



Discussion paper

Achtergrondkenmerken van als verdachte geregistreeerde *first offenders* van geweldsmisdrijven: Een LASSO logistische regressie

I. Stolwijk
K. Dijkstra
M. Puts
C. de Coo
F. Klaessen
C. Verkleij

Maart 2025

Inhoud

Samenvatting 3

Dankwoord 4

1. Inleiding 5

1.1 Aanleiding 5

1.2 Onderzoeksvraag 6

2. Achtergrondkenmerken 7

2.1 Eerder onderzoek 7

2.2 Van factoren naar bruikbare indicatoren 10

3. Methoden 14

3.1 Populatie 14

3.2 Analysemethoden 16

4. Resultaten 20

4.1 Jongeren 20

4.2 Volwassenen 30

4.3 Modelvergelijking van de jongeren en volwassenen 39

5. Discussie en conclusie 41

Literatuur 47

Bijlagen 49

Bijlage 1: Operationalisatie van kenmerken 49

Bijlage 2: Missende waarden 59

Bijlage 3: Resultaten basismodel voor de jongeren en de volwassenen 61

Bijlage 4: Robuustheidscontroles 73

Samenvatting

Misdaad en gevoelens van (on)veiligheid zijn van invloed op de kwaliteit van leven en hebben impact op diverse aspecten van het dagelijkse bestaan. Het versterken van de veiligheid is dan ook één van de grote opgaven waar ons land zich voor gesteld ziet. Het CBS wil bij deze opgave een rol spelen door nog meer data beschikbaar te stellen over diverse facetten van het veiligheidsdomein. Ten behoeve van toekomstige trendanalyses heeft het Centraal Bureau voor de Statistiek (CBS) in opdracht van de Korpsleiding van de Politie en het Ministerie van Justitie en Veiligheid (JenV) onderzocht welke achtergrondkenmerken, op basis van beschikbare registerdata, het sterkst samenhangen met door de politie geregistreerde verdachte van geweldscriminaliteit in 2017. Dit huidige rapport betreft de resultaten van een verkennend onderzoek waarin is gezocht naar een set kenmerken die kunnen dienen om op landelijk niveau ontwikkelingen op het gebied van geweldscriminaliteit te monitoren en vroeg te signaleren.

Twee verschillende groepen werden onderzocht: (1) jongeren, 12-25 jaar; en (2) volwassenen, 18 jaar en ouder. Met behulp van LASSO (*Least Absolute Shrinkage and Selection Operator*) logistische regressie werd de data geanalyseerd. Met de LASSO regressie werd op een data-gedreven statistische manier een beperktere selectie gemaakt van de belangrijkste kenmerken. Hiermee werden twee modellen samengesteld, één met zo min mogelijk kenmerken zonder in te leveren op de kwaliteit van het model (Model I) en één beperkter model met zo'n 10 kenmerken (Model II). Bij Model I werd het aantal kenmerken teruggebracht van 34 naar 29 en van 31 naar 28 voor respectievelijk jongeren en volwassenen. Bij Model II werd dit voor zowel jongeren als volwassenen teruggebracht tot 9 kenmerken, maar nam de kwaliteit van het model in beide gevallen af. Uit de modellen bleek dat persoonlijke en netwerkfactoren het sterkst samenhangen met door de politie geregistreerde verdachte *first offenders* van geweldsmisdrijven in 2017. Omgevingsfactoren leken daarentegen een relatief kleine samenhang te hebben met geweldsmisdrijven.

Het huidige onderzoek heeft bijgedragen aan kennis over achtergrondkenmerken op basis van beschikbare CBS-registerdata van als verdachte geregistreerde *first offenders* van geweldscriminaliteit. De resultaten van dit onderzoek kunnen gebruikt worden als leidraad bij het kiezen van kenmerken bij toekomstige (trend)analyses over geregistreerde verdachten en om ontwikkelingen in prevalentie op deze kenmerken in de samenleving te monitoren. Ten behoeve van de externe validiteit moet in vervolgonderzoek eerst bekeken worden hoe stabiel de kenmerken over de tijd zijn. Bij het gebruik van de resultaten is het van belang kanttekeningen en beperkingen van het onderzoek mee te nemen. Zo is het op basis van het huidige onderzoek niet mogelijk om causaliteit aan te tonen tussen de kenmerken en registratie als verdachte, of inzicht te bieden in de achterliggende mechanismen bij de ontwikkeling van crimineel gedrag. Tevens heeft het gebruik van registerdata enkele nadelen, bijvoorbeeld de mogelijke selectiebias die in de verdachtenregistratie zit. Door deze selectiebias kan het voorkomen dat er een bepaalde selectiviteit zit in wie wel en wie niet geregistreerd wordt als verdachte. Ook staan de registraties van verdachte niet een-op-een gelijk aan het aantal geweldsmisdrijven.

Dankwoord

Allereerst danken de auteurs iedereen die op enige manier betrokken is geweest bij de totstandkoming van dit onderzoek. Daarnaast dank aan S. Verschuren en L. van Leeuwen voor de uitvoerige bijdrage in het klaarzetten en bewerken van de data. Ook veel dank aan onze collega's van de afdeling Methodologie voor de kritische en constructieve review van het paper. Tot slot veel dank aan het Informatie-Analyse Team voor de prettige en constructieve samenwerking.

1. Inleiding

1.1 Aanleiding

Misdaad en gevoelens van (on)veiligheid zijn van invloed op de kwaliteit van leven en hebben impact op diverse aspecten van het dagelijkse bestaan. Het versterken van de veiligheid is dan ook één van de grote opgaven waar ons land zich voor gesteld ziet. Het CBS wil bij deze opgave een rol spelen door nog meer data beschikbaar te stellen over diverse facetten van het veiligheidsdomein. De politie speelt een belangrijke rol in het bewerkstelligen van veiligheid en het veiligheidsgevoel door het tegengaan (repressief) en voorkomen (preventief) van criminaliteit. De ontwikkelingen omtrent veiligheid en criminaliteit gaan snel. Om op deze ontwikkelingen grip te houden en daarmee de politie adequaat te kunnen besturen is betrouwbare data van groeiend belang.

In opdracht van de Korpsleiding van de Politie en de Minister van Justitie en Veiligheid (JenV) is in 2018 het Informatie-Analyseteam (IAT) opgericht. Dit betreft een samenwerking tussen Politie, het Ministerie van JenV, en het Centraal Bureau voor de Statistiek (CBS). De taak van het IAT bestaat uit het verrichten van thematische (kwantitatieve) analyses op veiligheidsvraagstukken en op het structureel monitoren van maatschappelijke en veiligheidsontwikkelingen. Beide taken geven bouwstenen voor strategische sturing aan het Korps en zijn input voor overleg tussen Korpsleiding en Minister(ie).

De afgelopen jaren is door het IAT gebouwd aan een generieke monitor voor criminaliteit, namelijk het interne dashboard *Galileo*. Met deze monitor worden lange en korte termijn ontwikkelingen in de criminaliteit in kaart gebracht en gerelateerd aan de werking van de politieorganisatie. Bij de huidige monitoring van criminaliteit wordt gebruik gemaakt van een beperkte set aan demografische en sociaaleconomische achtergrondkenmerken. Deze lijst van kenmerken is niet alleen beperkt; inmiddels is ook bekend dat de associaties met enkele van deze kenmerken (bijvoorbeeld herkomst; Jaarrapport Integratie, 2020) geheel of gedeeltelijk verklaard konden worden door andere kenmerken. Zo ontstond de wens om de huidige monitoring te verbeteren. Hiermee kan (vroegtijdig) geanticipeerd worden op ontwikkelingen in de maatschappij en kan de inzet van de politieorganisatie hierop aangepast worden. Door samenwerking met het CBS was er de mogelijkheid om grote hoeveelheden kwantitatieve registerdata voor de gehele bevolking te combineren en relevante kenmerken in samenhang te bekijken.

Het hoofddoel van de huidige verkennende studie was om te bepalen welke kenmerken het sterkst samenhangen met (door politie als zodanig geregistreerd) crimineel gedrag. Deze kenmerken kunnen gebruikt worden voor trendanalyses en

monitoring. Er is gezocht naar een set kenmerken die kunnen dienen om op landelijk en/of regionaal niveau ontwikkelingen te monitoren en vroeg te signaleren, en dus niet op individueel niveau. Onder andere omdat het gebruik van onderzoeksresultaten over verdachtenregistraties voor nieuwe verdachtenregistraties kan leiden tot een vertekend en/of zelfversterkend effect. Door kenmerken in relatie tot elkaar te bekijken en na te gaan in hoeverre achtergrondkenmerken samenhangen met crimineel gedrag wordt het mogelijk om een meer gefundeerde keuze te maken voor het meenemen van bepaalde kenmerken in het dashboard *Galileo* en daarnaast voor andere (vervolg)onderzoeken. Zo wordt de kwaliteit van het onderzoek verhoogd en wordt de maatschappelijke meerwaarde van data vergroot.

1.2 Onderzoeksvraag

Het huidige onderzoek omvatte een empirische studie waarin werd onderzocht welke achtergrondkenmerken het sterkst samenhangen met (door politie geregistreerd) crimineel gedrag. De analyse richtte zich specifiek op geweldsmisdrijven. De hoofdvraag van dit onderzoek luidde: Wat zijn op basis van beschikbare registerdata de belangrijkste achtergrondkenmerken van geregistreerde verdachten van geweldscriminaliteit? Hierbij werd gekeken naar twee onderzoeksgroepen, namelijk jongeren (12 t/m 25 jaar) en volwassenen (18 jaar en ouder).

In dit paper wordt allereerst een overzicht van de literatuur beschreven. Op basis daarvan werd een selectie gemaakt van relevante achtergrondkenmerken die middels beschikbare (register)data te meten waren. Vervolgens wordt de methode beschreven waarmee de onderzoeksvragen zijn beantwoord. Hierin worden de populatieafbakening, operationalisatie van de kenmerken, en analysemethoden beschreven. Daarna worden de resultaten van de analyses besproken. Er wordt afgesloten met een discussie en conclusie.

2. Achtergrondkenmerken

In dit hoofdstuk worden twee onderzoeken uitgelicht die het uitgangspunt vormden voor de huidige studie, en worden verschillende typen achtergrondkenmerken beschreven (2.1). Ten tweede wordt beschreven hoe de factoren uit de literatuur zijn omgezet naar bruikbare indicatoren voor het huidige onderzoek (2.2). Hierin wordt een korte beschrijving gegeven hoe er van de in de literatuur gevonden factoren is gekomen tot de indicatoren die worden meegenomen in dit onderzoek. Eerst wordt gekeken naar de voorwaarden, de indeling van de factoren in de domeinen en factoren die niet uit de literatuur afkomstig zijn maar mogelijk wel van belang zijn om mee te nemen. Vervolgens wordt een overzicht gegeven van de gebruikte indicatoren.

2.1 Eerder onderzoek

Binnen de literatuur wordt onderscheid gemaakt tussen risico-, beschermende, en versterkende factoren (bv Servaas, Weerman, en Fischer, 2021). Dit is een veelgebruikt onderscheid binnen de criminologie met een belangrijke kanttekening. De genoemde factoren zijn factoren die samenhangen met crimineel gedrag. Dit betekent echter niet dat het ook verklarende factoren zijn. Om te kunnen spreken van verklarende factoren is experimenteel onderzoek nodig. Dergelijk experimenteel onderzoek maakte geen deel uit van het huidige onderzoek. Criminogene risicofactoren zijn kenmerken en omstandigheden van mensen en hun omgeving die de kans vergroten dat iemand betrokken is of raakt bij crimineel gedrag. Daarnaast kunnen risicofactoren ook ten aanzien van recidive een voorspellende waarde hebben. Versterkende factoren zijn kenmerken van de personen en hun omgeving die gerelateerd zijn aan een lagere kans op crimineel gedrag en recidive. Beschermende factoren zijn kenmerken van personen en hun omgeving die bij aanwezigheid van een risicofactor de kans op crimineel gedrag en recidive verlagen.

Twee eerdere onderzoeken vormden het uitgangspunt van de huidige studie, namelijk de empirische studie van Boutellier, Scholte en Heijnen (2008) en een recente literatuurstudie van Servaas e.a. (2021). In de studie van Boutellier e.a. (2008) werd op geaggregeerd niveau (o.a., stad en stadsdelen van Amsterdam) gekeken naar relevante risicofactoren voor criminaliteit.

Servaas e.a. (2021) hebben op basis van eerdere overzichtsstudies en meta-analyses een uitgebreid literatuuronderzoek gedaan naar de belangrijkste factoren die samenhangen met crimineel gedrag. De factoren werden ingedeeld in vijf domeinen, namelijk: (1) persoonlijke factoren; (2) gezins- en factoren; (3) peerfactoren; (4) school- en werkgerelateerde factoren; en (5) omgevingsfactoren. Deze indeling vormde thans de basis voor het huidige onderzoek. Hieronder volgt een beschrijving van factoren per domein die uit bovengenoemde studies naar voren kwamen.

2.1.1 Persoonlijke factoren

Factoren in het persoonlijke domein hebben betrekking op kenmerken van het individu zelf. Uit de studie van Servaas e.a. (2021) blijken vooral factoren uit de jeugd en adolescentie belangrijke voorspellers van crimineel gedrag. Het gaat dan bijvoorbeeld om internaliserende (bv. depressie, angst) en externaliserende factoren (bv. hyperactiviteit), een gebrek aan schuldgevoel en een hoge mate van lef. Jongeren met (één of meer van) deze kenmerken hebben een grotere kans om crimineel gedrag te vertonen dan jongeren zonder deze kenmerken. Een hoge mate van intelligentie hangt daarentegen samen met een lagere kans op crimineel gedrag, met name voor hen die aan veel risicofactoren worden blootgesteld. Tijdens de adolescentie speelt het geloof in *the code of the street* een belangrijke rol, welke is ingegeven door “slechte omstandigheden in de buurt, een gebrek aan conventionele mogelijkheden om succesvol te worden, uitsluiting van sociale instellingen, hoge criminaliteitscijfers en een onjuiste behandeling door wetshandhavers” (Servaas e.a., 2021, p. 17). Hoe meer men gelooft in deze straatcode hoe sterker de samenhang met crimineel gedrag. Boutellier e.a. (2008) richtten zich uitsluitend op geaggregeerde gegevens en vonden een sterke samenhang met achtergrondkenmerken van individuen, zoals het aandeel mannen, jongeren en adolescenten, en personen met een niet-westerse herkomst in de gemeente Amsterdam. Daarnaast hingen verslavingsproblematiek, armoede en capaciteitsbeperkingen (o.a. lage intelligentie, laag opleidingsniveau, en een gebrek aan basisvaardigheden; Boutellier e.a., 2008) sterk samen met de ontwikkeling van crimineel gedrag.

2.1.2 Gezins- en familiefactoren

In het gezins- en familiedomein spelen de gezinsstructuur en problematische omstandigheden binnen het gezin een belangrijke rol bij de ontwikkeling van crimineel gedrag. Belangrijke factoren van de gezinsstructuur zijn het hebben van gescheiden ouders, het opgroeien in een éénoudergezin (veelal met de moeder), en ouderschap gedurende de adolescentie. Voorbeelden van problematische omstandigheden binnen het gezin zijn conflicten binnen het gezin, verwaarlozing en kindermishandeling, en crimineel gedrag van ouders met als mogelijk gevolg de betrokkenheid van hulpinstanties. Een gebrek aan sociale controle, emotionele problemen en een beperkt opvoedingsrepertoire hangen hier onder meer mee samen. De opvoedstijl van ouders speelt een belangrijke rol in de ontwikkeling van crimineel gedrag. Zo vergroot een autoritaire opvoedstijl en psychologische controle van het kind de kans op het ontwikkelen van crimineel gedrag, terwijl een hoge mate van steun, discipline en een gevoel van warmte tijdens de opvoeding de kans op crimineel gedrag juist verkleint.

2.1.3 Peerfactoren

Bij peerfactoren gaat het voornamelijk om de rol van vriendschappen. Ondanks dat aan dit thema in de wetenschappelijk literatuur veel belang wordt gehecht, zijn er weinig beschikbare studies die de samenhang van kenmerken uit de directe sociale omgeving in kaart hebben gebracht. Uit de beschikbare literatuur bleek dat de omgang met delinquente anderen samenhangt met een verhoogde kans dat

iemand zelf delicten pleegt. Het kan dan gaan om vrienden, maar ook om delinquente ouders of andere delinquente bekenden. Deze factoren zijn zowel van invloed tijdens de adolescentie als in de volwassenheid.

2.1.4 School- en werkgerelateerde factoren

Het vierde domein bestaat uit school- en werkgerelateerde factoren. Schoolfactoren in de jeugd hangen samen met crimineel gedrag in de volwassenheid. Zo hangen slechte schoolprestaties, leerplichtigen die verzuimen (i.e. voortijdig schoolverlaters), en een gebrek aan betrokkenheid bij school samen met crimineel gedrag. In het latere leven spelen ook factoren uit de werkomgeving een rol. Zo vergroot werkloosheid de kans op vermogenscriminaliteit. Dit komt niet alleen voort uit armoede maar ook omdat werklozen meer tijd tot hun beschikking hebben, de informele sociale controle van de werkomgeving afwezig is en omdat het de rem op het nastreven van een criminele carrière opheft. Het hebben van een baan betekent echter niet per definitie dat mensen geen crimineel gedrag kunnen ontwikkelen. Zo zijn banen die van korte duur zijn of van slechte kwaliteit belangrijke risicofactoren van crimineel gedrag. Het hebben van stabiel werk van voldoende kwaliteit hangt daarentegen wel samen met een verlaagde kans op criminaliteit.

2.1.5 Omgevingsfactoren

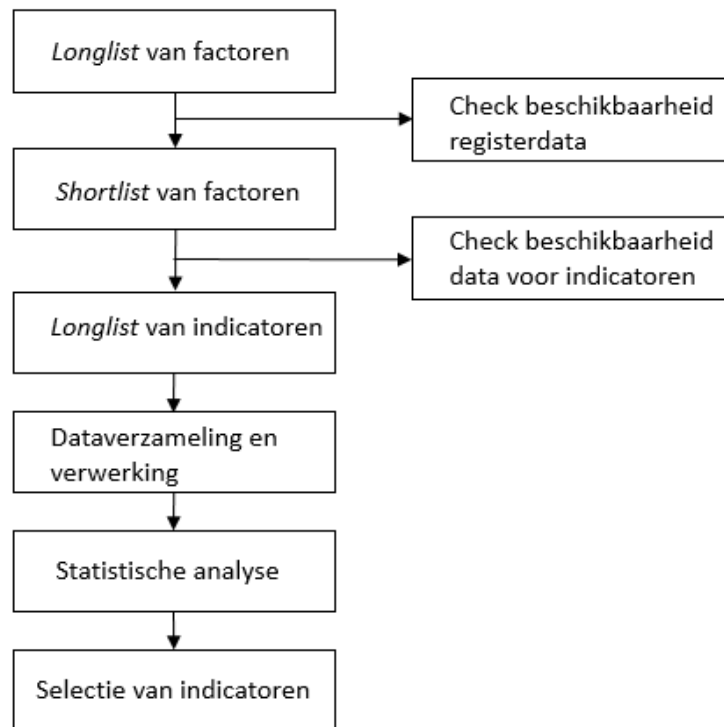
Tot slot spelen omgevingsfactoren een rol. De literatuur richtte zich hierbij voornamelijk op kenmerken van buurten en wijken, onderverdeeld naar sociaaleconomische kenmerken, fysieke kenmerken en sociale (des)organisatie. Bij sociaaleconomische risicofactoren blijken onder meer armoede en ongelijkheid in een buurt of wijk samen te hangen met de ontwikkeling van crimineel gedrag. Armoede en materiële ongelijkheid hangen negatief samen met de sociale controle in een woonbuurt, waardoor de kans op crimineel gedrag in de buurt toeneemt. Een hogere sociaaleconomische status in een buurt (gemeten op basis van inkomen) hangt juist samen met een kleinere kans op crimineel gedrag. Naast sociaaleconomische factoren hangt ook de fysieke structuur van een buurt samen met ontwikkelingen in crimineel gedrag. Voorbeelden hiervan zijn fysieke verloedering van de woonomgeving en slechte huisvesting. Deze kenmerken geven het signaal af dat niemand betrokken is bij de buurt en dat de sociale controle laag is. Hierdoor trekken goedwillende burgers zich terug en worden criminelen juist aangetrokken, hetgeen kan leiden tot een neerwaartse spiraal van criminaliteit. Bij sociale (des)organisatie verhogen tot slot vooral een hoge mate van mobiliteit en de aanwezigheid van lokale peergroepen de kans op crimineel gedrag. Het gebrek aan binding met de buurt en lokale organisatie speelt hierbij een belangrijke rol. De aanwezigheid van sterke (non-)profit organisaties, informele sociale controle en sociale ondersteuning van buurtbewoners hangen daarentegen samen met een kleinere kans op crimineel gedrag.

