragen aan discriminatie.

In een reactie hierop stelde de reclassering dat in de OXREC "geen etnische variabelen ten grondslag aan de factoren die uitgevraagd worden. In de Nederlandse versie wordt niet gevraagd naar etniciteit"⁴⁸. De minister voor Rechtsbescherming heeft in antwoord op Kamervragen ook een reactie gegeven over het mogelijk etnisch profileren door de OXREC. Hierin zegt de minister: "In het aangehaalde artikel wordt gesteld dat het opnemen van de buurtscore indirect kan leiden tot etnische profilering. Door de buurtscore in de OxRec op te nemen wordt niet op basis van ras, etniciteit, religie of afkomst geselecteerd maar op basis van de sociaal-economische status van de verdachte."⁴⁹

In de reactie van zowel de minister als de reclassering wordt verder niet onderbouwd waarom de buurtscore geen *proxy* zou zijn voor iemands etniciteit.

Het College voor de Rechten van de Mens (CRM) heeft in 2021 en in 2025 een toetsingskader voor risicoprofilering gepubliceerd.⁵⁰ Dit is een toetsingskader met normen voor het gebruik van risicoprofilering gericht op het voorkomen van discriminatie. In een interview met de Inspectie JenV geeft het CRM aan dat de normen uit het toetsingskader ook van toepassing zijn op de OXREC. De OXREC is immers een profilerend algoritme dat mogelijk nadelige gevolgen voor personen kan hebben. De Inspectie JenV stelt vast dat twee van de variabelen die gebruikt worden in de OXREC door het CRM geclassificeerd worden als waarschijnlijke *proxy* voor ras of nationaliteit. Dit gaat om de 'buurtscore' en de 'hoogte van het

⁴⁶ van Dijck, Gijs. "Algoritmische risicotaxatie van recidive. Over de Oxford Risk of Recidivism tool (OXREC), ongelijke behandeling en discriminatie in strafzaken." *Nederlands Juristenblad* 95.25 (2020): 1784-1790.

⁴⁷ Van Teeffelen, K. "Algoritmes gebruikt door reclassering zorgen voor etnisch profileren" *Trouw* (26 juni 2020).

⁴⁸ Maas, M., E. Legters, and S. Fazel. "Professional en risicotaxatieinstrument hand in hand: Hoe de reclassering risico's inschat." *Nederlands Juristenblad* 28 (2020): 2055-2058. Pagina 4.

⁴⁹ Tweede Kamer. "Antwoord op vragen van het lid Van Nispen over twijfels bij het gebruik van een bepaald risicotaxatie-instrument bij de reclassering vanwege het gevaar op etnisch profileren" (18 augustus 2020). Pagina 1.

⁵⁰ Het toetsingskader uit 2025 bevat dezelfde uitgangspunten die relevant zijn voor dit onderzoek als het eerdere toetsingskader uit 2021.

inkomen'. Dat betekent dat het zeer waarschijnlijk is dat deze criteria leiden tot indirect onderscheid op grond van ras of nationaliteit. Dit is slechts onder strikte voorwaarden gerechtvaardigd. Zo moet onder andere aangetoond worden dat deze variabelen de kwaliteit van de OXREC substantieel verhogen, is continue kwaliteitsmonitoring noodzakelijk, moeten er geen redelijke andere alternatieven beschikbaar zijn, moet het gebruik proportioneel zijn en moeten de voordelen opwegen tegen de nadelen voor individuen en de samenleving. Een volledig overzicht van de benodigde stappen zijn gepubliceerd in het Toetsingskader Risicoprofilering.⁵¹

De Inspectie JenV stelt vast dat niet aan deze voorwaarden voldaan is. Zo is er bijvoorbeeld geen vergelijkend onderzoek gedaan naar het verschil in de kwaliteit van de OXREC zonder deze twee *proxy*-variabelen. De reclassering zelf stelt overigens dat de buurtscore "minimale impact" heeft op de uiteindelijke voorspelling.⁵² Hiermee is het waarschijnlijk dat het gebruik van de buurtscore de kwaliteit van de OXREC niet substantieel verhoogt. De impact van het weglaten van de hoogte van het inkomen is voorsnog onbekend. Daar komt bij dat ook als deze variabelen wel een grotere mate van kwaliteit zouden toevoegen, het onzeker is dat het gebruik van deze variabelen vervolgens de strenge proportionaliteitstoets kan doorstaan. Het CRM stelt in haar Toetsingskader Risicoprofilering dat het gebruik van algoritmes die niet aan deze voorwaarden voldoen, verboden zijn op grond van de Grondwet artikel 1 en internationale verdragen.⁵³

De factoren op basis waarvan de STATIC-99R, STABLE-2007 en ACUTE-2007 tot een risico-inschatting komen, komen niet voor op de lijst van verboden variabelen van het CRM. Hierdoor constateert de Inspectie JenV een lager risico op discriminatie.

De Inspectie JenV constateert dat de reclassering sinds de publicatie van Van Dijk in 2020 geen stappen heeft ondernomen om eventuele discriminatie te voorkomen, te meten of te verminderen. Zo is er geen mensenrechtentoets uitgevoerd, is er geen DPIA uitgevoerd, is er geen onderzoek gedaan of er sprake is van etnisch profileren en zijn de twee verboden variabelen niet verwijderd uit de OXREC.

4.7 Menselijke controle

De OXREC wordt toegepast als onderdeel van de RISC waarmee medewerkers op gestructureerde wijze tot een risico-inschatting komen (zie inleiding). Vanuit het besef dat risico-inschattingen over recidive soms onbewust gebeuren vanuit 'onderbuikgevoelens' (heuristieken), acht de reclassering het van belang dat de

⁵¹ College voor de Rechten van de Mens. "Toetsingskader Risicoprofilering" (2025).

⁵² Maas, M., E. Legters, and S. Fazel. "Professional en risicotaxatieinstrument hand in hand: Hoe de reclassering risico's inschat." *Nederlands Juristenblad* 28 (2020): 2055-2058: "Deze factor [buurtscore] heeft minimale impact op de waarschijnlijkheidsscores: goede danwel slechtere buurtscores beïnvloeden deze waarschijnlijkheidsscores met gemiddeld 0-2%."

⁵³ College voor de Rechten van de Mens. "Toetsingskader Risicoprofilering" (2025). Pagina 18.

risico-inschattingen zo gestructureerd mogelijk tot stand komen en dat de risicotaxatie-algoritmes daarin ook een ondersteunende rol hebben.⁵⁴

Dit standpunt wordt ook duidelijk uitgedragen in handleidingen, externe onderzoeken⁵⁵ en in gesprekken met reclasseringsmedewerkers. In trainingsmateriaal wordt aangegeven dat de risicotaxatie-algoritmes uitsluitend ter aanvulling gebruikt mogen worden en niet het oordeel van de medewerkers vervangen. De medewerker wordt verteld dat afwijken van de algoritmes mogelijk is.

Desalniettemin ziet de Inspectie JenV wel risico's ten aanzien van betekenisvolle menselijke tussenkomst.

Hoewel afwijken van de OXREC mogelijk is, moet dit beargumenteerd gebeuren. De Raad voor Strafrechtstoepassing en Jeugdbescherming schrijft hier het volgende over: "De reclasseringswerker vergelijkt het eigen gestructureerde oordeel met de uitkomst van de OXREC, wanneer dit afwijkt kan de reclasseringswerker het eigen oordeel bijstellen of daar beredeneerd van afwijken".⁵⁶ Ook onderzoek van de VU gebruikt een soortgelijke uitdrukking: "Wanneer de reclasseringswerker een verschil constateert kan hij of zij het eigen oordeel bijstellen of, wanneer daar argumenten voor zijn, ongewijzigd laten (beredeneerd afwijken)".⁵⁷

In trainingsmateriaal wordt medewerkers verteld dat de OXREC betrouwbaar is en serieus genomen moet worden, en dat het oordeel van de medewerkers zonder de OXREC even betrouwbaar is als het opgooien van een muntje.⁵⁸ Ook worden in trainingsmateriaal reclasseringsmedewerkers in sommige situaties gestimuleerd om de uitkomst van de OXREC over te nemen in een situatie waarin deze afwijkt van hun eigen oordeel.⁵⁹ De Inspectie JenV ziet hierin een verhoogd risico op *automation bias* waarin het mogelijk is dat medewerkers de prestaties van algoritmes overschatten en de uitkomst te snel aannemen als waarheid.⁶⁰ Hiermee lijkt de OXREC een meer voorschrijvende dan puur ondersteunende rol te hebben.

