



# Vooringenomenheid voorkomen

## Addendum

## Inhoudsopgave

<b>1. Managementsamenvatting</b>	<b>3</b>
1.1 Aanleiding en opzet van onderzoek	3
1.2 Bevindingen	6
<b>2. Overzicht CUB-onderzoek</b>	<b>10</b>
2.1 Stappen in het CUB-proces waarvoor data zijn verkregen	10
2.2 Recente chronologie CUB-onderzoek	11
2.3 Andere onderzoeken naar het CUB-proces	12
<b>3. Onderzoeksmethodologie</b>	<b>13</b>
3.1 Doel van de analyse	13
3.2 Onderzoekspopulaties	13
3.3 Data verstrekt door het CBS	15
3.4 Onderzoeksvragen biasmeting met betrekking tot migratieachtergrond	18
<b>4. Resultaten</b>	<b>19</b>
4.1 Resultaten onderzoeksvraag 1 (startpopulatie)	19
4.2 Resultaten onderzoeksvraag 2 (risicoprofiel)	23
4.3 Resultaten onderzoeksvraag 2a (proxy-analyse)	28
4.4 Resultaten onderzoeksvraag 3 (handmatige selectie)	36
4.5 Resultaten onderzoeksvraag 3a (mate waarin voorspelling algoritme wordt gevolgd)	38
4.6 Resultaten onderzoeksvraag 4 (onrechtmatig gebruik)	43
4.7 Resultaten onderzoeksvraag 5 (bezwaarpopulaties)	46
4.8 Overzicht van het gehele CUB-proces	51
<b>5. Voorbehouden</b>	<b>53</b>
<b>6. Conclusie</b>	<b>54</b>
<b>Appendix A – Unsupervised bias detectie</b>	<b>57</b>

## Over Algorithm Audit

Algorithm Audit is een Europees kennisplatform voor AI bias testing en normatieve AI-standaarden. De doelen van de stichting zijn drieledig:



### Normatieve adviescommissies

Adviseren over ethische kwesties in concrete algoritmische toepassingen door het samenbrengen van deliberatieve, diverse adviescommissies, met [algorudentie](#) als resultaat



### Technische hulpmiddelen

Implementeren en testen van technische tools om bias in AI systemen te detecteren en mitigeren, zoals onze [bias detectie tool](#) en [synthetische data generatie](#)



### Kennisplatform

Samenbrengen van kennis en experts om collectief leerproces over de verantwoorde inzet van algoritmes aan te jagen, zie ons [AI beleidsobservatorium](#) en [white papers](#)

## 1. Managementsamenvatting

Hieronder volgen de aanleiding en opzet van dit onderzoek (§1.1) en de bevindingen (§1.2)

### 1.1 Aanleiding en opzet van onderzoek

Dit rapport is een addendum op het rapport Vooringenomenheid voorkomen dat op 1 maart 2024 aan de Tweede Kamer is gestuurd.<sup>1</sup> In dat rapport is het onafhankelijke externe onderzoek gepresenteerd dat DUO heeft laten uitvoeren naar aanleiding van berichten in de media over oververtegenwoordiging van studenten met een migratieachtergrond in het zogeheten controle uitwonendenbeurs (CUB) proces. Op het moment van publiceren van het rapport Vooringenomenheid voorkomen was er nog geen verrijkte data van het Centraal Bureau voor de Statistiek (CBS) beschikbaar over de herkomst van gecontroleerde studenten. Op 6 mei 2024 heeft het CBS op groepsniveau de herkomst van studenten in verschillende stappen van het CUB-proces in de jaren 2014, 2017, 2019, 2021 en 2022 aan DUO en Algorithm Audit verstrekt. In dit addendum staat de analyse van de CBS-data. Dit addendum moet worden begrepen in samenhang met het rapport Vooringenomenheid voorkomen.

Hieronder volgen een samenvatting van de opzet en de resultaten van het onderzoek.

#### 1.1.1 Opzet van de biasmeting

Dit rapport presenteert de uitkomsten van een biasmeting van het CUB-proces. De biasmeting is gebaseerd op door het CBS verstrekte geaggregeerde data over de herkomst van meer dan 300.000 ontvangers van de uitwonendenbeurs in 2014, 2017, 2019, 2021 en 2022.<sup>2</sup> Door de verdeling van studenten met een migratieachtergrond te meten in verschillende stappen van het CUB-proces kan worden onderzocht of, en zo ja, waar en in welke mate vooringenomenheid aanwezig is in de verschillende stappen van het CUB-proces. In [2.1 Stappen in het CUB-proces](#) wordt iedere stap die relevant is voor de biasmeting in detail beschreven.

Dit rapport presenteert ook het onderzoek naar de vraag of de criteria die gebruikt worden in het risicoprofiel een proxykenmerk vormen voor studenten met een migratieachtergrond. Dit gaat om de criteria onderwijsvorm, leeftijd en afstand tot het ouderlijk adres en de combinaties van deze criteria. Deze proxy-analyse kan verklaren hoe het risicomodel vooringenomen is ten aanzien van studenten met een migratieachtergrond, ook al is dit kenmerk zelf geen criterium in het profiel (zie [Box 1](#)).

<sup>1</sup> Intern onderzoek controle uitwonendenbeurs (rapport DUO), bijlage bij Kamerstukken II 2023/24, 24724, nr. 220.

<sup>2</sup> <https://www.cbs.nl/nl-nl/maatwerk/2024/21/ontvangers-uitwonendenbeurs-herkomst-2014-2017-2019-2021-en-2022>

#### Box 1

### Wat zijn proxykenmerken?

Met proxykenmerken worden ogenschijnlijk neutrale gegevens bedoeld die sterk gecorreleerd zijn met gevoelige persoonskenmerken of een beschermde grond uit het non-discriminatierecht, zoals etniciteit of religie. Door een sterke correlatie kan men door proxykenmerken te gebruiken (mogelijk onbedoeld) ook de gevoelige groep raken. In de context van het CUB-proces is relevant in hoeverre de kenmerken van het risicoprofiel (I. onderwijsvorm, II. leeftijd en III. afstand tot ouder(s)) en/of de verschillende mogelijke combinaties van deze criteria, een proxykenmerk zijn voor migratieachtergrond.

Dit rapport beschrijft daarnaast in hoeverre de stappen in het CUB-proces na toekenning van een risicoscore door het risicoprofiel bijdragen aan de vooringenomenheid ten aanzien van studenten met een migratieachtergrond. Zo is onderzocht wat de invloed is van de handmatige selectie van studenten voor controle. Ook wordt geanalyseerd of bij studenten met een migratieachtergrond vaker een huisbezoek heeft plaatsgevonden waaruit geen onrechtmatig gebruik van de uitwonendenbeurs bleek. Ten slotte wordt de mate van oververtegenwoordiging van studenten met een migratieachtergrond gemeten voor de groep studenten waarbij DUO onrechtmatig gebruik heeft vastgesteld, voor de groep die vervolgens bij DUO in bezwaar ging, en voor de groep waarbij dit bezwaar gegrond is verklaard.

Afhankelijk van de data die voor de individuele jaren beschikbaar zijn, worden bepaalde of al deze vragen beantwoord voor de jaren 2014, 2019, 2021 en 2022. In [3.2 Onderzoekspopulaties](#) wordt de samenstelling van alle onderzochte studentenpopulaties beschreven. In [3.4 Biasmeting](#) worden concrete onderzoeksvragen geïntroduceerd.

### 1.1.2 Invulling van de term migratieachtergrond

De in dit rapport gebruikte term migratieachtergrond hanteert het CBS zelf niet. In plaats daarvan onderscheidt het CBS studenten die zelf buiten Nederland zijn geboren (migranten) of waarvan een van de ouders buiten Nederland is geboren (kinderen van een migrant) van studenten met een Nederlandse herkomst. Binnen de twee migrantencategorieën onderscheidt het CBS verder Europese en niet-Europese migranten.

Ten behoeve van de duidelijkheid en de aansluiting op de onderzoeksvraag wordt in dit rapport de term migratieachtergrond gehanteerd en vormt daarmee een licht gewijzigde categorisering ten opzichte van het CBS. [Figuur 1](#) geeft een overzicht van de (sub)categorieën die in dit rapport worden gehanteerd. Waar wordt gesproken over studenten met een migratieachtergrond gaat het over studenten die zelf buiten Nederland zijn geboren of waarvan één van de ouders buiten Nederland is geboren. Waar dit voor de analyse nuttig is wordt de categorie migratieachtergrond uitgesplitst in Europese en niet-Europese migratieachtergrond. Studenten die niet tot een van deze groepen behoren worden aangeduid als studenten met een Nederlandse herkomst.



Figuur 1 – Gebruikte onderverdeling categorieën studenten

De groepen migratieachtergrond en Nederlandse herkomst vormen samen de gehele populatie van ontvangers van de uitwonendenbeurs.<sup>3</sup> Een uitgebreide toelichting en verantwoording van deze categorisering, met vermelding van de beperkingen ervan, wordt gegeven in [3.3 Data verstrekt door het CBS](#).

### 1.1.3 CBS-data

Door data uit de verschillende jaren van het CUB-proces met elkaar te vergelijken kan worden vastgesteld of eventuele vooringenomenheid eenmalig was of onderdeel van een structurele trend. Bij het CBS zijn aggregatiestatistieken over de migratieachtergrond van meer dan 300.000 uitwonende studenten opgevraagd die in 2014<sup>4</sup> en 2019<sup>5</sup> onderdeel zijn geweest van het CUB-proces, die in 2014 en 2017 zijn geselecteerd voor een aselechte steekproef of die in 2014, 2019, 2021 en 2022 bezwaar hebben gemaakt tegen een besluit van DUO dat zij onrechtmatig gebruik hebben gemaakt van de uitwonendenbeurs. Aggregatiestatistieken zijn niet herleidbaar naar individuen. Deze statistieken geven enkel kenmerken van groepen groter dan 10 personen weer. Tenzij anders vermeld corresponderen de in deze samenvatting gepresenteerde resultaten met de data uit 2014, omdat dit de meest representatieve dataset is voor wo-, hbo- en mbo-studenten. Algorithm Audit ziet geen bezwaar tegen het gebruik van tien jaar oude data omdat dit onderzoek de werking van het CUB-proces in het verleden onderzoekt. De geanalyseerde onderzoekspopulaties worden in meer detail beschreven in [3. Onderzoeksmethodologie](#).

<sup>3</sup> Studenten die vanuit het buitenland naar Nederland komen om te studeren ontvangen geen uitwonendenbeurs en maken dus geen deel van uit van de populatie die hier wordt onderzocht.

<sup>4</sup> 2014 was het laatste jaar voor introductie van het leenstelsel, de populatie bestaat uit zowel mbo-, hbo- als wo-studenten. Dit is relevant omdat vanaf studiejaar 2023-24 de basisbeurs voor alle studenten geherintroduceerd is.

<sup>5</sup> 2019 was het laatste jaar voordat het gangbare CUB-proces werd verstoord door de coronapandemie. Door de pandemie konden in 2020-2022 ca. 33% minder huisbezoeken worden uitgevoerd. Merk op dat in 2019 de uitwonendenbeurspopulatie voornamelijk bestaat uit mbo-studenten.

#### Box 2

## Verband met biasmeting rapport Vooringenomenheid voorkomen

In de biasmeting beschreven in *Voringenomenheid voorkomen* is gemeten of studenten in bepaalde onderwijs-, leeftijds- of afstandscategorieën vaker worden geselecteerd voor een huisbezoek dan men op basis van de door het risicoprofiel toegekende risicoscore kan verwachten. In de biasmeting die in dit addendum wordt gepresenteerd wordt ingegaan op vooringenomenheid ten opzichte van studenten met een migratieachtergrond. Zowel met betrekking tot het risicoprofiel als in de latere stappen van het CUB-proces wordt gemeten of studenten met een migratieachtergrond vaker worden geselecteerd voor controle dan studenten zonder migratieachtergrond. Deze biasmeting verdiept de meting in het eerste rapport omdat het één- en multidimensionale proxykarakter van de drie profileringscategorieën ten opzichte van migratieachtergrond wordt onderzocht.

## Box 3

## Vooringenomenheid

In dit rapport wordt met vooringenomenheid het volgende bedoeld: significante afwijkingen in de demografische verhoudingen ten opzichte van de bronpopulatie (de populatie van alle ontvangers van een uitwonendenbeurs) die ontstaan door onevenredige selectie van bepaalde studenten in de loop van het CUB-proces. Als er in verschillende stappen van het CUB-proces een aanzienlijke oververtegenwoordiging van studenten met een migratieachtergrond wordt gemeten, geeft dit aanleiding tot het oordeel dat het CUB-proces (en/of de verschillende stappen daarin) vooringenomen is ten aanzien van deze demografie. Vooringenomenheid impliceert hier geen bewuste vooringenomenheid, d.w.z. vooropgezette en doelbewuste selectie van een bepaalde demografie door DUO of individuele medewerkers. In dit rapport duidt vooringenomenheid op gemeten onevenredigheden in de data als (onbedoeld) effect van CUB-processtappen (wat ook wel *bias* wordt genoemd). Er is geen sprake van vooringenomenheid als de demografische verhoudingen nagenoeg gelijk blijven door de loop van het CUB-proces heen.