2.2 Van factoren naar bruikbare indicatoren

Op basis van de hierboven genoemde factoren en (interne) expertsessies¹ is getracht bijpassende indicatoren te vinden op basis van beschikbare (register)data van het CBS. Hiervoor werd een aantal voorwaarden gehanteerd, namelijk: het kenmerk komt in de literatuur naar voren als relevant voor criminaliteit; het kenmerk is te vertalen (operationalisatie) naar beschikbare (CBS-)registerdata; en het kenmerk is meetbaar over de tijd. Voor enkele kenmerken die uit de literatuur naar voren kwamen was het niet mogelijk om deze om te zetten naar bruikbare indicatoren. Dit omvatte onder meer kenmerken gerelateerd aan de opvoedstijl van de ouders, geloof in de straatcode, een gebrek aan betrokkenheid bij school, en informele sociale controle in de buurt. Daarnaast waren peerfactoren niet direct één-op-één te operationaliseren. Deze werden geoperationaliseerd als omgang met delinquente anderen. Dit zijn niet alleen vrienden, maar ook andere personen in het sociale netwerk van iemand. Om die reden is er in dit onderzoek voor gekozen om verschillende netwerkindicatoren op te nemen als peerfactoren (hierna netwerkfactoren).

Naast de indicatoren die genoemd werden in het literatuuronderzoek zijn enkele indicatoren toegevoegd op basis van (interne) expertsessies, waaronder: wonen in een institutioneel huishouden, een Halt-verwijzing voor een overtreding, het zijn van een scholier of student, en het ontvangen van jeugdhulp zonder verblijf. Een overzicht van de stappen die gezet zijn om tot een uiteindelijke selectie van indicatoren te komen, is weergegeven in figuur 1.

¹ Binnen deze expertsessies werd met inhoudelijk experts van binnen en buiten het CBS gekeken naar welke beschikbare registerdata (of combinatie van data) gebruikt kon worden voor de indicator. Daarnaast werd ook bekeken of er factoren waren die niet uit het literatuuronderzoek van Servaas e.a. (2021) en Bouttelier e.a. (2008) naar voren kwamen, maar op basis van de inhoud en ervaring wel als mogelijk belangrijk geacht werden.



Figuur 1. Overzicht selectieprocedure, gebaseerd op Bouttellier e.a. (2008).

Er werden 45 kenmerken geselecteerd, waarbij onderscheid werd gemaakt tussen indicatoren die relevant waren voor (1) jongeren, met een leeftijd tussen de 12 en 25 jaar; en (2) volwassenen, met een leeftijd vanaf 18 jaar. De leeftijdsgrenzen werden zo gekozen dat ze aansluiten op andere analyses van het IAT en daarmee goed toepasbaar zijn voor het IAT. Dit resulteerde in enige overlap in de onderzoeksgroepen. Voor de jongeren was sprake van 36 indicatoren en voor de volwassenen 31, zie tabel 1.

Een voorbeeld van een indicator die inhoudelijk minder relevant is voor jongeren, is het recente arbeidsverleden. Als een volwassene in de afgelopen vier jaar ‘maar’ twaalf maanden heeft gewerkt, kan dat een indicator zijn van armoede of andere problemen. Als een jongere eenzelfde aantal maanden heeft gewerkt, kan dit te maken hebben met het volgen van een opleiding. Vanwege risico op onthulling en ter waarborging van de robuustheid van de analyse werden kenmerken alleen meegenomen indien deze een minimaal aantal keer (100) voorkwam per onderzoeksgroep. Indien dit niet het geval was, werd het kenmerk niet meegenomen voor de betreffende onderzoeksgroep of werden categorieën (indien mogelijk) samengevoegd. Om deze reden werd voor het jongerenmodel de indicator medicijngebruik voor middelengebruik niet meegenomen, en werd voor de indicator geboren in Nederland i.c.m. herkomst een ingedikte variant gebruikt (drie i.p.v. tien categorieën).

Een deel van de kenmerken werd gepeild op 1 januari 2017 (nader genoemd peildatum), een deel van de kenmerken in de jaren vóór de peildatum, en voor een deel van de kenmerken werd bekeken of deze actief waren ten tijde van de peildatum (bijvoorbeeld jeugdbescherming en jeugdhulp met/zonder verblijf).

Hiermee werd het overgrote deel van de kenmerken gemeten vóór de registratie als verdachte *first offender* van een geweldsmisdrijf. Bijlage 1 beschrijft de operationalisatie van de kenmerken in meer detail (zie tabel A1).

Tabel 1. Overzicht van indicatoren per onderzoeksgroep

Indicator	Jongeren	Volwassenen
<i>Persoonlijke factoren</i>		
Slachtoffer van een ernstig misdrijf	X	X
Halt-verwijzing overtreding	X	
Medicijngebruik middelengebruik		X
Medicijngebruik hyperactiviteit	X	X
Medicijngebruik angststoornis	X	X
Medicijngebruik depressie	X	X
Negatieve levenservaringen	X	X
Geboren in Nederland (ja/nee) i.c.m. herkomstland	X	X
<i>Gezins- en familiefactoren</i>		
Tienerouder	X	
Institutioneel huishouden	X	X
Jeugdhulp met verblijf	X	
Jeugdhulp zonder verblijf	X	
Jeugdbescherming	X	
Iemand met problematische schulden in het huishouden	X	X
Eenouderhuishouden met (wel of niet thuiswonende) minderjarig(e) kind(eren)		X
Eenouderhuishouden, kind	X	
Slachtoffer van een ernstig misdrijf in huishouden	X	X
Ouder veroordeeld door OM of rechtbank	X	X
Aantal personen in huishouden	X	
Gestandaardiseerd besteedbaar inkomen huishouden	X	X
<i>Netwerfactoren</i>		
Veroordeelde huisgenoot	X	X
Veroordeelde in familienetwerk	X	X
Veroordeelde in klasgenotennetwerk	X	
Veroordeelde in colleganetwerk		X
Veroordeelde in burennetwerk	X	X
<i>School- en werk gerelateerde factoren</i>		
Scholier of student	X	X
Hoogst behaalde onderwijsniveau		X
Hoogst gevolgde onderwijsniveau	X	
Speciaal onderwijs	X	

Voortijdig schoolverlater	X	
Persoon is gezakt	X	
Licht verstandelijke beperking	X	X
Gemiddeld cijfer eindexamen	X	
Halt-verwijzing leerplichtwet	X	
Werkzaam als zelfstandige		X
Deeltijdfactor werk	X	X
Uitzend- of oproepkracht		X
Werkzaam met tijdelijk contract		X
Kwetsbare arbeidsmarktpositie		X
Persoonlijk inkomen		X
Recent arbeidsverleden		X
<i>Omgevingsfactoren</i>		
Misdrijven per 1000 inwoners buurt	X	X
Gemiddelde SES-totaalscore buurt	X	X
Stedelijkheidsgraad buurt	X	X
Diversiteit buurt	X	X

3. Methoden

In dit hoofdstuk wordt de populatie van het onderzoek beschreven, gevolgd door een beschrijving van de analysemethoden.

3.1 Populatie

In het huidige onderzoek werd ervoor gekozen om registratie als verdachte te gebruiken als proxy voor crimineel gedrag. De keuze voor het onderzoeken van de verdachtenpopulatie kent enkele kanttekeningen. De belangrijkste kanttekening zit in de selectiviteit van de registerdata over geregistreeerde verdachten; niet al het crimineel gedrag wordt geregistreerd en bij het geregistreeerde criminele gedrag wordt ook niet altijd een verdachte geregistreerd. Tevens zal de selectiviteit afhangen van het type delict. Desondanks is er gekozen voor deze populatie omdat deze integraal beschikbaar is, in tegenstelling tot bijvoorbeeld zelfgerapporteerd daderschap waarvan alleen steekproeven beschikbaar zijn. Daarnaast is niet gekeken naar veroordeelden, omdat daar personen ontbreken die een sanctie door bijvoorbeeld het OM opgelegd hebben gekregen, of om andere reden dan niet langer verdacht zijn, niet vervolgd worden. Tevens kan ook een veroordeeld persoon in hoger beroep (of cassatie) alsnog worden vrijgesproken.

Hieronder wordt de populatieafbakening voor de populaties verdachten (3.1.1) en de niet-verdachten (3.1.2) besproken.

3.1.1 Verdachten

De populatie verdachten bestond uit alle personen van 12 jaar en ouder die verdacht waren van een geweldsmisdrijf in 2017. Hiervoor is gebruik gemaakt van politieregistraties (de Basisvoorziening Handhaving; BVH). De BVH is het incidentenregistratiesysteem dat de Politie gebruikt. Politied medewerkers registreren hier incidenten en aangiften in. Personen worden hierin als verdachte van een misdrijf geregistreerd als er een redelijk vermoeden van schuld aan dat misdrijf bestaat. Van alle personen is opgenomen hoe vaak en van welk misdrijf ze werden verdacht. Zodra op een later moment de verdachtmaking wordt ingetrokken wordt deze persoon met terugwerkende kracht verwijderd uit de data.

Binnen de huidige studie werden alleen de personen geïnccludeerd die voor het eerst als verdachte geregistreerd waren, de zogeheten *first offenders*. Daarnaast werd enkel gekeken naar verdachten van geweldsmisdrijven. Naar verwachting is dit een distincte groep verdachten ten opzichte van andere type misdrijven, waardoor de achtergrondkenmerken zullen verschillen ten opzichte van andere type misdrijven of criminaliteit in het algemeen. Voor het operationaliseren van geweldsmisdrijven werd de door het CBS gemaakte en gebruikte standaardclassificatie misdrijven (SCM) 2010 aangehouden (Standaardclassificatie

Misdrijven, 2011). Hieruit werden de gewelds- en seksuele misdrijven (SCM 3) en diefstal en inbraak met geweld (SCM 1.1.1) geselecteerd.

Tot slot zijn alleen de verdachten meegenomen die op 1 januari van het peiljaar (2017) volgens de Basisregistratie Personen (BRP) in een gemeente stonden ingeschreven. Personen die niet in de BRP staan ingeschreven, komen in andere registraties ook nauwelijks voor en daardoor is het niet mogelijk om de achtergrondkenmerken van deze personen te bepalen. Voor personen die wel voorkomen in het BRP maar geen inwoner waren van Nederland zijn kenmerken vaak slechter of niet geregistreerd. Door alleen personen te includeren die stonden ingeschreven in een gemeente was sprake van selectieve non-respons. Dit betekent dat achtergrondkenmerken van de niet geïncludeerde personen mogelijk anders zijn dan die van de geïncludeerde personen (bv herkomst). Het was niet mogelijk om hiervoor in het huidige onderzoek te controleren. Tabel 2 beschrijft voor de jongeren- en de volwassenenpopulatie de gehele populatie, het totaal aantal verdachten, het totale aantal *first offenders*, en het aantal *first offenders* van geweldsmisdrijven.

Tabel 2. Beschrijving van de onderzoekspopulatie in 2017.

	Jongeren (12-25 jaar)		Volwassenen (18 jaar en ouder)	
	Aantal	%	Aantal	%
Totaal Nederland (incl. niet-verdachten)	2.924.280	-	13.670.840	-
Totaal verdachten	55.950	1,9	126.495	0,9
<i>First offenders</i>	29.110	1,0	52.295	0,4
<i>First offenders</i> van geweldsmisdrijven	6.625	0,2	13.295	0,1

Noot. Het percentage werd berekend ten opzichte van de gehele Nederlandse populatie.

3.1.2 Niet-verdachten

Naast de verdachtenpopulatie werd uit de Nederlandse bevolking (op 1 januari 2017) een groep personen samengesteld die niet in de verdachtenregistraties van 2017 en daarvoor voorkwamen (de niet-verdachten). Hiervoor is niet de hele beschikbare populatie van niet-verdachten gebruikt maar een steekproef getrokken. Dit heeft als voordeel dat analyses minder rekenkracht kosten. Het aantal niet-verdachten in Nederland is namelijk vele malen groter dan het aantal verdachten. Daarnaast is ervoor gekozen bij het trekken van de steekproef niet-verdachten te selecteren met dezelfde leeftijd en geslacht als de verdachten. Door te matchen op leeftijd werd de leeftijdsafhankelijke kans, zoals bij indicatoren als onderwijsniveau, op missende waarden voor verdachten en niet-verdachten gelijk. Mocht dit een vertekening opleveren dan is deze voor beide populaties gelijk. Voor jongeren en de volwassenen werden voor elke verdachte vier niet-verdachte personen uit de Nederlandse bevolking geselecteerd met dezelfde leeftijd en hetzelfde geslacht.

3.2 Analysemethoden

De populatie verdachten en niet-verdachten werden verrijkt met de data van de achtergrondkenmerken. Voor sommige personen ontbraken waarden op sommige indicatoren. Deze missende waarden werden geïmputeerd, wat betekent dat missende waarden worden vervangen door geschatte waarden. Om te onderzoeken wat de belangrijkste kenmerken zijn die samenhangen met registratie als verdachte van geweldsmisdrijven werd een lasso logistische regressieanalyse uitgevoerd. Hieronder worden de verschillende stappen beschreven. In hoofdstuk 4 volgen de resultaten.

3.2.1 Imputatie van missende waarden

Voor een deel van de indicatoren was sprake van missende waarden. Er zijn verschillende mogelijke verklaringen voor missende waarden, met elk eigen mogelijke consequenties voor de robuustheid van de data. In huidig onderzoek werd aangenomen dat de missende waarden *missing at random* (MAR) waren (Rubin, 1976; Van Buuren, & Groothuis-Oudshoorn, 2011). In deze gevallen is de kans dat een waarde ontbreekt afhankelijk van andere waargenomen gegevens. Een voorbeeld van een indicator die geïmputeerd is, is het behaalde onderwijsniveau. Voor personen die onderwijs in het buitenland of particulier onderwijs gevolgd hebben, en voor een groot deel van de personen die onderwijs volgden voor de start van verschillende onderwijsregisters, zijn geen onderwijsregistraties beschikbaar. Deze personen hebben echter in de meeste gevallen wel een onderwijsniveau. Het onderwijsniveau van een persoon hangt in de praktijk samen met andere factoren zoals inkomen, leeftijd, geslacht en gezinssamenstelling, die meegenomen kunnen worden bij het imputeren. Voor sommige indicatoren werd ervoor gekozen missende waarden niet te imputeren. Dit gold bijvoorbeeld voor het inkomen van personen in een institutioneel huishouden. Een inschatting van het inkomen van personen in institutionele huishoudens is complex en instabiel. Als alternatief werden deze personen ingedeeld in een middencategorie of grootste categorie.

Doordat aangenomen werd dat de missende waarden MAR waren, werd gebruikgemaakt van multiële imputatie. Hiermee worden de missende waarden meerdere keren geschat om onzekerheid in de schattingen expliciet mee te kunnen nemen. Iedere indicator met een missende waarde werd geïmputeerd met een model op basis van de geobserveerde waarden van een set van twaalf indicatoren als voorspellers, namelijk: leeftijd, geslacht, type huishouden, plaats in het huishouden, uitkering (ao, ww en/of bijstand), geboren in Nederland i.c.m. tiendeling herkomst, scholier of student, deeltijdfactor werk, gestandaardiseerd besteedbaar huishoudensinkomen, persoonlijk inkomen, gemiddelde sociaaleconomische status (SES) van de buurt, en stedelijkheidsgraad van de buurt. Deze indicatoren worden verder beschreven in bijlage 1.

Een overzicht van het aantal missende waarden per indicator is weergegeven in tabel A2 (bijlage 2). De missende waarden besloegen tussen de 0 en 3,8 procent van de indicatoren voor de jongeren, en tussen de 0 en 29 procent van de variabelen voor de volwassenen. Voor de volwassenen was voor het hoogst behaalde onderwijsniveau sprake van een relatief groot aandeel missende waarden. Dit had

te maken met de beperkte beschikbaarheid van registraties in het verleden. Zodoende werd ervoor gekozen de categorieën van de indicator in te dikken van acht categorieën (zoals bij de jongeren) naar drie categorieën, en deze vervolgens te imputeren. In totaal zijn drie geïmputeerde datasets gegenereerd aan de hand van een *random forest* of *predictive mean matching* methode, zodat kon worden omgegaan met zowel numerieke als categorische waarden. De imputaties zijn apart uitgevoerd voor de jongeren- en de volwassenen.

3.2.2 Hoofdanalyse

Het doel van dit onderzoek was om inzicht te krijgen in welke kenmerken het meest samenhangen met geregistreerd worden als verdachte van geweldscriminaliteit. Vanwege het grote aantal kenmerken dat werd afgeleid en het doel van dit onderzoek om tot de belangrijkste kenmerken te komen werd een LASSO (*Least Absolute Shrinkage and Selection Operator*) logistische regressie (hierna lasso regressie) uitgevoerd om tot beperktere sets van kenmerken te komen. Dit is een variant op een multivariate logistische regressie, een statistische methode om de kans op een bepaalde uitkomst (verdachte van een geweldsmisdrijf) te berekenen door meerdere factoren tegelijk mee te nemen.

Lasso regressie is een data-gedreven exploratieve selectiemethode waarmee variabelen geselecteerd worden op basis van hun bijdrage aan het model (Tibshirani, 1996; McNeish, 2015). Door een zogeheten strafterm (λ ; lambda) toe te voegen aan de coëfficiënten van de indicatoren bij het schatten van het model zorgt lasso ervoor dat sommige coëfficiënten naar nul worden teruggebracht waardoor deze in essentie worden uitgesloten. De grootte van de strafterm λ heeft invloed op het aantal coëfficiënten dat tot nul wordt teruggebracht. Een geschikte strafterm voor een dataset kan het beste empirisch worden bepaald met behulp van kruisvalidatie. Daarbij wordt de regressie geschat met een reeks steeds strengere straftermen en geëvalueerd op data die niet is meegenomen bij het schatten. Hoeveel kenmerken bij een bepaalde strafterm overblijven is afhankelijk van de data en dus niet vooraf te voorspellen. Bij een lasso regressie worden alle categorieën van de categorische indicatoren als individuele variabelen beschouwd. Het kan dus voorkomen dat van een categorisch kenmerk geen, of slechts enkele categorieën worden meegenomen. Het voordeel van de lasso aanpak is dat de kenmerken in samenhang worden beschouwd. Kenmerken die niet statistisch significant zijn in een reguliere regressie kunnen nog steeds van invloed kunnen zijn op de samenhang van de andere variabelen. Zodoende is de lasso regressie te verkiezen boven het maken van een selectie puur op basis van statistische significantie of effectgrootte.

De lasso regressies werden uitgevoerd voor zowel de jongeren als de volwassenen met als uitkomstmaat (afhankelijke variabele) de registratie als verdachte *first offender* van een geweldsmisdrijf (ja/nee), en de kenmerken als voorspellers (onafhankelijke variabelen). Zoals uitgelegd in paragraaf 3.1 werden per verdachte vier op leeftijd en geslacht gematchte niet-verdachten meegenomen. Om deze reden werden de regressie analyses gewogen uitgevoerd om bepaalde uitkomstmaten makkelijker te kunnen interpreteren. Uit de reeks lasso regressies werden twee modellen geselecteerd: (1) Model I, een model waarvan de

kruisgevalideerde afwijking, een maat van hoe ver de aannemelijkheid van een model afwijkt van die van een perfect model, maximaal 1 standaardfout verschilt van het lasso model met de laagste afwijking.; (2) Model II, het model met de strengste strafterm waarbij minstens tien variabelen overbleven.

