



Profilering van burgers tijdens wapencontroles in Amsterdam: een observationele analyse van praktijken en mogelijke biases

Marie Rosenkrantz Lindegaard
Peter Ejbye-Ernst
Mara van Dalen
Hans Myhre Sunde
Carlijn van Baak
Melissa Sexton
Virginia Pallante
Fabienne Thijs
Lea Echelmeyer
Steve van de Weijer
Laura Pighini
Gabriele Chlevickaite
Jo Thomas
Lasse Suonperä Liebst

Amsterdam, december 2022

Profilering van burgers tijdens wapencontroles in Amsterdam: een observationele analyse van praktijken en mogelijke biases

DECEMBER 2022

Marie Rosenkrantz Lindegaard ^{1, 2}

Peter Ejbye-Ernst ¹

Mara van Dalen ¹

Hans Myhre Sunde ^{1, 2}

Carlijn van Baak ^{1, 2}

Melissa Sexton ¹

Virginia Pallante ¹

Fabienne Thijs ¹

Lea Echelmeyer ¹

Steve van de Weijer ¹

Laura Pighini ¹

Gabriele Chlevickaite ¹

Jo Thomas ¹

Lasse Suonperä Liebst ^{1, 3}

¹ NSCR: Nederlands Studiecentrum Criminaliteit en Rechtshandhaving

² Afdeling Sociologie, Universiteit van Amsterdam

³ Afdeling Sociologie, Universiteit van Kopenhagen

Contactgegevens

Prof. Dr. Marie Rosenkrantz Lindegaard, mrlindegaard@nscr.nl

Acknowledgements

Met dank aan de projectcoördinatoren bij de gemeente Amsterdam en bij de politie. Wij danken ook Wim Bernasco voor zijn bijdragen aan het lanceren van de studie.

Bijdragende auteurs

Conceptualisatie: MRL, LSL, PEE. Dataconservering: MD. Formele analyse: LSL, PEE, VP, MD. Acquisitie van fondsen: MRL. Meetinstrument: MRL, CB, MR, PEE, MS. Veldobservaties: MRL, PEE, HMS, VP, MS, CB, FT, LE, SW, LP, GC, JT. Methodologie: MRL, LSL, PEE. Projectadministratie: MRL. Bronnen: MRL. Supervisie: MRL. Validering: LSL. Visualisatie: PEE. Schrijven – Opstellen originele conceptversie: MRL, PEE, LSL, VP, MR. Schrijven – Review en Editing: MRL, PEE, HMS, VP, MS, CB, FT, LE, SW, LP, GC, JT, LSL

Belangrijkste bevindingen

- Jongere mensen, vooral jongeren van ongeveer 25 tot 30 jaar, werden volgens onze bevindingen onevenredig vaak geselecteerd voor een wapencontrole, in vergelijking tot de personen die op dezelfde locatie aanwezig waren en niet geselecteerd werden.
- Wij hebben *niet* gevonden dat niet-witte mensen en mannen onevenredig vaak werden geselecteerd voor een wapencontrole, in vergelijking tot de personen die op dezelfde locatie aanwezig waren en niet geselecteerd werden.

Samenvatting

Voor de legitimiteit van politiewerk is het van essentieel belang dat zoekacties onbevooroordeeld worden uitgevoerd. Bestaande aanwijzingen, gebaseerd op meldingen en vragenlijsten, over zoekpraktijken van de politie in Nederland en elders wijzen erop dat agenten in de ogen van burgers bevooroordeeld zijn bij het selecteren van hun doelwitten tijdens zoekacties. In dit project hebben we een pilotonderzoek uitgevoerd waarin we hebben onderzocht of er bij wapencontroles in Amsterdam sprake is van eventuele vooroordelen op basis van gender, etnische afkomst en/of leeftijd. Hierbij gebruikten we een observationele, objectieve maat voor het selecteren van burgers. De pilot werd beëindigd na niet-geautoriseerde praktijken van de politie. Daarom werd ook dit onderzoek tijdens het verzamelen van de data stopgezet. We zijn er echter wel in geslaagd om een redelijke hoeveelheid kwantitatieve data en veldwaarnemingsdata te verzamelen, zodat we enkele voorlopige conclusies kunnen trekken. Uit onze statistische analyse blijkt dat jongeren van midden tot eind 20 jaar een onevenredig grote kans hadden om te worden uitgekozen voor een wapencontrole. Voor niet-witte mensen en voor mannen werden dergelijke tendensen niet gevonden. We willen hierbij aantekenen dat het niet duidelijk is of het feit dat geen profilering op basis van etnische afkomst of gender werd gevonden, komt doordat deze bias niet bestaat, of doordat de verzamelde dataset eenvoudig te klein en te vertroebeld is om deze mogelijke bias vast te stellen.