Om een beeld te krijgen over het gebruik van de OXREC in de praktijk zijn er door de Inspectie JenV ook interviews gevoerd met vijf reclasseringsmedewerkers. In

⁵⁴ Maas, M., E. Legters, and S. Fazel. "Professional en risicotaxatieinstrument hand in hand: Hoe de reclassering risico's inschat." *Nederlands Juristenblad* 28 (2020): 2055-2058.

⁵⁵ Zoals Van den Berg, Chantal, et al. "Validatiestudie risico-en beschermende factoren van de RISC: Een evaluatieonderzoek naar de leefgebieden van het risicotaxatie en adviesinstrument van de 3RO." (2021) en Raad voor Strafrechtstoepassing en Jeugdbescherming. "Risicotaxatie in de strafrechtstoepassing" (2021)

⁵⁶ Raad voor Strafrechtstoepassing en Jeugdbescherming. "Risicotaxatie in de strafrechtstoepassing" (2021)

⁵⁷ Van den Berg, Chantal, et al. "Validatiestudie risico-en beschermende factoren van de RISC: Een evaluatieonderzoek naar de leefgebieden van het risicotaxatie en adviesinstrument van de 3RO." (2021). Pagina 8.

⁵⁸ Citaat uit trainingspodcast reclasseringsacademie over de OXREC: "Om tot een goede inschatting te komen dan is het belangrijk om beide [de OXREC en het eigen oordeel] te gebruiken, schuif de OXREC dus niet zomaar aan de kant. Je eigen oordeel is net zo betrouwbaar als het opgooien van munt en of hij dan op kop of munt komt."

⁵⁹ Citaat uit trainingspodcast reclasseringsacademie over de OXREC: "Maar als ik je goed begrijp dan kan het dus zijn dat het instrument een lage recidivekans inschat, maar dat wij denken dat de kans op recidive gemiddeld is. Dan zou je dus naar de score [percentage] kunnen kijken of die misschien één of twee punten verwijderd is van de score gemiddeld en daar dus je professioneel oordeel op aanpassen."

⁶⁰ Autoriteit Persoonsgegevens "Betekenisvolle Menselijke Tussenkomst" (2025). Pagina 11. "*Automation bias* verwijst naar het idee dat mensen het prestatievermogen en de nauwkeurigheid van algoritmes vaak overschatten. We hebben te veel vertrouwen in de werking van algoritmes, zelfs wanneer deze fouten maken."

deze gesprekken wordt het voorschrijvende karakter deels bevestigd. Zo geven medewerkers aan dat de OXREC even zwaar meeweegt in het eindoordeel als hun eigen inschatting, en dat ze minder na hoeven te denken en te onderbouwen door het gebruik van de OXREC. Ook wordt aangegeven dat in twijfelgevallen de OXREC soms doorslaggevend kan zijn. Aan de andere kant geven sommige medewerkers aan dat een volledige onderbouwing van hun risico-inschatting toch gegeven moet worden en dat afwijken van de OXREC daarin zeker mogelijk is.

Uit cijfers opgevraagd door de Inspectie JenV blijkt dat in de periode 2023-2024 de OXREC ruim 43.500 keer per jaar wordt toegepast en dat het oordeel van de medewerkers in 78% van de gevallen overeenkwam met de uitkomst van de OXREC. Het is hierbij niet vast te stellen of de eindoordelen van de medewerkers anders geweest zouden zijn zonder het gebruik van de OXREC. In een kwantitatieve analyse van de VU is echter wel onderzocht in welke mate de OXREC bijdraagt aan het eindoordeel van de reclasseringsmedewerker. Hieruit blijkt dat de OXREC-score samen met het delictverleden de meeste invloed hebben op het eindoordeel van de medewerker.⁶¹ De VU stelt dat de OXREC hierbij geen doorslaggevende rol heeft op het eindoordeel. In de toepassing van de OXREC om geweldsrecidive in te schatten is de OXREC de meest bepalende factor.⁶²

Nadat de medewerkers informatie verzameld hebben over relevante zaken zoals werk, huisvesting en houding, wordt de OXREC ingevuld. De OXREC wordt voor elke verdachte of veroordeelde verplicht ingevuld. Zowel de percentuele uitkomst als de bijbehorende risicocategorie (laag, gemiddeld, hoog) wordt automatisch ingevuld in de rapportage. Medewerkers kunnen de uitkomst dus niet zomaar naast zich neerleggen.

Het invullen van de OXREC geldt als de laatste stap voordat de medewerker een eigen eindoordeel geeft. Medewerkers krijgen de uitslag van de OXREC te zien en in de reclasseringssoftware wordt direct daarna gevraagd om een eindoordeel te geven. Hiermee ziet de Inspectie JenV een risico op *anchoring bias*, omdat het eindoordeel van de medewerker gestuurd kan worden door de uitkomst van het algoritme en medewerkers eerder geneigd zouden kunnen zijn dit over te nemen.⁶³

Tot slot constateert de Inspectie JenV dat medewerkers niet voldoende uitgelegd wordt hoe de OXREC werkt. Dit is van belang voor betekenisvolle menselijke tussenkomst. Als de medewerker niet begrijpt hoe de OXREC werkt dan kan de medewerker bijvoorbeeld niet goed inschatten voor welke personen de OXREC minder betrouwbaar is en wat de beperkingen van de OXREC zijn. Dit is in lijn met

⁶¹ Van den Berg, Chantal, et al. "Validatiestudie risico- en beschermende factoren van de RISC: Een evaluatieonderzoek naar de leefgebieden van het risicotaxatie en adviesinstrument van de 3RO." (2021) pagina 33.

⁶² Van den Berg, Chantal, et al. "Validatiestudie risico- en beschermende factoren van de RISC: Een evaluatieonderzoek naar de leefgebieden van het risicotaxatie en adviesinstrument van de 3RO." (2021) pagina 34.

⁶³ Autoriteit Persoonsgegevens "Betekenisvolle Menselijke Tussenkomst" (2025). Pagina 14.

het advies van de Autoriteit Persoonsgegevens over betekenisvolle menselijke tussenkomst⁶⁴ en de Handvatten Betekenisvolle Menselijke Tussenkomst⁶⁵.

Ten aanzien van de STATIC-99R, STABLE-2007 en ACUTE-2007 gelden soortgelijke bevindingen. Het gebruik van deze algoritmes is bij de reclassering een verplicht onderdeel van de risico-inschatting bij mannen die verdacht van of veroordeeld zijn voor zedendelicten en worden toegepast voordat de medewerker een eindoordeel geeft. Ook bij deze algoritmes ziet de Inspectie JenV een risico op *automation bias* en *anchoring bias*. Het eindoordeel van de medewerkers komt in 82.1% overeen met de (samengestelde) uitkomst van de zedenrecidive-algoritmes, hoger dan bij de OXREC.