Bij Model I was het doel zo min mogelijk variabelen mee te nemen, zónder significant in te leveren op kwaliteit. Ter illustratie werd in Model II gekozen voor een strengere strafterm om zo uit te komen op ongeveer tien kenmerken.

De geschatte coëfficiënten van een lasso model zijn beïnvloed door de strafterm. Hierdoor kunnen ze niet op dezelfde wijze geïnterpreteerd worden als bij een reguliere logistische regressie. Om de uitkomsten beter te kunnen interpreteren, is ervoor gekozen om de beperkte set van kenmerken die uit de modellen I en II kwamen, te gebruiken als input voor een reguliere multivariate logistische regressie. Door het wegvallen van categorieën in de lasso was het hierbij niet meer nodig handmatig een referentiecategorie aan te wijzen, de categorieën die door de lasso werden geëlimineerd vormen samen de referentiecategorie van een categoriaal kenmerk.

Het model werd geschat op 80% van de data (schattingset). De resterende 20% werd gebruikt voor evaluatie van de kwaliteit van het model. De samenhang tussen de indicatoren en de uitkomstmaat zijn weergegeven met een *odds ratio* (*OR*), welke de mate beschrijft waarin de kans op een bepaalde uitkomst verschilt tussen twee groepen. Hierbij geeft een waarde groter dan 1 aan dat de relatieve kans op de uitkomst groter is in de onderzochte groep ten opzichte van de referentiecategorie, en een waarde kleiner dan 1 wijst op een kleinere kans op de uitkomst. De *OR*'s en het bijbehorende 95% betrouwbaarheidsinterval werden geschat met behulp van bootstrapping (1000 iteraties). Hierbij werden voor de logistische regressie personen uit de schattingset herhaaldelijk, met terugleggen, getrokken, gestratificeerd naar de uitkomstmaat. Voor elke trekking werd de regressie opnieuw geschat, en over deze schattingen werden een gemiddelde en betrouwbaarheidsintervallen voor de *OR*'s bepaald.

Op basis van de overige 20% van de data (evaluatieset) werden voorspellingen gegenereerd waarover enkele kwaliteitsmaten zijn berekend waarmee het logistische regressie model en de lasso modellen werden geëvalueerd. Dit zijn de *precision*, *recall*, *F1-score* en de *balanced accuracy*. *Precision* drukt uit hoeveel personen aangewezen door het model als verdachte ook verdachte waren. *Recall* drukt uit hoeveel verdachten van alle daadwerkelijke verdachten door het model zijn aangewezen als verdachte. De *F1-score* vat deze twee samen in een maat (harmonisch gemiddelde). De *balanced accuracy* wordt berekend als het percentage correcte voorspelling (zowel positief als negatief) ten opzichte van alle voorspellingen, en corrigeert voor de ratio verdachten:niet-verdachten (1:4). Hiernaast wordt voor de logistische regressie een *pseudo R²* gerapporteerd. Bij een lineaire regressie is er de *R²*, die het percentage van de variantie in de uitkomstmaat dat wordt verklaard door de indicatoren uitdrukt. Voor een logistisch model is het niet mogelijk om de traditionele *R²* te gebruiken, daarom zijn er verschillende alternatieven als *pseudo R²*. De volgende *pseudo R²*'s worden gerapporteerd: Tjur's *pseudo R²*, welke aangeeft hoe goed de twee groepen (verdachten/niet-verdachten) onderscheiden kunnen worden (Tjur, 2009);

McFadden's *pseudo R²*, welke is gebaseerd op hoe het model presteert ten opzichte van een nulmodel; en Nagelkerke's *pseudo R²*, wat een herschaling is van Cox en Snell's *pseudo R²* formule voor de verhouding tussen het nulmodel en het model. Deze variëren onderling van grootte, maar hebben allen een maximale waarde van 1 en kunnen enkel geïnterpreteerd worden bij het vergelijken van meerder modellen. Hierbij weergeeft een hogere waarde het betere model.

Voor een totaalbeeld is er ook een multivariate logistische regressie geschat op de complete set kenmerken voor de jongeren (tabel A3) en volwassenen (tabel A4; bijlage 3). Hiertoe wordt nader gerefereerd als het basismodel. Tot slot zijn de aannames van de logistische regressieanalyses gecontroleerd en zijn enkele sensitiviteitsanalyses uitgevoerd ter controle van de robuustheid van de resultaten. Deze worden uitgebreid beschreven in bijlage 4.

3.2.3 Software

De analyses zijn uitgevoerd met R (versie 4.4; R Core Team, 2024) en met behulp van verschillende packages voor de verschillende onderdelen van de analyse. Voor de (evaluatie van de) imputatie is het package mice gebruikt (versie 3.16.0 ; Van Buuren, & Groothuis-Oudshoorn, 2011). Voor de multivariate logistische regressie is (o.a.) de functie glm van het package stats gebruikt, en voor de lasso is de functie cv.glmnet uit het package glmnet gebruikt (Friedman, Hastie, & Tibshiriani, 2010)

4. Resultaten

Dit hoofdstuk beschrijft de resultaten van de jongeren (4.1) en volwassenen (4.2). Per onderzoeksgroep worden de beschrijvende statistieken en de resultaten van de reguliere logistische regressies met de variabelen uit het lasso model (Model I) en het beperkte lasso model (Model II) gepresenteerd. De resultaten van een reguliere logistische regressie (nader genoemd basismodel) op alle kenmerken per onderzoeksgroep is ter referentie opgenomen in bijlage 3. Vervolgens worden de resultaten van het basismodel, Model I, en Model II van de jongeren- en volwassenen vergeleken op enkele kwaliteitsmaten (4.3).

De beschrijvende statistieken in tabel 3 en tabel 4 zijn opgebouwd uit 20% verdachten en 80% op leeftijd en geslacht gematchte niet-verdachten. Het gaat dus niet om de totale omvang van de groep niet-verdachten in de bevolking. Dit betekent dat indien er geen sprake is van verschillen binnen een categorie (bijvoorbeeld het zijn van een tienerouder), de verhouding verdachten en niet-verdachten 20-80 is. Indien het aantal verdachten onder de 20% zit, dan is sprake van een relatief laag aantal geregistreerde verdachten. Indien het aantal verdachten boven de 20% ligt, dan is sprake van een relatief hoog aantal geregistreerde verdachten. De omvang van de groep per categorie is afgerond op tientallen om onthulling en schijnnaauwkeurigheid te voorkomen. De percentages verdachten en niet-verdachten per categorie zijn berekend aan de hand van de afgeronde aantallen.

De lasso regressie werd gebruikt voor kenmerkselectie waarna vervolgens deze set kenmerken opnieuw is geanalyseerd met een reguliere logistische regressie. De uitkomsten hiervan zijn weergegeven in tabel 3 en 4, waarbij de *odds ratio's* (*OR*), betrouwbaarheidsintervallen (*BI*) en kwaliteitsmaten uit de reguliere logistische regressie voortkomen. De samenhang van de indicatoren met de uitkomstmaat worden uitgedrukt in *OR's* per groep (categorie) en zijn te interpreteren ten opzichte van de referentiecategorie(ën). Dit zijn de categorieën die door de lasso regressie zijn geëlimineerd, en waarvoor de *OR's* in de tabel ontbreken. Bijvoorbeeld: bij een *OR* van 2 is de *odds* dat iemand als verdachte geregistreerd staat twee keer zo hoog voor een specifieke groep ten opzichte van de referentiecategorie, en bij een *OR* van 0.5 twee keer zo laag ten opzichte van de referentiecategorie, gecontroleerd voor alle andere kenmerken. Bij elke *OR* wordt het bijbehorende 95% betrouwbaarheidsinterval gegeven. Bij een *OR* van een zijn de *odds* voor verdachten en niet-verdachten gelijk en als de waarde 1 binnen het betrouwbaarheidsinterval valt, is er dus geen sprake van een significante samenhang.

4.1 Jongeren

Hier worden de beschrijvende statistieken en de uitkomsten van de lasso analyse beschreven voor de jongeren van 12 tot en met 25 jaar die al dan niet als *first offender* van een geweldsmisdrijf in 2017 geregistreerd staan. Er wordt eerst

gekeken naar de beschrijvende statistieken, gevolgd door de uitkomsten van de lasso analyse.

4.1.1 Beschrijvende statistieken voor de jongeren

De totale steekproefgrootte van de jongeren was 33.140. Van deze jongeren stond 20% geregistreerd als verdachte *first offender* van een geweldsmisdrijf ($n = 6.625$, zie tabel 2), en 80% bestond uit op leeftijd en geslacht gematchte jongeren die niet geregistreerd stonden als verdachte ($n = 26.515$). De jongeren bestonden voor 80,8% uit mannen en de gemiddelde leeftijd was 17,9 ($SD = 3,4$). Het jongerenmodel bevatte 34 kenmerken, bestaande uit 98 variabelen². De beschrijvende statistieken van alle kenmerken en de selectie van variabelen op basis van de lasso regressie (Model I en II) van de jongeren zijn per domein weergegeven in tabel 3.

Hieronder worden per domein de kenmerken uitgelicht met variabelen waarbij sprake is van een relatief groot aandeel verdachten, namelijk een verhouding van minstens 35% verdachten en 65% niet-verdachten of een relatief klein aandeel verdachten, met maximaal 10% verdachten en 90% niet-verdachten (en daarmee afwijken van de verdeling verdachten/niet-verdachten in de steekproef 20/80). Jongeren die in de zeven jaar voor het pleegjaar slachtoffer zijn geweest van een ernstig misdrijf (en als zodanig bij de politie staan geregistreerd) waren relatief vaak geregistreerd als verdachte (48%). Bij jongeren die geen slachtoffer zijn geweest, is dit aandeel veel lager (18%). Andere persoonlijke factoren met een relatief hoog aantal verdachten waren: een Halt-verwijzing voor een overtreding, en jongeren met een Buiten-Europese herkomst die niet geboren zijn in Nederland. Binnen de gezins- en familiefactoren waren jongeren die op de peildatum een actieve jeugdbeschermingsmaatregel hadden relatief vaak geregistreerd als verdachte *first offender* van een geweldsmisdrijf (68%). Bij jongeren die geen actieve jeugdbeschermingsmaatregel hadden op de peildatum lag dit aandeel lager (20%). Andere kenmerken met een relatief hoog aantal verdachten waren: tienerouder, institutioneel huishouden, jeugdhulp met- en zonder verblijf, iemand met problematische schulden in het huishouden, een ouder veroordeeld door het OM of de rechtbank. Binnen de school- en werkgerelateerde factoren hadden de volgende kenmerken een relatief hoog aantal verdachten: speciaal onderwijs, voortijdig schoolverlater, licht verstandelijke beperking, en een Halt-verwijzing voor de leerplichtwet. Binnen het onderwijsniveau waren enkele categorieën met een relatief laag aantal verdachten, namelijk: havo of vwo, en een hbo- of wo-bachelor of master. Ook binnen de jongeren met een zeer hoog gemiddeld eindexamencijfer was sprake van relatief weinig verdachten. Binnen de omgevingsfactoren waren er geen kenmerken met een verhouding van minstens 35/65 verdachten/niet-verdachten.

De netwerkfactoren zijn niet uitgedrukt in aandeel van (niet-) verdachte met een bepaald netwerk, maar een gemiddelde van het aandeel van de netwerken dat bestond uit personen die eerder veroordeeld waren voor een misdrijf. Voor al

² . Indien er wordt gesproken over variabelen dan worden de onderliggende categorieën van een kenmerk als los van elkaar beschouwd.

deze netwerkfactoren gold dat voor als verdachte geregistreerde jongeren een groter deel van het netwerk uit veroordeelden bestond, dan bij niet-verdachte jongeren. Binnen het familienetwerk was dit verschil het grootst. Bij jongeren die als *first offender* bij de politie geregistreerd stonden voor een geweldsmisdrijf was gemiddeld 14,8% van het familienetwerk in de afgelopen zestien jaar veroordeeld, bij niet-verdachte jongeren ging het om gemiddeld 6,5%.

4.1.2 Uitkomsten van de lasso regressie voor de jongeren

De uitkomsten van de regressie op basis van de kenmerken uit Model I en Model II worden gepresenteerd in tabel 3. Door de lasso analyse werden bij Model I in totaal 29 van de 34 kenmerken geselecteerd, met in totaal 46 variabelen. Hieronder worden de kenmerken met de grootste *odds* beschreven per domein. In het persoonlijke domein werden vijf kenmerken geselecteerd, waaronder: slachtoffer van een ernstig misdrijf ($OR = 2,741$; 95% *BI* [2,415-3,139]) en een Halt-verwijzing voor een overtreding ($OR = 2,412$; 95% *BI* [2,050-2,857]). Daarnaast speelde herkomst een rol. Jongeren die geboren zijn in Nederland met een Nederlandse herkomst ($OR = 0,865$, 95% *BI* [0,757-1,007]), hadden een lagere kans om geregistreerd te staan als verdachte *first offender* van een geweldsmisdrijf – ongeacht of ze geboren zijn in Nederland – dan jongeren met een Europese herkomst. Jongeren met een Buiten-Europese herkomst hadden een hogere kans om geregistreerd te staan als verdachte, ongeacht of ze geboren zijn in Nederland ($OR = 1,347$, 95% *BI* [1,151-1,605]) of niet ($OR = 2,242$; 95% *BI* [1,865-2,716]), ten opzichte van jongeren met een Europese herkomst. Binnen de gezins- en familiefactoren werden bijna alle kenmerken geselecteerd, behalve jeugdbescherming en het aantal personen in een huishouden. De kenmerken met de sterkste samenhang met registratie als verdachte waren jeugdhulp met verblijf ($OR = 3,293$, 95% *BI* [2,530-4,383]) en het zijn van een tienerouder ($OR = 2,274$, 95% *BI* [1,302-3,975]). Daarnaast speelde het gestandaardiseerde besteedbare inkomen van het huishouden een rol. Jongeren met een gestandaardiseerd besteedbaar huishoudensinkomen in het 4e kwintiel ($OR = 0,890$, 95% *BI* [0,801-0,981]) hadden een lagere kans om als verdachte geregistreerd te staan, ten opzichte van jongeren met een inkomen in het 2e of 3e kwintiel.

Alle netwerkfactoren werden geselecteerd binnen Model I van de lasso regressie en toonden significante samenhang met registratie als verdachte. Voor elke stijging van één procentpunt van het aandeel veroordeelden in het familie netwerk nam de *odds* op registratie als verdachte toe met 1,020 (95% *BI* [1,017-1,024]). De andere netwerkfactoren toonden een vergelijkbare samenhang. Binnen de school- en werkgerelateerde factoren werd het overgrote deel van de kenmerken geselecteerd in Model I, alleen een Halt-verwijzing voor de leerplicht niet. De kenmerken met de hoogste *odds* waren voor de jongeren met als hoogst gevolgd onderwijsniveau het basisonderwijs ($OR = 1,803$, 95% *BI* [1,432-2,269]) en vmbo-b/k of mbo 1 ($OR = 1,757$, 95% *BI* [1,583-1,947]), ten opzichte van jongeren met een hoogst gevolgd niveau van mbo 4, vmbo-g/t, of de havo- of vwo-onderbouw. Daarnaast hing het hoogst gevolgd onderwijsniveau ook samen met een lagere kans op registratie als verdachte, specifiek binnen het volgen van havo of vwo ($OR = 0,528$, 95% *BI* [0,455-0,612]), en een hbo- of wo-bachelor ($OR = 0,498$, 95% *BI* [0,435-0,564]) of masteropleiding ($OR = 0,279$, 95% *BI* [0,186-0,400]), ten opzichte

van jongeren met een hoogst gevolgd niveau van mbo 4, vmbo-g/t, of de havo- of vwo-onderbouw. Ook het gemiddelde eindexamencijfer speelde een rol, waarbij de kans op registratie als verdachte afnam naarmate het eindexamencijfer toenam, ten opzichte van jongeren met een zeer laag eindexamencijfer, die nog geen eindexamen hebben gedaan, of waarvan het eindexamencijfer onbekend was. De omgevingsfactoren lieten over het algemeen een kleinere samenhang zien met registratie als verdachte, dan de andere domeinen. Omgevingsfactoren met een lagere kans op registratie als verdachte waren jongeren die vielen onder het 1e kwintiel van het aantal misdrijven per 1000 inwoners ($OR = 0,836$, 95% BI [0,754-0,918]) en die in een niet stedelijke buurt wonen ($OR = 0,895$, 95% BI [0,801-0,995]). De *balanced accuracy* van Model I was 0,71, dit betekent dat het model gemiddeld 71% van zowel de positieve als negatieve uitkomsten correct heeft geclassificeerd, ongeacht de ongelijke verdeling van de groepen verdachten versus niet-verdachten (1:4).

Naast Model I werd een tweede, beperktere selectie uitgevoerd (Model II) waarin 9 van de 34 kenmerken werden geselecteerd, met een totaal van 11 variabelen. Hierin werden zowel kenmerken met een lagere als een hogere *odds* op registratie als verdachte geselecteerd. Kenmerken met de sterkste samenhang met registratie als verdachte waren: slachtoffer van een ernstig misdrijf ($OR = 3,240$, 95% BI [2,874-3,674]), en het volgen van speciaal onderwijs ($OR = 2,141$, 95% BI [1,875-2,410]). Daarnaast waren er kenmerken die samenhangen met een lagere kans op registratie als verdachte, waaronder een hoogst gevolgd onderwijsniveau van havo of vwo ($OR = 0,375$, 95% BI [0,327-0,426]), of een hbo- of wo-bacheloropleiding ($OR = 0,384$, 95% BI [0,344-0,429]), ten opzichte van jongeren met een ander onderwijsniveau. Ook jongeren die geboren zijn in Nederland met een herkomst uit Nederland hadden een lagere kans op registratie als verdachte ($OR = 0,518$, 95% BI [0,484-0,556]), dan jongeren met een andere herkomst. Binnen de netwerkfactoren werden drie kenmerken geselecteerd, namelijk: een veroordeelde huisgenoot, en veroordeelden binnen het familie- en burennetwerk. Omgevingsfactoren werden in Model II niet meer geselecteerd. De *balanced accuracy* van Model II was 0,68 wat betekent dat het model gemiddeld 68% van zowel de positieve als negatieve uitkomsten correct heeft geclassificeerd.

Tabel 3. Beschrijvende statistieken en uitkomsten van de logistische regressies op de kenmerken van lasso Model I en het beperktere lasso Model II voor de jongeren, gecontroleerd voor leeftijd en geslacht.