## 1.2 Bevindingen

In deze sectie worden de belangrijkste bevindingen van de biasmeting gepresenteerd.

### **Bevinding 1 – Het CUB-proces is als geheel vooringenomen geweest ten aanzien van studenten met een niet-Europese migratieachtergrond**

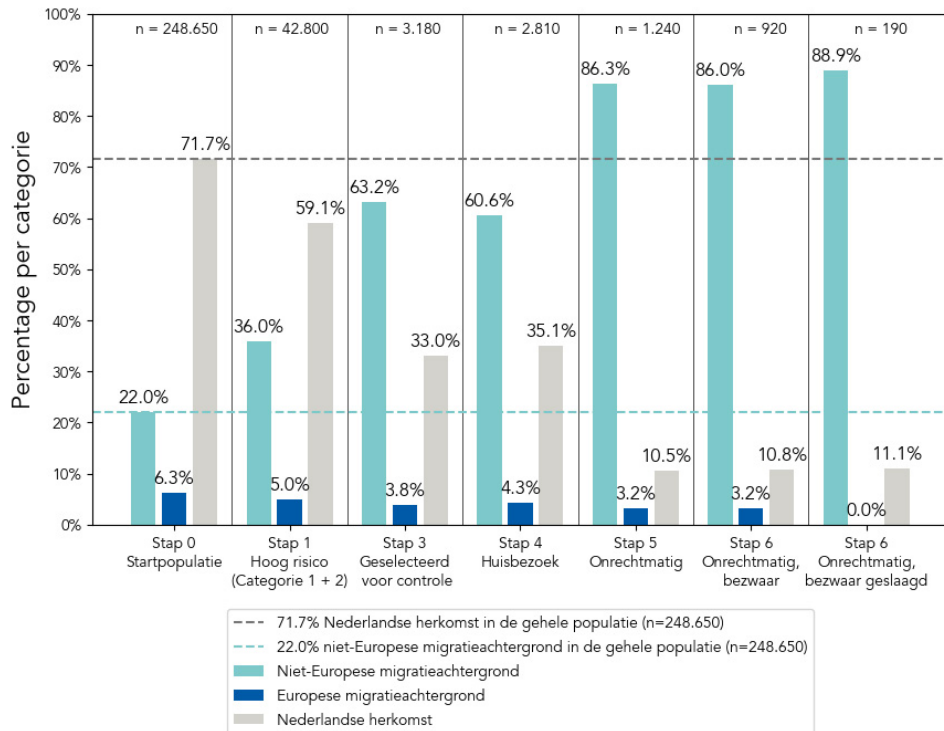
In het CUB-proces bestond een sterke vooringenomenheid ten aanzien van studenten met een niet-Europese migratieachtergrond. Studenten met een niet-Europese migratieachtergrond werden 2.0x vaker bestempeld als hoog risico door het risicoprofiel dan studenten met een Nederlandse herkomst. De groep werd 6.2x vaker handmatig geselecteerd voor een huisbezoek. Uiteindelijk hadden studenten met een niet-Europese migratieachtergrond een 3.0x grotere kans om onterecht een huisbezoek te krijgen dan studenten met een Nederlandse herkomst.

Rekening houdend met de wezenlijk verschillende populaties in de verschillende jaren (in belangrijke mate als gevolg van de invoering van het leenstelsel<sup>6</sup>), is er sprake van een duidelijke structurele trend in de vooringenomenheid ten aanzien van studenten met een niet-Europese migratieachtergrond.

[Figuur 2](#) geeft een overzicht van de verdeling studenten van verschillende herkomst per stap van het CUB-proces. [Figuur 2](#) wordt nader toegelicht in onderstaande bevindingen en in [4. Resultaten](#).

<sup>6</sup> In de periode 2015-2023 hadden alleen studenten die aan het mbo gingen studeren recht op een uitwonendenbeurs. Meer details kunnen worden gevonden in [3.2 Onderzoekspopulaties](#).

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per stap van het CUB-proces in de uitwonendenbeurspopulatie-2014 (n=248.650)



Figuur 2 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per stap van het CUB-proces in de uitwonendenbeurspopulatie-2014 (n=248.650)<sup>7</sup>

<sup>7</sup> Omdat de aggregatiestatistieken zoals samengesteld door het CBS afgerond worden op tientallen kan het voorkomen dat percentages niet exact tot 100% opsommen. Hetzelfde geldt voor de som van aantallen.

## **Bevinding 2 – Het risicoprofiel dat in het CUB-proces werd gebruikt was vooringenomen ten aanzien van studenten met een niet-Europese migratieachtergrond. Het toekennen van een hoger risico aan mbo-studenten en studenten die dicht bij het ouderlijk adres stonden ingeschreven was hier de oorzaak van.**

In het gehanteerde risicoprofiel (dat gebruik maakt van de kenmerken onderwijsvorm, leeftijd en afstand tussen het ingeschreven adres en het ouderlijk adres) ontstond een vooringenomenheid ten aanzien van migratieachtergrond. Studenten met een niet-Europese migratieachtergrond werden 2.0x vaker bestempeld als hoog risico door het risicoprofiel dan studenten met een Nederlandse herkomst. Dit is te herleiden naar het proxykarakter van met name onderwijsvorm en afstand. De kenmerken mbo 1-2 en een kleine afstand tot ouder(s) correleren sterk met de groep studenten met een niet-Europese migratieachtergrond. Door te profileren op deze kenmerken ontstaat vooringenomenheid ten aanzien van deze demografie. Een *unsupervised bias detectiemethode* bevestigt deze resultaten.

De proxy-analyse wordt nader toegelicht in [4.2 Resultaten onderzoeksvraag 2 \(risicoprofiel\)](#) en [4.3 Resultaten onderzoeksvraag 2a \(proxy-analyse\)](#). De resultaten van de *unsupervised bias detectiemethode* worden toegelicht in [Appendix A](#).

## **Bevinding 3 – Handmatige selectie versterkte de vooringenomenheid van het CUB-proces.**

Na indeling in risicocategorieën door het risicoprofiel vond handmatige selectie van studenten plaats voor controle en huisbezoek. Bij die handmatige selectie speelden schriftelijk onderzoek, werkinstructies, bepaalde uitsluitingsgronden en andere procedurele aspecten een rol. De vooringenomenheid die door het risicoprofiel ontstaat wordt in deze stap versterkt. Er zijn uit het onderzoek geen aanwijzingen naar voren gekomen dat deze vooringenomenheid door menselijke vooringenomenheid van individuele ambtenaren komt. Vooringenomenheid kan ook voortkomen uit de aard van de werkinstructies, zoals de uitsluiting van studentenhuizen van huisbezoek en zorginstellingen. Selectiegronden binnen de geselecteerde groep om wel een huisbezoek af te leggen waren onder meer de verhouding tussen oppervlakte en het aantal bewoners op het adres, inschrijving bij familie en de afstand van de inschrijving tot het adres van de ouders. De vraag hoe de vooringenomenheid in de handmatige selectie ontstond, valt buiten de reikwijdte van dit onderzoek.

Studenten met een niet-Europese migratieachtergrond die als hoog risico zijn geclassificeerd door het risicoprofiel hadden 5.5x meer kans om geselecteerd te worden voor controle dan studenten met Nederlandse herkomst en dezelfde risicocategorie. Studenten met een niet-Europese migratieachtergrond die als laag risico zijn geclassificeerd door het risicoprofiel hadden 1.8x meer kans om alsnog geselecteerd te worden voor controle dan studenten met een Nederlandse herkomst en dezelfde risicocategorie. Hierbij moet worden meegenomen dat de toegekende risicoscores zelf al vooringenomen zijn.

De analyse van de handmatige selectie van studenten wordt nader toegelicht in [4.4 Resultaten onderzoeksvraag 3 \(handmatige selectie\)](#) en [4.5 Resultaten onderzoeksvraag 3a \(mate waarin voorspelling algoritme wordt gevolgd\)](#).



**Bevinding 4 – Omdat het CUB-proces vooringenomen is geweest ten aanzien van studenten met een niet-Europese migratieachtergrond, is bij deze groep veel onrechtmatig gebruik vastgesteld. Dit is grotendeels het gevolg van overmatige controle van deze demografie.**

Door de vooringenomenheid van het gehele CUB-proces zijn studenten met een niet-Europese migratieachtergrond relatief vaker geselecteerd voor een huisbezoek dan studenten met een Nederlandse herkomst. Dit is in het geval van deze groep ook vaker een onterechte selectie: studenten met een niet-Europese migratieachtergrond hadden 3.0x meer kans op een huisbezoek waaruit later geen onrechtmatig gebruik bleek dan studenten met een Nederlandse herkomst.

Ook in de groep studenten waarbij onrechtmatig gebruik uiteindelijk wordt vastgesteld is het aandeel studenten met een niet-Europese migratieachtergrond groot. Er ontstaat door de stappen van het CUB-proces heen een vergrootglaseffect met betrekking tot de groep met een niet-Europese migratieachtergrond. Studenten met een Nederlandse herkomst werden relatief minder vaak gecontroleerd en onrechtmatigheid wordt in deze groep minder vaak opgespoord.

Of studenten met een migratieachtergrond ook vaker onrechtmatig gebruik maken van de uitwonendenbeurs kan niet worden vastgesteld op basis van de beschikbare data. Dat komt doordat deze verhouding niet kan worden geïsoleerd van de vooringenomenheid van het CUB-proces, en omdat de aselecte steekproef te klein is om dit onafhankelijk van het CUB-proces te meten. Of er naast het genoemde vergrootglaseffect verdere vooringenomenheid schuilt in het proces van de huisbezoeken zelf kon eveneens niet worden onderzocht.

De oververtegenwoordiging van deze groep studenten in de onrechtmatig-gebruik-populatie wordt nader toegelicht in [4.6 Resultaten onderzoeksvraag 4 \(onrechtmatig gebruik\)](#).

**Bevinding 5 – De groep studenten die in bezwaar gaat tegen vaststelling van onrechtmatig gebruik bestaat voor een overgroot deel uit studenten met een niet-Europese migratieachtergrond. Er is geen vooringenomenheid vastgesteld in de bezwaarprocedure zelf.**

Uit de verschillende peiljaren (2014, 2019, 2021 en 2022) treedt een consistent beeld naar voren dat de bezwaarpopulatie voor 79-85% uit studenten met een niet-Europese migratieachtergrond bestaat. Dit bevestigt globaal het beeld dat uit journalistiek onderzoek naar voren kwam omtrent de sterke oververtegenwoordiging van studenten met een migratieachtergrond die bezwaar maken tegen een besluit van DUO.<sup>8</sup>

Aan de andere kant wordt in de bezwaarstap in het CUB-proces geen vooringenomenheid vastgesteld. Bezwaar wordt even vaak gegrond verklaard voor alle studenten ongeacht hun herkomst.

De oververtegenwoordiging van studenten met een niet-Europese migratieachtergrond in de bezwaarpopulaties en gelijke behandeling tijdens de bezwaarprocedure wordt nader toegelicht in [4.7 Resultaten onderzoeksvraag 5 \(bezwaarpopulaties\)](#).

<sup>8</sup> B. Belleman, B. Heilbron & A. Kootstra. De discriminerende fraudecontroles van DUO. Investico Onderzoeksjournalisten, 2023

## 2. Overzicht CUB-onderzoek

Hieronder volgt een overzicht van het CUB-proces waarvoor data zijn verkregen (§2.1), een overzicht van stappen die zijn ondernomen na publicatie van eerdere onderzoeken naar het CUB-proces (§2.2) en een overzicht van deze onderzoeken (§2.3).