4.8 Governance

De rollen en verantwoordelijkheden voor het gebruik en onderhoud van de risicotaxatie-algoritmes bij de reclassering zijn niet belegd bij één persoon of gremium. Inhoudelijke kennis over algoritmes is versnipperd aanwezig binnen de drie reclasseringsorganisaties. Adviezen ten aanzien van het gebruik of onderhoud van algoritmes worden ad-hoc gegeven. Besluiten ten aanzien van deze adviezen worden uiteindelijk vastgesteld door de directies van de drie reclasseringsorganisaties.

De Inspectie JenV heeft onvoldoende structuur en waarborgen aangetroffen om te komen tot een goed *life-cycle management* voor algoritmes. Zo worden er geen periodieke validaties⁶⁶ en updates van de risicotaxatie-algoritmes uitgevoerd die waarborgen dat de kwaliteit gemonitord wordt en op peil blijft. Ook zijn er onvoldoende softwaretesten geïmplementeerd om de onjuiste software-implementatie van de OXREC op te kunnen sporen. De reclassering heeft daarnaast geen standaardprocedure om algoritmes te toetsen op mogelijke discriminatie door middel van een mensenrechtentoets (zoals IAMA) of andere methoden. Tot slot constateert de Inspectie JenV een gebrek aan juridische waarborgen zoals het opstellen van een verplichte DPIA en juridische kennis ten aanzien van de AI-verordening.

Hiermee constateert de Inspectie JenV dat duidelijke sturing door de reclassering op het gebruik en onderhoud van de risicotaxatie-algoritmes ontbreekt.

⁶⁴ Autoriteit Persoonsgegevens. Advies artikel 22 AVG en geautomatiseerde selectie-instrumenten: "Om betekenisvol te zijn en daadwerkelijk gevolgen zo nodig te kunnen voorkomen ofwijzigen, moet degene die 'tussenkomt' voldoende kunnen beoordelen of selectie in een bepaald geval terecht is. De behandelend ambtenaar zal daarom moeten weten hoe het geautomatiseerde proces (selectieregel/algorithmetechnologie) werkt en moeten begrijpen hoe en op welke wijze deze (de totstandkoming) van het uiteindelijke besluit vormt en beïnvloedt."

⁶⁵ Autoriteit Persoonsgegevens "Betekenisvolle Menselijke Tussenkomst" (2025).

⁶⁶ De OXREC is gevalideerd voordat het in gebruik is genomen voor de reclassering, maar wordt niet periodiek opnieuw gevalideerd om te bepalen of het algoritme nog de juiste voorspellingen maakt. De STATIC-99R, STABLE-2007 en ACUTE-2007 zijn nog nooit door de reclassering zelf gevalideerd.

5. Conclusies

Eén van de kerntaken van de reclassering is onderzoek te doen naar en advies te geven over personen die worden verdacht van of zijn veroordeeld voor een strafbaar feit. De reclassering stelt het advies op aan de hand van relevante factoren als de aard van het gepleegde delict, het delictverleden, de persoonlijke omstandigheden van de verdachte/veroordeelde en het recidiverisico. De adviezen die de reclassering opstelt worden in de regel gegeven aan het Openbaar Ministerie (OM) of de rechtbank. Zo kan het advies van invloed zijn op de strafeis van het OM, en kan de rechter het advies van de reclassering meewegen om verdachten een gevangenisstraf op te leggen of een gedetineerde juist vervroegd vrij te laten. Uiteindelijk beslist de rechter of een strafmaatregel wordt opgelegd.

Om tot gestructureerde adviezen te komen heeft de reclassering de RISC ontwikkeld. Dit is een gestructureerd risicotaxatie- en adviesinstrument dat door reclasseringsmedewerkers wordt gebruikt om in te schatten wat de kans is dat een verdachte of veroordeelde opnieuw de fout in zal gaan, oftewel zal recidiveren. Met de RISC beoogt de reclassering factoren in kaart te brengen die van invloed kunnen zijn op het toekomstige gedrag van een verdachte of veroordeelde.

Als onderdeel van de RISC worden algoritmes gebruikt om de kans op recidive in te schatten. De OXREC is zo'n algoritme, en is een vast onderdeel van de RISC. Daarnaast gebruikt de reclassering nog andere algoritmes om de kans op recidive in te schatten bij mannelijke zedendelinquenten. Dit zijn de STATIC-99R, STABLE-2007 en ACUTE-2007.

De deelconclusies van de Inspectie JenV hebben betrekking op het algoritmegebruik door de reclassering zelf. De Inspectie doet geen uitspraken over het gebruik en/of de (wetenschappelijke) kwaliteit van de algoritmes *an sich*, en heeft ook niet onderzocht hoe de voorspellingen van de algoritmes feitelijk hebben doorgewerkt in concrete adviezen van de reclassering aan het OM en/of een rechtbank.

5.1 Deelconclusies

De reclassering maakt gebruik van risicotaxatie-algoritmes die niet goed aansluiten op de situaties waarin ze worden toegepast. Hoewel de algoritmes wetenschappelijk gevalideerd zijn, sluiten deze validaties niet aan op de wijze waarop de reclassering de algoritmes toepast. De algoritmes zijn ontwikkeld met sterk verouderde data (uit de periode 1957-2009) die in een andere context zijn verzameld. Zo zijn de modelparameters van de OXREC vastgesteld met verouderde gegevens van Zweedse gedetineerden, maar wordt het onder andere toegepast op Nederlandse personen die nog niet veroordeeld zijn. De STATIC-99R, STABLE-2007 en ACUTE-2007 zijn ontwikkeld met zeer verouderde voornamelijk Noord-Amerikaanse data en niet in een Nederlandse context. De STATIC-99R is gevalideerd met zeer verouderde Nederlandse data. Het risico bestaat hierdoor dat de algoritmes tot een afwijkende risico-inschatting komen omdat de algoritmes toegepast worden op situaties waar ze niet voor ontworpen zijn.

Er is geen eenduidig beleid om de risicotaxatie-algoritmes te valideren en te evalueren. De reclassering heeft de OXREC laten valideren voordat het in gebruik is genomen, maar heeft de zedenrecidive-algoritmes niet zelf laten valideren. De STATIC-99R is in het verleden wel gevalideerd in een Nederlandse populatie, maar de STABLE-2007 en ACUTE-2007 niet. Hierdoor is het onduidelijk of deze algoritmes überhaupt goede voorspellingen geven in een Nederlandse reclasseringspopulatie. De reclassering had ten tijde van het onderzoek geen plannen om zelf de algoritmes periodiek te (laten) valideren of te evalueren.

De reclassering is transparant over het gebruik van risicotaxatie-algoritmes. De reclassering vermeldt in risicorapportages dat het advies mede tot stand is gekomen door middel van risicotaxatie-algoritmes. Uitkomsten van de algoritmes worden ook vermeld in deze rapportages. De algoritmes die gebruikt worden zijn relatief eenvoudig en makkelijk te interpreteren. Hoewel de algoritmes nog niet geregistreerd zijn in het algoritmeregister, geeft de reclassering aan dit wel te gaan doen voor het einde van 2025.

De reclassering gebruikt sinds de ingebruikname de OXREC op een verkeerde manier door een onjuiste software-implementatie. Door het weglaten van twee variabelen en gebreken in de eigen software-implementatie heeft de reclassering de OXREC nog nooit correct toegepast. Dit leidt ertoe dat de berekende kans op recidive altijd afwijkt van de bedoelde voorspelling. Deze gebreken zijn dusdanig groot dat de Inspectie JenV schat dat bij correct gebruik de OXREC bij algemene recidive in ruim 21% van de gevallen zelfs tot een andere risicocategorie zou zijn gekomen. Dit zouden vaak hogere risico-inschattingen zijn geweest. Er is niet onderzocht in welke gevallen dit ook feitelijk is gebeurd.