	Categorieën	Omvang groep	Verdachten/ niet-verdachten	Model I		Model II	
		<i>n</i>	%	<i>OR</i>	<i>95% BI</i>	<i>OR</i>	<i>95% BI</i>
<i>Intercept</i>	-	33.140	20/80	0,488	0,405 - 0,592	0,861	0,796 - 0,93
Persoonlijke factoren							
Slachtoffer van ernstig misdrijf	Ja	1.860	48/52	2,741	2,415 - 3,139	3,240	2,874 - 3,674
	Nee	31.280	18/82	-	-	-	-
Halt-verwijzing overtreding	Ja	1.190	43/57	2,412	2,050 - 2,857	-	-
	Nee	31.950	19/81	-	-	-	-
Medicijngebruik hyperactiviteit	Ja	3.790	33/67	1,769	1,587 - 1,960	1,999	1,812 - 2,198
	Nee	29.340	18/82	-	-	-	-
Medicijngebruik angststoornis	Ja	830	30/70	-	-	-	-
	Nee	32.300	20/80	-	-	-	-
Medicijngebruik depressie	Ja	910	31/69	1,251	1,001 - 1,538	-	-
	Nee	32.220	20/80	-	-	-	-
Negatieve levenservaringen	Ja	950	28/72	-	-	-	-
	Nee	32.190	20/80	-	-	-	-
Geboren in Nederland i.c.m. herkomst							
Geboren in Nederland	Herkomst Nederland	23.780	16/84	0,865	0,757 - 1,007	0,518	0,484 - 0,556
	Herkomst Europa (excl. Nederland)	1.160	21/79	-	-	-	-

	Herkomst buiten Europa	5.020	34/66		1,347	1,151 - 1,605	-	-
Niet geboren in Nederland	Herkomst Europa (excl. Nederland)	1.100	15/85		-	-	-	-
	Herkomst buiten Europa	2.080	35/65		2,242	1,865 - 2,716	-	-
Gezins- en familiefactoren								
Tienerouder	Ja	100	60/40		2,274	1,302 - 3,975	-	-
	Nee	33.030	20/80		-	-	-	-
Institutioneel huishouden	Ja	460	54/46		1,727	1,326 - 2,233	-	-
	Nee	32.670	19/81		-	-	-	-
Jeugdhulp met verblijf	Ja	450	67/33		3,293	2,530 - 4,383	-	-
	Nee	32.690	19/81		-	-	-	-
Jeugdhulp zonder verblijf	Ja	1.720	44/56		1,699	1,474 - 1,957	-	-
	Nee	31.420	19/81		-	-	-	-
Jeugdbescherming	Ja	340	68/32		-	-	-	-
	Nee	32.790	20/80		-	-	-	-
Iemand met problematische schulden in het huishouden	Ja	2.530	43/57		1,435	1,268 - 1,604	1,710	1,523 - 1,903
	Nee	30.600	18/82		-	-	-	-
Eenouderhuishouden, kind	Ja	5.990	30/70		1,266	1,152 - 1,382	-	-
	Nee	27.150	18/82		-	-	-	-
Slachtoffer van een ernstig misdrijf in huishouden	Ja	3.800	32/68		1,418	1,283 - 1,558	-	-
	Nee	29.330	18/82		-	-	-	-

Ouder veroordeeld door OM of rechtbank	Ja	5.080	40/60		1,172	1,056 - 1,311	1,225	1,102 - 1,356
	Nee	28.050	16/84		-	-	-	-
Aantal personen in huishouden	1	3.630	22/78		-	-	-	-
	2	3.640	25/75		-	-	-	-
	3	6.640	21/79		-	-	-	-
	4	11.240	16/84		-	-	-	-
	5 of meer	7.990	21/79		-	-	-	-
Gestandaardiseerd besteedbaar inkomen huishouden	1 t/m 20%	7.230	30/70		1,174	1,076 - 1,287	-	-
	21 t/m 40%	5.050	24/76		-	-	-	-
	41 t/m 60%	6.880	21/79		-	-	-	-
	61 t/m 80%	7.110	14/86		0,890	0,801 - 0,981	-	-
	81 of meer	6.870	12/88		0,917	0,819 - 1,024	-	-
Netwerfactoren								
			Verdachten/ niet-verdachten		Model I		Model II	
		<i>n</i>	<i>M (SD)</i>		<i>OR</i>	<i>95% BI</i>	<i>OR</i>	<i>95% BI</i>
Veroordeelde huisgenoot	-	33.140	14,8 (27,1)/ 5,1 (16,3)		1,004	1,002 - 1,006	1,005	1,004 - 1,007
Veroordeelde in het familienetwerk	-	33.140	14,8 (17,3)/ 6,5 (10,5)		1,020	1,017 - 1,024	1,024	1,020 - 1,027
Veroordeelde in het klasgenotennetwerk	-	33.140	3,5 (9,3)/ 2,0 (5,8)		1,012	1,007 - 1,018	-	-

Veroordeelde in het burennetwerk	-	33.140	9,3 (7,2)/ 6,8 (5,9)		1,012	1,006 - 1,019	1,026	1,020 - 1,031
School- en werkgerelateerde factoren								
Scholier of student	Ja	26.690	19/81		0,743	0,669 - 0,824	-	-
	Nee	6.450	25/75		-	-	-	-
Hoogst gevolgde onderwijsniveau	Basisonderwijs	620	34/66		1,803	1,432 - 2,269	-	-
	Vmbo-b/k, mbo 1	5.000	31/69		1,757	1,583 - 1,947	-	-
	Vmbo-g/t, havo-, vwo-onderbouw	6.630	24/76		-	-	-	-
	Mbo 2 en mbo 3	4.310	31/69		1,571	1,407 - 1,753	-	-
	Havo, vwo	4.360	8/92		0,528	0,455 - 0,612	0,375	0,327 - 0,426
	Mbo 4	5.150	19/81		-	-	-	-
	Hbo-, wo-bachelor	6.270	9/91		0,498	0,435 - 0,564	0,385	0,344 - 0,429
	Hbo-, wo-master, doctor	800	5/95		0,279	0,186 - 0,400	-	-
Speciaal onderwijs	Ja	1.870	47/53		1,717	1,459 - 1,978	2,141	1,875 - 2,410
	Nee	31.260	18/82		-	-	-	-
Voortijdig schoolverlater	Ja	1.930	38/62		1,365	1,174 - 1,574	-	-
	Nee	31.200	19/81		-	-	-	-
Persoon is gezakt	Ja	1.210	21/79		1,235	1,003 - 1,523	-	-
	Nee	31.930	20/80		-	-	-	-
Licht verstandelijke beperking	Ja	230	48/52		1,499	1,040 - 2,098	-	-
	Nee	32.860	20/80		-	-	-	-
Gemiddeld eindexamencijfer	Zeer laag	130	31/69		-	-	-	-

	Laag	1.010	26/74		1,386	1,111 - 1,770	-	-
	Onder gemiddeld	6.950	19/81		1,186	1,061 - 1,335	-	-
	Boven gemiddeld	4.810	13/87		0,823	0,723 - 0,932	-	-
	Hoog	1.710	11/89		0,679	0,554 - 0,836	-	-
	Zeer hoog	570	4/96		0,254	0,137 - 0,418	-	-
	Onbekend/nog geen eindexamen gedaan	17.950	23/77		-	-	-	-
Halt-verwijzing leerplichtwet	Ja	270	52/48		-	-	-	-
	Nee	32.860	20/80		-	-	-	-
Deeltijdfactor werk	Geen baan	17.730	22/78		-	-	-	-
	0,01 t/m 0,29	6.190	15/85		-	-	-	-
	0,30 t/m 0,59	3.160	20/80		1,286	1,126 - 1,454	-	-
	0,60 t/m 0,79	1.280	22/78		-	-	-	-
	0,80 t/m 0,94	1.120	22/78		1,270	1,045 - 1,543	-	-
	0,95 of meer	3.650	19/81		-	-	-	-
Omgevingsfactoren								
Gemiddelde SES-totaalscore in de buurt	1e kwintiel	6.370	30/70		1,126	1,008 - 1,246	-	-
	2e kwintiel	6.050	23/77		-	-	-	-
	3e kwintiel	6.490	19/81		-	-	-	-
	4e kwintiel	6.750	16/84		-	-	-	-
	5e kwintiel	7.480	14/86		0,950	0,862 - 1,053	-	-
Misdrijven per 1000 inwoners in de buurt	1e kwintiel	7.400	14/86		0,836	0,754 - 0,918	-	-
	2e kwintiel	6.950	17/83		-	-	-	-

	3e kwintiel	6.700	20/80	-	-	-	-
	4e kwintiel	6.670	24/76	-	-	-	-
	5e kwintiel	5.420	28/72	1,048	0,948 - 1,161	-	-
Stedelijkheidsgraad buurt	Zeer sterk	7.660	23/77	-	-	-	-
	Sterk	8.330	22/78	-	-	-	-
	Matig	6.200	20/80	1,206	1,096 - 1,324	-	-
	Weinig	5.430	17/83	-	-	-	-
	Niet	5.520	15/85	0,895	0,801 - 0,995	-	-
			Verdachten/niet verdachten	Model I		Model II	
		<i>n</i>	<i>M (SD)</i>	<i>OR</i>	<i>95% BI</i>	<i>OR</i>	<i>95% BI</i>
Diversiteit buurt	-	33.140	0,43 (0,22)/ 0,36 (0,20)	-	-	-	-

Noot. *N* = 33.140 (20% verdachten/80% niet-verdachten). *OR* = *odds ratio*; *BI* = betrouwbaarheidsinterval; *M* = gemiddelde; *SD* = standaardafwijking. De *odds ratio*'s moeten geïnterpreteerd worden ten opzichte van de (groep) categorieën zonder *odds ratio*'s (deze vormen samen de referentiecategorie). Voor de numerieke variabelen werden in plaats van de omvang van de groepen, het gemiddelde en de standaardafwijking gegeven.

4.2 Volwassenen

Na de jongeren in de vorige paragraaf, wordt hier gekeken naar volwassenen van 18 jaar en ouder die al dan niet als *first offender* van een geweldsmisdrijf in 2017 geregistreerd staan. Er wordt eerst gekeken naar de beschrijvende statistieken, gevolgd door de uitkomsten van de lasso analyse.

4.2.1 Beschrijvende statistieken voor de volwassenen

De totale steekproefgrootte van de volwassenen was 66.480. Van deze volwassenen stond 20% geregistreerd als *first offender* van een geweldsmisdrijf ($n = 13.295$, zie tabel 2), en 80% bestond uit op leeftijd en geslacht gematchte volwassenen die niet geregistreerd stonden als verdachte ($n = 53.175$). De volwassenen bestonden voor 79,9% uit mannen en de gemiddelde leeftijd was 39,2 ($SD = 15,4$). Het volwassenenmodel bevatte 31 kenmerken, bestaande uit 99 variabelen. De beschrijvende statistieken van alle kenmerken en de selectie van variabelen op basis van de lasso regressie (Model I en II) van de volwassenen zijn per domein weergegeven in tabel 4.

Hieronder worden per domein de kenmerken uitgelicht met variabelen waarbij sprake is van een van een relatief groot aandeel verdachten, namelijk een verhouding van minstens 35% verdachten en 65% niet-verdachten (en daarmee afwijken van de verdeling verdachten/niet-verdachten in de steekproef 20/80). Volwassenen die medicijnen krijgen voorgeschreven voor middelengebruik waren relatief vaak geregistreerd als verdachte (50%). Bij volwassenen die deze medicijnen niet krijgen voorgeschreven is dit aandeel veel lager (20%). Binnen de persoonlijke factoren speelde daarnaast slachtofferschap van een ernstig misdrijf en herkomst een rol. Bij volwassenen met een herkomst uit Marokko, Suriname of overige Afrikaanse landen – ongeacht of zij geboren zijn in Nederland – was sprake van een relatief groot aandeel verdachten. Dit gold ook voor volwassenen die niet geboren zijn in Nederland met een Nederlands-Caribische herkomst.

Binnen de gezins- en familiefactoren waren volwassenen met iemand met problematische schulden in het huishouden relatief vaak als verdachte geregistreerd (47%). Bij volwassenen zonder iemand met problematische schulden in het huishouden, lag dit aandeel veel lager (18%). Andere kenmerken met een relatief hoog aantal verdachten waren: behoren tot een institutioneel huishouden, een eenouderhuishouden met (wel of niet thuiswonende) minderjarige kinderen, en een ouder hebben die veroordeeld is door het OM of de rechtbank. Binnen de school- en werkgerelateerde factoren hadden volwassenen met een licht verstandelijke beperking een relatief hoog aantal verdachten (37%). Bij volwassenen zonder een licht verstandelijke beperking lag dit aandeel lager (20%). Binnen de omgevingsfactoren waren er geen kenmerken met een verhouding van minstens 35/65 verdachten/niet-verdachten.

Voor alle netwerkfactoren gold dat voor als verdachte geregistreerde volwassenen het aandeel veroordeelden in het netwerk gemiddeld hoger lag, dan voor niet-verdachte volwassenen.

4.2.2 Uitkomsten van de lasso regressie voor de volwassenen

De uitkomsten van de lasso regressie van Model I en Model II worden gepresenteerd in tabel 4. Door de lasso analyse werden in totaal 28 van de 31 kenmerken geselecteerd voor Model I, bestaande uit 44 variabelen. Hieronder worden de kenmerken met de grootst *odds* beschreven per domein. In het persoonlijke domein werden zeven kenmerken geselecteerd, waaronder: slachtoffer van een ernstig misdrijf ($OR = 2,286$, 95% *BI* [2,10-2,476]) en medicijngebruik voor middelengebruik ($OR = 2,124$, 95% *BI* [1,777-2,485]). Ook bij de volwassenen speelde herkomst een rol. Volwassenen die geboren zijn in Nederland hadden overwegend een lagere kans op een registratie als verdachte – ten opzichte van volwassenen met een herkomst uit de categorieën die door de lasso zijn geëlimineerd – met uitzondering van volwassenen met een Marokkaanse herkomst. Volwassenen die niet geboren zijn in Nederland hadden overwegend een hogere kans op registratie als verdachte, ten opzichte van volwassenen met een herkomst uit de geëlimineerde categorieën. Binnen de gezins- en familiefactoren werden alle kenmerken geselecteerd, waarbij de sterkste samenhang werd gevonden voor: volwassenen in een eenouderhuishouden met (wel of niet thuiswonende) minderjarige kinderen ($OR = 2,103$, 95% *BI* [1,896-2,358]) en onderdeel zijn van een huishouden met iemand met problematische schulden ($OR = 1,909$, 95% *BI* [1,762-2,064]).

Alle netwerkfactoren werden geselecteerd binnen Model I van de lasso regressie en toonden significante samenhang met registratie als verdachte, met de grootste samenhang voor veroordeelden in het colleganetwerk. Voor elke stijging van één procentpunt van het aandeel veroordeelden in het colleganetwerk nam de *odds* op registratie als verdachte toe met 1,023 (95% *BI* [1,020-1,026]). De andere netwerkfactoren toonden een vergelijkbare samenhang. Binnen de school- en werkgerelateerde factoren hing het werkzaam zijn als zelfstandige het sterkst samen met een hogere kans op registratie als verdachte ($OR = 1,954$, 95% *BI* [1,807-2,122]), ten opzichte van volwassenen die niet werkzaam zijn als zelfstandige. Daarnaast hingen enkele kenmerken samen met een lagere kans op registratie als verdachte, waaronder: het zijn van een scholier of student, een hbo- of wo-diploma, een persoonlijk inkomen in het laagste kwintiel, of voortdurend werkzaam zijn geweest in de 4 jaar voor het peiljaar. De omgevingsfactoren lieten overwegend kleine samenhang zien. In tegenstelling tot de jongeren, hing voor de volwassenen een zeer sterke stedelijkheidsgraad samen met een verlaagde kans op registratie als verdachte ($OR = 0,867$, 95% *BI* [0,815-0,924]), ten opzichte van de andere categorieën. De *balanced accuracy* van Model I was 0,70, dit betekent dat het model gemiddeld 70% van zowel de positieve als negatieve uitkomsten correct heeft geclassificeerd, ongeacht de ongelijke verdeling van de groepen verdachten versus niet-verdachten (1:4).

Naast Model I werd een tweede, beperktere selectie uitgevoerd (Model II) waarin 9 van de 31 kenmerken werden geselecteerd, met een totaal van 10 variabelen. Kenmerken met de sterkste samenhang met een verhoogde kans op registratie als verdachte waren: slachtoffer van een ernstig misdrijf ($OR = 2,586$, 95% BI [2,389-2,795]), en iemand met problematische schulden in het huishouden ($OR = 2,315$, 95% BI [2,146-2,494]). Daarnaast hingen enkele kenmerken samen met een verlaagde kans op registratie als verdachte, waaronder: volwassenen die geboren zijn in Nederland met een Nederlandse herkomst ($OR = 0,657$, 95% BI [0,622-0,696]) en volwassenen met een hbo- of wo-diploma ($OR = 0,649$, 95% BI [0,609-0,688]). Binnen de netwerkfactoren werden veroordeelden in het familienetwerk en in het burennetwerk geselecteerd in Model II. Er werden geen omgevingsfactoren geselecteerd. De *balanced accuracy* van Model II was 0,67 wat betekent dat het model gemiddeld 67% van zowel de positieve als negatieve uitkomsten correct heeft geclassificeerd.

Tabel 4. Beschrijvende statistieken en uitkomsten van de logistische regressies op de kenmerken van lasso Model I en het beperktere lasso Model II voor de volwassenen, gecontroleerd voor leeftijd en geslacht.

	Categorieën	Omvang groep	Verdachte/ Niet-verdachte	Model I		Model II	
		<i>n</i>	%	<i>OR</i>	<i>95% BI</i>	<i>OR</i>	<i>95% BI</i>
<i>Intercept</i>	-	66.480	20/80	0,489	0,434 - 0,551	0,571	0,534 - 0,608
Persoonlijke factoren							
Slachtoffer van ernstig misdrijf	Ja	4.650	42/58	2,286	2,100 - 2,476	2,586	2,389 - 2,795
	Nee	61.840	18/82	-	-	-	-
Medicijngebruik middelengebruik	Ja	1.020	50/50	2,124	1,777 - 2,485	-	-
	Nee	65.460	20/80	-	-	-	-
Medicijngebruik hyperactiviteit	Ja	2.690	34/66	1,789	1,602 - 2,002	-	-
	Nee	63.800	19/81	-	-	-	-
Medicijngebruik angststoornis	Ja	7.580	34/66	1,420	1,323 - 1,525	-	-
	Nee	58.910	18/82	-	-	-	-
Medicijngebruik depressie	Ja	8.390	34/66	1,461	1,361 - 1,566	1,830	1,720 - 1,948
	Nee	58.100	18/82	-	-	-	-
Negatieve levenservaringen	Ja	3.340	31/69	1,432	1,292 - 1,587	-	-
	Nee	63.140	19/81	-	-	-	-
Geboren in Nederland i.c.m. herkomst							
Geboren in Nederland	Herkomst Nederland	48.600	17/83	0,730	0,673 - 0,788	0,657	0,622 - 0,696
	Herkomst Europa (excl. Nederland)	1.770	20/80	0,755	0,647 - 0,880	-	-

	Herkomst Nederlands-Caribisch gebied	250	32/68		-	-	-	-
	Herkomst Turkije	880	33/67		-	-	-	-
	Herkomst Marokko	670	45/55		1,674	1,377 - 2,031	-	-
	Herkomst Suriname	740	38/62		-	-	-	-
	Herkomst Indonesië	1.250	18/82		0,797	0,665 - 0,967	-	-
	Herkomst overig Azië	380	18/82		0,675	0,467 - 0,916	-	-
	Herkomst overig Afrika	280	36/64		-	-	-	-
	Herkomst Amerika en Oceanië	280	21/79		-	-	-	-
Niet geboren in Nederland	Herkomst Europa (excl. Nederland)	3.510	22/78		-	-	-	-
	Herkomst Nederlands-Caribisch gebied	480	38/62		1,375	1,069 - 1,756	-	-
	Herkomst Turkije	1.040	33/67		-	-	-	-
	Herkomst Marokko	910	40/60		-	-	-	-
	Herkomst Suriname	780	38/62		1,418	1,177 - 1,727	-	-
	Herkomst Indonesië	290	17/83		-	-	-	-
	Herkomst overig Azië	2.520	34/66		1,441	1,274 - 1,640	-	-
	Herkomst overig Afrika	1.160	42/58		1,678	1,408 - 1,962	-	-
	Herkomst Amerika en Oceanië	670	22/78		-	-	-	-
Gezins- en factoren								
Institutioneel huishouden	Ja	920	40/60		1,345	1,124 - 1,600	-	-
	Nee	65.570	20/80		-	-	-	-