INLEIDING

Voor de legitimiteit van politiewerk is het van essentieel belang dat zoekacties a-selectief worden uitgevoerd (Bradford & Loader, 2016). Uit onderzoeken die in verschillende landen werden uitgevoerd, ook in Nederland (Hesseling & Wilde, 2022), blijkt echter dat burgers het gevoel hebben dat agenten tijdens zoekacties bevooroordeeld zijn bij het selecteren van hun doelwitten. Daarbij gaat het vooral om een onevenredig grote focus op jonge mannen van kleur (voorbeeld: Denison & Finkeldey, 2021). Een probleem van deze onderzoeken – en van onderzoeken naar discriminatie in het algemeen – is dat die vaak afhankelijk zijn van de subjectieve ervaringen van de betrokkenen, en dat die ervaringen voor meerdere interpretaties vatbaar kunnen zijn. Hierdoor is er behoefte aan objectievere maten voor profilering van overtredders ontstaan, maar het is niet eenvoudig om een nauwkeurige methode te vinden om dit probleem te evalueren (Farrell & McDevitt, 2010). Er is wel een veelbelovende manier om dit te doen, maar die wordt helaas weinig gebruikt: het direct ter plaatse observeren van de zoekacties. Een van de zeldzame keren dat deze methode wel werd toegepast, was in Parijs. De resultaten van dat onderzoek in termen van gedrag bevestigden de subjectieve ervaringen: tijdens de zoekacties selecteerden agenten onevenredig vaak jonge mannen van kleur – ondanks de gegeven instructies om neutrale profileringstrategieën te gebruiken (Jobard & Lévy, 2011). In dit rapport hebben we deze veelbelovende gedragsbenadering gevolgd door preventieve wapencontroles in Amsterdam te observeren.

Vanwege de stijging in wapengerelateerde misdrijven in de gemeente Amsterdam heeft de burgemeester van de stad besloten om een pilot naar wapencontroles te laten uitvoeren. De gemeenteraad heeft jarenlang wapencontroles vermeden, omdat de impact geïnterpreteerd werd als disproportioneel – waarbij het potentiële probleem van bias door etnische profilering in de selectieprocedures ook een rol speelde. Hierdoor bestaat er een spanning tussen de potentiële negatieve impact van controles op mensen die worden onderzocht aan de ene kant, en aan de andere kant de potentiële voordelen van het vinden van wapens en het voorkomen van wapengerelateerde misdrijven. Als onderdeel van de pilot liet de burgemeester de negatieve impact onderzoeken door burgers te vragen om zich op te geven als externe ‘waarnemers’ van de politie tijdens controles. In totaal gaven 37 burgers zich op om tijdens controles met de politie samen te werken. Zij hebben hun indruk van de procedures vastgelegd in een rapport (Hesseling & Wilde, 2022). Slechts zes procent van de waarnemers merkte dat de politie etnische profilering toepaste in hun selectieprocedures. Er werd slechts één officiële klacht ingediend. Aan de andere kant bleek uit een enquête waarin burgers werd gevraagd naar hun algemene mening over de controles, dat ongeveer 25 procent van de respondenten verwachtte dat etnische profilering plaatsvond. In het rapport van de eerste pilot wordt bij de vraag of er sprake is van etnische profilering alleen aandacht besteed aan deze subjectieve beoordelingen van externe waarnemers en burgers.

Aangezien de vraag of er sprake is van etnische profilering in de eerste pilot alleen gebaseerd was op meldingen en vragenlijsten, blijft het dus onduidelijk of de politie onevenredig vaak mensen van kleur had geselecteerd voor een controle. Daarnaast liet de evaluatie van de eerste

pilot zien dat er meer mannen dan vrouwen werden geselecteerd voor een controle. Dit kan wijzen op een bias naar het selecteren van mannen. Bovendien liet het onderzoek ook zien dat er meer jongeren tot 30 jaar werden geselecteerd dan mensen ouder dan 30 jaar. Dit kan wijzen op een bias naar het selecteren van jongeren. Dit resultaat hoeft echter geen aanwijzing te zijn voor een bias naar mannen of jongeren in de selectiepraktijk. Het kan ook eenvoudig worden toegeschreven aan het feit dat er op de gecontroleerde openbare plaatsen meer mannen dan vrouwen aanwezig waren, en meer jongeren dan ouderen. Uit de evaluatie van de eerste pilot is het ook niet duidelijk wat de etniciteit was van de geselecteerde mensen, aangezien de politie deze categorie niet registreert.

Na de evaluatie van de eerste pilot concludeerde de gemeente Amsterdam dat de pilot kon worden voortgezet. De gemeente besloot dat in de tweede pilot geen burgers als waarnemers moesten worden benoemd, maar onafhankelijke onderzoekers de rol van waarnemer op zich moesten nemen – deze taak hebben wij in het huidige project uitgevoerd. Onze rol was specifiek het onderzoeken of de selectieprocedures van de politie onbevooroordeeld plaatsvonden – dat wil zeggen dat iedereen die op de locatie waar de controles werden gehouden aanwezig was, een gelijke kans moest hebben om te worden geselecteerd voor een controle.

Dit onderzoek richt zich op de tweede pilot naar wapencontroles in Amsterdam, waarbij 25 controles gepland werden tussen oktober 2022 en januari 2023. De pilot werd echter al na dat wij zes controles konden observeren beëindigd vanwege niet-geautoriseerde praktijken van de politie. Daarom zijn er in dit project alleen data beschikbaar van deze zes uitgevoerde controles. Voordat de controles werden uitgevoerd had de politie vijf stadsdelen aangewezen waar – volgens hun misdaadcijfers – de kans op wapengerelateerde misdrijven extra groot was. Deze zogeheten ‘veiligheidsrisicogebieden’ waren: Noord, Zuid, Zuid-Oost, Noord-Nieuw-West en Zuid-Nieuw-West. In elk van deze gebieden had de politie een aantal ‘hotspots’ aangewezen waar zij verwachtte meer wapens aan te treffen. Aan het publiek werd bekendgemaakt dat de politie soms gebruik zou maken van handscanners (om wapens te detecteren), een detectiepoort, en een random-selectiepaal (een slimme paal waarbij de burgers op een knop moeten drukken om te mogen passeren, waarbij een deel van de passanten automatisch, geheel willekeurig, worden geselecteerd voor een controle).