De reclassering heeft signalen ontvangen over mogelijke discriminatie door de OXREC. Hier zijn geen concrete maatregelen op getroffen. Het algoritme bevat verboden variabelen die mogelijk kunnen leiden tot discriminatie. Na kritische geluiden in 2020 die stelden dat de OXREC kan leiden tot discriminatie, maakte de reclassering de inschatting dat er geen sprake was van discriminatie. Hierdoor zijn er vijf jaar na dato geen stappen ondernomen om mogelijke discriminatie door de OXREC te voorkomen, te meten of te verminderen.

De Inspectie JenV constateert dat de OXREC niet voldoet aan het Toetsingskader Risicoprofilering van het College voor de Rechten van de Mensen. Het huidige gebruik van de OXREC is verboden op basis van het discriminatieverbod vanwege de variabelen 'buurtscore' en 'hoogte van het inkomen'.

Naar inzicht van de Inspectie Justitie en Veiligheid is de OXREC een hoog-risico AI-systeem waar de AI-verordening vanaf 2030 op van toepassing is. De OXREC valt in de categorie hoog-risico AI-systemen waarvoor strenge voorwaarden gelden. De reclassering is tot een andere conclusie gekomen en heeft de inschatting gemaakt dat de OXREC geen AI-systeem is en daarmee niet aan de AI-verordening hoeft te voldoen. Hierdoor zijn er geen maatregelen getroffen om te (gaan) voldoen aan de AI-verordening.

De reclassering voldoet niet aan de wettelijke vereisten en richtlijnen die gesteld worden voor het gebruik van risicotaxatie-algoritmes. Zowel het gebruik van de OXREC als de STATIC-99R, STABLE-2007 en ACUTE-2007 voldoen niet aan de AVG omdat een verplichte DPIA ontbreekt. De reclassering voldoet nog niet aan de wettelijke vereisten vanuit de AI-verordening voor de OXREC, maar heeft hiervoor nog tot 2 augustus 2030. Daarnaast voldoen de risicotaxatie-algoritmes niet aan het Algoritmekader voor de overheid, de richtlijn Betekenisvolle Menselijke Tussenkost en het Toetsingskader Risicoprofilering.

De Inspectie JenV ziet risico's ten aanzien van betekenisvolle menselijke tussenkost bij het gebruik van risicotaxatie-algoritmes. De rol van de risicotaxatie-algoritmes wordt groot gemaakt door de reclassering door het trainingsmateriaal, de wijze waarop het geïmplementeerd is in het IT-systeem en door het gebruik van de algoritmes te verplichten. De uitkomsten van de risicotaxatie-algoritmes zijn belangrijke factoren in het eindoordeel van de medewerkers en komen in circa 80% van de gevallen hiermee overeen. In trainingsmateriaal wordt deels gestimuleerd om de uitkomst van het algoritme over te nemen. Medewerkers leren onvoldoende hoe de OXREC werkt zodat ze kunnen inschatten wanneer de uitkomst minder betrouwbaar is.

De reclassering heeft geen adequate sturing georganiseerd voor het gebruik van risicotaxatie-algoritmes. Bij de reclassering ontbreekt een duidelijke sturing op het gebruik van risicotaxatie-algoritmes. Zo is er onvoldoende inzicht en bewustzijn bij de reclassering om te voldoen aan de geldende wet- en regelgeving. Bij de implementatie van de OXREC was er onvoldoende inzicht in de methodologische tekortkomingen van de algoritmes. Ook zijn er geen goede controles op de software uitgevoerd waardoor er afwijkende voorspellingen gemaakt zijn. Er zijn geen afspraken over het periodiek valideren en evalueren van de algoritmes. Signalen over mogelijk discriminerende variabelen in de OXREC zijn niet opgevolgd.

5.2 Hoofdconclusies

De Inspectie JenV ziet risico's in het algoritmegebruik van de reclassering die kunnen leiden tot negatieve gevolgen voor zowel de maatschappij als individuele verdachten en veroordeelde personen. De reclassering geeft meer dan 43.500 adviezen per jaar waarin de OXREC toegepast wordt. Door methodologische problemen krijgen o.a. personen met een ernstige psychische aandoening zoals psychoses een te lage risico-inschatting van de OXREC. Het is van belang om de risico's op een juiste wijze en zo nauwkeuring mogelijk in te schatten. Een te lage risico-inschatting kan leiden tot adviezen waarin de maatschappij mogelijk niet voldoende beschermd wordt tegen criminelen. Andersom kan een te hoge risico-inschatting juist leiden tot mogelijke nadelige gevolgen voor verdachten en veroordeelden.

De reclassering gebruiken de STATIC-99R, STABLE-2007 en ACUTE-2007 niet op een aantoonbaar verantwoorde wijze. De drie algoritmes zijn niet getoetst met recente data uit de Nederlandse reclasseringspopulatie, waardoor het onduidelijk is of er goede voorspellingen gemaakt worden. De algoritmes voldoen niet aan de AVG en er zijn risico's geconstateerd op het gebied van betekenisvolle menselijke tussenkomst. Aan de andere kant is de software-implementatie van de algoritmes correct uitgevoerd en is er een verlaagd risico op discriminatie.

De reclassering gebruikt de OXREC niet op een verantwoorde wijze. De OXREC voldoet op bijna geen enkel vlak aan de normen voor overheidsalgoritmes. Er zijn meerdere tekortkomingen geconstateerd op het vlak van technische robuustheid, software-implementatie, wet- en regelgeving, non-discriminatie en betekenisvolle menselijke tussenkomst.

5.3 Opvolging van conclusies

De Inspectie JenV heeft in juli 2025 het bestuur van de drie reclasseringsorganisaties op de hoogte gebracht van haar eerste bevindingen. De Inspectie JenV heeft daarbij aandacht gevraagd voor de risico's van het gebruik van de OXREC. Het bestuur van de 3RO heeft op basis van de melding van de Inspectie JenV de eerste verbetermaatregelen in gang gezet, zoals het opstellen van verplichte DPIA's, het aanpassen van de software-implementatie en het opzetten van een programma om de algoritme-governance te verbeteren. Hiervoor heeft de reclassering ook een claim ingediend bij het ministerie van Justitie en Veiligheid. Deze verbetermaatregelen zijn geen onderwerp van het onderzoek geweest.

6. Aanbevelingen

Aan het bestuur van de drie reclasseringsorganisaties:

In het licht van haar bevindingen acht de Inspectie het logisch dat alle implementatiegebreken van de OXREC zo snel mogelijk worden verholpen, en dat - wanneer dit niet op korte termijn mogelijk blijkt - de OXREC (tijdelijk) wordt stilgelegd. De Inspectie heeft echter geen onderzoek gedaan wat de impact hiervan zal zijn voor de taakuitvoering van de reclassering, en begrijpt van de reclassering dat een dergelijke aanbeveling zeer nadelige operationele gevolgen zou hebben.

De Inspectie beveelt daarnaast het volgende aan:

1. Richt een alomvattende governance-structuur in voor de ontwikkeling, toepassing en het onderhoud van algoritmes, en voer deze ook uit;
2. Valideer en evalueer periodiek de risicotaxatie-algoritmes op een wijze die aansluit op het gebruik van deze algoritmes door de reclassering. Hertrain hierbij AI-systemen zoals de OXREC periodiek;
3. Verifieer de correctheid van de software-implementatie van de algoritmes, ook na aanpassingen aan de algoritmes;
4. Voldoe bij het gebruik van risicotaxatie-algoritmes in ieder geval aan de geldende wettelijke normen en richtlijnen die gesteld zijn in het algoritmekader voor de overheid, de AVG, de richtlijn betekenisvolle menselijke tussenkomst, het toetsingskader risicoprofilering en de AI-verordening;
5. Rapporteer periodiek, iedere zes maanden, over de opvolging van de aanbevelingen.

Bijlage A: begrippen

AenM – Asiel en Migratie

AI – Artificial Intelligence/Kunstmatige Intelligentie

AP – Autoriteit Persoonsgegevens

AUC – Area Under the Curve, een getal dat het voorspellend vermogen van een algoritme meet.