Iemand met problematische schulden in het huishouden	Ja	4.500	47/53	1,909	1,762 - 2,064	2,315	2,146 - 2,494
	Nee	61.990	18/82	-	-	-	-
Eenouderhuishouden met (wel of niet thuiswonende) minderjarige kinderen	Ja	2.990	42/58	2,103	1,896 - 2,358	-	-
	Nee	63.500	19/81	-	-	-	-
Slachtoffer van een ernstig misdrijf in huishouden	Ja	4.390	32/68	1,564	1,428 - 1,714	-	-
	Nee	62.100	19/81	-	-	-	-
Ouder veroordeeld door OM of rechtbank	Ja	4.420	35/65	1,308	1,189 - 1,433	-	-
	Nee	62.070	19/81	-	-	-	-
Gestandaardiseerd besteedbaar inkomen huishouden	1 t/m 20%	11.700	34/66	1,283	1,184 - 1,388	1,330	1,255 - 1,408
	21 t/m 40%	9.570	25/75	1,116	1,026 - 1,207	-	-
	41 t/m 60%	13.930	20/80	-	-	-	-
	61 t/m 80%	15.110	15/85	0,946	0,876 - 1,024	-	-
	81 of meer	16.180	12/88	0,873	0,805 - 0,945	-	-
Netwerfactoren							
		Omvang groep	Verdachte/ Niet-verdachte	Model I		Model II	
		<i>n</i>	<i>M (SD)</i>	<i>OR</i>	<i>95% BI</i>	<i>OR</i>	<i>95% BI</i>
Veroordeelde huisgenoot	-	66.480	10,2 (25,8)/ 3,9 (16,4)	1,004	1,003 - 1,005	-	-

Veroordeelde in het familienetwerk	-	66.480	11,2 (15,5)/ 6,2 (10,3)	1,015	1,013 - 1,017	1,021	1,019 - 1,023
Veroordeelde in het colleganetwerk	-	66.480	6,9 (10,8)/ 5,4 (7,6)	1,023	1,020 - 1,026	-	-
Veroordeelde in het burennetwerk	-	66.480	9,2 (7,2)/ 7,0 (5,9)	1,018	1,014 - 1,022	1,025	1,021 - 1,028
School- en werkgerelateerde factoren							
Scholier of student	Ja	11.230	17/83	0,894	0,811 - 0,990	-	-
	Nee	55.250	21/79	-	-	-	-
Hoogst behaalde onderwijsniveau (3 categorieën)	Basisonderwijs, vmbo, mbo1	16.700	32/68	1,391	1,310 - 1,476	1,404	1,324 - 1,489
	Havo, vwo, mbo2-4	28.540	19/81	-	-	-	-
	Hbo, wo	21.240	11/89	0,696	0,650 - 0,741	0,649	0,609 - 0,688
Licht verstandelijke beperking	Ja	510	37/63	-	-	-	-
	Nee	65.970	20/80	-	-	-	-
Werkzaam als zelfstandige	Ja	7.520	22/78	1,954	1,807 - 2,122	-	-
	Nee	58.960	20/80	-	-	-	-
Deeltijdfactor werk	Geen baan	18.110	29/71	1,250	1,152 - 1,353	-	-
	0,01 t/m 0,29	4.170	19/81	-	-	-	-
	0,30 t/m 0,59	4.700	20/80	-	-	-	-
	0,60 t/m 0,79	4.950	20/80	-	-	-	-
	0,80 t/m 0,94	5.370	19/81	-	-	-	-
	0,95 of meer	29.170	15/85	-	-	-	-
Uitzend- of oproepkracht	Ja	5.700	25/75	1,230	1,125 - 1,356	-	-
	Nee	60.780	20/80	-	-	-	-

Werkzaam met tijdelijk contract	Ja	10.110	18/82	-	-	-	-
	Nee	56.370	20/80	-	-	-	-
Kwetsbare arbeidsmarktpositie	Ja	22.120	31/69	1,253	1,175 - 1,337	1,566	1,488 - 1,649
	Nee	30.020	14/86	-	-	-	-
	Startkwalificatie onbekend	14.350	15/85	-	-	-	-
Persoonlijk inkomen	1 t/m 20%	14.770	22/78	0,775	0,719 - 0,835	-	-
	21 t/m 40%	10.180	31/69	-	-	-	-
	41 t/m 60%	11.780	24/76	1,050	0,984 - 1,120	-	-
	61 t/m 80%	13.640	16/84	-	-	-	-
	81 of meer	16.100	11/89	-	-	-	-
Recent arbeidsverleden (werkzaam in de 4 jaar voor peiljaar)	Niet werkzaam	10.210	30/70	-	-	-	-
	Minimaal 1 maand tot maximaal 1/4 werkzaam	4.180	29/71	1,140	1,042 - 1,262	-	-
	Meer dan 1/4 werkzaam	4.720	25/75	-	-	-	-
	Meer dan 2/4 werkzaam	5.940	23/77	-	-	-	-
	Meer dan 3/4 werkzaam	7.940	21/79	-	-	-	-
	Voortdurend werkzaam	33.490	14/86	0,796	0,740 - 0,854	-	-
Omgevingsfactoren							
Gemiddelde SES-totaalscore buurt	1e kwintiel	12.570	29/71	-	-	-	-
	2e kwintiel	13.180	23/77	1,081	1,019 - 1,146	-	-
	3e kwintiel	13.510	18/82	-	-	-	-
	4e kwintiel	13.570	16/84	-	-	-	-

	5e kwintiel	13.650	14/86	-	-	-	-
Misdrijven per 1000 inwoners buurt	1e kwintiel	13.580	13/87	0,800	0,741 - 0,860	-	-
	2e kwintiel	13.900	17/83	0,916	0,859 - 0,974	-	-
	3e kwintiel	13.740	20/80	-	-	-	-
	4e kwintiel	13.590	24/76	-	-	-	-
	5e kwintiel	11.670	27/83	-	-	-	-
Stedelijkheidsgraad buurt	Zeer sterk	16.840	23/77	0,867	0,815 - 0,924	-	-
	Sterk	16.730	22/78	-	-	-	-
	Matig	12.060	19/81	1,064	0,996 - 1,137	-	-
	Weinig	10.400	16/84	-	-	-	-
	Niet	10.460	16/84	-	-	-	-
		Omvang groep	Verdachte/ Niet-verdachte	Model I		Model II	
		<i>n</i>	<i>M (SD)</i>	<i>OR</i>	<i>95% BI</i>	<i>OR</i>	<i>95% BI</i>
Diversiteit buurt	-	66.480	0,43 (0,22)/ 0,37 (0,21)	-	-	-	-

Noot. *N* = 66.480 (20% verdachten/ 80% niet-verdachten). *OR* = *odds ratio*; *BI* = betrouwbaarheidsinterval; *M* = gemiddelde; *SD* = standaardafwijking. De *odds ratio*'s moeten geïnterpreteerd worden ten opzichte van de (groep) categorieën zonder *odds ratio*'s (deze vormen samen de referentiecategorie). Voor de numerieke variabelen werden in plaats van de omvang van de groepen, het gemiddelde en de standaardafwijking gegeven.

4.3 Modelvergelijking van de jongeren en volwassenen

Een overzicht van de kenmerken en kwaliteitsmaten van de verschillende modellen voor zowel de jongeren als de volwassenen is weergegeven in tabel 5. Het basismodel refereert aan het multivariate logistische regressiemodel waarin alle van tevoren vastgestelde kenmerken zijn meegenomen (zie bijlage 3). Voor de jongeren en de volwassenen laat de modelvergelijking vergelijkbare resultaten zien. Voor beide populaties zijn de kwaliteitsmaten van Model I overwegend vergelijkbaar met die van het basismodel. Dit betekent dat het weglaten van de kenmerken weinig tot geen effect heeft op de voorspelkracht van het model. Zoals verwacht werd voor beide populaties bij Model II wel ingeleverd op de voorspelkracht en het onderscheidende vermogen. Onder meer de hogere *F1-scores* van Model I suggereren dat het geschikter is in het correct identificeren van verdachten en het vermijden van fout-positieven, dan Model II.

Daarnaast werden modelprestaties vergeleken op basis van verschillende *pseudo R²*. Net als bij de bovenstaande kwaliteitsmaten, waren de verschillen tussen het basismodel en Model I voor beide populaties minimaal. Verder was de Tjur's *R²* hoger voor Model I dan voor Model II, wat suggereert dat Model I beter is in staat is om de verschillende groepen juist te classificeren. Ook Nagelkerke's *R²* en McFadden's *R²* lieten een hogere waarde zien bij Model I dan Model II wat suggereert dat Model I beter aansluit bij de data, dan Model II.

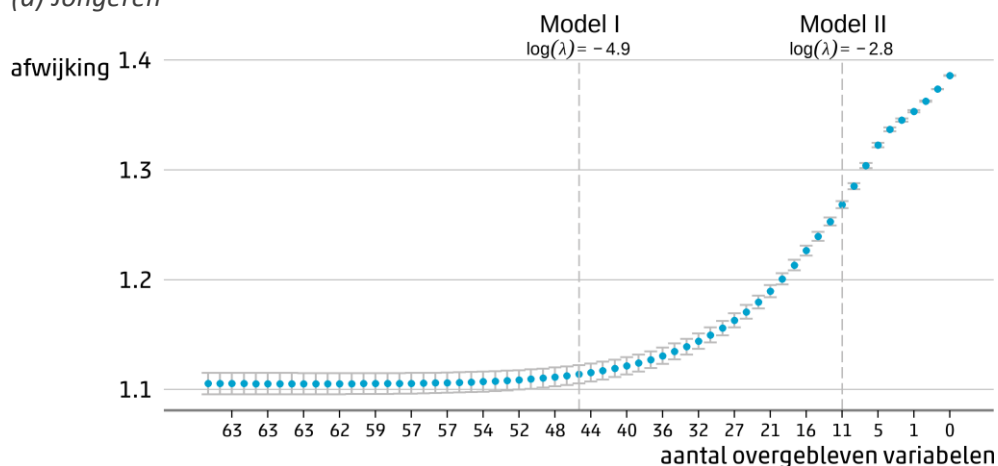
Tabel 5. Overzicht van de kwaliteitsmaten van het basismodel, Model I, en Model II voor de jongeren en volwassenen.

	Jongeren				Volwassenen		
	<i>Basismodel</i>	<i>Model I</i>	<i>Model II</i>		<i>Basismodel</i>	<i>Model I</i>	<i>Model II</i>
Aantal kenmerken	34	29	9		31	28	9
<i>Balanced accuracy</i>	0,71	0,71	0,68		0,70	0,70	0,67
<i>Recall</i>	0,68	0,68	0,64		0,64	0,65	0,60
<i>Precision</i>	0,40	0,40	0,37		0,39	0,39	0,60
<i>F1-score</i>	0,50	0,51	0,47		0,48	0,48	0,45
<i>Pseudo R²</i>							
McFadden	0,22	0,21	0,16		0,17	0,16	0,13
Tjur	0,27	0,26	0,21		0,21	0,20	0,16
Nagelkerke	0,42	0,41	0,34		0,34	0,33	0,27

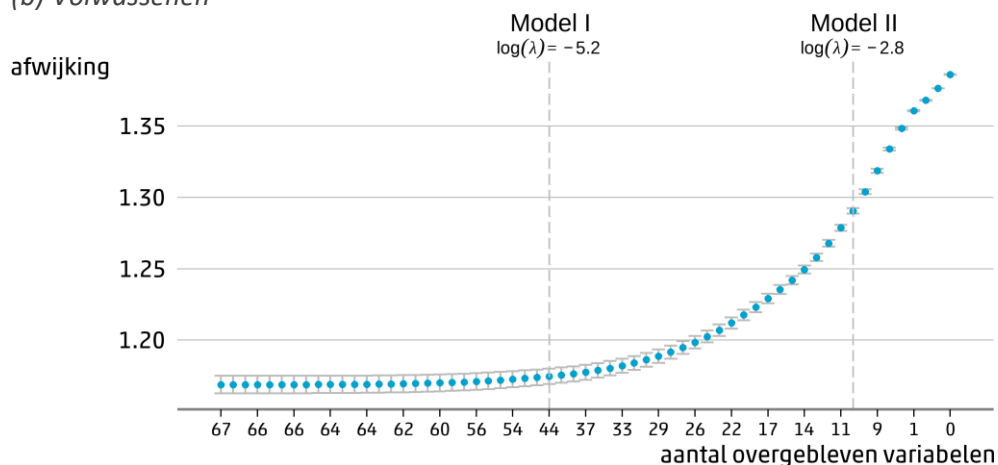
Figuur 2 geeft een grafische weergave van hoe de modelkwaliteit afneemt naarmate er minder kenmerken worden geïncludeerd (vanwege een strengere lasso strafterm) voor jongeren en volwassenen. De afwijking is weergegeven als het gemiddelde $\pm$ de standaardfout, en is een maat van de aannemelijkheid van een model ten opzichte van een model wat de data perfect past. De afwijking is

hoger wanneer een model de data minder goed beschrijft. Voor zowel jongeren als volwassenen neemt de afwijking tot Model I weinig toe. Het laatste datapunt in de grafieken geeft een model zonder variabelen (nulmodel) weer.

(a) Jongeren



(b) Volwassenen



Figuur 2. De afwijking (deviance) afgezet tegen de log van de verschillende straftermen λ , uitgedrukt in het aantal variabelen dat overblijft bij die strafterm, bij (a) de jongeren en (b) de volwassenen (NB: Het aantal variabelen verwijst naar de kenmerken inclusief de eventuele onderliggende categorieën. Dit wijkt dus af van het aantal kenmerken). Hoe hoger de afwijking, hoe slechter het model de data modelleert. De afwijking is weergegeven als het gemiddelde ($\bullet$) $\pm$ de standaardfout ($\mathbf{I}$) van de kruisvalidatie. Voor Model I en Model II is de log van de strafterm (λ) weergegeven.

5. Discussie en conclusie

In het huidige verkennende onderzoek werd met behulp van lasso logistische regressie onderzocht hoe achtergrondkenmerken samenhangen met als verdachte geregistreerde *first offenders* van geweldscriminaliteit. Daarbij werd onderscheid gemaakt tussen twee onderzoeksgroepen, namelijk jongeren (12 t/m 25 jaar) en volwassenen (18 jaar en ouder). In het huidige onderzoek werd een groot aantal kenmerken afgeleid met als doel te komen tot een beperkte set van kenmerken die gebruikt kan worden voor monitoringsdoeleinden in het dashboard *Galileo*. Op basis van de lasso regressie zijn voor elke onderzoeksgroep twee modellen samengesteld, één met de beste balans tussen het elimineren van kenmerken met behoud van modelkwaliteit (Model I) en ter illustratie één model die het aantal kenmerken tot een behapbaar aantal voor monitoring beperkt (Model II). De belangrijkste kenmerken – gecontroleerd voor leeftijd en geslacht – op basis van de twee modellen zijn weergegeven in tabel 7.

Bij zowel de jongeren als de volwassenen kwamen factoren uit elk domein naar voren als belangrijk. Daarnaast was sprake van een grote mate van overlap tussen de factoren die belangrijk waren voor zowel de jongeren als de volwassenen. Dit kan deels te maken hebben met de overlap in leeftijd tussen de onderzoeksgroepen. Binnen de persoonlijke factoren bleek geregistreerd slachtofferschap van een ernstig misdrijf een belangrijke indicator te zijn voor beide populaties. Binnen de literatuur is dit een bekend fenomeen, ook wel bekend als de *offender/victim overlap* (Berg & Schreck, 2022). Ook het medicijngebruik voor verschillende soorten problematiek is terug te zien in de resultaten van de lasso analyse. Een andere persoonlijke factor met samenhang met geregistreerd verdachtenschap van geweldscriminaliteit was herkomst. Bij beide werd deze in Model II samengevat als een verminderde kans om geregistreerd worden te als verdachte ten opzichte van de overige categorieën. In Model I van de jongeren was het beeld minder eenduidig samen te vatten, deels door de verschillen in de operationalisatie. Bij de volwassenen, waar veel categorieën werden uitgesplitst, waren er zowel groepen met een niet-Nederlandse herkomst met een verlaagde als een verhoogde kans op registratie als verdachte. Hier kunnen meerdere zaken een rol spelen, o.a. culturele dissonantie (Jennissen, 2015), etnische profilering, maar vooral ook de samenhang met andere kenmerken. Eerder onderzoek van het CBS liet zien dat als rekening werd gehouden met individuele factoren als opleidingsniveau, en gezinsfactoren als huishoudensinkomen, dat de oververtegenwoordiging verdween of tenminste sterk werd verminderd (o.a., CBS, 2020). Het feit dat voor zowel jongeren als volwassenen de categorie geboren in Nederland, met herkomst Nederland, de *OR* dicht bij de 1 komt te liggen wanneer meer kenmerken in beschouwing worden genomen (Model I vs. Model II), suggereert mogelijk iets vergelijkbaars.

Binnen de gezins- en familiefactoren speelden met name factoren gerelateerd aan armoede en andere vormen van stress in het huishouden een rol. Hierbij ging het bijvoorbeeld om iemand met problematische schulden in het huishouden en het onderdeel zijn van een eenouderhuishouden (als ouder of kind). Ook het hebben

van veroordeelden in het netwerk speelt een rol in de kans om als verdachte te worden geregistreerd als *first offender* van een geweldsmisdrijf, waarbij het familienetwerk het belangrijkste leek. Binnen het school- en werkdomein verschilden de gebruikte indicatoren het meest tussen de twee onderzoeksgroepen. Hier zaten ook meer verschillen tussen de onderzoeksgroepen. Bij de jongeren waren er meer schoolgerelateerde factoren die samenhangen met geregistreerd verdachtenschap (o.a., het volgen van speciaal onderwijs en het gemiddelde eindexamencijfer), en bij de volwassenen speelde het type en stabiliteit in werk een rol (o.a., een kwetsbare arbeidsmarktpositie en het recente arbeidsverleden). Echter, ook hier zaten overeenkomsten. Voor zowel de jongeren als volwassenen speelde het onderwijsniveau een rol. Omgevingsfactoren die voor zowel de jongeren als de volwassenen samenhangen met geregistreerd verdachtenschap waren de gemiddelde SES-totaalscore van de buurt, en het aantal geregistreerde misdrijven per 1000 inwoners in de buurt. In het beperkte lasso-model (Model II) van zowel de jongeren als de volwassenen komen deze factoren echter niet meer voor. Mogelijk wordt de samenhang van deze factoren met geweldscriminaliteit (genoeg) ondervangen door andere factoren in het model.

Tabel 7. Overzicht van de belangrijkste kenmerken per domein per onderzoeksgroep uit Model I en Model II van de lasso regressie voor jongeren en volwassenen, gecontroleerd voor leeftijd en geslacht.

Domein	Jongeren	Volwassenen
Persoonlijke factoren	Slachtoffer van een ernstig misdrijf, medicijngebruik hyperactiviteit, geboren in Nederland i.c.m. herkomst, Halt-verwijzing voor een overtreding, medicijngebruik depressie	Slachtoffer van een ernstig misdrijf, medicijngebruik depressie, geboren in Nederland i.c.m. herkomst, medicijngebruik middelengebruik, medicijngebruik hyperactiviteit, medicijngebruik angststoornis, negatieve levenservaringen,
Gezins- en familiefactoren	Iemand met problematische schulden in het huishouden, ouder veroordeeld door OM of rechtbank, tienerouder, institutioneel huishouden, jeugdhulp met verblijf, jeugdhulp zonder verblijf, eenouderhuishouden (kind), slachtoffer van een ernstig misdrijf in	Iemand met problematische schulden in het huishouden, gestandaardiseerd besteedbaar inkomen in het huishouden, institutioneel huishouden, eenouderhuishouden met (wel of niet thuiswonende) minderjarige kinderen,

	het huishouden, gestandaardiseerd besteedbaar inkomen in het huishouden	slachtoffer van een ernstig misdrijf in het huishouden, ouder veroordeeld door OM of rechtbank
Netwerfactoren	Veroordeelde huisgenoot, veroordeelde in het familienetwerk, veroordeelde in het burennetwerk, veroordeelde in het klasgenotennetwerk	Veroordeelde in het familienetwerk, veroordeelde in het burennetwerk, veroordeelde huisgenoot, veroordeelde in het colleganetwerk
School- en werkgerelateerde factoren	Hoogst gevolgde onderwijsniveau, speciaal onderwijs, scholier of student, voortijdig schoolverlater, persoon is gezakt, licht verstandelijke beperking, gemiddeld eindexamencijfer, deeltijdfactor werk	Hoogst behaalde onderwijsniveau, kwetsbare arbeidsmarktpositie, werkzaam als zelfstandige, uitzend- of oproepkracht, recent arbeidsverleden
Omgevingsfactoren	Gemiddelde SES-totaalscore van de buurt, misdrijven per 1000 inwoners in de buurt, stedelijkheidsgraad buurt	Gemiddelde SES-totaalscore van de buurt, misdrijven per 1000 inwoners in de buurtstedelijkheidsgraad buurt.