Het huidige onderzoek richt zich specifiek op het statistisch onderzoeken of er bij de politie sprake was van bias naar het selecteren op basis van etniciteit, gender of leeftijdscategorieën tijdens wapencontroles. Om deze vraag te onderzoeken moesten we eerst begrijpen hoe de selectieprocedure van de politie in de praktijk werkt. We werden vooraf geïnformeerd dat er twee manieren werden gebruikt om mensen te selecteren: afzonderlijke burgers werden ofwel geselecteerd door de agenten, of door de random-selectiepaal – wij spreken hier verder van respectievelijk ‘gewone selectie’ en ‘paal-selectie’. Bij het observeren van hun werkprocedures bleek echter dat er ook een derde selectiemethode werd gebruikt. Bij deze methode werd een bepaald gebied afgesloten, waarbij niemand dat gebied mocht verlaten voordat ze waren onderzocht. Deze methode zullen wij ‘gebied-selectie’ noemen. In onze analyse van mogelijke bias in de selectieprocedures onderscheiden we deze drie methoden.

METHODE

We besloten om de politiecontroles ter plaatse te observeren, omdat we hiermee ecologisch zeer valide en kwantitatief fijnmazige data konden verkrijgen over hoe de controles in werkelijkheid plaatsvonden. Aan de andere kant kennen observaties ter plekke een mogelijk nadeel, namelijk het probleem van ‘reactiviteit’: agenten zouden hun praktijken kunnen aanpassen omdat zij zich bewust zijn van het feit dat ze worden geobserveerd. We hebben hierover gesproken met de politie en met de gemeente en daarnaast ethisch advies over wetenschappelijk onderzoek ingewonnen. Daarop hebben we besloten om de observaties ter plaatse uit te voeren vanuit een verdeckte positie. De ethische code van de *American Sociological Association* (1999) en vergelijkbare verenigingen benadrukken dat onderzoekers op openbare plaatsen natuurlijke observaties mogen doen, zoals in dit geval gebeurt. Hiervoor is geen toestemming nodig. Omdat de observaties vanuit een verdeckt positie werden gedaan, wisten de agenten die de controles uitvoerden niet of ze wel of niet direct geobserveerd werden. Dat betekent, in onze ervaring, dat we ons gewoon tussen de mensen konden begeven zonder enige aanwijzing dat we ontdekt waren.

Het moet echter gezegd worden dat agenten in de praktijk wel enig idee van onze mogelijke aanwezigheid konden hebben. Het was immers aangekondigd – zowel intern in de politieorganisatie als in de openbare media – dat een onderzoeksgroep de politiepraktijken zou evalueren. Bovendien zijn we ook tweemaal ontdekt tijdens het observeren en werd ons verteld dat de agenten in een paar andere gevallen hadden vermoed dat ze werden geobserveerd (terwijl we in die gevallen juist niet aanwezig waren geweest). Onze opzet heeft het risico van reactiviteit dus niet helemaal uitgesloten, maar we hebben deze potentiële bias waarschijnlijk geminimaliseerd door onzekerheid te creëren over de vraag of de uitgevoerde controle wel of niet werd geobserveerd. Onze voorkeur zou zijn geweest dat de agenten volledig geblindeerd zouden zijn voor het project, en geobserveerd zouden zijn met veiligheidsgegevens van video-opnamen, zonder enige mogelijkheid van reactiviteit. Om praktische redenen was dit echter niet haalbaar (de politie vond bijvoorbeeld dat video-waarnemingen te veel inbreuk maakten in vergelijking tot observaties ter plaatse).

In totaal hebben we ter plaatse geobserveerd tijdens zes uitgevoerde controles: drie in Zuid, één in Noord, één in Zuid-Oost en één in Nieuw-West. We werkten in een team van 12 onderzoekers die instructie hadden gehad over het gebruik van het meetinstrument en we observeerden de controles per toerbeurt. Doorgaans werkten er vier onderzoekers tijdens één politiedienst. Zij observeerden gedurende ongeveer vier uur. Bij elke controle kregen we op de dag zelf van de verantwoordelijke politiecoördinator te horen waar de controle zou plaatsvinden. Wij vertelden deze politiecoördinator niet of we daadwerkelijk aanwezig zouden zijn. Verder had de politie van ons de instructie gekregen om de agenten van dienst niets te zeggen over onze mogelijke aanwezigheid. Nadat de onderzoekers waren aangekomen op de plaats van de controle observeerden zij eerst het normale gedrag op die plaats, om uit te vinden hoe zij zich tijdens hun observaties onopvallend tussen de mensen konden begeven. Omdat mensen in de publieke ruimte zelden lang hetzelfde doen, hadden de onderzoekers de instructie om hun activiteiten

regelmatig te veranderen om niet op te vallen. Typische activiteiten om niet op te vallen waren eten, koffiedrinken, telefoneren, wachten op openbaar vervoer, wachten met een koffertje en kletsen met een andere waarnemer. Verder moesten de waarnemers de observaties beëindigen op het moment dat de politie hen leek te herkennen.

Steekproeven

Tijdens de controles zou de politie moeten afwisselen tussen het controleren van voetgangers, fietsers en automobilisten. Vanwege de haalbaarheid, en om onze resultaten vergelijkbaar te maken met eerder onderzoek naar profilering (Jobard & Lévy, 2011), hebben we alleen controles van voetgangers geregistreerd. Daarnaast hebben de waarnemers resultaten vastgelegd met een mobiele telefoon, met gebruik van een online enquête-tool. De selectie van personen voor codering hing af van de controlemethoden die de politie gebruikte. Ten eerste, wanneer afzonderlijke burgers geselecteerd werden door een agent of door de random-selectiepaal, selecteerden wij willekeurig ongeveer elke derde persoon die een denkbeeldige lijn passeerde terwijl hij/zij naar het controlegebied liep. In situaties waar het niet zo druk was en dit praktisch haalbaar was, benaderden we alle aanwezigen.