### 2.1 Stappen in het CUB-proces waarvoor data zijn verkregen

In dit addendum is data verkregen over vijf stappen in het CUB-proces. In het rapport Vooringenomenheid voorkomen is het CUB-proces in zeven stappen opgedeeld. Van vijf van deze stappen is data verkregen; van de twee overige stappen was het niet relevant om data op te vragen.<sup>9</sup>

Stappen uit het CUB-proces die relevant zijn voor het onderzoek:

- > **Bronpopulatie:** Alle ontvangers van een uitwonendenbeurs voor een gegeven peildatum. Dit zijn alle studenten die ingeschreven staan als student aan het mbo, hbo of wo en ingeschreven staan op een adres dat niet het adres van hun ouder(s) is. Het is mogelijk dat later duidelijk wordt dat de student geen recht had op de uitwonendenbeurs. Naar deze groep wordt gerefereerd als de *uitwonendenbeurspopulatie*.
- > **Risicoprofiel:** Toekennen van een risicoscore tussen de 0-180 aan alle studenten in de uitwonendenbeurspopulatie aan de hand van een risicoprofiel. Voor de precieze werking van het risicoprofiel wordt verwezen naar het rapport Vooringenomenheid voorkomen. Voor dit addendum volstaat de constatering dat het risicoprofiel op basis van drie criteria tot een risicoscore kwam:
  - > leeftijd
  - > opleidingsniveau
  - > afstand tot het ouderlijk huis.
- > **Handmatige selectie:** Doorlopen van een *risicoscore-populatie per regio* door een medewerker van DUO. Een DUO-medewerker loopt de lijst met studenten door op volgorde van risicoscore. De medewerker selecteert handmatig, met inachtneming van werkinstructies, welke studenten wel of niet worden gecontroleerd. Meer details over deze stap kunnen worden gevonden in het rapport Vooringenomenheid voorkomen.
- > **Onrechtmatig gebruik:** Verwerken van uitkomsten huisbezoek. Vaststellen van rechtmatig en onrechtmatig gebruik. Naar de populatie die onrechtmatig gebruik heeft gemaakt van de uitwonende beurs wordt gerefereerd als *onrechtmatig-gebruik-populatie*.

<sup>9</sup> Opdeling per regio (stap 2 uit het rapport Vooringenomenheid voorkomen) wordt enkel vanuit praktisch oogpunt gemaakt. Studenten worden in deze stap niet door DUO gecontroleerd. Het extern uitgevoerde huisbezoek (stap 4 uit het rapport Vooringenomenheid voorkomen) is omwille van de duidelijkheid van het stappenoverzicht ingevoegd. De uitkomst van deze stap wordt echter in stap 5 (terugkoppeling, afhandeling en vervolgstappen) verwerkt. Gegevens over stap 5 zijn wel bij het CBS opgevraagd.

- > **Bezwaarprocedures:** Studenten waarbij DUO onrechtmatig gebruik van de uitwonendenbeurs heeft vastgesteld (*de onrechtmatig-gebruik-populatie*) kunnen in bezwaar gaan tegen die beschikking. Naar deze populatie wordt gerefereerd als de *bezwaarpopulatie*.

In het rapport Vooringenomenheid voorkomen worden alle stappen van het CUB-proces in detail beschreven.

## 2.2 Recente chronologie CUB-onderzoek

Voor een beschrijving van de chronologie van het CUB-proces tot 1 maart 2024 wordt verwezen naar het rapport Vooringenomenheid voorkomen. Hieronder volgt een korte opsomming van relevante gebeurtenissen sindsdien.

- > 1 maart 2024 – Het rapport Vooringenomenheid voorkomen is als onderdeel van het interne onderzoek van DUO met de Tweede Kamer gedeeld. Ook het externe onderzoek dat het Ministerie van Onderwijs Cultuur en Wetenschap (OCW) heeft laten uitvoeren door PricewaterhouseCoopers (PwC) is met de Tweede Kamer gedeeld.
- > 1 maart 2024 – Het kabinet heeft een reactie gegeven op het onderzoek naar het controleproces uitwonendenbeurs. In haar reactie noemt de regering onder meer:

*“DUO heeft hierbij onafhankelijke onderzoekers van de Stichting Algorithm Audit ingeschakeld. DUO laat het CBS de komende tijd nader onderzoek doen naar oververtegenwoordiging van studenten met een migratieachtergrond in het controleproces, met behulp van data die voor PwC niet beschikbaar was binnen de looptijd van het onderzoek”<sup>10</sup>*

- > 1 maart 2024 – Directeur-generaal van DUO en minister Dijkgraaf van OCW hebben hun excuses aangeboden voor indirecte discriminatie in het CUB-proces.<sup>11</sup>
- > 21 maart 2024 – De Tweede Kamer heeft een motie van kamerlid Soepboer aangenomen waarin de kamer de regering verzoekt

*“Het CBS bij dit onderzoek ook expliciet te laten kijken naar de correlatiecijfers tussen de gebruikte parameters en iemands nationaliteit. Waarbij in het vervolgtraject ook de gebruikte (statistische) modellen en de bewustheid van de keuzes voor het gebruik van deze parameters en dit model worden meegenomen”<sup>12</sup>;*

- > 6 mei 2024 – Het CBS deelt de door DUO opgevraagde data. Die data zijn op 21 mei 2024 openbaar gemaakt.<sup>13</sup>

<sup>10</sup> Kamerstukken II 2023/24 p.6, 24724, nr. 220

<sup>11</sup> Kamerstukken II 2023/24, 24724, nr. 220

<sup>12</sup> Kamerstukken II, 2023/24, 24724, nr. 237 ter vervanging van de op 21 maart 2024 ingediende motie Kamerstukken II, 2023/24, 24724, nr. 234.

<sup>13</sup> Zie <https://www.cbs.nl/nl-nl/maatwerk/2024/21/ontvangers-uitwonendenbeurs-herkomst-2014-2017-2019-2021-en-2022>

## 2.3 Andere onderzoeken naar het CUB-proces

Naast het onderzoek van Algorithm Audit zijn ook andere onderzoeken naar mogelijke vooringenomenheid in het CUB-proces uitgevoerd of aangekondigd. PwC heeft in opdracht van het Ministerie van OCW onder meer onderzocht hoe het CUB-proces tot stand is gekomen en of het CUB-proces kon leiden tot discriminatie.<sup>14</sup> Op basis van een datastudie waarin migratiestatistieken van uitwonende studenten op postcodeniveau zijn geanalyseerd concludeerde PwC dat er niet met volle zekerheid kan worden vastgesteld dat in het CUB-proces gelijke gevallen gelijk zijn behandeld in de periode 2012-2023

De Autoriteit Persoonsgegevens onderzoekt het CUB-proces ook. Het resultaat van dat onderzoek wordt verwacht in juni 2024.

Het voornaamste verschil tussen dit addendum en de andere onderzoeken is de beschikbaarheid van CBS-data over de migratieachtergrond van de onderzochte studentenpopulaties. Hierdoor kan in dit addendum kwantitatief worden vastgesteld of er in het CUB-proces vooringenomenheid ten aanzien van studenten met een bepaalde migratieachtergrond was.

---

<sup>14</sup> Het gehele onderzoek bestond uit meer vragen. *Onderzoek misbruik uitwonendenbeurs* PwC, 21 januari 2024 V1F6 <https://www.rijksoverheid.nl/documenten/rapporten/2024/03/01/eindrapport-pwc-rapportage-onderzoek-misbruik-uitwonendenbeurs>

## 3. Onderzoeksmethodologie

Het doel van dit vervolgonderzoek wordt toegelicht (§3.1). De onderzoekspopulaties (§3.2) en opgevraagde data bij het CBS (§3.3) worden beschreven. Daarnaast wordt de methodologie toegelicht voor de biasmeting van het CUB-proces met betrekking tot migratieachtergrond (§3.4).

### 3.1 Doel van de analyse

Het doel van dit onderzoek is het traceren van mogelijke (multidimensionale) vooringenomenheid ten aanzien van studenten met een migratieachtergrond in verschillende stappen van het CUB-proces. Door per stap een meting uit te voeren kan gericht vervolgonderzoek worden gedaan naar waar in het CUB-proces mogelijke vooringenomenheid zich voordoet. Deze analyse wordt een *biasmeting* genoemd. Een overzicht van relevante stappen in het CUB-proces kan worden gevonden in [2.1 Stappen in het CUB-proces waarvoor data zijn verkregen](#).

### 3.2 Onderzoekspopulaties

Voor dit onderzoek zijn acht populaties geselecteerd om nader te onderzoeken. Twee populaties bestaan uit studenten waarop het CUB-proces van toepassing was in 2014 en 2019. Twee populaties hebben betrekking tot de studenten die geselecteerd zijn in de aselechte steekproeven 2014 en 2017. De overige vier populaties bestaan uit de studenten die bezwaarprocedures zijn gestart in 2014, 2019, 2021 en 2022. De samenstelling van de populaties en de reden dat zij in dit onderzoek worden gebruikt, worden hieronder toegelicht.

#### CUB-populaties

De eerste CUB-populatie betreft studenten die voor de introductie van het leenstelsel een uitwonendenbeurs hebben ontvangen. Voor deze populatie geldt de peildatum 01-02-2014. Naar deze populatie, bestaande uit wo-, hbo- en mbo-studenten, wordt gerefereerd als *uitwonendenbeurspopulatie-2014* (n=248.649). De tweede CUB-populatie betreft studenten die na de introductie van het leenstelsel een uitwonendenbeurs hebben ontvangen. Voor deze populatie, bestaande uit mbo-studenten en uitfaserende hbo- en wo-studenten die vóór 2015 recht hadden op een uitwonendenbeurs, geldt peildatum 01-02-2019. Naar deze populatie wordt gerefereerd als *uitwonendenbeurspopulatie-2019* (n=50.233). Indien mogelijk wordt de analyse voor 2019 volledig toespitst op mbo-studenten (n=36.630), aangezien deze studenten tot de primaire CUB-doelgroep behoren.

#### Box 4

### Statistische significantie

In deze analyse zijn aggregatiestatistieken gebruikt over de herkomst van de gehele populatie uitwonende studenten die zijn samengesteld door het CBS. Er vinden in deze analyse daarom geen schattingen plaats. De cijfers zijn een weergave van de daadwerkelijke situatie op populatieniveau. De populatie betreft in dit geval studenten die op 1 februari 2014 of 2019 een beroep hebben gedaan op de uitwonendenbeurs en onderworpen zijn geweest aan het CUB-proces. In dit rapport wordt de term statistische significantie alleen gebruikt in verband met de aselechte steekproeven.

Uitwonendenbeurspopulatie-2014 en uitwonendenbeurspopulatie-2019 worden gevolgd voor de relevante stappen van het CUB-proces. Een overzicht van alle relevante stappen kan worden gevonden in [2.1 Stappen in het CUB-proces waarvoor data zijn verkregen](#).

### Aselecte steekproefpopulaties

De aselechte-populaties betreffen studenten die in 2014 en 2017 zijn getrokken voor een aselechte steekproef. Naar deze populaties wordt gerefereerd als de *aselechte steekproefpopulatie-2014* en *aselechte steekproefpopulatie-2017*. Zowel de CUB-populaties als de aselechte steekproefpopulaties worden onderzocht omdat deze populaties verschillende informatie opleveren. Het verschil tussen de twee soorten populaties bestaat uit het feit dat bij de CUB-populaties het risicoprofiel en de handmatige selectie zijn toegepast. Bij de aselechte steekproefpopulaties is het risicoprofiel en handmatige selectie niet toegepast omdat studenten willekeurig werden geselecteerd voor controle. De gegevens uit de aselechte steekproef geven informatie over het onrechtmatigheid gebruik-percentages per groep onafhankelijk van toepassing van het risicoprofiel en handmatige selectie voor controle uit het CUB-proces. Resultaten van de analyse naar een verband tussen de gebruikte criteria uit het risicoprofiel en onrechtmatig gebruik van de uitwonendenbeurs aan de hand van de aselechte steekproeven kan worden gevonden in het rapport Vooringenomenheid voorkomen. Alternatieve steekproeven uit andere jaren zijn niet voorhanden.

### Bezwaarpopulaties

Naast de CUB-populaties en de aselechte steekproefpopulaties worden vier extra populaties onderzocht om de bezwaarprocedures nauwkeurig te analyseren. Uitwonendenbeurspopulatie-2014 en uitwonendenbeurspopulatie-2019, vastgesteld aan de hand van de respectievelijke peildata 01-02-2014 en 01-02-2019, bevatten minder dan 50% van de bezwaarprocedures die plaatsvinden naar aanleiding van een selectiemoment en huisbezoek in kalenderjaar 2014 of 2019. De reden hiervoor is dat veel studenten na 01 februari een uitwonendenbeurs aanvragen en vervolgens in het desbetreffende kalenderjaar voor een huisbezoek worden geselecteerd, een onrechtmatigheidsbesluit ontvangen en in bezwaar gaan tegen de beschikking. Om een compleet beeld te krijgen van de gehele bezwaarpopulatie is besloten ook studenten die in het gehele kalenderjaar 2014 of 2019 een bezwaarprocedure zijn gestart te analyseren. Naar deze groepen wordt gerefereerd als *bezwaarpopulatie-2014* en *bezwaarpopulatie-2019*. Daarnaast zijn de *bezwaarpopulatie-2021* en *bezwaarpopulatie-2022* opgevraagd. CUB populatie-2021 en CUB populatie-2022 zijn niet opgevraagd aangezien in deze jaren het CUB-proces is verstoord door coronamaatregelen<sup>5</sup>.