Noot. Dikgedrukt zijn de kenmerken die zowel in Model I als Model II terugkwamen. Dit zijn de belangrijkste kenmerken.

Een van de krachten van het huidige onderzoek lag in de mogelijkheid om een grote hoeveelheid kenmerken in samenhang te onderzoeken, terwijl in andere onderzoeken veelal gekeken wordt naar één of enkele kenmerken. Ook waren door het gebruik van beschikbare registerdata voor een groot deel van de populatie de kenmerken in beeld. Het gebruik van registerdata heeft echter ook enkele nadelen. Niet alle kenmerken die uit het literatuuronderzoek naar voren kwamen, waren meetbaar met registerdata (o.a. de opvoedstijl van de ouders). Hierdoor is het mogelijk dat enkele (belangrijke) factoren missen. Daarnaast was bij het omzetten van de kenmerken uit de literatuur naar indicatoren op basis van de beschikbare data het niet altijd mogelijk om deze te operationaliseren zoals deze in de literatuur omschreven staat. Dit geldt bijvoorbeeld voor geregistreerd medicijngebruik voor middelengebruik, wat gebruikt wordt als indicator voor verslavingsproblematiek. Niet iedereen met verslavingsproblematiek krijgt daarvoor medicijnen voorgeschreven. Daarnaast neemt niet iedereen die de medicijnen voorgeschreven krijgt, deze ook daadwerkelijk in. Ook kan enerzijds gesteld worden dat sprake was van de desbetreffende problematiek, maar anderzijds ook dat de problematiek werd aangepakt. Mogelijk gaat het om een

beschermende factor (i.e., risicoverlagend in aanwezigheid van een risicofactor), maar dat kon in huidig onderzoek niet worden vastgesteld. Daarnaast zijn er enkele kanttekeningen bij de netwerkindicatoren (Van der Laan, 2022). Deze netwerkindicatoren moeten worden gezien als proxy van het ‘echte’ sociale netwerk. Hierin is namelijk onbekend of en in welke mate de personen binnen het netwerk daadwerkelijk contact met elkaar hebben. Het gebruik van de beschikbare netwerkindicatoren heeft echter als voordeel dat deze op grotere schaal gekoppeld kunnen worden aan andere kenmerken. Het CBS beschikt ook over niet-registerdata zoals enquêtes, maar deze zijn niet beschikbaar voor de gehele Nederlandse bevolking.

Een andere kracht van het huidige onderzoek lag in de selectiemethode om te komen tot een beperkte set aan kenmerken. Een veelgebruikte manier om te komen tot een beperkte set van indicatoren is selectie op basis van bijvoorbeeld de hoogte van de *odds ratio's* en de statistische significantie van de kenmerken. Deze methode garandeert echter niet dat de overgebleven kenmerken in samenhang het meest verklarend zijn, en selectiemethoden zoals *stepwise selection* hebben hun beperkingen (McNeish, 2015). Met de lasso regressie was het mogelijk om op een data-gedreven statistische manier een beperktere selectie te maken van de kenmerken, zonder in te leveren op de kwaliteit van het model (Model I). Naarmate er minder kenmerken worden geselecteerd zal de verklarende capaciteit van het model achteruit gaan. De keuze voor een model is dus een *trade-off* tussen kwaliteit en het aantal geselecteerde kenmerken.

Bij interpretatie van de uitkomsten van de lasso regressie is het belangrijk de invloed van de keuzes bij de operationalisatie van de kenmerken in acht te nemen. Het niveau van aggregatie van de categorieën van het kenmerk kan een rol spelen in welke categorieën in de beperkte set kenmerken overblijven. In het huidige onderzoek werd het kenmerk (hoogst gevolgde/behaalde) onderwijsniveau op lager niveau geaggregeerd voor de jongeren (8 categorieën) dan voor de volwassenen (3 categorieën), vanwege het hoge aantal ontbrekende waarden bij volwassenen. Bij de volwassenen werd in Model I hbo of wo meegenomen terwijl bij jongeren hbo- of wo-bachelor en hbo, wo-master, doctor worden meegenomen. In Model II werd bij de volwassenen hbo of wo nog steeds meegenomen en bij de jongeren alleen nog hbo, wo-bachelor. Dergelijke keuzes in de operationalisatie spelen ook een rol bij een reguliere logistische regressie.

Een andere kanttekening van het huidige onderzoek ligt in het gebruik van data van geregistreerde verdachten als proxy voor crimineel gedrag, namelijk selectiebias. Personen die crimineel gedrag vertonen die niet worden geregistreerd, of ander crimineel gedrag van dezelfde persoon dat niet wordt geregistreerd, blijven buiten beeld. De kenmerken van deze groep kunnen afwijken van de kenmerken van de geregistreerde verdachten. Bovendien hangt de mate van selectiviteit af van het type delict. Verdachten van zwaardere delicten zullen relatief vaker in beeld komen van de politie dan verdachten van relatief lichte delicten. Belangrijk is ook om in acht te nemen dat niet alle verdachten uiteindelijk daders (blijken te) zijn. Daarnaast is bekend dat (gewelds)criminaliteit niet altijd wordt aangegeven bij de politie. Hierdoor is sprake van onderregistratie. Dit heeft verschillende oorzaken, waaronder angst voor represailles, normalisatie

van geweld, schaamte en stigma (vooral bij seksuele misdrijven) en onvoldoende bewijs voor of ernst van het incident (Van de Weijer & Bernasco, 2016). Desondanks is gekozen om deze data te gebruiken, omdat dit het meest volledige beeld van de populatie geeft en integraal beschikbaar is. Een andere kanttekening van het onderzoek is dat het op basis van de huidige opzet niet mogelijk was om causaliteit aan te tonen tussen de kenmerken en crimineel gedrag of inzicht te bieden in de achterliggende mechanismen bij de ontwikkeling van crimineel gedrag. Er kan ook sprake zijn van een schijnverband (*spurious association*) of een indirect verband en het is aannemelijk dat het risico daarop toeneemt hoe verder een model wordt beperkt door de lasso regressie.

Samengevat heeft het huidige onderzoek bijgedragen aan kennis over achtergrondkenmerken op basis van beschikbare CBS-registerdata van als verdachte geregistreerde *first offenders* van geweldscriminaliteit. De belangrijkste kenmerken (gecontroleerd voor leeftijd en geslacht) zijn weergegeven in tabel 7. De resultaten van dit onderzoek kunnen gebruikt worden als leidraad bij het kiezen van kenmerken bij nieuwe analyses over geregistreerd verdachten (zie bv. Model I). Ten tweede kunnen de resultaten van dit onderzoek gebruikt worden om ontwikkelingen in prevalentie op deze kenmerken in de samenleving te monitoren. Ten behoeve van de externe validiteit moet in vervolgonderzoek eerst bekeken worden hoe stabiel de kenmerken over de tijd zijn. Als deze stabiel zijn, kunnen ze gerelateerd worden aan een verwachte toename of afname van geweldsmisdrijven en de daarbij benodigde politie-inzet. Hiervoor hoeven de kenmerken niet een direct causaal verband te hebben, maar kan de gevonden samenhang gebruikt worden als proxy. Voor dergelijke monitoring kan het wenselijk zijn om een beperkte(re) set van kenmerken te hanteren (zie bv. Model II). Bij gebruik moeten de eerder genoemde kanttekeningen in acht genomen worden:

- **Causaliteit.** Op basis van het huidige onderzoek was het niet mogelijk om causaliteit aan te tonen tussen de kenmerken en registratie als verdachte, of inzicht te bieden in de achterliggende mechanismen bij de ontwikkeling van crimineel gedrag.
- **Trade-off kwaliteit model versus aantal kenmerken.** Bij een keus voor een beperkt aantal kenmerken moet goed gekeken worden naar de bijbehorende *trade-off* in kwaliteit van het model. Hier moet per specifieke vraag/monitoringsdoel een afweging gemaakt worden.
- **Selectiebias in verdachtendata.** In het gebruik van data van geregistreerde verdachten zit selectiebias. Hierdoor kan het voorkomen dat er een bepaalde selectiviteit zit in wie wel en wie niet geregistreerd wordt als verdachte. Ook staan de registraties van verdachte niet een-op-een gelijk aan het aantal geweldsmisdrijven. Hierdoor kan een ontwikkeling op een specifiek kenmerken leiden tot een verandering in geregistreerde verdachten van geweldscriminaliteit, maar niet in het aantal geweldsmisdrijven. Mogelijk spelen bij personen die niet worden gepakt bij geweldsmisdrijven andere kenmerken een rol.
- **Gebruik beschikbare registerdata.** De kenmerken in het huidige onderzoek zijn geselecteerd op beschikbaarheid en de mogelijkheid om ze integraal te monitoren. Hierdoor missen potentieel belangrijke kenmerken waar geen integrale data voor beschikbaar was.

- **Ontwikkelingen door de tijd.** De samenhang tussen de kenmerken en registratie als verdachte kan veranderen over de tijd. De kenmerken uit 2017 hoeven niet dezelfde samenhang te hebben met de kans op registratie als verdachte in andere jaren.

Literatuur

Berg, M. T., & Schreck, C. J. (2022). The meaning of the victim-offender overlap for criminological theory and crime prevention policy. *Annual Review of Criminology*, 5, 277–297. <https://doi.org/10.1146/annurev-criminol-030920-120724>

Boutellier, H., Scholte, R., & Heijnen, M. (2008). Kijken achter de cijfers. De ontwikkeling van een criminogeniteitsmonitor voor Amsterdam. Vrije Universiteit: Amsterdam.

Centraal Bureau voor de Statistiek. (2020). Jaarrapport integratie 2020. <https://www.cbs.nl/nl-nl/publicatie/2020/46/jaarrapport-integratie-2020>

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87(417), 178–883. <https://doi.org/10.1080/01621459.1992.10475190>

Friedman, J. H., Hastie, T., & Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, 33(1), 1–22. <https://doi.org/10.18637/jss.v033.i01>

Jennissen, R. (2015). Gewelddadige Marokkaanse jeugdcriminaliteit: Een kwestie van culturele dissonantie? *Tijdschrift voor Criminologie*, 57(1), 133–139. DOI:10.5553/TvC/0165182X2015057001008

McNeish, D. M. (2015). Using Lasso for predictor selection and to assuage overfitting: A method long overlooked in behavioral sciences. *Multivariate Behavioral Research*, 50(5), 471–484. <http://dx.doi.org/10.1080/00273171.2015.1036965>

Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(3), 581–592. <https://doi.org/10.1093/biomet/63.3.581>

Servaas, L., Weerman, F., & Fischer, T. (2021). Risico-, versterkende en beschermende factoren voor crimineel gedrag. Een literatuuronderzoek naar de wetenschappelijke stand van zaken.

Standaard Classificatie Misdrijven 2010 (2011). <https://www.cbs.nl/nl-nl/onze-diensten/methoden/classificaties/misdrijven/standaardclassificatie-misdrijven-2010>.

Tibshirani, R. (1996). *Regression shrinkage and selection via the Lasso*. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288. <https://doi.org/10.1111/j.2517-6161.1996.tb02080.x>

Tjur, T. (2009). Coefficients of determination in logistisch regression models – a new proposal: The coefficient of discrimination. *The American Statistician*, 63(4), 366–372. <https://doi.org/10.1198/tast.2009.08210>

Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>

Van der Laan, D. J. (2022). A person network of the Netherlands. Centraal Bureau voor de Statistiek. <https://www.cbs.nl/nl-nl/achtergrond/2022/20/a-person-network-of-the-netherlands>

Van de Weijer, S., & Bernasco, W. (2016). Aangifte- en meldingsbereidheid: Trends en determinanten. Wetenschappelijk Onderzoek- en Documentatiecentrum (WODC).

Bijlagen

Bijlage 1: Operationalisatie van kenmerken

In deze bijlage wordt het omzetten van kenmerken uit de literatuur naar bruikbare indicatoren gedefinieerd en verder geoperationaliseerd. De meeste kenmerken – met uitzondering van de indicator licht verstandelijke beperking – zijn voor zowel de verdachten als de niet-verdachten gepeild op 1 januari 2017 (hierna peildatum), of vóór het peiljaar 2017. De kenmerken met een * erachter zijn niet meegenomen in de analyse, maar dit zijn bijvoorbeeld kenmerken die gebruikt zijn voor de matching van de verdachten aan de niet-verdachte referenten (seks en leeftijd), of voor de imputatie van de missende waarden (o.a., type huishouden, uitkering).

Tabel A1. Operationalisatie van de kenmerken op basis van beschikbare CBS-data.

Factor	Indicator	Operationalisatie
Persoonlijke factoren		
Sekse*	Geslacht	Het geslacht dat in de Basisregistratie Personen (BRP) op de peildatum staat geregistreerd.
Leeftijd*	Leeftijd	De leeftijd in jaren op de peildatum
Eerder slachtofferschap	Slachtoffer van een ernstig misdrijf	Persoon is in zeven jaar vóór de peildatum slachtoffer geworden volgens politieregistratie van minimaal één misdrijf met een ernstscore van 1 of hoger. Deze ernstscore is geconstrueerd op basis van ernst van delicten naar de wettelijke maximum straf die op overtreding daarvan staat. Vervolgens is er binnen het jaar gekeken hoe vaak deze delicten voor komen en op basis daarvan is een rangscore gemaakt per ondervonden delict. Door deze rangorde te standaardiseren ontstaat de ernstscore.
Eerder deviant gedrag	Halt-verwijzing overtreding	Persoon is in de twaalf jaar vóór peiljaar verwezen naar bureau Halt voor een overtreding.
Middelengebruik	Medicijngebruik middelengebruik	Persoon heeft in de elf jaar vóór pleegjaar GGZ-medicatie voorgeschreven gekregen voor middelengebruik. Hierbij werd gekeken naar geneesmiddelen binnen de ATC4-code(s): N07B Middelen bij verslaving.
Hyperactiviteit en impulsiviteit	Medicijngebruik hyperactiviteit	Persoon heeft in de elf jaar vóór pleegjaar GGZ-medicatie voorgeschreven gekregen

		voor hyperactiviteit. Hierbij werd gekeken naar geneesmiddelen binnen de ATC4-code(s): N06B ADHD-middelen .
Angststoornis	Medicijngebruik angststoornis	Persoon heeft in de elf jaar vóór pleegjaar GGZ-medicatie voorgeschreven gekregen voor een angststoornis. Hierbij werd gekeken naar geneesmiddelen binnen de ATC4-code(s): N05B Anxiolytica .
Depressie	Medicijngebruik depressie	Persoon heeft in de elf jaar vóór pleegjaar GGZ-medicatie voorgeschreven gekregen voor een depressie. Hierbij werd gekeken naar geneesmiddelen binnen de ATC4-code(s): N06A Antidepressiva .
Stress in de familie	Negatieve levenservaringen	Persoon heeft in het jaar vóór pleegjaar minimaal één negatieve levenservaring meegemaakt of ondergaan. Hieronder valt: – persoon is recent gescheiden – ouders van persoon zijn recent gescheiden – minstens één ouder van persoon is overleden – minstens één kind van persoon is overleden partner van persoon is overleden Bij een scheiding wordt zowel gekeken naar een huwelijk als naar een geregistreerde partnerschap en geregistreerd samenwonen. Onder dit laatste valt bijvoorbeeld het gezamenlijk verhuizen, het hebben van een gezamenlijk kind of het als partner in een paar verkrijgen van zorgtoeslag, huurtoeslag of een gezamenlijke voorlopige aanslag voor de Belastingdienst. Om de ouders te bepalen, is gekeken naar de juridische ouders van een persoon. Om een partner te bepalen, is gekeken naar zowel huwelijk als geregistreerd partnerschap en samenwonen. Om kinderen te bepalen, is gekeken naar juridische kinderen.
Herkomst	Geboren in Nederland i.c.m. herkomst	De herkomst van een persoon op de peildatum. Het betreft het al dan niet geboren zijn in Nederland in combinatie met het herkomstland. Op basis van de BRP is gekeken of de persoon Nederland als geboorteland heeft en naar het

		geboorteland van de (juridische) ouder(s). Op die manier is het herkomstland van de persoon bepaald, en ingedikt naar een drie- of tiendeling van herkomstlanden, conform de <u>herkomstindelingen van het CBS</u> . Als één van beide ouders in het buitenland is geboren, is het geboorteland van deze ouder het herkomstland van de persoon. Als beide ouders in het buitenland zijn geboren wordt gekeken naar het geboorteland van de moeder. Indien het geboorteland van de moeder onbekend is wordt gekeken naar het geboorteland van de vader.
Gezins- en familiefactoren		
Ouderschap tijdens adolescentie	Tienerouder	Persoon is zelf tienerouder (jonger dan 20 jaar) bij geboorte van het oudste (juridische) kind
	Institutioneel huishouden	Persoon woont op peildatum in een institutioneel huishouden: Eén of meer personen die samen een woonruimte bewonen en daar bedrijfsmatig worden voorzien in dagelijkse levensbehoeften. Ook de huisvesting vindt bedrijfsmatig plaats. Het gaat om instellingen zoals verpleeg-, en verzorgingshuizen, instellingen voor geestelijke gezondheidszorg, forensische centra, instellingen voor verstandelijk, lichamelijk en zintuiglijk gehandicapten, instellingen voor verslavingszorg en daklozenopvang, internaten, kloosters, gevangenissen, kazernes, en asielzoekerscentra, waarin de personen in principe voor langere tijd (zullen) verblijven.
Gescheiden van ouders opgroeien/aanwezigheid van pleegzorg	Jeugdhulp met verblijf	Persoon heeft op peildatum jeugdhulp met verblijf, zoals bedoeld en beschreven in de Jeugdwet. Het betreft hulp en zorg aan jongeren en hun ouders bij psychische, psychosociale en/of gedragsproblemen, een verstandelijke beperking van de jongere, of opvoedingsproblemen van de ouders. De jongere verblijft in een pleeggezin, gezinshuis, leef- of behandelgroep, gesloten afdeling, GGZ-instelling of soortgelijke locaties waar jeugdhulp

		geleverd wordt. Het gaat hier alleen om de verblijfsvormen waarbij sprake is van een overnachting en de jongere dus formeel niet thuis slaapt. Ook verblijf in logeerhuizen, alleen tijdens weekenden of juist door de week, vallen onder jeugdhulp met verblijf.
Aanwezigheid hulp(instanties)	Jeugdhulp zonder verblijf	Persoon heeft op peildatum jeugdhulp zonder verblijf. Het betreft hulp en zorg zoals deze bedoeld en beschreven is in de Jeugdwet, en voor zover deze in natura is geleverd door de zorgaanbieder (dus exclusief PGB). Het gaat om hulp en zorg aan jongeren en hun ouders bij psychische, psychosociale en/of gedragsproblemen, een verstandelijke beperking van de jongere, of opvoedingsproblemen van de ouders. De jongere verblijft thuis, in het eigen gezin en slaapt formeel gezien thuis. Het kan zijn dat de jongere bij opa en oma slaapt of bij iemand anders, echter dit is dan niet formeel zo geregeld.
Aanwezigheid van pleegzorg/aanwezigheid hulp(instanties)	Jeugdbescherming	Persoon heeft op peildatum jeugdbescherming. Een jeugdbeschermingsmaatregel wordt door de rechter dwingend opgelegd. Het doel van de jeugdbeschermingsmaatregel is het opheffen van de bedreiging voor de veiligheid en ontwikkeling van het kind. Een kind of jongere wordt dan 'onder toezicht gesteld' of 'onder voogdij geplaatst'.
Stress in de familie/schulden/lage sociaaleconomische status van de familie	Iemand met problematische schulden in het huishouden	Dit betreft de smalle definitie van geregistreerde problematische schulden. In het CBS-onderzoek Schuldenproblematiek in Beeld worden meer registraties meegenomen. Bij institutionele huishoudens wordt gekeken of de persoon zelf problematische schulden heeft. Er is sprake van problematische schulden als ten minste één persoon in het huishouden voldoet aan ten minste één van de volgende criteria op peildatum: – Volgt een WSNP-traject (Wet schuldsanering natuurlijk personen) – Heeft ten minste zes maanden de zorgpremie niet betaald.