Ten tweede veranderde deze procedure voor steekproefselectie in de loop van het onderzoek, omdat we merkten dat de politie soms alle aanwezigen in een bepaald gebied selecteerde voor controle – in plaats van afzonderlijke personen te controleren, zoals bij de gewone selecties en paal-selecties. Voor deze gebied-selecties pasten wij een casus-controle-benadering voor de steekproefselectie toe, waarbij we eerst zoveel mogelijk mensen registreerden binnen het gebied waar de politie iedereen onderzocht (de casussen). Vervolgens verzamelden wij een vergelijkingsgroep van mensen uit niet-onderzochte gebieden die direct rond het doorzochte gebied lagen (de controles). Deze procedure voor steekproefselectie volgde de aanbeveling dat vergelijkbare controlepersonen in dezelfde periode en ruimte als de casussen kunnen worden geselecteerd (Grimes & Schulz, 2005).

In de praktijk lukte het echter niet om een voldoende grote subset van controles te verzamelen – vaak wordt een verhouding van controles tot casussen van 4:1 aanbevolen. Om deze verhouding tussen controles en casussen toch te benaderen, hebben we besloten om één week nadat de politiecontrole had plaatsgevonden, controles te verzamelen uit het niet-onderzochte gebied. We moeten toegeven dat deze controles (dat wil zeggen de meeste daarvan) mogelijk minder goed vergelijkbaar zijn dan wanneer ze in dezelfde periode en ruimte als de politiecontroles waren verzameld (zo was bijvoorbeeld het weer in de week erop aanmerkelijk slechter en was het kouder). We hebben echter aangenomen dat deze controlepersonen niet verschillen van de personen die een week eerder werden geselecteerd, in termen van leeftijd, gender en etniciteit, en op die basis hebben we besloten om deze controlepersonen in de analyse te gebruiken. We hebben deze beslissing ook genomen omdat we zo een dataset konden bouwen met een betere statistische power, zelfs nadat de burgemeester had besloten om de pilot naar politiecontroles te beëindigen. Op deze manier konden wij met een gevoeligheidsanalyse van de dataset

van gebied-selecties van in totaal 393 observaties, zelfs een klein effect ($f^2 = 0,02$) vaststellen met een power van 80%.

Ter vergelijking hadden de datasets van de gewone selecties en vooral die van de waarnemingen bij paal-selecties een substantieel lagere power. We konden met name met de dataset van 185 gewone controles een klein-tot-medium effect ($f^2 = 0,04$) vaststellen, en met de dataset van paal-selecties van slechts 30 gevallen een medium-tot-groot effect ($f^2 = 0,28$). Bij het analyseren van de resultaten dient men in gedachte te houden dat de lagere power van de datasets het risico van fout-negatieven, en door ruis verstoorde schattingen in het algemeen, verhoogt.

Codering en maatstaven

Voor de systematische codering van bijvoorbeeld gender en etniciteit moeten duidelijke afbakeningen van deze sociale groepen worden geformuleerd. Waar mogelijk zijn wij uitgegaan van eerder onderzoek waarin vergelijkbare maatstaven in de praktijk werden getoetst. Dit konden we doen bij de definiëring van etniciteit, die vergelijkbaar was met de definiëring die werd gebruikt in het hierboven genoemde Parijse onderzoek (Jobard & Lévy, 2011). Daarnaast is uit eerder onderzoek gebleken dat in de praktijk leeftijd, gender en relatieverbanden nauwkeurig kunnen worden gecodeerd in openbare settings (bijv. Liebst *et al.*, 2022). Om deze maatstaven echter te kunnen generaliseren naar de context van dit onderzoek hebben we een intercoder-betrouwbaarheidstest voor de gebruikte maatstaven uitgevoerd. Dit deden we door de overeenstemming te beoordelen tussen twee onafhankelijke beoordelingen van dezelfde, willekeurig geselecteerde personen op een openbare plaats. De scores van de betrouwbaarheidstest tussen beoordelaars staan hieronder vermeld. We onderkennen dat onze betrouwbaarheidstests tussen beoordelaars geen garantie geven dat onze maatstaven geldig zijn vanuit het perspectief van de beoordeelde personen. Het is bijvoorbeeld mogelijk dat we systematisch verkeerde beoordelingen geven (de beoordelaars kunnen bijvoorbeeld de werkelijke leeftijd van oudere mensen systematisch te laag schatten). Ook kunnen onze beoordelingen afwijken van de identiteit die de personen zelf voelen.

Alle maatstaven waren gebaseerd op alle in die situatie beschikbare zichtbare factoren (bijvoorbeeld kleding, haardracht of huidskleur): Etniciteit werd vastgelegd in vijf categorieën: Wit, Arabisch, Zwart, Indo-Pakistaans en Aziatisch. Deze maatstaf had een ‘redelijke’ intercoder-betrouwbaarheid (Krippendorff’s alfa = 0,68) bij beoordeling met standaard grenswaarden voor betrouwbaarheid tussen beoordelaars (Fleiss, 1981). Wanneer echter opnieuw werd gecodeerd in een dichotoom systeem van wit of niet-wit, was de betrouwbaarheidsscore ‘goed’ (0,80). *Gender* werd geregistreerd als man of vrouw en had een ‘bijna perfecte’ score tussen beoordelaars (1,0). *Leeftijd* werd gemeten op een continue schaal en bereikte een ‘bijna perfecte’ score tussen beoordelaars (0,94). *Familierelaties* legde vast of de personen in gezelschap waren van jonge kinderen (0-12 jaar) en daarom naar verwachting niet gecontroleerd zouden worden. Deze maatstaf had een ‘goede’ score tussen beoordelaars (0,80).