Algorithm Audit ziet geen bezwaar tegen het gebruik van tien jaar oude data omdat dit onderzoek de werking van het CUB-proces in het verleden onderzoekt.

### 3.3 Data verstrekt door het CBS

Voor de uitwonendenbeurspopulatie-2014, uitwonendenbeurspopulatie-2019, aselechte steekproefpopulatie-2014, aselechte steekproefpopulatie-2017, bezwaarpopulatie-2014, bezwaarpopulatie-2019, bezwaarpopulatie-2021 en bezwaarpopulatie-2022 zijn geaggregeerde data over de herkomstland en geboorteland van studenten opgevraagd bij het CBS. Het CBS heeft voor iedere populatie de aggregatiestatistieken samengesteld voor de volgende twee variabelen:

- > **Geboorteland**<sup>15</sup> – Het land waar een persoon geboren is. Waarden: *Geboren in Nederland, geboren buiten Nederland.*
- > **Herkomstland**<sup>16</sup> (nieuwe indeling) – Kenmerk dat weergeeft in welk land iemand geboren is of waar een van diens ouders geboren zijn. Het herkomstland van personen die in het buitenland zijn geboren wordt bepaald door hun eigen geboorteland. Bij personen die in Nederland geboren zijn, wordt het herkomstland bepaald door het geboorteland van de ouders. Wanneer beide ouders in het buitenland zijn geboren, is het geboorteland van de moeder leidend in het bepalen van het herkomstland. De geboortegegevens van de moeder zijn vaker bekend dan die van de vader. Wanneer de moeder in Nederland is geboren of het geboorteland van de moeder onbekend is, dan wordt het geboorteland van de vader gebruikt. Waarden: *Nederland, Europa (excl. Nederland) en buiten Europa.*<sup>17</sup>

In dit onderzoek worden derde generatie migranten niet meegenomen. Dit zijn studenten die zelf in Nederland zijn geboren, waarvan hun beide ouders in Nederland zijn geboren, maar van wie één of meerdere van de grootouders in het buitenland is geboren. Het CBS houdt deze specifieke groep niet bij. Mogelijke vooringenomenheid van het CUB-proces ten aanzien van deze groep is daarom niet onderzocht.

Sinds 2022 gebruikt het CBS het begrip migratieachtergrond niet meer. Er wordt sindsdien een nieuwe indeling gebruikt die voorheen bekend stond als 'bevolking met een westerse of niet-westerse migratieachtergrond'. In de nieuwe indeling is meer bepalend waar iemand zelf geboren is, en minder bepalend waar iemands ouders geboren zijn. De nieuwe indeling vindt plaats op basis van werelddelen (Nederland, Europa exclusief Nederland en buiten Europa) en niet meer op basis van westerse en niet-westerse landen.<sup>18</sup>

<sup>15</sup> Zie ook <https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/geboorteland>

<sup>16</sup> Zie ook <https://www.cbs.nl/nl-nl/onze-diensten/methoden/begrippen/herkomstland>

<sup>17</sup> Welke landen als Europa en buiten Europa worden aangemerkt kan hier worden gevonden: <https://www.cbs.nl/nl-nl/maatwerk/2024/08/landenindeling-van-de-variabele-herkomstland-2022>

<sup>18</sup> Zie ook <https://www.cbs.nl/nl-nl/nieuws/2022/07/cbs-introduceert-nieuwe-indeling-bevolking-naar-herkomst>

#### Box 5

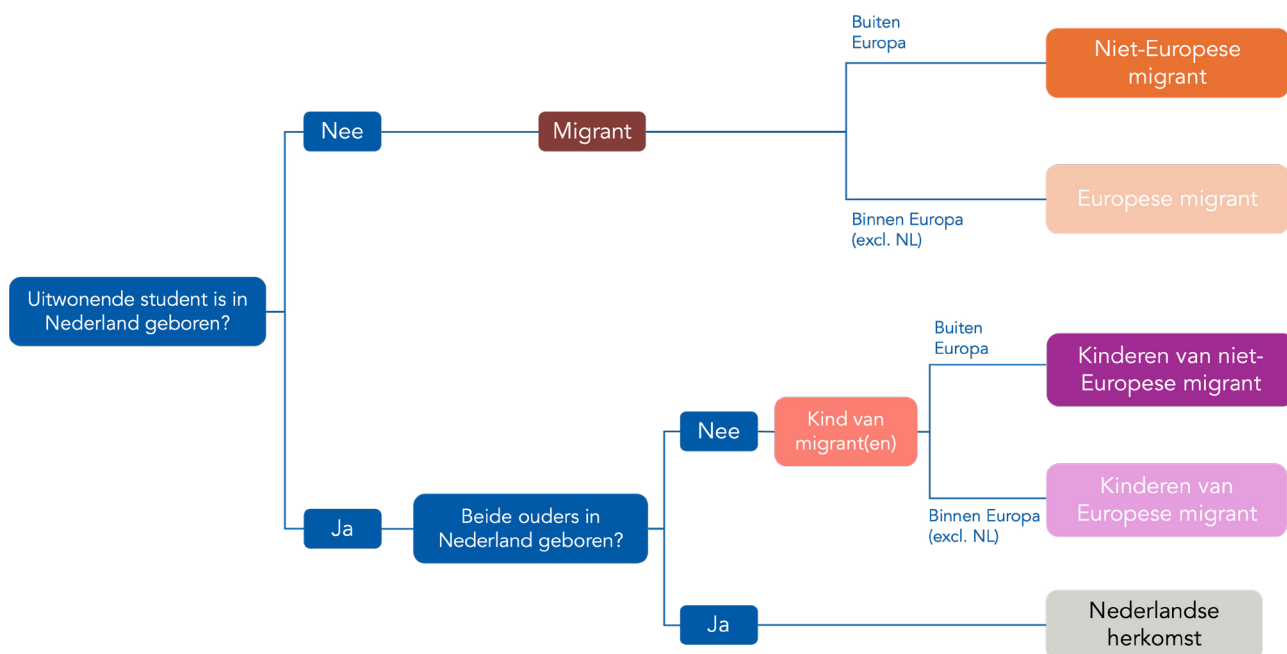
### Herkomst onbekend

Er zijn varianten denkbaar waarin niet kan worden vastgesteld in welke categorie een student valt. Bijvoorbeeld omdat niet bekend is waar de ouders geboren zijn of omdat de term ouder naar een vreemd rechtssysteem anders is dan naar het Nederlandse rechtssysteem. Voorzieningen opnemen voor alle mogelijke nuances gaat de reikwijdte van dit onderzoek te buiten. Het CBS kan specifieke vragen over de omgang met dergelijke situaties beantwoorden.

Op basis van de kruising van de variabele geboorteland en herkomstland heeft het CBS studenten in vijf migrantengroepen onderscheiden:

1. **Niet-Europese migranten:** deze studenten zijn buiten Europa geboren.
2. **Europese migranten:** deze studenten zijn binnen Europa maar buiten Nederland geboren.
3. **Kinderen van niet-Europese migrant:** deze studenten zijn zelf in Nederland geboren, maar tenminste één van hun ouders is buiten Europa geboren.
4. **Kinderen van Europese migrant:** deze studenten zijn zelf in Nederland geboren, maar tenminste één van hun ouders is in Europa, maar niet buiten Nederland geboren.
5. **Nederlandse herkomst:** deze studenten en hun beide ouders zijn in Nederland geboren. 3e generatie migranten behoren tot deze groep.

De combinatie 'niet geboren in Nederland x herkomstland Nederland' kan op logische gronden niet voorkomen (iemand met herkomstland Nederland is per definitie in Nederland geboren) en is buiten de analyse gehouden. Een schematisch overzicht van bovenstaande groepen wordt gegeven in [Figuur 3](#). Een vergelijkbaar overzicht waarin bovenstaande termen worden gelinkt aan de term migratieachtergrond kan worden gevonden in [Figuur 1](#).



Figuur 3 – Overzicht van de door het CBS onderscheiden categorieën studenten



Geboorteland en herkomstland, en de kruising van deze variabelen, is voor de volgende (sub)populaties opgevraagd:

- > Uitwonendenbeurspopulatie-2014, uitwonendenbeurspopulatie-2019, aselectesteekproefpopulatie-2014, aselecte steekproefpopulatie-2017, bezwaarpopulatie-2014, bezwaarpopulatie-2019, bezwaarpopulatie-2021 en bezwaarpopulatie-2022
- > Onderwijsvorm (univariaat, 4 waarden: mbo 1-2, mbo 3-4, hbo, wo)
- > Leeftijd (univariaat, 5 waarden: 15-18, 19-20, 21-22, 23-24, 25-50)
- > Afstand (univariaat, 9 waarden: 0km, 1m-1km, 1-2km, 2-5km, 5-10km, 10-20km, 20-50km, 50-500km, onbekend)
- > Onderwijsvorm x leeftijd: (bivariaat,  $4 \times 5 = 20$  waarden)
- > Onderwijsvorm x afstand (bivariaat:  $4 \times 9 = 36$  combinaties)
- > Leeftijd x afstand (bivariaat  $5 \times 9 = 45$  combinaties)
- > Onderwijsvorm x leeftijd x afstand (trivariaat:  $4 \times 5 \times 9 = 180$  combinaties)
- > Risicocategorie (univariaat, 6 waarden: Risicocategorie 1, 2, 3, 4, 5 en 6)
- > Geselecteerd voor controle (univariaat, 2 waarden: wel/niet geselecteerd voor controle)
- > Uitkomst huisbezoek (univariaat, 4 waarden: rechtmatig, onrechtmatig, huisbezoek heeft niet kunnen plaatsvinden, niet geselecteerd voor controle)
- > Bezwaar (univariaat, 2 waarden wel/geen bezwaar)
- > Uitkomst bezwaar (univariaat, 3 waarden: geslaagd, deels geslaagd, niet geslaagd)
- > Risicocategorie x geselecteerd voor controle (bivariaat:  $6 \times 2 = 12$  combinaties)
- > Risicocategorie x uitkomst huisbezoek (bivariaat:  $6 \times 2 = 12$  combinaties).

Verdeeld over vijf verschillende jaren zijn in totaal aggregatiestatistieken over de migratieachtergrond van 1.401 groepen studenten opgevraagd. Voor de uitwonendenbeurspopulatie-2014 en -2019 zijn aggregatiestatistieken over de migratieachtergrond van bovenstaande 341 groepen opgevraagd. Voor de aselecte steekproefpopulatie-2014 zijn aggregatiestatistieken over migratieachtergrond voor 146 groepen opgevraagd (alle bovenstaande groepen zonder trivariaat, geselecteerd voor controle en risicocategorie x geselecteerd voor controle). Voor de aselecte steekproefpopulatie-2019 zijn aggregatiestatistieken over migratieachtergrond voor 116 groepen opgevraagd (dezelfde als voor de aselecte steekproefpopulatie-2014 maar zonder de waarden hbo en wo voor onderwijsvorm). Voor de bezwaarpopulatie-2014, -2019, -2021 en -2022 zijn aggregatiestatistieken over migratieachtergrond voor respectievelijk 129, 114, 99 en 115 groepen opgevraagd. De exacte datalevering kan op de website van het CBS worden gevonden.<sup>4</sup>

### 3.4 Onderzoeksvragen biasmeting met betrekking tot migratieachtergrond

Met behulp van bovenstaande gegevens kan voor alle stappen van het CUB-proces en voor de verschillende (sub)populaties de verhouding studenten per combinatie van herkomstland en geboorteland worden gemeten. Aan de hand van de volgende onderzoeksvragen wordt per stap van het CUB-proces mogelijke vooringenomenheid onderzocht (zie [2.1 Stappen in het CUB-proces waarvoor data zijn verkregen](#)).