		– Een betalingsachterstand van een Wet Mulder-boete bij het CJIB (Centraal Justitieel Incasso Bureau) waarvan de tweede aanmaning ten minste twee maanden openstaat, of zich al in een ernstigere wanbetalersfase bevindt. Daarnaast moet het openstaande bedrag in totaal minimaal 50 euro zijn. – Heeft langer dan 27 maanden een toeslagschuld van totaal minimaal 50 euro openstaan bij de Belastingdienst. – Heeft langer dan 15 maanden een schuld van totaal minimaal 50 euro voor overige belastingaanslagen openstaan bij de Belastingdienst. – Heeft een belastingschuld die in de 12 maanden voor het peilmoment oninbaar is geleden. – Heeft een betalingsachterstand bij de DUO (Dienst Uitvoering Onderwijs) van drie maanden of langer en van minimaal 270 euro.
Alleen wonen maar wel een kind hebben	Eenouderhuishouden met (wel of niet thuiswonende) minderjarig(e) kind(eren)	Persoon is op peildatum woonachtig in een particulier huishouden bestaande uit één persoon of bestaande uit één ouder met een of meer kinderen onder de 18 jaar (die al dan niet in hetzelfde huishouden wonen). Een particulier huishouden ontstaat als één of meer personen samen een woonruimte bewonen en zichzelf, dus niet-bedrijfsmatig, voorzien in de dagelijkse levensbehoeften.
Opgroeien in eenoudergezin	Eenouderhuishouden , kind	Persoon is op peildatum woonachtig in een eenouderhuishouden als kind.
Stress in de familie	Slachtoffer van een ernstig misdrijf in huishouden	Persoon heeft op peildatum iemand in het huishouden wonen die in zeven jaar vóór peiljaar slachtoffer is geworden volgens politieregistratie van minimaal één ernstig misdrijf. Daarbij wordt dezelfde definitie gehanteerd als bij eigen slachtofferschap (zie indicator onder persoonlijke factoren)
Ouders met criminele historie	Ouder veroordeeld door OM of rechterbank	Persoon heeft een juridische ouder die in de zestien jaar vóór pleegjaar is veroordeeld door het Openbaar Ministerie (OM) of de rechter in eerste aanleg. Bij een veroordeling gaat het om een transactie, strafbeschikking,

		voorwaardelijk sepot of een schuldigverklaring.
Groot gezin	Aantal personen in huishouden	Het aantal personen in het huishouden op de peildatum.
	Gestandaardiseerd besteedbaar inkomen huishouden	Het besteedbaar inkomen van een huishouden gecorrigeerd voor de omvang en samenstelling van een huishouden in het jaar vóór het peiljaar. Het gestandaardiseerd inkomen is gelijk aan het besteedbaar huishoudensinkomen gedeeld door de equivalentiefactor van het huishouden.
	Type huishouden*	Het type huishouden van de persoon op peildatum.
	Plaats in het huishouden*	De plaats van de persoon in het huishouden op peildatum. Het gaat bijvoorbeeld om partner, ouder of kind in het huishouden
Netwerkfactoren		
Blootstelling aan delinquentie	Veroordeelde huisgenoot	Persoon heeft op peildatum een <u>huisgenoot</u> die in de zestien jaar vóór peiljaar is veroordeeld door het Openbaar Ministerie (OM) of de rechter in eerste aanleg. Een huisgenoot is iemand die op peildatum op hetzelfde adres woont. Zowel de persoon als de huisgenoot dient op peildatum ingeschreven te staan in de Basisregistratie Personen (BRP) en niet te wonen in een institutioneel huishouden.
Blootstelling aan delinquentie/crimineel familielid	Veroordeelde in familienetwerk	Persoon heeft op peildatum een <u>familielid</u> die in de zestien jaar vóór peiljaar is veroordeeld door het Openbaar Ministerie (OM) of de rechter in eerste aanleg. Onder familie wordt verstaan de kernfamilie (personen die biologisch en/of via adoptie een familierelatie hebben plus personen die een gezamenlijk kind hebben), partners, partners van kernfamilie, schoonfamilie en stieffamilie.
Blootstelling aan delinquent gedrag van peers	Veroordeelde in klasgenotennetwerk	Persoon heeft op peildatum een persoon in het <u>klasgenotennetwerk</u> die in de zestien jaar vóór peiljaar is veroordeeld door het Openbaar Ministerie (OM) of de rechter in eerste aanleg. Onder klasgenotennetwerk wordt verstaan de personen die op 1 oktober vóór peiljaar staan ingeschreven voor dezelfde opleiding. Het is belangrijk om te

		benadrukken dat het in dit netwerk niet mogelijk is om af te leiden welke personen daadwerkelijk bij elkaar in de klas zitten. Dit komt omdat de noodzakelijke informatie om deze indeling te maken voor een aantal onderwijsniveaus onvoldoende beschikbaar is en de beschikbare informatie soms onvoldoende betrouwbaar is. De omvang van het netwerk wordt hierdoor vaak overschat.
Blootstelling aan delinquent gedrag van peers	Veroordeelde in colleganetwerk	Persoon heeft op peildatum een persoon in het <u>colleganetwerk</u> die in de zestien jaar vóór peiljaar is veroordeeld door het Openbaar Ministerie (OM) of de rechter in eerste aanleg. Als collega's worden beschouwd personen die op 1 januari van het betreffende jaar staan ingeschreven in de BRP en met een dienstverband werkzaam zijn bij hetzelfde bedrijf als de persoon in kwestie.
Blootstelling aan delinquent gedrag van peers	Veroordeelde in burennetwerk	Persoon heeft op peildatum een persoon in het <u>burennetwerk</u> die in de zestien jaar vóór peiljaar is veroordeeld door het Openbaar Ministerie (OM) of de rechter in eerste aanleg. Iemands netwerk bestaat uit een combinatie van burens en buurtgenoten. Als buur wordt gedefinieerd een persoon op een van de 10 dichtstbijzijnde adressen. Hierbij geldt geen maximale afstand. Alle personen woonachtig op die adressen worden als buur geregistreerd. Om buurtgenoten af te leiden wordt eerst cirkel met een straal van 200 meter rondom het adres van de persoon bepaald. Vervolgens worden willekeurig 20 personen die wonen op een adres in die cirkel geselecteerd als buurtgenoten.
School- en werkgerelateerde factoren		
	Scholier of student	Persoon heeft in de maand van de peildatum als sociaaleconomische categorie 'schoolgaand'.
Opleidingsniveau	Hoogst behaalde onderwijsniveau	Het hoogst behaalde onderwijsniveau van de persoon op 1 oktober van het jaar vóór het peiljaar. Het hoogst behaalde onderwijsniveau is het hoogste onderwijsniveau waarvoor een diploma is behaald, óf het hoogste onderwijsniveau

		dat volledig is doorlopen (als voor het behalen van de opleiding een diploma niet van toepassing is). Als een opleiding (nog) niet is afgerond, gaat het om het onderwijsniveau dat wordt vereist om deze opleiding te kunnen volgen. N.B. de personen die een opleiding hebben gevolgd voordat de registers beschikbaar waren, staan niet integraal geregistreerd. Voor slechts een deel van hen is middels steekproefgegevens de opleiding wel bekend.
Opleidingsniveau	Hoogst gevolgde onderwijsniveau	Het hoogst gevolgde onderwijsniveau op 1 oktober van het jaar vóór peiljaar. Deze opleiding hoeft (nog) niet met een diploma te zijn afgerond. De bij het hoogst behaalde onderwijsniveau genoemde kanttekening over het opleidingsniveaubestand is ook hier van toepassing.
Speciaal onderwijs	Speciaal onderwijs	Persoon heeft in de negen jaar voor peiljaar op enig moment speciaal onderwijs gevolgd waarbij alleen wordt gekeken naar kinderen met een lichamelijke en verstandelijke beperking en kinderen met een ernstige ontwikkelingsstoornis. Kinderen met een visuele beperking of slechthorende/dove kinderen worden voor deze indicator niet meegenomen als zij speciaal onderwijs volgen.
Voortijdig schoolverlaten	Voortijdig schoolverlater	Persoon was in de twaalf jaar voor peiljaar ten minste eenmaal aangemerkt als voortijdig schoolverlater. Iemand wordt aangemerkt als voortijdig schoolverlater als deze het onderwijs verlaat zonder startkwalificatie. Een leerling beschikt over een startkwalificatie met een havo- of vwo-diploma of met een mbo2-diploma. Personen die 23 jaar of ouder zijn en niet beschikken over een startkwalificatie worden niet aangemerkt als voortijdig schoolverlater.
Motivatied voor school/betrokkenheid bij school/spijbelen	Halt-verwijzing leerplichtwet	Persoon is in de twaalf jaar vóór peiljaar verwezen naar bureau Halt voor een overtreding van de leerplichtwet.
Schoolprestaties	Persoon is gezakt	Persoon is gezakt voor het eindexamen voortgezet onderwijs in de elf jaar voor peiljaar

Schoolprestaties	Gemiddeld eindexamencijfer	Het gemiddelde cijfer berekend op basis van de cijferlijst voor het laatst bekende eindexamen voortgezet onderwijs voor het peiljaar. Er wordt tot elf jaar terug gekeken naar de behaalde cijfers voor het eindexamen.
Schoolprestaties/opleidingsniveau	Licht verstandelijke beperking	Persoon staat op 31 december van het peiljaar ingeschreven in de BRP en voldoet in het betreffende jaar aan een van de voorwaarden om als licht verstandelijk beperkte te worden aangemerkt. Binnen het CBS wordt een licht verstandelijke beperking geoperationaliseerd als een persoon die voldoet aan ten minste één van de volgende voorwaarden: – ontvangt een arbeidsongeschiktheidsuitkering met een diagnose met betrekking tot een licht verstandelijke beperking; – heeft een Wlz-indicatie (Wet langdurige zorg) in het zorgprofiel licht verstandelijk gehandicapt/beperkt; – heeft een WSW-indicatie (WSW = werken in de sociale werkvoorziening) of; – heeft praktijkonderwijs als hoogst behaalde opleiding.
Aanwezigheid goede, stabiele baan	Werkzaam als zelfstandige	Persoon heeft in de maand van de peildatum als sociaaleconomische categorie zelfstandige zonder personeel, zelfstandige met personeel of zelfstandige, meewerkend gezinslid.
Aanwezigheid goede, stabiele baan	Deeltijdfactor werk	De deeltijdfactor van werkzame personen in de maand van de peildatum. Het betreft het aantal uren van alle banen bij elkaar.
Aanwezigheid goede, stabiele baan	Uitzend- of oproepkracht	De positie in de voornaamste werkkring van personen met inkomen uit arbeid (werkzame personen) in de maand van de peildatum is werknemer zijnde uitzendkracht en werknemer zijnde oproepkracht.
Aanwezigheid goede, stabiele baan	Werkzaam met tijdelijk contract	De positie in de voornaamste werkkring van personen met inkomen uit arbeid (werkzame personen) in de maand van de peildatum is werknemer met een tijdelijk dienstverband.
Aanwezigheid goede, stabiele baan	Kwetsbare arbeidsmarktpositie	Persoon heeft een kwetsbare arbeidsmarktpositie. Dit is het geval

		wanneer iemand voldoet aan ten minste één van de volgende voorwaarden: – persoon heeft een ww-, bijstands-, of ao-uitkering in het jaar vóór peiljaar; – persoon heeft geen startkwalificatie in het jaar voor peiljaar (zie eerdergenoemde ‘Voortijdig schoolverlater’ voor een definitie van startkwalificatie); – persoon is financieel afhankelijk op de peildatum waarbij gekeken is of het basisloon lager is dan 130% van het wettelijke minimumloon of (NB: hierin is geen onderscheid gemaakt met het minimumjeugdloon); – persoon stond in de vijf jaar voorafgaand aan peiljaar in het doelgroepenregister voor arbeidsbeperkten. Personen die ingeschreven staan als scholier of student zijn hiervan uitgesloten.
Sociaaleconomische status	Persoonlijk inkomen	Persoonlijk inkomen van een persoon in het jaar voor peiljaar in kwintielen. Inkomens van personen in institutionele huishoudens ontbreken.
Aanwezigheid goede, stabiele baan	Recent arbeidsverleden	Het aantal maanden in de vier jaar voorafgaande aan peiljaar dat een persoon werkzaam is geweest. Hierbij worden zowel werknemers als zelfstandigen meegenomen.
	Uitkering*	Persoon heeft in het jaar voor peildatum minimaal één ww-, bijstands-, of ao-uitkering gehad.
Omgevingsfactoren		
	Misdrijven per 1000 inwoners buurt	Het aantal misdrijven zijnde diefstal uit woning/schuur/e.d., vernieling, misdrijf tegen openbare orde en gewelds- en seksuele misdrijven per 1000 inwoners in de buurt waarin de persoon woont op peildatum.
Ongelijkheid	Gemiddelde SES-totaalscore in de buurt	De sociaaleconomische status (SES-WOA) van de buurt waarin de persoon woont op peeldatum/peildatum. Deze status wordt beschreven in termen van de financiële welvaart, het opleidingsniveau en het recente arbeidsverleden van particuliere

		huishoudens in de buurt op 1 januari van het verslagjaar.
Bevolkingsdichtheid	Stedelijkheidsgraad buurt	De stedelijkheidsgraad van de buurt waarin de persoon woont op de peildatum. Stedelijkheid is een maatstaf voor de concentratie van menselijke activiteiten gebaseerd op de gemiddelde omgevingsadressendichtheid (OAD) van de buurt. Er zijn vijf stedelijkheidsklassen, gebaseerd op het aantal adressen per km ² : zeer sterk stedelijk (OAD van 2500 of meer), sterk stedelijk (OAD van 1500 tot 2500), matig stedelijk (OAD van 1000 tot 1500), weinig stedelijk (OAD van 500 tot 1000) en niet stedelijk (OAD van minder dan 500).
Etnische homogeniteit	Diversiteit buurt	De Herfindahl-Hirschman-index (HHI) bepaald op basis van de buurt waarin de persoon woont op peildatum. De HHI geeft de kans weer dat twee willekeurige personen uit de populatie tot verschillende groepen behoren. Dit wordt bepaald op basis van het geboorteland van de persoon zelf en het geboorteland van de ouders. Op basis van het herkomstland wordt er een indeling gemaakt in achttien categorieën (hierbij wordt de indeling naar herkomstgroepen van de Wetenschappelijke Raad voor het Regeringsbeleid gevolgd). De categorieën vormen de basis voor het berekenen van de HHI. Een HHI van 0 betekent dat er geen diversiteit in herkomst is in de buurt (alle personen in de buurt hebben dezelfde herkomst) en een HHI van 1 dat er veel diversiteit is (iedereen in de buurt heeft een andere herkomst).

Bijlage 2: Missende waarden

Tabel A2. Missende waarden per indicator per onderzoeksgroep.

Indicator	Jongeren (N = 33.140)		Volwassenen (N = 66.480)	
	Missende waarden (n)	Missende waarden (%)	Missende waarden (n)	Missende waarden (%)
Iemand met problematische	510	1,5	940	1,4

schulden in het huishouden					
Gestandaardiseerd besteedbaar huishoudinkomen	1.245	3,8		2.140	3,2
Hoogst gevolgde onderwijsniveau	880	2,7		-	-
Hoogst behaalde onderwijsniveau	-	-		19.275	29
Persoonlijk inkomen	-	-		4.170	6,3
Recent arbeidsverleden	-	-		5	0
Kwetsbare arbeidsmarktpositie	-	-		90	0,1
Stedelijkheidsgraad buurt	-	-		5	0
Gemiddelde SES-totaalscore per buurt	5	0		10	0

Bijlage 3: Resultaten basismodel voor de jongeren en de volwassenen

Tabel A3. Beschrijvende statistieken en resultaten multivariate logistische regressieanalyse voor de verschillende domeinen voor de jongeren.

	Categorieën	Omvang groep	Verdachte/ niet-verdachte			
		<i>n</i>	<i>M (SD)</i>		<i>OR</i>	<i>95% BI</i>
<i>Intercept</i>	-	33.140	20/80		0,79	0,598 - 1,051
Persoonlijke factoren						
Slachtoffer van ernstig misdrijf	Ja	1.860	48/52		2,737	2,409 - 3,132
	Nee (ref.)	31.280	18/82		-	-
Halt-verwijzing overtreding	Ja	1.190	43		2,51	2,09 - 3,01
	Nee (ref.)	31.950	19		-	-
Medicijngebruik hyperactiviteit	Ja	3.790	33/67		1,775	1,593 - 1,970
	Nee (ref.)	29.340	18/82		-	-
Medicijngebruik angststoornis	Ja	830	30/70		1,034	0,837 - 1,256
	Nee (ref.)	32.300	20/80		-	-
Medicijngebruik depressie	Ja	910	31/69		1,267	1,006 - 1,571
	Nee (ref.)	32.220	20/80		-	-
Negatieve levenservaringen	Ja	950	28/72		1,143	0,935 - 1,399
	Nee (ref.)	32.190	20/80		-	-

Geboren in Nederland i.c.m. herkomst					
Geboren in Nederland	Herkomst Nederland (.ref)	23.780	16/84	-	-
	Herkomst Europa (excl. Nederland)	1.160	21/79	1,175	0,972 - 1,429
	Herkomst buiten Europa	5.020	34/66	1,590	1,435 - 1,748
Niet geboren in Nederland	Herkomst Europa (excl. Nederland)	1.100	15/85	1,182	0,927 - 1,492
	Herkomst buiten Europa	2.080	35/65	2,641	2,286 - 3,057
Tienerouder	Ja	100	60/40	2,276	1,332 - 4,020
	Nee (ref.)	33.030	20/80	-	-
Gezins- en familiefactoren					
Institutioneel huishouden	Ja	460	54/46	1,749	1,310 - 2,298
	Nee (ref.)	32.670	19/81	-	-
Jeugdhulp met verblijf	Ja	450	67/33	3,030	2,211 - 4,132
	Nee (ref.)	32.690	19/81	-	-
Jeugdhulp zonder verblijf	Ja	1.720	44/56	1,702	1,482 - 1,966
	Nee (ref.)	31.420	19/81	-	-
Jeugdbescherming	Ja	340	68/32	1,277	0,888 - 1,848
	Nee (ref.)	32.790	20/80	-	-
Iemand met problematische schulden in het huishouden	Ja	2.550	44/56	1,429	1,262 - 1,601
	Nee (ref.)	30.590	18/82	-	-
Eenouderhuishouden, kind	Ja	5.990	30/70	1,312	1,176 - 1,448
	Nee (ref.)	27.150	18/82	-	-

Slachtoffer van een ernstig misdrijf in huishouden	Ja	3.800	32/68		1,411	1,270 - 1,556
	Nee (ref.)	29.330	18/82		-	-
Ouder veroordeeld door OM of rechtbank	Ja	5.080	40/60		1,161	1,045 - 1,297
	Nee (ref.)	28.050	16/84		-	-
Aantal personen in huishouden	1	3.630	22/78		1,072	0,919 - 1,249
	2	3.640	25/75		0,900	0,788 - 1,036
	3 (ref.)	6.640	21/79		-	-
	4	11.240	16/84		1,025	0,921 - 1,134
	5 of meer	7.990	21/79		1,027	0,916 - 1,153
Gestandaardiseerd besteedbaar inkomen huishouden	1 t/m 20%	7.040	30/70		1,244	1,115 - 1,386
	21 t/m 40%	5.080	24/76		1,099	0,981 - 1,232
	41 t/m 60% (ref.)	6.940	21/79		-	-
	61 t/m 80%	7.160	14/86		0,920	0,818 - 1,032
	81 of meer	6.910	12/88		0,945	0,837 - 1,060
Peerfactoren						
		<i>n</i>	<i>M (SD)</i>		<i>OR</i>	<i>95% BI</i>
Veroordeelde huisgenoot	-	33.140	14,8 (27,1)/ 5,1 (16,3)		1,004	1,003 - 1,006
Veroordeelde in het familienetwerk	-	33.140	14,8 (17,3)/ 6,5 (10,5)		1,020	1,016 - 1,024
Veroordeelde in het klassennetwerk	-	33.140	3,5 (9,3)/ 2,0 (5,8)		1,012	1,007 - 1,018