RESULTATEN

In dit gedeelte beschrijven we onze kwantitatieve resultaten. In tabel 1 staat beschreven hoe de data werden verzameld bij de drie verschillende selectieprocedures: paal-selectie, gewone selectie en gebied-selectie. We verzamelden gegevens van 30 personen bij controles waarbij de politie gebruikmaakte van paal-selectie. Van deze personen werden er drie gecontroleerd op wapens. We verzamelden gegevens van 185 personen bij controles waarbij de politie een gewone selectieprocedure gebruikte. Van deze personen werden er 58 gecontroleerd. We verzamelden gegevens van 393 personen voor het analyseren van gebied-selectie, onder wie 81 personen die werden gecontroleerd¹.

Tijdens het gebruik van de random-selectiepaal was ongeveer een op de drie geobserveerde personen niet-wit, net iets minder dan de helft was man, en de gemiddelde leeftijd was 35 jaar. Van de personen die tijdens gewone selecties werden geobserveerd was iets meer dan 40% niet-wit, ongeveer de helft was man en de gemiddelde leeftijd was 30 jaar. Van de personen die tijdens gebied-selecties werden geobserveerd was ongeveer een kwart niet-wit, net iets meer dan de helft was man en de gemiddelde leeftijd was ongeveer 35 jaar. Daardoor varieerde de genderverdeling en de gemiddelde leeftijd slechts weinig over de drie selectiemethoden. Het grootste verschil tussen de drie steekproeven zat in de proportie niet-witte mensen die werden geobserveerd. Dit verschil is mogelijk toe te schrijven aan de variatie in de etnische samenstelling van de buurten waar de verschillende soorten zoekacties werden geobserveerd. De verschillende selectiemethoden werden in verschillende delen van de stad geobserveerd: de random-selectiepaal werd alleen in Amsterdam-Zuid geobserveerd. De gewone selectie werd in Amsterdam-Noord en Zuid-Oost geobserveerd. De gebied-selectie werd in Amsterdam-Nieuw-West en -Zuid geobserveerd. Hieruit volgt dat niet-witte personen nooit een even grote kans hadden om willekeurig geselecteerd te worden voor onze gegevensverzameling over de verschillende methoden, als de gebieden een andere etnische samenstelling hebben.

¹ Voor de hieronder weergegeven analyse werden families met kinderen jonger dan 12 jaar en personen van wie de leeftijd werd geschat op ouder dan 65 jaar uitgesloten. Beide groepen werden uitgesloten omdat de politie ze ook niet controleert op wapens.

Tabel 1. Beschrijvende statistiek van de mensen in de steekproeven tijdens de verschillende selectieprocedures

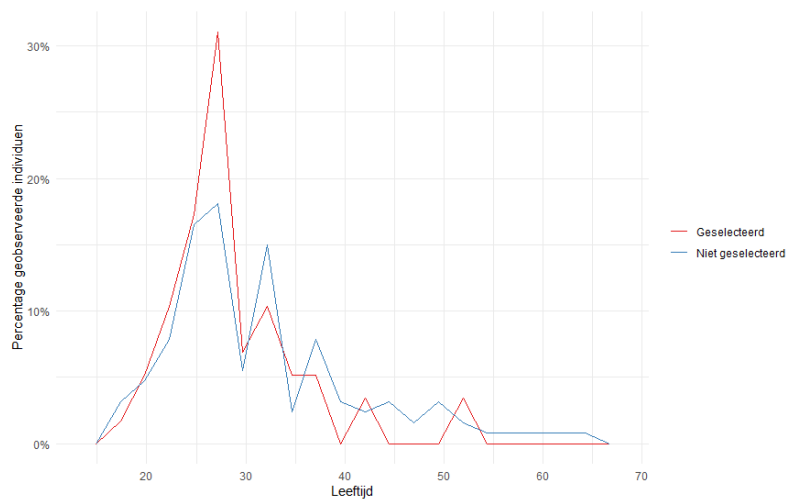
	Paal-selectie	Gewone selectie	Gebied-selectie
Totaal aantal mensen in de steekproef	30	185	393
Aantal personen in de steekproef die op wapens werden gecontroleerd	3	58	81
Niet-wit	30%	42,7%	26,7%
Mannen	43,3%	49,7%	54,2%
Gemiddelde leeftijd	35,0	30,6	35,4
Stadsdeel van Amsterdam	Zuid	Noord en Zuid-Oost	Nieuw-West en Zuid

Helaas konden we maar een klein aantal personen observeren bij de wapencontroles waarbij de politie de random-selectiepaal gebruikte ($n = 30$), daarom hebben we deze specifieke selectiemethode niet verder geanalyseerd. Voor de andere twee selectiemethoden hebben we onderzocht hoe de demografische kenmerken van etniciteit, gender en leeftijd geassocieerd waren met de waarschijnlijkheid om te worden geselecteerd voor een wapencontrole.