#### Onderzoeksvraag 1 (Stap 0 – startpopulatie)

*Wat was de verdeling van studenten met een migratieachtergrond in de uitwonendenbeurspopulatie-2014 en -2019?*

#### Onderzoeksvraag 2 (Stap 1 – risicoprofiel)

*Zijn studenten met een migratieachtergrond in 2014 en 2019 oververtegenwoordigd geweest in de hogere risicocategorieën die volgen uit het risicoprofiel?*

#### Onderzoeksvraag 2a (Proxy-analyse risicoprofiel)

*Is er een verband tussen criteria gebruikt in het risicoprofiel en studenten met een migratieachtergrond?*

#### Onderzoeksvraag 3 (Stap 3 – handmatige selectie)

*Zijn studenten met een migratieachtergrond in 2014 en 2019 vaker geselecteerd voor controle tijdens handmatige selectie dan studenten met een Nederlandse herkomst?*

#### Onderzoeksvraag 3a (volgen of afwijken van algoritme)

*Wat is de mate waarin de uitkomst van het risicoprofiel en de handmatige selectie voor huisbezoek overeenkomt, in het bijzonder voor migranten en kinderen van een migrant?*

#### Onderzoeksvraag 4 (Stap 5 – onrechtmatig gebruik)

*Wat was in 2014 en 2019 het aandeel studenten met een migratieachtergrond in de onrechtmatig gebruik-populatie?*

#### Onderzoeksvraag 5 (Stap 6 – bezwaarprocedure)

*Wat is in 2014, 2019, 2021 en 2022 de verdeling van studenten met een migratieachtergrond geweest voor selectie voor bezwaarpopulatie?*

## 4. Resultaten

Het aandeel studenten met een migratieachtergrond in de gehele uitwonendenbeurspopulatie wordt gepresenteerd (§4.1). Dit wordt gevolgd door de biasmeting van het risicoprofiel (§4.2), proxy-analyse van het risicoprofiel (§4.3), biasmeting van de handmatige controle (§4.4), een analyse van de mate waarin de voorspelling van het algoritme tijdens handmatige selectie werd gevolgd (§4.5), vaststelling van het aandeel studenten met een migratieachtergrond in de onrechtmatig gebruik-populatie (§4.6) en in de bezwaarpopulatie (§4.7). De code waarmee onderstaande analyse tot stand is gekomen is online beschikbaar.<sup>19</sup>

### 4.1 Resultaten onderzoeksvraag 1 (startpopulatie)

Deze paragraaf presenteert de resultaten van de biasmeting voor stap 0 van het CUB-proces. Een precieze beschrijving van de uitwonendenbeurspopulatie-2014 en -2019 wordt gegeven in 3.2 Onderzoekspopulaties. In het rapport Vooringenomenheid voorkomen worden histogrammen weergegeven die de verdeling van variabelen die relevant zijn in het CUB-proces weergeven.

#### Onderzoeksvraag 1

*Wat was de verdeling van studenten met een migratieachtergrond in de uitwonendenbeurspopulatie-2014 en -2019?*

#### Antwoord op onderzoeksvraag 1

Van de 248.650 studenten die op 01-02-2014 aanspraak maakten op de uitwonendenbeurs (uitwonendenbeurspopulatie-2014) had 22% een niet-Europese migratieachtergrond. Daarvan waren

- > 9.5% niet-Europese migranten en
- > 12.5% kinderen van een niet-Europese migrant.

6.3% van de uitwonendenbeurspopulatie-2014 behoorde tot de groep Europese migratieachtergrond. Daarvan waren

- > 3.1% Europese migranten en
- > 3.2% kinderen van een Europese migrant.

Oftewel, 28.3% van de studenten in de uitwonendenbeurspopulatie-2014 had een migratieachtergrond en 71.7% van de studenten had Nederlandse herkomst. Zie [Figuur 4-6](#).

Van de 36.630 ontvangers van een uitwonendenbeurs op 01-02-2019 die aan het mbo studeerden (mbo-studenten binnen de uitwonendenbeurspopulatie-2019<sup>20</sup>) behoorde 47.7% van de studenten tot de groep niet-Europese migratieachtergrond, waarvan:

- > 27.8% niet-Europese migranten waren en
- > 19.9% kinderen van een niet-Europese migrant.

<sup>19</sup> <https://github.com/NGO-Algorithm-Audit>

<sup>20</sup> Een toelichting waarom alleen naar mbo-studenten in 2019 wordt gekeken kan worden gevonden in 3.2 Onderzoekspopulaties

5.2% van de uitwonendenbeurspopulatie-2019 behoorde tot de groep Europese migratieachtergrond, waarvan:

- > 2.2% Europese migrant waren en
- > 2.9% kinderen van een Europese migrant.

Oftewel: 52.8% van de studenten in de uitwonendenbeurspopulatie-2019 had een migratieachtergrond en 47.2% van de studenten had Nederlandse herkomst. Zie [Figuur 7-9](#).

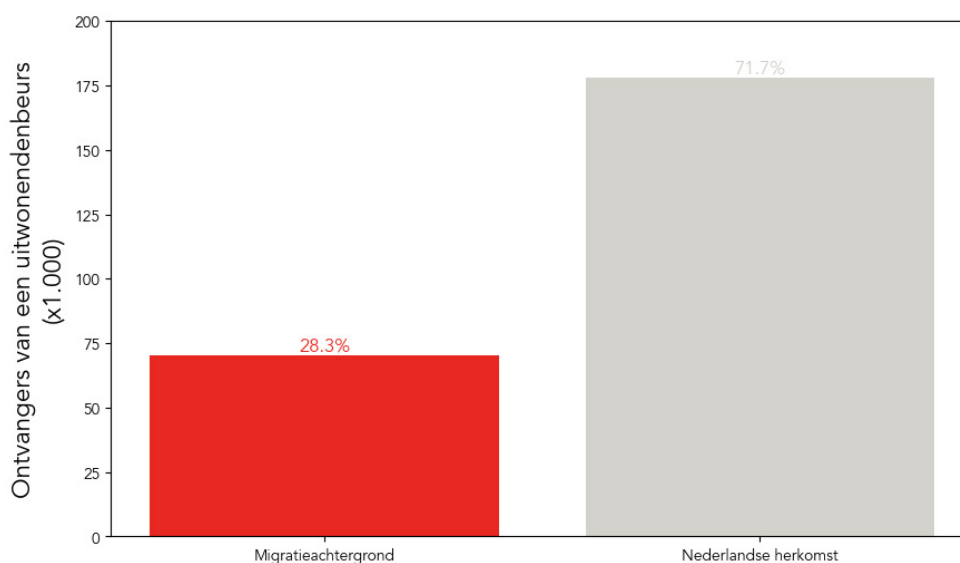
### Toelichting onderzoeksvraag 1

Eerst wordt de verdeling van alle migrantengroepen gepresenteerd voor de uitwonendenbeurspopulatie-2014. Daarna voor mbo-studenten binnen de uitwonendenbeurspopulatie-2019.

### Diagrammen en cijfers 2014

In [Figuur 4](#) wordt de verdeling van studenten met een migratieachtergrond en studenten met een Nederlandse herkomst gegeven in de uitwonendenbeurspopulatie-2014.

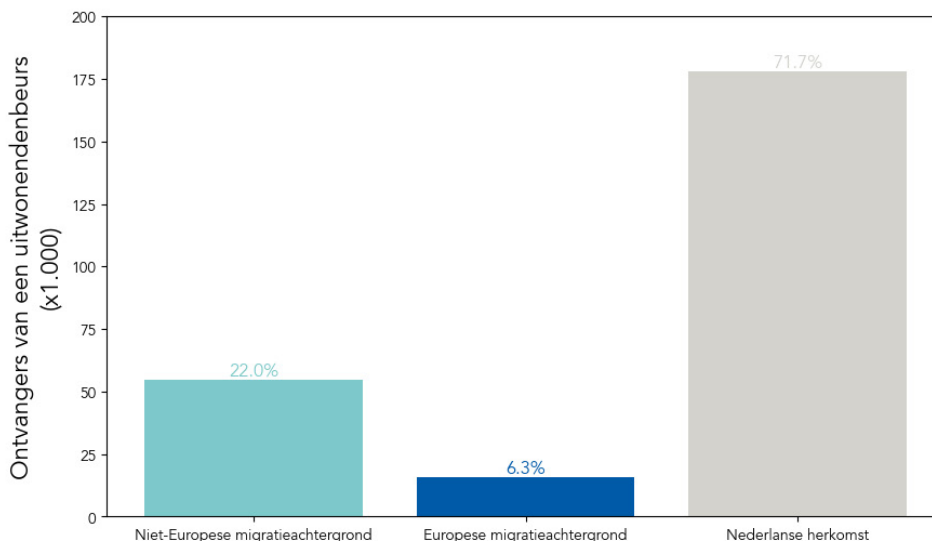
Verdeling van studenten met een migratieachtergrond en met Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)



[Figuur 4](#) – Verdeling studenten met een migratieachtergrond en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)

In [Figuur 5](#) wordt de verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst gegeven in de uitwonendenpopulatie-2014.

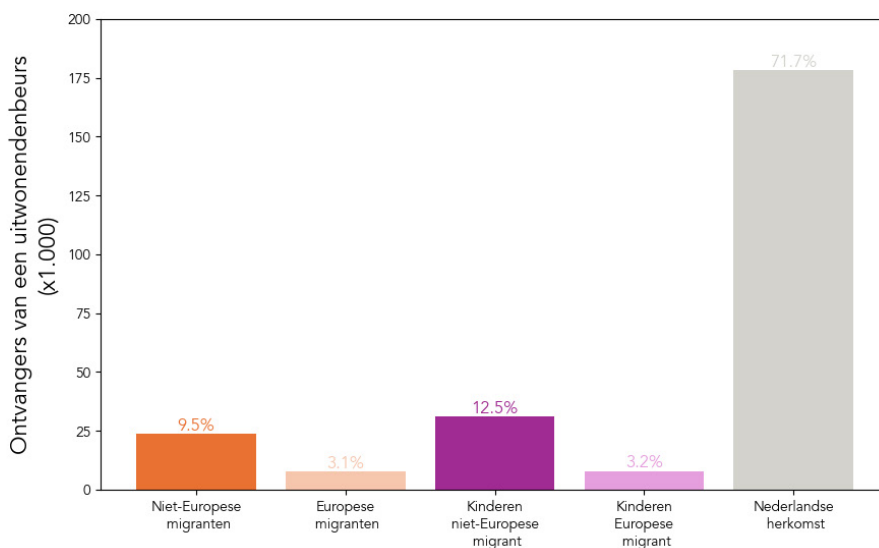
Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)



[Figuur 5](#) – Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)

In [Figuur 6](#) wordt de verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst gegeven in de uitwonendenpopulatie-2014.

Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en Nederlandse herkomst in uitwonendenbeurspopulatie-2014 (n=248.650)

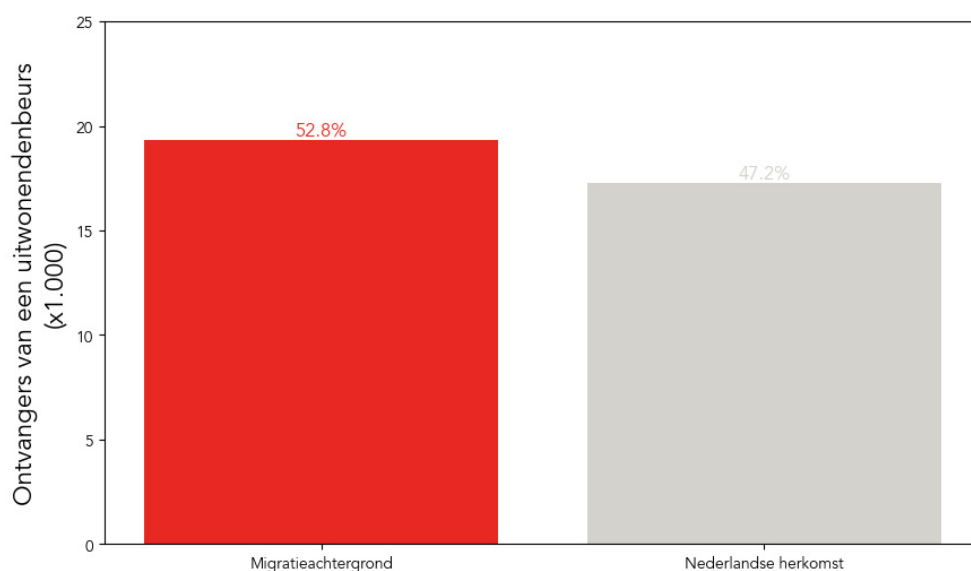


[Figuur 6](#) – Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)

### Diagrammen en cijfers 2019

In [Figuur 7](#) wordt de verdeling van mbo-studenten met een migratieachtergrond en mbo-studenten met een Nederlandse herkomst gegeven in de uitwonendenpopulatie-2019.