AVG – Algemene verordening gegevensbescherming

CDO – Chief Data Officer

CRM – College voor de Rechten van de Mens

DPIA – Data Protection Impact Assessment, ook wel Gegevensbeschermingseffectbeoordeling genoemd

EU – Europese Unie

IAMA – Impact Assessment Mensenrechten en Algoritmes

Inspectie JenV – de Inspectie Justitie en Veiligheid

JenV – Justitie en Veiligheid

OM – Openbaar Ministerie

Reclassering – de drie organisaties die in Nederland het reclasseringswerk uitvoeren: Reclassering Nederland, Stichting Verslavingsreclassering GGZ en Leger des Heils Jeugdbescherming & Reclassering.

Bijlage B: opzet van het onderzoek

De Inspectie JenV heeft onderzoek gedaan of de reclassering algoritmes op een verantwoorde manier gebruiken.

De centrale onderzoeksvraag bij het onderzoek naar de reclassering luidde: "Worden algoritmes bij de reclassering op een verantwoorde manier gebruikt?". De Inspectie JenV heeft de volgende deelvragen geformuleerd om deze vraag te kunnen beantwoorden:

- Welke algoritmes gebruikt de reclassering?
- Voldoen de algoritmes aan de normen die gesteld zijn voor overheidsalgoritmes?
- Welke gevolgen kunnen de algoritmes hebben op mensen in een kwetsbare positie?
- In hoeverre is de reclassering voorbereid op de nieuwe wettelijke vereisten van de AI-verordening?

De Inspectie JenV volgt hierin het plan van aanpak voor het toezicht op algoritmes.⁶⁷ De algoritmes zijn beoordeeld op grond van het Toetsingskader Algoritmes⁶⁸ van de Inspectie Justitie en Veiligheid op de thema's:

- Menselijke controle;
- Technische robuustheid;
- Privacy en datagovernance;
- Transparantie;
- Diversiteit, non-discriminatie en rechtvaardigheid;
- Governance.

De Inspectie JenV heeft dit onderzoek uitgevoerd in de periode van januari tot en met oktober 2025. Om te onderzoeken of de reclassering algoritmes op een verantwoorde manier gebruikt heeft de Inspectie JenV de algoritmes getoetst aan het toetsingskader door de volgende stappen te ondernemen:

- Het analyseren van wetenschappelijke studies over de ontwikkeling en validaties van de algoritmes die gebruikt worden bij de reclassering;
- Het analyseren van de governancestructuur ten aanzien van algoritmes. Hiervoor zijn beleidsdocumenten geanalyseerd en interviews gehouden met de *Chief Data Officer*, betrokken beleidsmedewerkers, directieleden, een functionaris gegevensbescherming en andere privacyfunctionarissen;
- Het analyseren van documentatie over de werking en software-implementatie van de algoritmes. Dit betreft onder andere het analyseren

⁶⁷ Inspectie Justitie en Veiligheid. Plan van aanpak toezicht op algoritmes (2025) (<https://www.inspectie-jenv.nl/documenten/2025/01/16/plan-van-aanpak-toezicht-op-algoritmes>).

⁶⁸ Inspectie Justitie en Veiligheid. Toetsingskader algoritmes (2025) (<https://www.inspectie-jenv.nl/site/binaries/site-content/collections/documents/2025/01/16/plan-van-aanpak-toezicht-op-algoritmes/Bijlage+B+-+Toetsingskader+Algoritmes.pdf>).

van de broncode en functioneel ontwerp van de algoritmes en interviews met medewerkers van de afdeling informatisering;

- Het analyseren van aanvullende documentatie, voor zover aanwezig, zoals DPIA's en mensenrechtentoetsen (IAMA) en beoordelingen ten aanzien van de AI-verordening;
- Het analyseren van instructies en gebruik van algoritmes in de praktijk. Hiervoor is trainingmateriaal van de reclassering geanalyseerd, de werking van de algoritmes binnen het IT-systeem van de reclassering onderzocht en zijn interviews gehouden met medewerkers die algoritmes toepassen en met medewerkers van de reclasseringsacademie;
- Het analyseren van gegevens over hoe vaak de algoritmes toegepast worden en in welke mate het eindoordeel van de medewerkers overeenkomt met de uitkomsten van de risicotaxatie-algoritmes;
- Aanvullende interviews met externe partijen: de directie coördinatie algoritmes van de Autoriteit Persoonsgegevens, het College voor de Rechten van de Mens en een interview met prof. Gijs van Dijck.

Computersimulaties OXREC

Omdat de Inspectie JenV geconstateerd heeft dat de reclassering twee variabelen van de OXREC niet toepast en er gebreken zijn aangetroffen in de software-implementatie (zie hoofdstuk 4.1), zijn er computersimulaties uitgevoerd om het effect van deze gebreken in kaart te brengen. Hiervoor is een dataset gecreëerd met alle mogelijke (ruim 2.9 miljoen) combinaties van inputvariabelen voor de OXREC. Voor elk van deze combinaties is de kans op recidive berekend met de correcte modelparameters inclusief Nederlandse kalibratieparameters en formules zoals beschreven in de validatiestudie van de OXREC⁶⁹ en de ontwerpstudie⁷⁰. Daarnaast is voor elk van deze combinaties ook de kans op recidive berekend volgens de implementatie van de reclassering, inclusief het weglaten van de twee variabelen, onjuiste modelparameters en verwisseling van de formules. Hiermee kan een beeld gevormd worden over de grootte van de voorspellingsfouten door de implementatie van de OXREC in het systeem van de reclassering.

Deze kans op recidive zijn ingedeeld in risicocategorieën (laag, gemiddeld, hoog) volgens de instructies van de ontwikkelaar van de OXREC. Voor elke combinatie van inputvariabelen is gekeken of de onjuiste implementatie van de OXREC heeft geleid tot een andere risicocategorie.

⁶⁹ Fazel, Seena, et al. "Prediction of violent reoffending in prisoners and individuals on probation: A Dutch validation study (OxRec)." *Scientific Reports* 9.1 (2019): 841.

⁷⁰ Fazel, Seena, et al. "Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool." *The Lancet Psychiatry* 3.6 (2016): 535-543.

Bijlage C: de OXREC

Wat is de OXREC?

De OXREC is een zogeheten risicotaxatie-algoritme. Het algoritme schat de kans in dat een verdachte of veroordeelde opnieuw de fout in zal gaan (recidiveren). In de Nederlandse versie van de OXREC gebeurt dit aan de hand van de volgende dertien persoonskenmerken:

- Geslacht (man/vrouw);
- leeftijd;
- detentieduur (< 6 maanden, 6 tot 12 maanden, 12-24 maanden, >=24 maanden);
- veroordeling voor een gewelddadig misdrijf;
- veroordeling in het verleden voor een gewelddadig misdrijf;
- burgerlijke stand (getrouwd vs. niet-getrouwd);
- opleidingsniveau (<9 jaar opleiding, 9-11 jaar opleiding, >=12 jaar opleiding);
- het hebben van een baan;
- inkomen (negatief, geen inkomen, laag, midden of hoog inkomen);
- buurtscore;
- alcoholmisbruik;
- drugsmisbruik;
- aanwezigheid van psychische problematiek.

Door deze kenmerken in te vullen met de gegevens van een persoon voorspelt de OXREC de kans dat de hij of zij binnen één of twee jaar recidiveert. Hierbij maakt het algoritme onderscheid tussen gewelddadige recidive (waarbij iemand recidiveert met een geweldsdelict) en 'algemene' recidive (waarbij iemand überhaupt recidiveert). De OXREC levert een percentage op, dit is de kans op recidive. Hier wordt ook een risicocategorie (laag, gemiddeld of hoog) aan gekoppeld afhankelijk van de hoogte van het percentage.