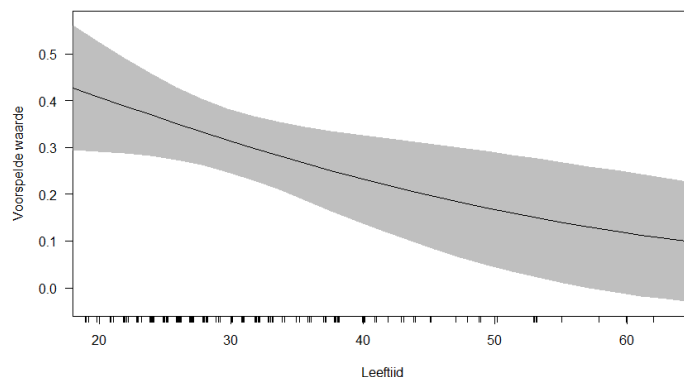
Gewone selectie

In figuur 1 staat de leeftijdsverdeling van de geselecteerde personen en van degenen die niet werden geselecteerd toen de politie de gewone selectieprocedure gebruikte. De figuur laat zien dat mensen van ongeveer 25 tot 30 jaar, in vergelijking met andere leeftijdsgroepen, een onevenredig grote kans hadden om te worden geselecteerd. De figuur laat ook zien dat personen ouder dan 35 jaar in de groep niet-gecontroleerde personen een grotere proportie vormen dan in de groep wel-gecontroleerde personen. Een regressiemodel gaf aan dat dit verband tussen leeftijd en waarschijnlijkheid van selectie statistisch duidelijk was, hoewel slechts marginaal ($OR = 0,96$; $p = 0,033$). Met dit voorbehoud in gedachte wijst de oddsratio (OR) erop dat iemand voor elk jaar dat hij of zij jonger is, een $1/0,96 = 1,04$ keer grotere kans heeft om te worden geselecteerd. Het verband wordt zichtbaar gemaakt in figuur 2, waarbij een lagere leeftijd een substantiële stijging voorspelt in de kans om geselecteerd te worden – personen rond de 20 jaar hadden een kans van ongeveer 40% om geselecteerd te worden, terwijl die kans terugvalt naar ongeveer 20% voor personen rond de 50 jaar. We willen benadrukken dat de resultaten van deze regressieanalyse geïnterpreteerd kunnen worden als statistisch niet duidelijk, als op de drempel van de p-waarde een Bonferroni-correctie naar 1% (ten opzichte van een traditionele grenswaarde van 5%) was toegepast vanwege een meervoudig vergelijkingsprobleem (omdat we dezelfde hypothese testen met meerdere tests). Opmerking: tenzij anders vermeld, werden alle gerapporteerde regressieanalyses gespecificeerd als bivariate Logit-modellen met robuuste standaardfouten.

Figuur 1. Leeftijdsverdeling van mensen die wel en niet werden geselecteerd via gewone selectie

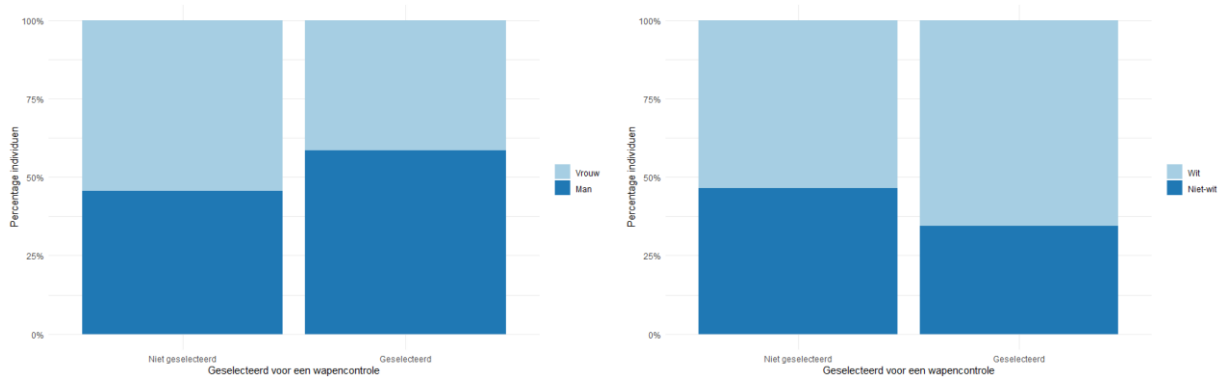


Figuur 2. Het geschatte verband tussen leeftijd en de voorspelde waarschijnlijkheid om geselecteerd te worden tijdens gewone zoekacties, waarbij het grijze gebied het 95%-betrouwbaarheidsinterval aangeeft



Figuur 3 laat de verdeling zien tussen mannen en vrouwen die geselecteerd werden en tussen witte en niet-witte personen die geselecteerd werden voor een wapencontrole. In het linkerpaneel van figuur 3 wordt het percentage mensen weergegeven die tijdens de controles werden geselecteerd tijdens gewone zoekacties. Deze grafiek laat zien dat mannen in de geselecteerde groep een grotere proportie uitmaken dan in de niet-geselecteerde groep. Dit patroon was echter statistisch niet duidelijk ($OR = 0,59$; $p = 0,110$). In het rechterpaneel van figuur 3 is te zien dat de proportie witte mensen die tijdens de gewone selectie werden geselecteerd, onevenredig groot lijkt. Dit patroon was echter ook niet statistisch duidelijk ($OR = 0,61$; $p = 0,130$).