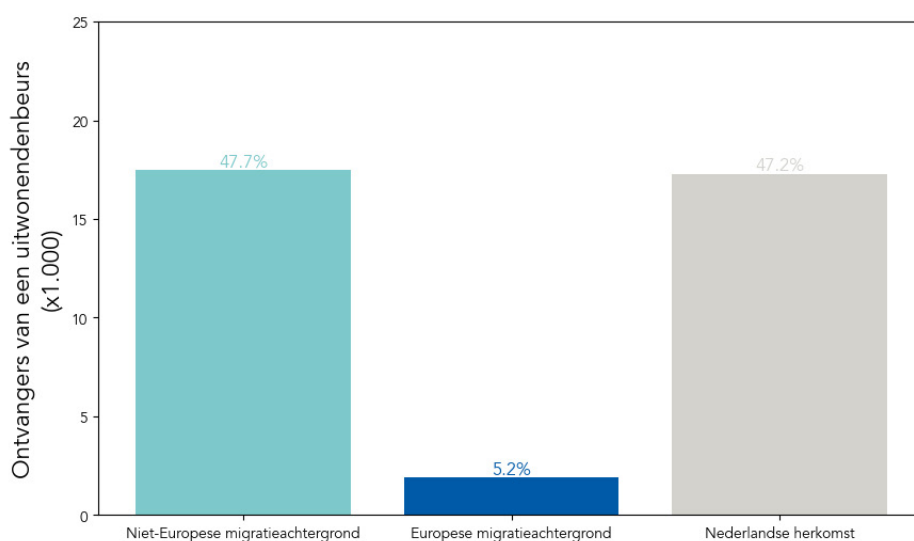
#### Verdeling van mbo-studenten met een migratieachtergrond en met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=36.630)



[Figuur 7](#) – Verdeling van mbo-studenten met een migratieachtergrond en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=36.630)

In [Figuur 8](#) wordt de verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en mbo-studenten met een Nederlandse herkomst gegeven in de uitwonendenpopulatie-2019.

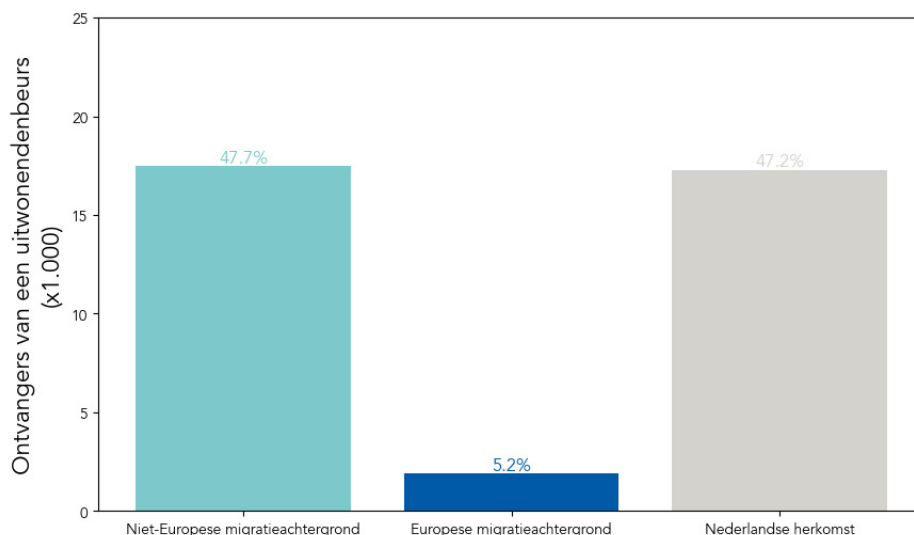
#### Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=36.630)



[Figuur 8](#) – Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=36.630)

In [Figuur 9](#) wordt de verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst gegeven, voor alleen mbo-studenten.

Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=36.630)



[Figuur 9](#) – Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en Nederlandse herkomst in de uitwonendenbeurspopulatie-2019, alleen mbo-studenten (n=36.630)

## 4.2 Resultaten onderzoeksvraag 2 (risicoprofiel)

Onderzoeksvraag 2 ziet op mogelijke oververtegenwoordiging van studenten met een migratieachtergrond in stap 1 van het CUB-proces (toepassing van het risicoprofiel). Een beschrijving van de werking van het risicoprofiel wordt gegeven in [2.1 Stappen in het CUB-proces](#).

### Onderzoeksvraag 2

*Zijn studenten met een migratieachtergrond in 2014 en 2019 oververtegenwoordigd geweest in de hogere risicocategorieën die volgen uit het risicoprofiel?*

### Antwoord op onderzoeksvraag 2

Ja. Studenten met een migratieachtergrond waren in 2014 en 2019 oververtegenwoordigd in de hogere risicocategorieën die volgen uit het risicoprofiel. Voor beide jaren zijn specifiek studenten met een niet-Europese migratieachtergrond en dan met name kinderen van een niet-Europese migrant oververtegenwoordigd geweest in de hogere risicocategorieën.

### Toelichting onderzoeksvraag 2

Van oververtegenwoordiging is sprake als de verhouding studenten met een migratieachtergrond in een risicocategorie zoals toebedeeld door het risicoprofiel groter is dan in de startpopulatie ([Box 3](#)). Uit het antwoord op onderzoeksvraag 1 volgt dat in de uitwonendenbeurs-2014 22.0% van de studenten een niet-Europese migratieachtergrond heeft. In 2019 was dat 47.7%.

De verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst in de door het risicoprofiel toegekende risicocategorieën voor de uitwonendenbeurspopulaties-2014 en -2019 wordt gegeven in [Figuur 10](#) en [Figuur 12](#). Dezelfde verdeling maar dan uitgesplitst naar (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met Nederlandse herkomst wordt gegeven in [Figuur 11](#) en [Figuur 13](#).

In 2014 is er in risicocategorie 1 (zeer hoog) een oververtegenwoordiging van 21.5 absolute procentpunten van studenten met een niet-Europese migratieachtergrond (43.5% ten opzichte van het gemiddelde 22.0%, zie [Figuur 10](#)). Voor de risicocategorie hoog is er een oververtegenwoordiging van 11.2 absolute procentpunten (31.2% ten opzichte van 22.0%)<sup>21</sup>. In 2014 wordt voor de drie laagste risicocategorieën een ondervertegenwoordiging van studenten met een niet-Europese migratieachtergrond waargenomen. Het omgekeerde effect is dientengevolge zichtbaar voor studenten met een Nederlandse herkomst. Gebaseerd op de data uit 2014 hadden studenten met een niet-Europese migratieachtergrond een 2.0x grotere kans om in een hoge risicocategorie (categorie 1 of 2) te worden ingedeeld dan studenten met een Nederlandse herkomst.<sup>22</sup>

Uit [Figuur 11](#) volgt dat in 2014 binnen de groep studenten met een migratieachtergrond vooral de groep kinderen van een niet-Europese migrant sterk oververtegenwoordigd is in de hoogste risicocategorieën. Zo behoort 32.5% van de studenten in de risicocategorie 1 (zeer hoog) tot de groep kinderen van een niet-Europese migrant. Dit neemt stapsgewijs af tot 8.2% in risicocategorie 5 (zeer laag). Dit effect bestaat niet in dezelfde mate voor de groep niet-Europese migranten. Voor deze groep valt de sterke oververtegenwoordiging op in risicocategorie 6 (onbekend) van 26.4%.

Het hoge aantal niet-Europese migranten met onbekende risicocategorie kan als volgt worden verklaard: een uitwonende student wordt de onbekend risicocategorie toebedeeld als de afstandscategorie onbekend is. De afstandscategorie is onbekend als van beide ouders het adres onbekend is. Dat kan het geval zijn als 1) wel bekend is wie de ouders zijn, maar het adres van de ouders onbekend is of 2) de ouders onbekend zijn. Bij niet-Europese migranten, onder andere vluchtelingen en immigranten, is over het algemeen vaker onbekend wie de ouders van een student zijn. Er zijn daarom relatief veel niet-Europese migranten met afstandscategorie onbekend, en daarom risicocategorie onbekend.

In 2019 is de tendens vergelijkbaar voor lagere risicocategorieën: studenten met een niet-Europese migratieachtergrond zijn ondervertegenwoordigd en studenten met een Nederlandse herkomst zijn oververtegenwoordigd in lage risicocategorieën ([Figuur 12](#)).<sup>23</sup> Gebaseerd op de data uit 2019 hadden studenten met een niet-Europese migratieachtergrond een 0.9x grotere kans om in een hoge risicocategorie (categorie 1 of 2) te worden ingedeeld dan studenten met een Nederlandse herkomst. Niet-Europese

<sup>21</sup> Deze afwijkingen hoeven niet getoetst te worden op statistisch significantie aangezien voor de analyse data over de gehele populatie beschikbaar was. Afwijkingen worden dus niet geschat maar feitelijk vastgesteld. Zie ook [Box 4](#).

<sup>22</sup> Deze kans wordt bepaald door de kans dat een student met een niet-Europese migratieachtergrond geclassificeerd wordt als hoog risico te vergelijken met de kans dat een student met Nederlandse herkomst wordt geclassificeerd als hoog risico. De kans is vastgesteld op basis van de uitwonendenbeurspopulatie-2014.

<sup>23</sup> Merk op: de toegekende risicocategorieën per onderwijsvorm zijn niet beschikbaar in de CBS-data. [Figuur 13-14](#) zijn daarom gebaseerd op alle studenten (mbo, hbo en wo) in de uitwonendenbeurspopulatie-2019, in plaats van alleen mbo-studenten.

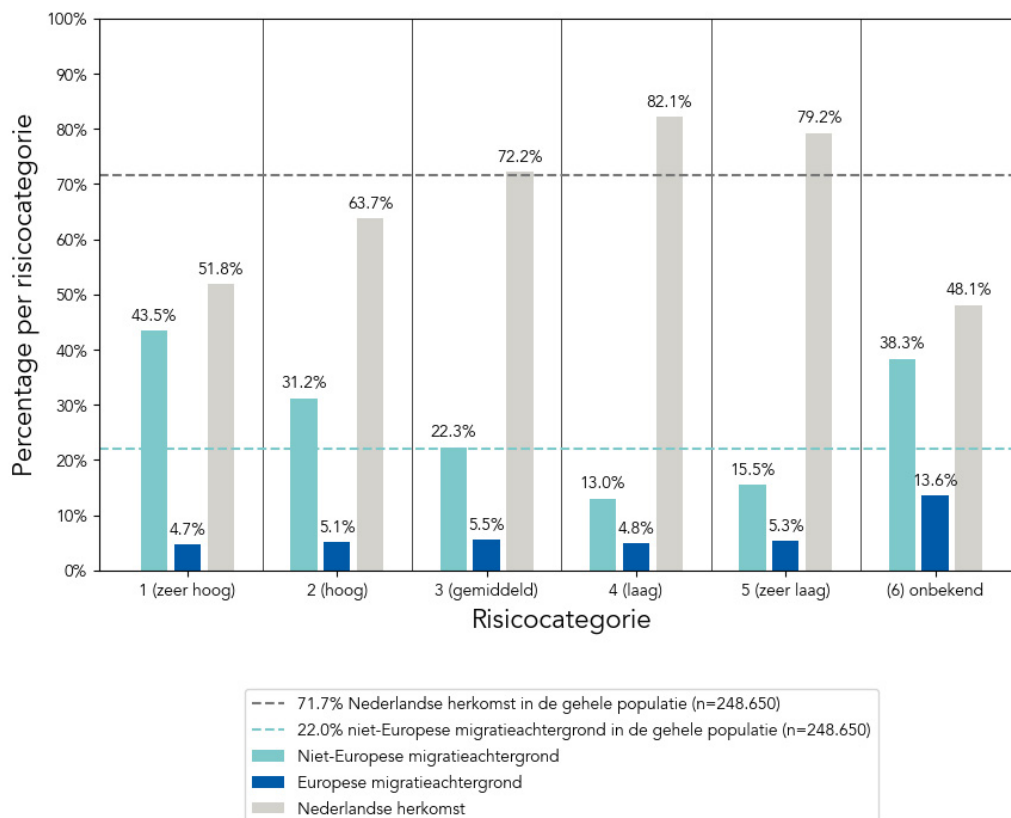


migranten waren in 2019 echter nog sterker oververtegenwoordigd in risicocategorie 6 (onbekend) dan in 2014 (zie bovenstaande verklaring). Voor kinderen van een niet-Europese migrant is een vergelijkbare oververtegenwoordiging zichtbaar in risicocategorieën 1-2 als in de data over 2014 (Figuur 13).

Samenvattend is duidelijk dat het risicoprofiel in 2014 en 2019, met uitzondering van de onbekend categorie, meer risico toebedeelt aan studenten met een niet-Europese migratieachtergrond. Dit betekent dat deze studenten door gebruik van het risicoprofiel in verhouding tot studenten met een Nederlandse herkomst hoger op de lijsten kwamen voor handmatige selectie voor controle in stap 3 van het CUB-proces. Of studenten met een niet-Europese migratieachtergrond daadwerkelijk ook vaker werden geselecteerd voor controle wordt besproken in 4.4 Resultaten onderzoeksvraag 3 (handmatige selectie). Eerst volgt een proxy-analyse van de criteria die in het risicoprofiel zijn gebruikt in 4.3 Resultaten onderzoeksvraag 2a (proxy-analyse).