Bij algemene recidive geldt dat wanneer de kans op recidive minder is dan 30% dat de risicocategorie 'laag' is, tussen 30% en 50% wordt de kans ingeschat op 'gemiddeld', en daarboven op 'hoog'. Bij gewelddadige recidive liggen de grenzen anders: 'laag' is tot 10% kans op recidive, tussen de 10% en 30% wordt de kans ingeschat op 'gemiddeld' en daarboven op 'hoog'.

Een fictief voorbeeld van een mogelijke output van de OXREC die reclasseringsmedewerkers te zien krijgen:

Risico op algemene recidive binnen 1 jaar: 25%
Risico op algemene recidive binnen 2 jaar: 36%
Risico op gewelddadige recidive binnen 1 jaar: 5%
Risico op gewelddadige recidive binnen 2 jaar: 7%
Algemeen risico: gemiddeld
Risico op geweld: laag

Ontwikkeling en ingebruikname van de OXREC

De OXREC is in 2016 ontwikkeld door de University of Oxford.⁷¹ Het algoritme is getraind op een dataset met de kenmerken van 37.100 Zweedse gevangenen die op enig moment tussen 2001 en 2009 gedetineerd waren. In het algoritme waren naast de hierboven genoemde dertien persoonskenmerken ook twee andere kenmerken opgenomen. Dit zijn migratieachtergrond en de aanwezigheid van ernstige psychische problematiek, zoals psychoses.

Voordat de reclassering het algoritme in 2018 in gebruik heeft genomen, is er aan de ontwikkelaar gevraagd om de OXREC te valideren met Nederlandse data. Deze validatie⁷² is uitgevoerd met de gegevens van 15.401 Nederlandse personen uit 2011-2012. Bij deze validatie is er voor gekozen om migratieachtergrond niet mee te nemen als voorspeller omdat dat kan leiden tot discriminatie. Omdat er geen gegevens beschikbaar waren over de aanwezigheid van ernstige psychische problematiek (zoals psychose), is deze variabele weggelaten uit het algoritme. Het weglaten van deze variabele houdt in dat er bij de berekening vanuit wordt gegaan dat niemand een ernstige psychische aandoening heeft. Omdat de aanwezigheid van ernstige psychische problematiek een risicoverhogende factor is, zorgt het weglaten van deze variabele ervoor dat de OXREC het risico op recidive onderschat voor mensen met een ernstige psychische aandoeningen.

Uit de validatie kwam naar voren dat de kans op recidive in de Nederlandse populatie fors werd overschat, waarop het algoritme opnieuw gekalibreerd is. Dit betekent dat het algoritme niet opnieuw getraind is met Nederlandse data, maar dat er voor alle personen de voorspellingen met een constante factor vermenigvuldigd zijn zodat de voorspellingen van de complete groep gemiddeld gezien overeenkomen met de geobserveerde kans op recidive.

Na deze validatie is de gerekalibreerde OXREC in gebruik genomen door de reclassering en wordt het sindsdien toegepast bij risico-inschattingen van reclasseringsmedewerkers. Uit opgevraagde cijfers blijkt dat in de periode 2023-2024 het algoritme ruim 43.500 keer per jaar is toegepast.

Hoe werkt de OXREC?

De OXREC is getraind met een dataset met kenmerken van Zweedse gedetineerden. Door middel van een machine-learning techniek (in dit geval *Cox proportional hazard regression*) heeft het algoritme geleerd wat het verband is tussen kenmerken als leeftijd en drugsmisbruik, en het wel of niet recidiveren.

Voor elk individueel persoonskenmerk krijgt een verdachte of veroordeelde een verhoging of verlaging van de score. Is een persoon een vrouw, dan gaat er 1,364

⁷¹ Fazel, Seena, et al. "Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool." *The Lancet Psychiatry* 3.6 (2016): 535-543.

⁷² Fazel, Seena, et al. "Prediction of violent reoffending in prisoners and individuals on probation: a Dutch validation study (OxRec)." *Scientific Reports* 9.1 (2019): 841.

van de score af. Misbruikt de persoon drugs, dan komt er juist 0,4117 bij de score bij.⁷³ Voor de overige variabelen geldt een soortgelijke verhoging of verlaging van de score. Deze getallen worden ook wel modelparameters genoemd, en zijn door het algoritme geleerd uit de dataset met kenmerken van Zweedse gedetineerden.

Nadat de reclasseringsmedewerker de persoonskenmerken heeft ingevuld, worden de modelparameters bij elkaar opgeteld en ingevuld in een formule. Deze formule staat beschreven in OXREC-validatiestudie. Deze formule levert uiteindelijk de gevraagde kans op recidive op. Dit is een percentage tussen de 0% en 100%.

⁷³ Deze getallen gelden voor de voorspelling van gewelddadige recidive.

Bijlage D: de STATIC-99R, STABLE-2007 en ACUTE-2007

Wat zijn de STATIC-99R, STABLE-2007 en ACUTE-2007?

De STATIC-99R, STABLE-2007 en ACUTE-2007 zijn drie risicotaxatie-algoritmes die ingezet worden om het risico op recidive in te schatten. In tegenstelling tot de OXREC kunnen deze niet ingezet worden bij alle personen, maar zijn ze ontworpen om het risico op recidive in te schatten bij mannelijke verdachten/veroordeelden van zedendelicten (zoals verkrachting of kindermisbruik).

De algoritmes bestaan elk uit meerdere vragen, bij elke vraag kunnen een geheel aantal punten gescoord worden. Bij de STATIC-99R worden vragen gesteld over zogeheten 'onveranderbare' factoren die samenhangen met recidive. Het gaat hierbij om de volgende tien factoren:

- Leeftijd;
- Ooit minstens twee jaar samengewoond;
- Veroordeeld voor seksueel geweld;
- Eerdere veroordeling voor seksueel geweld ;
- Aantal aanklachten en veroordelingen seksueel geweld;
- Aantal eerdere veroordelingen;
- Enige veroordeling voor non-contact seksueel delict;
- Enig slachtoffer buiten eigen familie;
- Enig onbekend slachtoffer;
- Enig mannelijk slachtoffer.

Voor elke vraag kan er 0 of 1 punt gescoord worden. Bij leeftijd kan er ook -1 of -3 gescoord worden voor personen die ouder dan 40 jaar zijn. Alle punten worden vervolgens bij elkaar opgeteld. Een score van maximaal 1 levert de risicocategorie 'laag' op, tot en met 3 punten 'laag-matig', tot en met 5 punten 'matig-hoog' en daarboven levert de categorie 'hoog' op. Medewerkers krijgen naast deze categorie, in tegenstelling tot de OXREC, geen inzicht in de percentuele kans op recidive. Wel krijgen ze het aantal punten per vraag te zien.

In aanvulling op de STATIC-99R kan ook de STABLE-2007 afgenomen worden. Dit is een zeer vergelijkbaar algoritme van dezelfde ontwikkelaar. Het verschil is dat de vragen gaan over 'veranderbare' factoren. Het gaat hierbij om de volgende dertien factoren:

- Belangrijke sociale contacten;
- Vermogen tot het vormen van een stabiele relatie;
- Emotionele identificatie met kinderen;
- Vrouwvrijandigheid;
- Eenzaamheid;
- Desinteresse in het welzijn van anderen;
- Impulsief gedrag;

- Ontoereikende probleemoplossingsvaardigheden;
- Vijandigheid;
- Seksuele preoccupatie;
- Seks als coping;
- Seksuele deviatie;
- Samenwerking met reclasseringsmedewerkers.

Voor elke vraag kan er 0, 1 of 2 punten gescoord worden. Alle punten worden bij elkaar opgeteld. Tot en met 3 punten levert de risicocategorie 'laag' op, tot en met 11 punten de categorie 'gemiddeld' en daarboven 'hoog'. Ook bij de STABLE-2007 krijgen medewerkers geen inzicht in de percentuele kans op recidive, maar in wel het aantal punten per vraag.