Figuur 3. Het linker- en rechterpaneel geven respectievelijk gender en etniciteit weer van mensen die wel en niet geselecteerd werden via gewone selectie

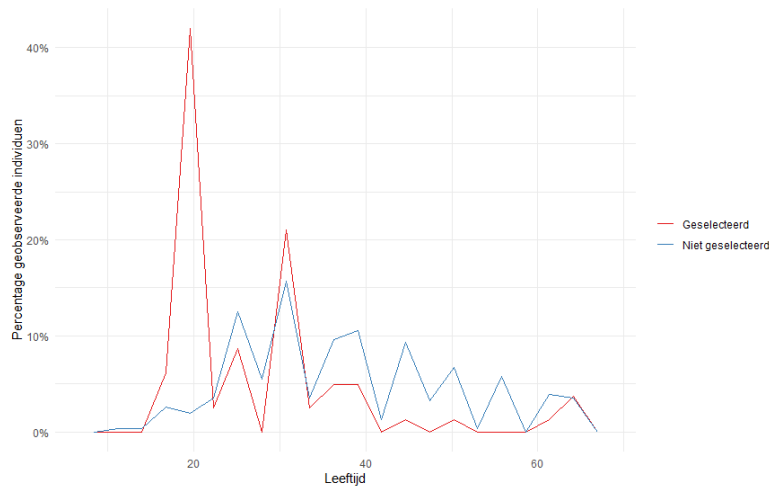


Vervolgens voerden we een multivariate analyse uit om te zien of de bovenstaande bevindingen hetzelfde bleven wanneer we alle drie de demografische voorspellers (leeftijd, gender en etniciteit) tegelijkertijd in één model gebruikten. Alleen de leeftijd liet enig verband zien met de waarschijnlijkheid om te worden geselecteerd voor een wapencontrole ($OR = 0,96$; $p = 0,020$), hoewel het verband wederom alleen statistisch duidelijk was op 5%-niveau en niet op 1%-niveau.

Gebied-selectie

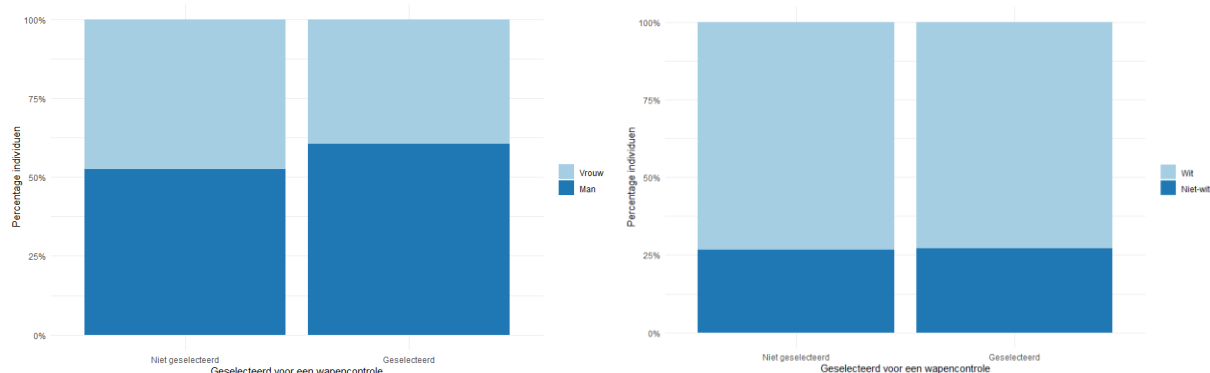
Figuur 4 laat de leeftijdsverdeling zien van de personen die werden gecontroleerd op wapens en van degenen die niet werden geselecteerd tijdens de gebied-selectieprocedure. In deze figuur is te zien dat een onevenredig grote proportie personen van rond de 20 jaar werd geselecteerd. Daarnaast was een grotere proportie van de personen die niet werden gecontroleerd ouder dan 35 jaar, ten opzichte van degenen die wel werden gecontroleerd. Dit verband tussen leeftijd en waarschijnlijkheid van selectie was statistisch zeer duidelijk ($OR = 0,91$; $p < 0,001$). Deze oddsratio wijst erop dat iemand voor elk jaar dat hij of zij jonger is, een $1/0,91 = 1,10$ keer grotere kans heeft om te worden geselecteerd. Dit patroon was vergelijkbaar, maar duidelijker, bij vergelijking met het leeftijdspatroon dat werd waargenomen tijdens de gewone selectieprocedure.

Figuur 4. Leeftijdsverdeling van mensen die wel en niet werden geselecteerd via gebied-selectie



Figuur 5 laat de verdeling zien van gender en etniciteit van de personen die werden gecontroleerd en van degenen die niet werden gecontroleerd tijdens gebied-selectie. In het linkerpaneel van figuur 5 worden de proporties mannen en vrouwen weergegeven in de groep die werd gecontroleerd en in de groep die niet werd gecontroleerd. Het blijkt dat meer mannen dan vrouwen werden geselecteerd voor wapencontroles, maar dit verschil was niet statistisch duidelijk ($OR = 0,72$; $p = 0,204$). In het rechterpaneel van figuur 5 worden de proporties witte en niet-witte personen weergegeven die wel en niet werden gecontroleerd. Er was geen visueel of statistisch duidelijk ($OR = 1,03$; $p = 0,920$) verschil tussen de proporties geselecteerde personen in deze twee groepen.

Figuur 5. Het linker- en rechterpaneel geven respectievelijk gender en etniciteit weer van mensen die wel en niet geselecteerd werden via gebied-selectie



Om de robuustheid van bovenstaande bevindingen te controleren, hebben we een multivariate logistische regressieanalyse uitgevoerd met zowel leeftijd, gender als etniciteit. Ook hier werd gevonden dat de leeftijd een statistisch zeer duidelijk verband had met de

waarschijnlijkheid van selectie ($OR = 0,91$; $p < 0,001$). Bovendien gaf deze analyse van robuustheid aan dat mannen een grotere waarschijnlijkheid hadden om te worden geselecteerd dan vrouwen ($OR = 0,55$; $p = 0,036$), hoewel deze schatting slechts marginaal statistisch duidelijk was bij een 5%-drempel voor de p-waarde, en niet duidelijk was als op de drempel van de p-waarde een Bonferroni-correctie werd toegepast. Merk op dat dit laatste resultaat verschilde van de bivariate analyse en van de bovenstaande visualisatie van de data. Ten slotte stellen we vast dat de bovenstaande casus-controle-resultaten hetzelfde bleven als de niet-geautoriseerde controles werden uitgesloten van de steekproef.