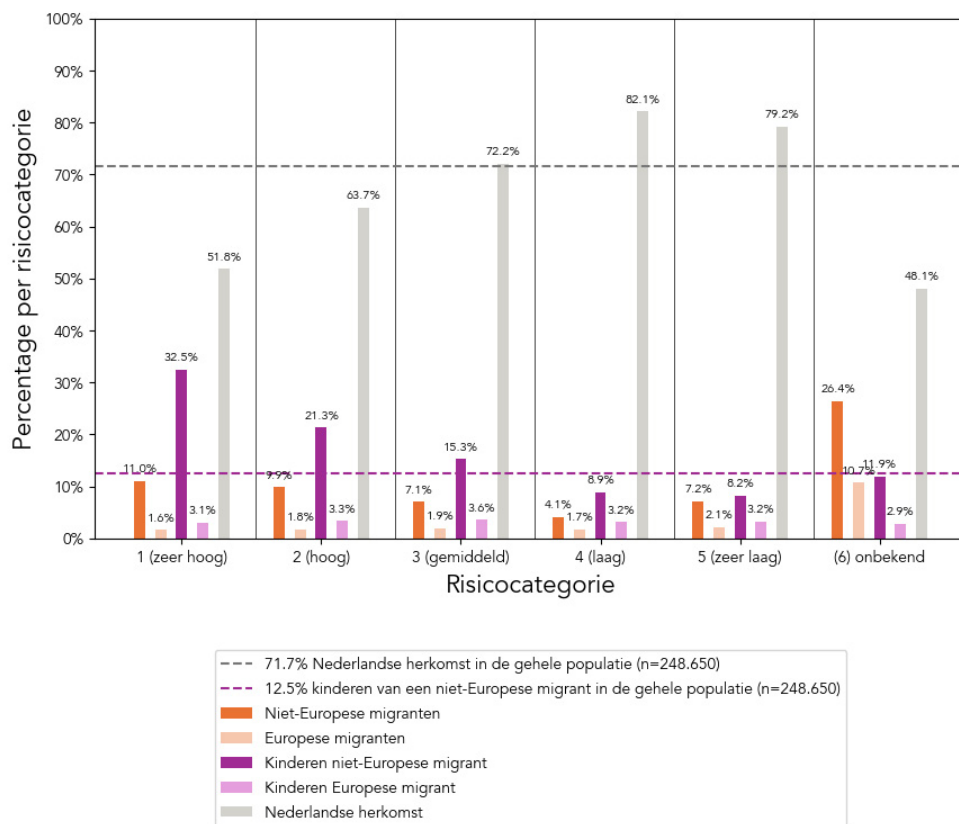
### Diagrammen en cijfers 2014

Toegekende risicocategorie aan studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)



Figuur 10 – Toegekende risicocategorieën aan studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)

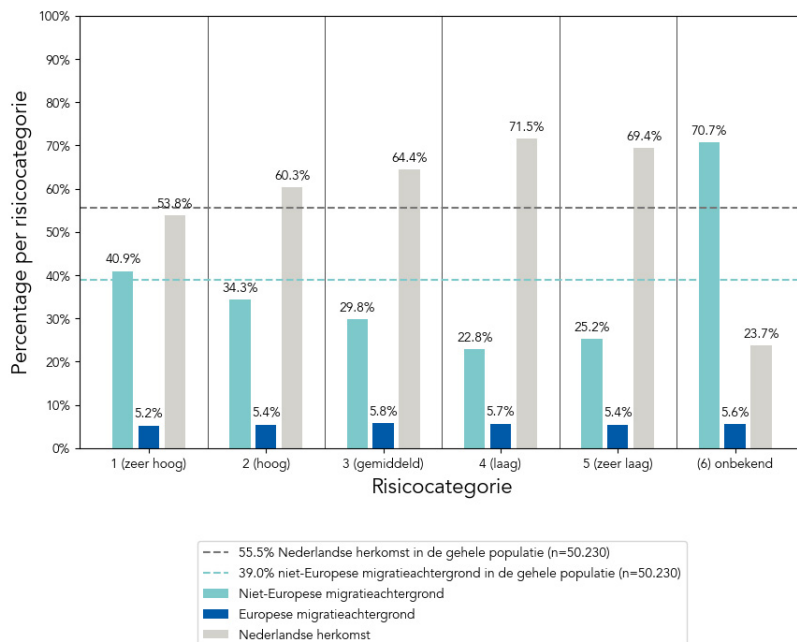
Toegekende risicocategorie aan (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)



Figuur 11 – Toegekende risicocategorieën aan (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2014 (n=248.650)

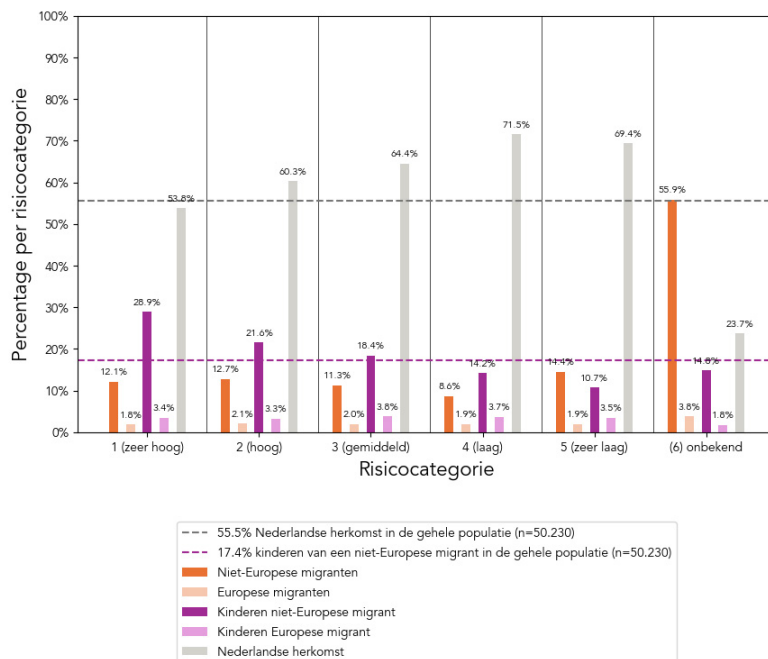
## Diagrammen en cijfers 2019

Toegekende risicocategorie aan studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=50.230)



Figuur 12 – Toegekende risicocategorieën aan studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=50.230)

Toegekende risicocategorie aan (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=50.230)



Figuur 13 – Toegekende risicocategorieën aan (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met Nederlandse herkomst in de uitwonendenbeurspopulatie-2019 (n=50.230)

### 4.3 Resultaten onderzoeksvraag 2a (proxy-analyse)

Onderzoeksvraag 2a ziet op het verband tussen criteria gebruikt in het risicoprofiel en studenten met een migratieachtergrond.

Er wordt onderzocht of de kenmerken die in het risicoprofiel worden gebruikt (onderwijsvorm, leeftijd en afstand tot ouder(s), en een combinatie van deze factoren) een proxykenmerk zijn voor studenten met een migratieachtergrond.

#### Onderzoeksvraag 2a

*Is er een verband tussen criteria gebruikt in het risicoprofiel en studenten met een migratieachtergrond?*

#### Antwoord op onderzoeksvraag 2a

Voor alle kenmerken en combinaties daarvan is er een verband met studenten met een niet-Europese migratieachtergrond, hoewel de sterkte en de vorm van dit proxykarakter verschillen. De kenmerken onderwijsvorm en afstand tot ouder(s) dragen via hun proxykarakter bij aan vooringenomenheid van het risicoprofiel ten aanzien van studenten met een niet-Europese migratieachtergrond. Voor het kenmerk leeftijd geldt het omgekeerde. Per criteria wordt het proxykarakter toegelicht:

- > **Onderwijsvorm:** sterk proxykarakter voor studenten met een niet-Europese migratieachtergrond. Mbo-studenten hebben aanzienlijk vaker een niet-Europese migratieachtergrond dan studenten met Nederlandse herkomst. Aangezien mbo-studenten door het risicoprofiel een hoger risico werd toebedeeld, werd aan studenten met een niet-Europese migratieachtergrond ook een hoger risico toebedeeld.
- > **Leeftijd:** invers proxykarakter voor studenten met een niet-Europese migratieachtergrond. Oudere studenten, voornamelijk in de leeftijdscategorie 25-50, hebben vaker een niet-Europese migratieachtergrond dan studenten met Nederlandse herkomst. Aangezien jongere studenten door het risicoprofiel een hoger risico werd toebedeeld, werd aan studenten met een niet-Europese migratieachtergrond een lager risico toebedeeld.
- > **Afstand tot ouder(s):** sterk proxykarakter voor studenten met een niet-Europese migratieachtergrond. Uitwonende studenten met een adres dicht bij het ouderlijk adres hebben vaker een niet-Europese migratieachtergrond dan studenten met Nederlandse herkomst. Aangezien studenten die dicht bij het ouderlijk adres stonden ingeschreven door het risicoprofiel een hoger risico werd toebedeeld, werd aan studenten met een niet-Europese migratieachtergrond ook een hoger risico toebedeeld.

Een tweedimensionale analyse, door te kijken naar onder andere het proxykarakter van jonge mbo-studenten (onderwijsvorm en leeftijd gecombineerd) en wo-studenten die dicht bij hun ouders wonen (onderwijs en afstand gecombineerd), bevestigen bovenstaand beeld. Het inverse proxykarakter van leeftijd heft het sterke proxykarakter van onderwijsvorm of afstand niet op.

#### Toelichting onderzoeksvraag 2a

Een kenmerk heeft een proxykarakter als er een verband is tussen het kenmerk en de groep studenten met een niet-Europese migratieachtergrond.

Uit [Figuur 14](#) en [Figuur 20](#) blijkt een duidelijk verband tussen het kenmerk onderwijsvorm en studenten met een niet-Europese migratieachtergrond. Zo zijn in 2014 en 2019 op het mbo 1-2 respectievelijk 63.3% en 72.1% studenten met een niet-Europese migratieachtergrond. Dat neemt in 2014 stapsgewijs af tot respectievelijk 13.2% op het wo. In 2019 neemt dit percentage af tot 38.5% in het mbo 3-4. Het kenmerk onderwijsvorm is dus een proxykenmerk voor het kenmerk niet-Europese migratieachtergrond.

Voor het kenmerk leeftijd is er uitsluitend een verband met studenten met een niet-Europese migratieachtergrond in de leeftijdscategorie 25-50 jaar. In de leeftijdscategorieën 15-18, 19-20 en 21-22 is het aandeel studenten met een niet-Europese migratieachtergrond in 2014 stabiel ([Figuur 15](#)). In de leeftijdsgroep 23-24 jaar loopt dit licht op. In de leeftijdsgroep 25-50 jaar stijgt dit aandeel sterk naar 43.2%. In 2019 is dit verband minder sterk; daar is de stijging van 54.5% naar 67.6% tussen leeftijdscategorie 23-24 en 25-50. Op basis van de data uit 2014 en 2019 volgt dat studenten met een niet-Europese migratieachtergrond ouder zijn dan studenten met een Nederlandse herkomst ([Figuur 21](#)). Het risicoprofiel gaf jongere studenten een hogere risicoscore. Een lage leeftijd is dus geen proxykenmerk ten nadele van studenten met een niet-Europese migratieachtergrond. Wat het kenmerk leeftijd betreft kregen studenten met een niet-Europese migratieachtergrond een lagere risicoscore dan studenten met een Nederlandse herkomst. Hoe dit effect zich dit verhoudt tot de andere kenmerken wordt hieronder nader uitgewerkt.

Er is een duidelijk verband tussen het kenmerk afstand tot ouder(s) en studenten met een niet-Europese migratieachtergrond. Zo is in 2014 en 2019 42.1% á 44.1% van de studenten die 1 meter tot 1 kilometer van hun ouder(s) wonen een student met een niet-Europese migratieachtergrond en loopt dit percentage stapsgewijs af tot 11.1% voor de afstandscategorie 50-500 kilometer in 2014 ([Figuur 16](#)) en 27.9% in 2019 ([Figuur 22](#)). Het kenmerk afstand tot ouder(s) is dus een proxykenmerk voor het kenmerk studenten met een niet-Europese migratieachtergrond.

In [Figuur 17-19](#) (2014) en [Figuur 23-25](#) (2019) zijn de percentages studenten met een niet-Europese migratieachtergrond gegeven voor de verschillende combinaties van twee van de drie kenmerken uit het risicoprofiel. Deze figuren bevestigen het beeld dat voortkomt uit het proxykarakter van de individuele kenmerken. Hierna volgen een aantal voorbeelden uit 2014. Voor 2019 geldt hetzelfde beeld.