Tot slot kan ook de ACUTE-2007 worden toegepast, welke vragen over 'acuut dynamische' factoren bevat. Ook dit algoritme is van dezelfde ontwikkelaar. Dit algoritme bevat vragen over de volgende zeven factoren:

- Toegang tot potentiële slachtoffers;
- Vijandigheid;
- Seksuele preoccupatie;
- Afwijzing van begeleiding;
- Emotionele inzinking;
- Veranderingen in sociale steun;
- Middelengebruik;

Voor elke vraag kan er 0, 1, 2 of 3 punten gescoord worden. Bij een totaalscore van 0 wordt de risicocategorie 'lage prioriteit', bij 1-2 'matig' en daarboven 'hoog'. Ook bij de ACUTE-2007 krijgen medewerkers geen inzicht in de percentuele kans op recidive, maar in wel het aantal punten per vraag.

Omdat de drie algoritmes tot drie verschillende risico-inschattingen kunnen leiden, geeft de handleiding van de reclassering instructies om hiermee om te gaan. De risicocategorie van de STATIC-99R is hierbij leidend, maar als de STABLE-2007 'laag' of juist 'hoog' is dan wordt de risicocategorie van de STATIC-99R één niveau verlaagd dan wel verhoogd. Een soortgelijke aanpassing van de risicocategorie vindt plaats met een lage of hoge score op de ACUTE-2007. De software-implementatie van de reclassering rekent deze gecombineerde score overigens niet uit.

Reclasseringsmedewerkers worden speciaal opgeleid om risico-inschattingen te doen over mannelijke zedenverdachten en/of -delinquenten. Dit omvat een opleiding van circa zes maanden waarin de algoritmes ook uitgebreid aan bod komen. De handleidingen van de STATIC-99R, STABLE-2007 en ACUTE-2007 omvatten samen ruim 560 bladzijden.

Ontwikkeling van de STATIC-99R, STABLE-2007 en ACUTE-2007

De STATIC-99R⁷⁴ is grotendeels identiek aan een ouder algoritme uit 2000, de STATIC-99 (zonder de achtervoegsel 'R'). De enige wijziging aan het algoritme is in het puntenaantal dat toegekend wordt aan de factor leeftijd. De STATIC-99 is op zijn beurt weer een samenvoeging van eerdere algoritmes. Uiteindelijk zijn deze algoritmes relatief eenvoudig tot stand gekomen door te bepalen welke factoren er het meest samenhangen met zedenrecidive en daar een vast aantal punten aan toe te kennen. De STATIC-99R is tot stand gekomen met een mengeling van (grotendeels) Noord-Amerikaanse gegevens uit 1957-2007. In 2014 is de STATIC-99R gevalideerd in een Nederlandse populatie⁷⁵. Uit deze validatie bleek dat het vermogen van het algoritme om in te schatten of een persoon wel of niet recidiveert 'redelijk'⁷⁶ is in deze populatie. De STATIC-99R is sinds 2018 opgenomen in de RISC en werd in de jaren 2023-2024 circa 1500 keer per jaar toegepast door de reclassering.

De STABLE-2007 en de ACUTE-2007 zijn op een vergelijkbare wijze ontwikkeld als doorontwikkeling van de eerdere STABLE-2000 en ACUTE-2000.⁷⁷ Ook deze algoritmes zijn gebaseerd op een Noord-Amerikaanse populatie, al vermeldt de wetenschappelijke publicatie niet in welke periode(s) de gegevens verzameld zijn. De STABLE-2007 en ACUTE-2007 zijn voor zover bekend niet gevalideerd in een Nederlandse populatie. In de periode 2023-2024 paste de reclassering de STABLE-2007 circa 1100 keer per jaar toe, en de ACUTE-2007 circa 400 keer per jaar.

⁷⁴ Helmus, Leslie, et al. "Improving the predictive accuracy of Static-99 and Static-2002 with older sex offenders: Revised age weights." *Sexual Abuse* 24.1 (2012): 64-101.

⁷⁵ Smid, Wineke J., et al. "A comparison of the predictive properties of nine sex offender risk assessment instruments." *Psychological assessment* 26.3 (2014): 691.

⁷⁶ De AUC in de Nederlandse validatiestudie was 0.74 (tussen 0.7-0.8 wordt deze geclassificeerd als 'redelijk').

⁷⁷ Hanson, R. Karl, et al. *Assessing the risk of sexual offenders on community supervision: The Dynamic Supervision Project*. Vol. 5. No. 6. Ottawa, Ontario: Public Safety Canada, 2007.

Bijlage E: onderbouwing dat de OXREC een AI-systeem is conform artikel 3(1) AI-verordening

Beschrijving van de OXREC

De OXREC is een algoritme dat ontwikkeld is om algemene en gewelddadige recidive te voorspellen van ex-gedetineerden. Het algoritme is gepubliceerd op 13 april 2016 in *Lancet Psychiatry*⁷⁸, en wordt sinds 2018 gebruikt door de reclassering. De OXREC is ontwikkeld met een *multivariable Cox proportional hazard regression*. Dit is een type regressiemodel dat een lineair verband veronderstelt tussen de predictoren en een binaire uitkomst (in dit geval het wel of niet gewelddadig recidiveren), evenals het moment waarop deze uitkomst plaatsvindt. Het algoritme zoals het gebruikt wordt door de reclassering bevat de volgende dertien variabelen: geslacht, leeftijd, duur van gevangenschap, gewelddadig misdrijf, voorafgaand gewelddadig misdrijf, burgerlijke staat, opleiding, werkloosheid, inkomen, buurtscore, alcoholverslaving, drugsverslaving, en aanwezigheid van psychische problematiek.

Het algoritme inclusief modelparameters is getraind met de gegevens van 37.100 Zweedse personen die op enig moment na 1 januari 2000 gedetineerd zijn geweest en vóór 31 december 2009 vrijgelaten zijn.

Voor een uitgebreide beschrijving van het algoritme, de modelparameters en de populatie waar het algoritme op getraind is, wordt verwezen naar de originele studie.⁷⁹

Definitie van AI-systeem

De definitie van een AI-systeem waar de AI-verordening⁸⁰ op van toepassing is, is gedefinieerd in artikel 3, lid 1 van de AI-verordening. Deze luidt als volgt, inclusief merkingen van zeven belangrijke elementen:

"een op een [1] machine gebaseerd systeem [2] dat is ontworpen om met verschillende niveaus van autonomie te werken en [3] dat na het inzetten ervan aanpassingsvermogen kan vertonen, en dat, [4] voor expliciete of impliciete doelstellingen, [5] uit de ontvangen input afleidt hoe output te genereren [6] zoals

⁷⁸ Fazel, S., Chang, Z., Fanshawe, T., Långström, N., Lichtenstein, P., Larsson, H., & Mallett, S. (2016). Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool. *The Lancet Psychiatry*, 3(6), 535-543.

⁷⁹ Fazel, Seena, et al. "Prediction of violent reoffending on release from prison: derivation and external validation of a scalable tool." *The Lancet Psychiatry* 3.6 (2016): 535-543.

⁸⁰ <http://data.europa.eu/eli/reg/2024/1689/oj>

voorspellingen, inhoud, aanbevelingen of beslissingen [7] die van invloed kunnen zijn op fysieke of virtuele omgevingen.”

Toepassing van de definitie van AI-systeem op de OXREC

Voor de toepassing van de definitie van AI-systeem op de OXREC wordt gebruik gemaakt van de AI-verordening inclusief bijbehorende overwegingen en de richtlijnen die de Europese Commissie gepubliceerd heeft op 6 februari 2025 over de definitie van AI-systemen⁸¹.