CONCLUSIE

In deze analyse hebben we onderzocht of en waarom er tekenen zijn die wijzen op een selectie-bias op basis van etniciteit, leeftijd of gender tijdens de wapencontroles. De data gaven geen robuuste aanwijzingen dat gender of etniciteit van burgers geassocieerd was met een onevenredig grote kans om geselecteerd te worden voor een wapencontrole. In vergelijking daarmee boden de data meer aanwijzingen dat jongere personen, met name jongeren van ongeveer 25 tot 30 jaar, een onevenredig grote kans hadden om geselecteerd te worden. Dit resultaat is overtuigend, omdat het in twee afzonderlijke steekproeven werd gerepliceerd: tijdens gewone selecties en tijdens gebied-selecties. Deze replicatie verzacht het niet-definitieve resultaat in de data-analyse voor de gewone selectie: een net wel of net niet statistisch duidelijke (data van gewone selectie) en een statistisch zeer duidelijke (data van gebied-selectie) schatting die in dezelfde richting wijzen, zouden over het algemeen eerder moeten worden geïnterpreteerd als convergerende dan als conflicterende aanwijzingen. Het verschil in statistische duidelijkheid tussen de twee datasets zou tot op zekere hoogte kunnen worden veroorzaakt door het relatief kleine aantal mensen dat werd geobserveerd in de controles met gewone selectie, vergeleken met gebied-selectie (minder observaties betekent een lagere statistische power om werkelijke effecten vast te stellen).

Er moet echter worden opgemerkt dat de data het risico niet uitsluiten dat etnische profilering zou kunnen plaatsvinden op het niveau van de te controleren buurten, wanneer de politie bepaalt waar de wapencontroles zullen worden uitgevoerd. Hiervan kan sprake zijn als de politie zich richt op buurten met relatief veel niet-witte mensen op straat. Dit zou op zijn beurt de telling van niet-witte personen die worden geselecteerd kunstmatig verhogen, zelfs als niet-witte en witte mensen dezelfde waarschijnlijkheid hebben om geselecteerd te worden. We willen ook benadrukken dat het, hoewel de analyse van de data bepaalde trends liet zien, niet zeker is dat deze patronen gegeneraliseerd kunnen worden naar andere contexten. Wij konden in totaal slechts zes wapencontroles observeren voordat de controles werden stopgezet. Daarom zijn alle drie de selectiemethoden gebaseerd op observaties gedurende slechts een paar dagen. Het is dus niet duidelijk of de bevindingen iets zeggen over de specifieke dagen dat we hebben geobserveerd, of over de algemene praktijken van de politie. Ten slotte willen we benadrukken dat het feit dat er geen aanwijzingen werden gevonden voor profilering op basis van etnische afkomst of gender, kan zijn veroorzaakt door te weinig en te vertroebelde data om een dergelijke mogelijke bias vast te stellen.

LITERATUUR

- American Sociological Association. (1999). *American Sociological Association code of ethics*. <https://www.asanet.org/sites/default/files/savvy/images/asa/docs/pdf/CodeofEthics.pdf>
- Bradford, B., & Loader, I. (2016). Police, Crime and Order: The Case of Stop and Search. In B. Bradford, B. Jauregui, I. Loader, & J. Steinberg (Eds.), *The SAGE Handbook of Global Policing* (pp. 241–260).
- Dennison, C. R., & Finkeldey, J. G. (2021). Self-reported experiences and consequences of unfair treatment by police. *Criminology*, 59(2), 254–290. <https://doi.org/10.1111/1745-9125.12269>
- Farrell, A., & McDevitt, J. (2010). Identifying and Measuring Racial Profiling by the Police. *Sociology Compass*, 4(1), 77–88.
- Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. Wiley.
- Grimes, D. A., & Schulz, K. F. (2005). Compared to what? Finding controls for case-control studies. *The Lancet*, 365(9468), 1429–1433.
- Hesseling, N., & Wilde, S. de. (2022). Directie Openbare Orde en Veiligheid. <https://openresearch.amsterdam/nl/page/79166/onderzoek-pilot-wapencontroles>
- Holmberg, L. (2003). *Policing stereotypes: A qualitative study of police work in Denmark*. Galda + Wilch.
- Jobard, F., & Lévy, R. (2011). Racial profiling: The Parisian police experience. *Canadian Journal of Criminology and Criminal Justice*, 53(1), 87–93.
- Liebst, L. S., Baggesen, L., Dausel, K. L., & Lindegaard, M. R. (2022). *Human observers have a high accuracy in age estimating strangers in public settings*. PsyArXiv. <https://doi.org/10.31234/osf.io/zbpkw>



Het NSCR is
onderdeel van de
institutenorganisatie
van de Nederlandse
Organisatie voor
Wetenschappelijk
Onderzoek (NWO)

Bezoekadres:
De Boelelaan 1077
1081 HV Amsterdam

Postadres:
Postbus 71304
1008 BH Amsterdam

T 020 598 5239
E nscr@nscr.nl
W www.nscr.nl

nscr

Nederlands Studiecentrum
Criminaliteit en Rechtshandhaving