Veroordeelde in het buurtnetwerk	-	33.140	9,3 (7,2)/ 6,8 (5,9)		1,012	1,005 - 1,018
School- en werkgerelateerde factoren						
Scholier of student	Ja	26.690	19/81		0,769	0,680 - 0,858
	Nee (ref.)	6.450	25/75		-	-
Hoogst gevolgde onderwijsniveau	Basisonderwijs	610	34/66		1,782	1,407 - 2,251
	Vmbo-b/k, mbo 1	4990	31/69		1,736	1,555 - 1,935
	Vmbo-g/t, havo-, vwo- onderbouw (ref.)	6.610	24/76		-	-
	Mbo2 en mbo3	4.250	31/69		1,451	1,241 - 1,695
	Havo, vwo	4.360	8/92		0,517	0,440 - 0,602
	Mbo4	5.140	19/81		0,921	0,782 - 1,094
	Hbo-, wo-bachelor	6.350	9/91		0,475	0,399 - 0,567
	Hbo-, wo-master, doctor	820	5/95		0,264	0,169 - 0,381
Speciaal onderwijs	Ja	1.870	47/53		1,712	1,453 - 1,982
	Nee (ref.)	31.260	18/82		-	-
Voortijdig schoolverlater	Ja	1.930	38/62		1,390	1,196 - 1,605
	Nee (ref.)	31.200	19/81		-	-
Persoon is gezakt	Ja	1.210	21/79		1,176	0,944 - 1,483
	Nee (ref.)	31.930	20/80		-	-
Licht verstandelijke beperking	Ja	230	48/52		1,502	1,036 - 2,113
	Nee (ref.)	32.900	20/80		-	-
Gemiddeld eindexamencijfer	Zeer laag: < 5.50	130	31/69		1,069	0,621 - 1,812
	Laag: 5.50 tot 6.00	1.010	26/74		1,181	0,940 - 1,497

	Onder gemiddeld: 6.00 tot 6.60 (ref.)	6.950	19/81		-	-
	Boven gemiddeld: 6.60 tot 7.10	4.810	13/87		0,692	0,602 - 0,788
	Hoog: 7.10 tot 7.60	1.710	11/89		0,574	0,466 - 0,702
	Zeer hoog: > 7.60	570	4/96		0,219	0,116 - 0,358
	Onbekend/nog geen eindexamen gedaan	17.950	23/77		0,837	0,712 - 0,968
Halt-verwijzing leerplichtwet	Ja	270	52/48		0,818	0,556 - 1,207
	Nee	32.860	20/80		-	-
Deeltijdfactor werk	Geen baan	17.730	22/78		0,720	0,630 - 0,831
	0,01 t/m 0,29	6.190	15/85		0,777	0,668 - 0,907
	0,30 t/m 0,59 (ref.)	3.160	20/80		-	-
	0,60 t/m 0,79	1.280	22/78		0,920	0,739 - 1,145
	0,80 t/m 0,94	1.120	22/78		1,021	0,825 - 1,253
	0,95 of meer	3.650	19/81		0,857	0,724 - 1,000
Omgevingsfactoren						
Gemiddelde SES-totaalscore in de buurt	1e kwintiel	6.370	30/70		1,163	1,003 - 1,337
	2e kwintiel	6.050	23/77		1,027	0,905 - 1,156
	3e kwintiel (ref.)	6.490	19/81		-	-
	4e kwintiel	6.750	16/84		0,993	0,885 - 1,118
	5e kwintiel	7.480	14/86		0,955	0,847 - 1,075
Misdrijven per 1000 inwoners in de buurt	1e kwintiel	7.400	14/86		0,795	0,701 - 0,898
	2e kwintiel	6.950	17/83		0,913	0,824 - 1,027

	3e kwintiel(ref.)	6.700	20/80		-	-
	4e kwintiel	6.670	24/76		1,002	0,899 - 1,126
	5e kwintiel	5.420	28/72		1,047	0,921 - 1,190
Stedelijkheidsgraad buurt	Zeer sterk	7.660	23/77		0,758	0,662 - 0,864
	Sterk	8.330	22/78		0,835	0,744 - 0,933
	Matig (ref.)	6.200	20/80		-	-
	Weinig	5.430	17/83		0,877	0,779 - 0,999
	Niet	5.520	15/85		0,740	0,652 - 0,839
		<i>n</i>	<i>M (SD)</i>		<i>OR</i>	<i>95% BI</i>
Diversiteit buurt	-	33.140	0,43 (0,22)		0,943	0,714 -1,274

Noot. *N* = 33.140 (20% verdachten/80% niet-verdachten). *OR* = *odds ratio*; *BI* = betrouwbaarheidsinterval; *M* = gemiddelde; *SD* = standaardafwijking. De *odds ratio's* moeten geïnterpreteerd worden ten opzichte van de (groep) categorieën zonder *odds ratio's*. Voor de numerieke variabelen werden in plaats van de omvang van de groepen, het gemiddelde en de standaardafwijking gegeven.

Tabel A4. Beschrijvende statistieken en resultaten multivariate logistische regressieanalyse voor de verschillende domeinen voor de volwassenen.

	Categorieën	Omvang groep	Verdachte/ Niet-verdachte			
		<i>n</i>	<i>M (SD)</i>		<i>OR</i>	<i>95% BI</i>
<i>Intercept</i>	-	66.480	20/80		0,336	0,275 - 0,409
Persoonlijke factoren						
Slachtoffer van ernstig misdrijf	Ja	4.650	42/58		2,282	2,097 - 2,474
	Nee (ref.)	61.840	18/82		-	-
Medicijngebruik middelengebruik	Ja	1.020	50/50		2,122	1,783 - 2,478
	Nee (ref.)	65.460	20/80		-	-
Medicijngebruik hyperactiviteit	Ja	2.690	34/66		1,777	1,592 - 1,984
	Nee (ref.)	63.800	19/81		-	-
Medicijngebruik angststoornis	Ja	7.580	34/66		1,425	1,325 - 1,533
	Nee (ref.)	58.910	18/82		-	-
Medicijngebruik depressie	Ja	8.390	34/66		1,472	1,372 - 1,578
	Nee (ref.)	58.100	18/82		-	-
Negatieve levenservaringen	Ja	3.340	31/69		1,434	1,293 - 1,590
	Nee (ref.)	63.140	19/81		-	-
Geboren in Nederland i.c.m. herkomst						
Geboren in Nederland	Herkomst Nederland (ref.)	48.600	17/83		-	-
	Herkomst Europa (excl. Nederland)	1.770	20/80		1,036	0,894 - 1,206

	Herkomst Nederlands-Caribisch gebied	250	32/68		1,965	1,379 - 2,721
	Herkomst Turkije	880	33/67		1,430	1,193 - 1,726
	Herkomst Marokko	670	45/55		2,344	1,926 - 2,846
	Herkomst Suriname	740	38/62		1,706	1,401 - 2,109
	Herkomst Indonesië	1.250	18/82		1,100	0,916 - 1,340
	Herkomst overig Azië	380	18/82		0,937	0,658 - 1,265
	Herkomst overig Afrika	280	36/64		1,837	1,339 - 2,521
	Herkomst Amerika en Oceanië	280	21/79		1,276	0,845 - 1,775
Niet geboren in Nederland	Herkomst Europa (excl. Nederland)	3.510	22/78		1,215	1,086 - 1,355
	Herkomst Nederlands-Caribisch gebied	480	38/62		1,875	1,476 - 2,391
	Herkomst Turkije	1.040	33/67		1,229	1,023 - 1,481
	Herkomst Marokko	910	40/60		1,595	1,338 - 1,879
	Herkomst Suriname	780	38/62		1,959	1,618 - 2,396
	Herkomst Indonesië	290	17/83		1,119	1,618 - 2,396
	Herkomst overig Azië	2.520	34/66		1,947	1,736 - 2,176
	Herkomst overig Afrika	1.160	42/58		2,277	1,927 - 2,666
	Herkomst Amerika en Oceanië	670	22/78		1,258	0,980 - 1,588
Gezins- en familiefactoren						
Institutioneel huishouden	Ja	920	40/60		1,263	1,052 - 1,515
	Nee (ref.)	65.570	20/80		-	-
Iemand met problematische schulden in het huishouden	Ja	4.520	47/53		1,898	1,754 - 2,053

	Nee (ref.)	61.960	18/82	-	-
Eenouder- of eenpersoonshuishouden met minderjarig(e) kind(eren)	Ja	2.990	42/58	2,125	1,915 - 2,383
	Nee (ref.)	63.500	19/81	-	-
Slachtoffer van een ernstig misdrijf in huishouden	Ja	4.390	32/68	1,558	1,418 - 1,707
	Nee (ref.)	62.100	19/81	-	-
Ouder veroordeeld door OM of rechtbank	Ja	4.420	35/65	1,293	1,173 - 1,417
	Nee (ref.)	62.070	19/81	-	-
Gestandaardiseerd besteedbaar inkomen huishouden	1 t/m 20%	11.730	34/66	1,296	1,192 - 1,403
	21 t/m 40%	9.550	25/75	1,117	1,025 - 1,208
	41 t/m 60% (ref.)	13.900	20/80	-	-
	61 t/m 80%	15.110	15/85	0,939	0,867 - 1,018
	81 of meer	16.200	12/88	0,862	0,794 - 0,940
Peerfactoren					
		<i>n</i>	<i>M (SD)</i>	<i>OR</i>	<i>95% BI</i>
Veroordeelde huisgenoot	-	66.480	10,2 (25,8)/ 3,9 (16,4)	1,004	1,003 - 1,005
Veroordeelde in het familienetwerk	-	66.480	11,2 (15,5)/ 6,2 (10,3)	1,015	1,012 - 1,017
Veroordeelde in het colleganetwerk	-	66.480	6,9 (10,8)/ 5,4 (7,6)	1,022	1,019 - 1,025
Veroordeelde in het buurtnetwerk	-	66.480	9,2 (7,2)/ 7,0 (5,9)	1,017	1,013 - 1,022

School- en werkgerelateerde factoren					
Scholier of student	Ja	11.230	17/83	0,941	0,842 - 1,059
	Nee (ref.)	55.250	21/79	-	-
Hoogst behaalde onderwijsniveau (3 categorieën)	Basisonderwijs, vmbo, mbo1	16.680	32/68	1,377	1,296 - 1,464
	Havo, vwo, mbo2-4 (ref.)	28.550	20/80	-	-
	Hbo, wo	21.250	11/89	0,697	0,650 - 0,744
Licht verstandelijke beperking	Ja	510	37/63	1,012	0,815 - 1,282
	Nee (ref.)	65.970	20/80	-	-
Werkzaam als zelfstandige	Ja	7.520	22/78	2,069	1,913 - 2,260
	Nee (ref.)	58.960	20/80	-	-
Deeltijdfactor werk	Geen baan	18.110	29/71	1,365	1,197 - 1,566
	0,01 t/m 0,29	4.170	19/81	0,942	0,817 - 1,071
	0,30 t/m 0,59 (ref.)	4.700	20/80	-	-
	0,60 t/m 0,79	4.950	20/80	0,940	0,827 - 1,069
	0,80 t/m 0,94	5.370	19/81	1,125	0,986 - 1,280
	0,95 of meer	29.170	15/85	1,074	0,948 - 1,201
Uitzend- of oproepkracht	Ja	5.700	25/75	1,388	1,244 - 1,548
	Nee (ref.)	60.780	20/80	-	-
Werkzaam met tijdelijk contract	Ja	10.110	18/82	1,210	1,109 - 1,313
	Nee (ref.)	56.370	20/80	-	-
Kwetsbare arbeidsmarktpositie	Ja	22.120	31/69	1,336	1,225 - 1,456
	Nee (ref.)	30.020	14/86	-	-
	Startkwalificatie onbekend	14.350	15/85	1,086	1,005 - 1,178
Persoonlijk inkomen	1 t/m 20%	14.830	22/78	0,745	0,676 - 0,823

	21 t/m 40%	10.140	31/69		0,934	0,854 - 1,028
	41 t/m 60% (ref.)	11.820	24/76		-	-
	61 t/m 80%	13.570	16/84		0,958	0,886 - 1,043
	81 of meer	16.110	11/89		0,966	0,879 - 1,063
Recent arbeidsverleden (werkzaam in de 4 jaar voor peiljaar)	Niet werkzaam (ref.)	10.220	30/70		-	-
	Minimaal 1 maand tot maximaal 1/4 werkzaam	4.170	29/71		1,109	0,988 - 1,241
	Meer dan 1/4 werkzaam	4.720	25/75		0,988	0,879 - 1,103
	Meer dan 2/4 werkzaam	5.940	23/77		0,953	0,853 - 1,063
	Meer dan 3/4 werkzaam	7.940	21/79		0,928	0,822 - 1,047
	Voortdurend werkzaam	33.500	14/86		0,785	0,698 - 0,889
Omgevingsfactoren						
Gemiddelde SES-totaalscore buurt	1e kwintiel	12.570	29/71		0,955	0,698 - 0,889
	2e kwintiel	13.180	23/77		1,074	0,993 - 1,158
	3e kwintiel	13.500	18/82		-	-
	4e kwintiel	13.570	16/84		0,999	0,925 - 1,079
	5e kwintiel	13.660	14/86		1,006	0,926 - 1,088
Misdrijven per 1000 inwoners buurt	1e kwintiel	13.580	13/87		0,810	0,747 - 0,886
	2e kwintiel	13.900	17/83		0,942	0,874 - 1,012
	3e kwintiel	13.740	20/80		-	-
	4e kwintiel	13.590	24/76		1,072	0,994 - 1,158
	5e kwintiel	11.670	27/73		1,086	1,002 - 1,175
Stedelijkheidsgraad buurt	Zeer sterk	16.840	23/77		0,787	0,722 - 0,860
	Sterk	16.730	22/78		0,893	0,821 - 0,968

	Matig (ref.)	12.060	19/81		-	-
	Weinig	10.400	16/84		0,973	0,894 - 1,060
	Niet	10.460	16/84		0,993	0,907 - 1,088
		<i>n</i>	<i>M (SD)</i>		<i>OR</i>	<i>95% BI</i>
Diversiteit buurt	-	66.480	0,43 (0,22)		1,096	0,901 - 1,341

Noot. *N* = 66.480 (20% verdachten/ 80% niet-verdachten). *OR* = *odds ratio*; *BI* = betrouwbaarheidsinterval; *M* = gemiddelde; *SD* = standaardafwijking. De *odds ratio*'s moeten geïnterpreteerd worden ten opzichte van de (groep) categorieën zonder *odds ratio*'s. Voor de numerieke variabelen werden in plaats van de omvang van de groepen, het gemiddelde en de standaardafwijking gegeven.

Bijlage 4: Robuustheidscontroles

In deze bijlage wordt beschreven hoe de aannames die horen bij een logistische regressie zijn gecontroleerd. Daarnaast zijn enkele sensitiviteitsanalyses uitgevoerd, die worden ook beschreven.

Bij een logistische regressie moet ten eerste voldaan worden aan de aanname dat er een lineaire relatie moet bestaan tussen elke continue onafhankelijke variabele en de (logistische) *odds* van de afhankelijke variabele (verdachte). Dit werd getest door het visualiseren van de relatie tussen de voorspelde waarden en de residuen. Ten tweede werd gecontroleerd op multicollineariteit. Multicollineariteit treedt op wanneer de onafhankelijke variabelen (te) sterk samenhangen en kan leiden tot vertekeningen van het regressiemodel. De *Variance Inflation Factor* (VIF) is een maat die aangeeft hoeveel de variantie van een regressiecoëfficiënt toeneemt door multicollineariteit. Een VIF van 5 of hoger duidt op mogelijke problemen. In het regressiemodel zaten zowel categorische als continue variabelen, zodoende werd de *Generalized Variance Inflation Factor* (GVIF) gebruikt. Voor interpretatie van de GVIF werd een correctie toegepast ($GVIF (1/(2 \cdot df))$), zodat de waarde bij categorische variabelen op dezelfde manier te interpreteren was als de VIF (Fox & Monette, 1992). Ten derde werden uitbijters (i.e., *outliers*) geïdentificeerd door het berekenen van de *Inter Quartile Range* (IQR). Tot slot werd de onafhankelijkheid van residuen bekeken door het visualiseren van de residuen. Geen van de aannames werd geschonden, behalve de aannames van de uitbijters. Hoe daar mee om is gegaan word hieronder beschreven.

Er werden sensitiviteitsanalyses uitgevoerd op vier onderdelen: (1) peilvarianten (verschillende peilmomenten); (2) imputatievarianten (verschillende geïmputeerde datasets gebruikt); (3) matchingsvarianten (verschillende mate van matching van verdachten met referenten); en (4) uitbijtervarianten (verschillende varianten van omgaan met uitbijters). Bij elk onderdeel werden minstens twee varianten vergeleken. De uitkomsten van de lasso regressie vormde het startpunt, waarna verschillen nader werden bekeken in de beschrijvende statistieken en de uitkomsten van de logistische regressieanalyse. De beoordeling van de verschillen tussen de varianten werd door twee onafhankelijke onderzoekers uitgevoerd. Eventuele verschillen in beoordeling werden voorgelegd aan een derde onderzoeker waarna in samenspraak tot consensus werd gekomen.

Ten eerste werd gekeken naar verschillende peilmomenten van de indicatoren voor de verdachtenpopulatie. Voor één variant werden de kenmerken van de verdachten zo dicht mogelijk bij de pleegdatum gepeild, overeenkomstig met andere analyses die uitgevoerd worden door het IAT. Voor de andere variant werden deze kenmerken op hetzelfde moment gepeild als voor de niet-verdachten populatie (1 januari 2017). Tussen deze varianten werden minimale verschillen gevonden in de resultaten. Op basis van inhoudelijke gronden werd ervoor gekozen om verder te gaan met de peilvariant waarbij de kenmerken van de verdachten en de niet-verdachten op hetzelfde moment te peilen (1 januari 2017).

Dit om mogelijke inconsistentie tussen kenmerken te voorkomen door het peilen op verschillende momenten door het jaar heen.

Ten tweede werden de verschillende geïmputeerde datasets gebruikt om te kijken of er verschillen in resultaten naar voren kwamen. Hierin werden minimale verschillen gevonden en zodoende werd één imputatiedataset gekozen. Ten derde werd gekeken naar verschillende methoden van matching met de referenten. Voor één variant werd gekeken naar vier gebalanceerde modellen, waarbij elke verdachte werd gematcht op leeftijd en geslacht met één van de vier referenten. Voor de andere variant werd een analyse uitgevoerd met vier op leeftijd- en geslacht gematchte niet-verdachte referenten voor elke verdachte. Hierin werden wel verschillen gevonden en leek het uit te maken met welke referent de verdachte gematcht werd. Zodoende werd ervoor gekozen om verder te gaan met een niet gebalanceerd model, waarbij elke verdachte werd gekoppeld aan vier niet-verdachte op leeftijd- en geslacht gematchte referenten.

Tot slot werd gekeken naar verschillende methoden voor het omgaan met uitbijters. De uitbijters speelden bij de indicatoren misdrijven per 1000 inwoners in de buurt, en de gemiddelde SES-totaalscore van de buurt (totaalscore van de sociaaleconomische status). Bij de eerste variant zijn de oorspronkelijke waarden en indeling van de variabelen gebruikt. Bij de tweede variant werden de indicatoren ingedeeld in kwintielen. Hierbij werden minimale verschillen gevonden en werd ervoor gekozen om de variant met kwintielen mee te nemen in het eindmodel.

Colophon

Publisher

Centraal Bureau voor de Statistiek
Henri Faasdreef 312, 2492 JP Den Haag
www.cbs.nl

Prepress

Statistics Netherlands, CCN Creation and visualisation

Design

Edenspiekermann

Information

Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contactform: www.cbs.nl/information

© Statistics Netherlands, The Hague/Heerlen/Bonaire **Fout! Verwijzingsbron niet gevonden.**Maart Maart .

Reproduction is permitted, provided Statistics Netherlands is quoted as the source.