Voor de beoordeling van de OXREC als AI-systeem worden de hiervoor genoemde zeven elementen toegepast op de OXREC. De richtlijnen maken duidelijk dat deze zeven elementen van toepassing dienen te zijn bij de ontwikkeling van het AI-systeem, of bij de ingebruikname van het AI-systeem maar dat het niet noodzakelijk is dat de elementen in beide fases tot uiting komen.⁸² Bij de interpretatie van de definitie zal dan ook gekeken worden of de zeven elementen bij in ieder geval de ontwikkeling dan wel het gebruik van de OXREC van toepassing zijn.

1. Machine gebaseerd

De OXREC is een machine gebaseerd systeem omdat het op een machine draait⁸³, in dit geval het systeem voor het primaire proces van de reclassering, genaamd IRIS. Ook bij de ontwikkeling van de OXREC is gebruik gemaakt van een machine, namelijk de machine waarin het *Cox proportional hazard regression* model gemaakt is met behulp van Stata 12 software.

2. Autonomie

Bij de OXREC is er sprake van een systeem met enig niveau van autonomie, omdat de output van de OXREC niet volledig bepaald wordt door de mens. Gegeven de dertien inputs van de OXREC, bepaalt het systeem zonder menselijke betrokkenheid op een rekenkundige wijze wat de kans op recidive is. Hiermee is voldaan aan de autonomie vereiste conform overweging 12 van de AI-verordening⁸⁴. Overweging 18 van de richtlijnen verduidelijkt dit nog verder door aan te geven dat ook wanneer de inputs van het systeem door de mens gespecificeerd worden,

⁸¹ Guidelines on the definition of an artificial intelligence system established by AI Act: <https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application>

⁸² Overweging 10 van 'Guidelines on the definition of an artificial intelligence system established by AI Act': "The definition of an AI system adopts a lifecycle-based perspective encompassing two main phases: the pre-deployment or 'building' phase of the system and the post-deployment or 'use' phase of the system. The seven elements set out in that definition are not required to be present continuously throughout both phases of that lifecycle. Instead, the definition acknowledges that specific elements may appear at one phase, but may not persist across both phases."

⁸³ Overweging 12 van de AI-verordening: "De term 'op een machine gebaseerd' verwijst naar het feit dat AI-systemen op machines draaien."

⁸⁴ Overweging 12 van de AI-verordening: "AI-systemen worden zodanig ontworpen dat zij in verschillende mate autonoom kunnen functioneren, wat betekent dat zij een zekere mate van onafhankelijkheid van menselijke betrokkenheid bezitten en zonder menselijke tussenkomst kunnen functioneren."

er nog steeds sprake is van een autonoom systeem omdat de output (de kans op recidive) niet volledig bepaald wordt door de mens.⁸⁵

3. Vertonen van aanpassingsvermogen na inzet

De OXREC vertoont geen aanpassingsvermogen na ingebruikname door de reclassering. Dat wil zeggen dat de modelparameters hetzelfde zijn als bij de publicatie van het systeem in 2016. Echter, het vertonen van aanpassingsvermogen na inzet is een optioneel en niet-verplicht element in de bepaling of de OXREC een AI-systeem is. Dit wordt benadrukt door het woord 'kan' (of het woord 'may' in de oorspronkelijke Engelse tekst) van artikel 3(1) AI-verordening. De richtlijnen geven hier verdere verduidelijking over: het aanpassingsvermogen is facultatief en niet noodzakelijk voor de bepaling of de OXREC een AI-systeem is.⁸⁶ Derhalve sluit het ontbreken van aanpassingsvermogen de beoordeling van de OXREC als AI-systeem niet uit.

4. Doelstellingen

De doelstelling van de OXREC bij de reclassering is het voorspellen van recidive van verdachten of veroordeelden. Hiermee heeft de OXREC een expliciete doelstelling.

De doelstelling van de reclassering wijkt enigszins af van de oorspronkelijke doelstelling van de ontwikkelaar van de OXREC, namelijk het voorspellen van recidive bij personen die vrijgelaten worden uit de gevangenis. Dit neemt niet weg dat de doelstelling nog steeds expliciet is, ondanks dat de OXREC voor een ander doel wordt ingezet.⁸⁷

5. Inferentievermogen

De OXREC is ontwikkeld door middel van *supervised machine learning*, waarmee het systeem een inferentievermogen vertoont om de kans op recidive af te leiden uit de dertien inputvariabelen.⁸⁸ Er is sprake van *supervised machine learning*⁸⁹, omdat het algoritme bij de ontwikkeling getraind is op de gelabelde data van 37.100 Zweedse gedetineerden. De labels hierbij zijn of de gedetineerden wel of niet gerecidiveerd hebben en wanneer dit plaatsvond. Er is door middel van Cox

⁸⁵ Overweging 18 van 'Guidelines on the definition of an artificial intelligence system established by AI Act': "For example, a system that requires manually provided inputs to generate an output by itself is a system with 'some degree of independence of action'. Because the system is designed with the capability to generate an output without this output being manually controlled, or explicitly and exactly specified by a human."

⁸⁶ Overweging 23 van 'Guidelines on the definition of an artificial intelligence system established by AI Act': "a system's ability to automatically learn, discover new patterns, or identify relationships in the data beyond what it was initially trained on is a facultative and thus not a decisive condition for determining whether the system qualifies as an AI system."

⁸⁷ Overweging 12 van de AI-verordening: "De doelstellingen van een AI-systeem kunnen verschillen van het beoogde doel van het AI-systeem in een specifieke context."

⁸⁸ Overweging 12 van de AI-verordening: "De technieken die inferentie mogelijk maken bij de opbouw van een AI-systeem, omvatten benaderingen op basis van machinaal leren waarbij aan de hand van data wordt geleerd hoe bepaalde doelstellingen kunnen worden bereikt."

⁸⁹ Overweging 33 van 'Guidelines on the definition of an artificial intelligence system established by AI Act': "In the case of supervised learning, the AI system learns from annotations (labelled data), whereby the input data is paired with the correct output. The system uses those annotations to learn a mapping from inputs to outputs and then generalises this to new, unseen data."

proportional hazards regression bepaald wat de relatie is tussen de input variabelen en de kans op recidive. Hiermee vertoont het systeem het vermogen om te infereren.

6. Output

De output van de OXREC is de voorspelde kans op recidive, en daarmee genereert de OXREC een output zoals een voorspelling.

7. Invloed op de omgeving

De OXREC heeft invloed op de omgeving, en heeft tenminste invloed op de registraties van de uitkomsten van de OXREC in het softwaresysteem van de reclassering. Hiermee heeft het in ieder geval invloed op een virtuele omgeving.⁹⁰

Conclusie

De Inspectie JenV concludeert dat de OXREC voldoet aan alle elementen die bepalend zijn voor een AI-systeem, en dat de OXREC derhalve een AI-systeem is conform artikel 3(1) AI-verordening.

⁹⁰ Overweging 60 van 'Guidelines on the definition of an artificial intelligence system established by AI Act': "Reference to 'physical or virtual environments' indicates that the influence of an AI system may be both to tangible, physical objects (e.g. robot arm) and to virtual environments, including digital spaces, data flows, and software ecosystems."

Inspectie Justitie en Veiligheid

Toezicht, omdat rechtvaardigheid en veiligheid
niet vanzelfsprekend zijn.

Dit is een uitgave van:

Inspectie Justitie en Veiligheid
Ministerie van Justitie en Veiligheid
Hoge Nieuwstraat 8 | 2514 EL Den Haag
Postbus 20301 | 2500 EH Den Haag
Contactformulier | **www.inspectie-jenv.nl**

Januari 2026

*Aan deze publicatie kunnen geen rechten worden ontleend.
Vermenigvuldigen van informatie uit deze publicatie is toegestaan,
mits deze uitgave als bron wordt vermeld.*