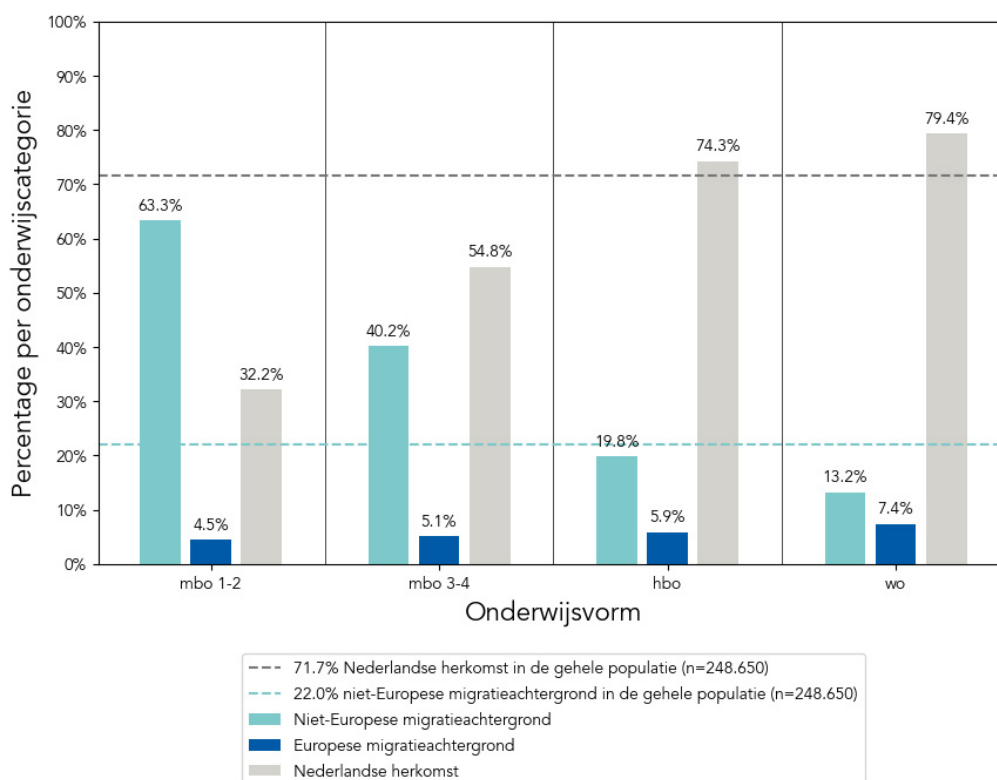
In 2014 is slechts 11.3% van de 19-20-jarige wo-studenten een student met een niet-Europese migratieachtergrond, terwijl 65.8% van de 23-24-jarige mbo 1-2-studenten tot deze migrantengroep behoort ([Figuur 17](#)). Ook bij de combinatie tussen de kenmerken onderwijsvorm en afstand tot ouder(s) valt op dat de gecombineerde kenmerken grotendeels het verwachte proxykarakter van de eendimensionale variant laten zien: van de wo-studenten die 50-500 kilometer van hun ouders wonen heeft slechts 9.5% een niet-Europese migratieachtergrond. Voor mbo 1-2-studenten die 1 meter tot 1 kilometer tot hun ouder(s) wonen behoort 57.5% tot deze groep ([Figuur 18](#)). Opvallend is daarbij wel dat het aandeel studenten met een niet-Europese migratieachtergrond onder mbo 1-2-studenten weinig verandert als de afstand tot ouder(s) toeneemt. Er is weinig verschil tussen deze groep studenten die dicht bij hun ouders wonen (57.5%) en studenten die ver van hun ouders wonen (53.5%) en alle afstandscategorieën daar tussenin. Dit aandeel verschuift wel naar mate studenten verder van hun ouderlijk huis wonen bij alle andere onderwijsvormen. Als leeftijd wordt afgezet tegen afstand tot ouder(s) ontstaat een wat meer gefragmenteerd beeld ([Figuur 19](#)). Over het algemeen is

het aandeel studenten met een niet-Europese migratieachtergrond laag onder jonge studenten die ver van hun ouders wonen (tussen de 8.0-10.6%) en hoog onder studenten die dicht bij hun ouders wonen (tussen de 44.1-55.2%). Voor de afstandscategorie springt het grote aandeel studenten met een niet-Europese migratieachtergrond in het oog in de onbekende categorie. Er wordt een verklaring gegeven waarom dit aandeel zo hoog is bij beantwoording van onderzoeksvraag 2. Uit onderstaande figuren volgt een vergelijkbaar beeld voor mbo-studenten uit de CUB-populatie van 2019. Voor oudere mbo 1-2-studenten is het aandeel studenten met een niet-Europese migratieachtergrond hoog (Figuur 23). Voor mbo 1-2 studenten speelt de afstand tot ouder(s) een beperkte rol om het aandeel studenten met een niet-Europese migratieachtergrond te duiden (Figuur 24). Behalve voor de onbekende afstandscategorie, waarvoor bij het beantwoorden van onderzoeksvraag 2 een verklaring is gegeven. Voor de gehele mbo-studentenpopulatie geldt dat oudere studenten die verder van het ouderlijk adres wonen minder vaak een niet-Europese migratieachtergrond hebben. Juist jongere studenten die dichtbij hun ouders wonen behoren doorgaans vaker tot de groep studenten met een niet-Europese migratieachtergrond (Figuur 25).

## Diagrammen en cijfers 2014

### Onderwijsvorm

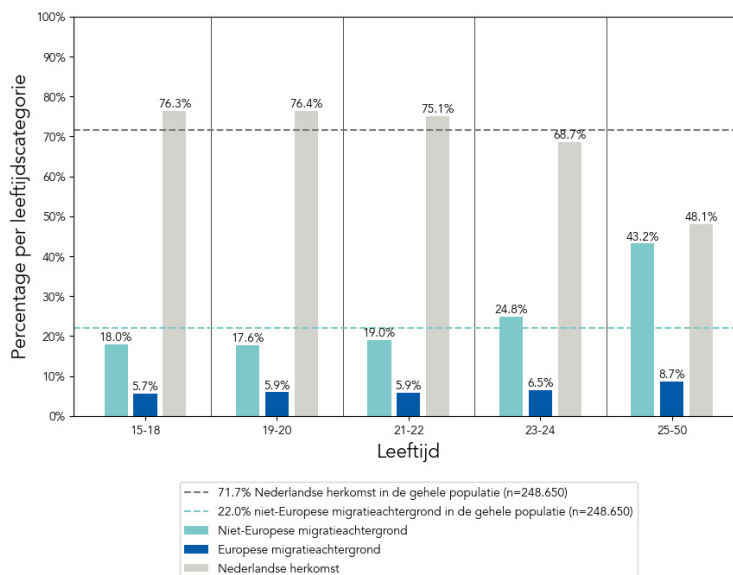
Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per onderwijsvorm in de uitwonendenbeurspopulatie-2014 (n=248.650)



Figuur 14 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per onderwijsvorm in de uitwonendenbeurspopulatie-2014 (n=248.650)

## Leeftijd

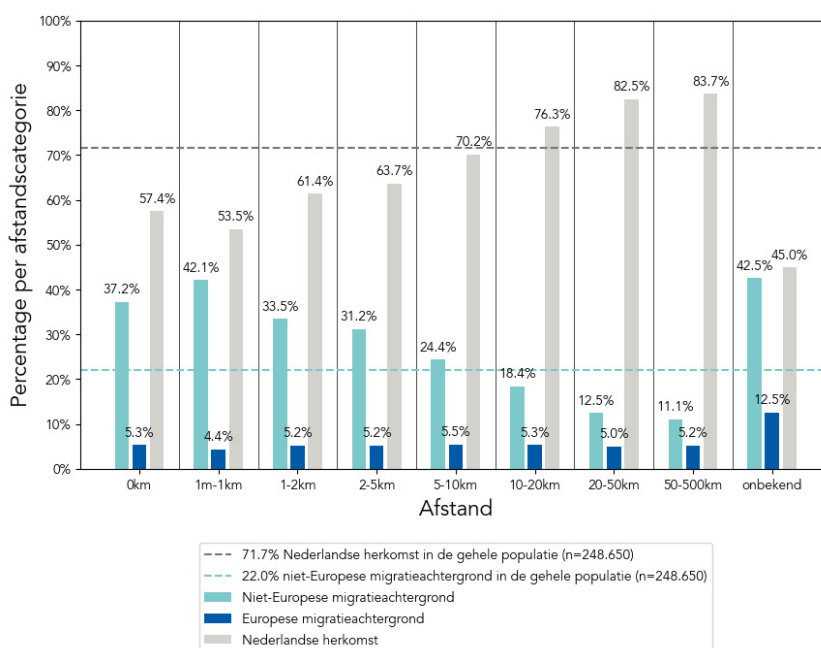
Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per leeftijdscategorie in de uitwonendenbeurspopulatie-2014 (n=248.650)



Figuur 15 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per leeftijdscategorie in de uitwonendenbeurspopulatie-2014 (n=248.650)

## Afstand tot ouder(s)

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per afstandscategorie in de uitwonendenbeurspopulatie-2014 (n=248.650)



Figuur 16 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per afstandscategorie in de uitwonendenbeurspopulatie-2014 (n=248.650)

### Onderwijsvorm in combinatie met leeftijd

Verdeling van studenten met een niet-Europese migratieachtergrond per combinatie onderwijs- en leeftijdscategorie in de uitwonendenbeurspopulatie-2014 (n=248.650)

	15-18	19-20	21-22	23-24	25-50	gemiddelde
mbo 1-2	49.1%	57.5%	63.3%	65.8%	79.5%	63.3%
mbo 3-4	31.8%	33.3%	39.0%	46.3%	58.4%	40.2%
hbo	11.9%	13.5%	17.4%	22.9%	36.8%	19.8%
wo	10.2%	11.3%	12.2%	16.5%	29.0%	13.2%
gemiddelde	18.0%	17.6%	19.0%	24.8%	43.2%	

Figuur 17 – Verdeling van studenten met een niet-Europese migratieachtergrond per combinatie onderwijs- en leeftijdscategorie in de uitwonendenbeurspopulatie-2014 (n=248.650)

### Onderwijsvorm in combinatie met afstand tot ouder(s)

Verdeling van studenten met een niet-Europese migratieachtergrond per combinatie onderwijs- en afstandscategorie in de uitwonendenbeurspopulatie-2014 (n=248.650)

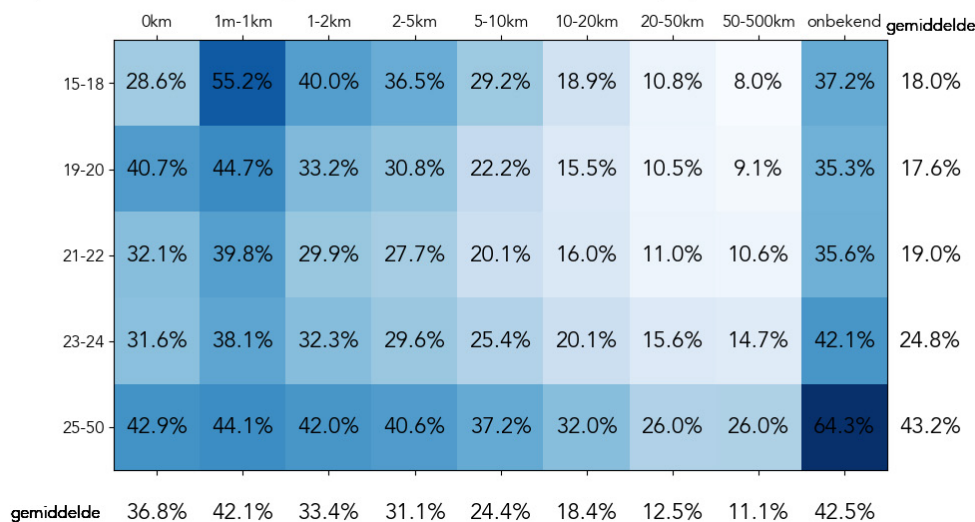
	0km	1m-1km	1-2km	2-5km	5-10km	10-20km	20-50km	50-500km	onbekend	gemiddelde
mbo 1-2	50.0%	57.5%	52.6%	56.4%	50.5%	45.5%	45.5%	53.5%	85.3%	63.3%
mbo 3-4	46.2%	49.9%	42.7%	42.6%	36.5%	30.4%	24.2%	23.1%	62.4%	40.2%
hbo	34.1%	36.2%	28.5%	26.0%	20.2%	16.2%	10.8%	10.4%	40.2%	19.8%
wo	25.0%	32.5%	21.5%	19.1%	15.1%	12.0%	10.2%	9.5%	24.6%	13.2%
gemiddelde	36.8%	42.1%	33.4%	31.1%	24.4%	18.4%	12.5%	11.1%	42.5%	

Figuur 18 – Verdeling van studenten met een niet-Europese migratieachtergrond per onderwijs- en afstandscategorie in de uitwonendenbeurspopulatie-2014 (n=248.650)



### Leeftijd in combinatie met afstand tot ouder(s)

Verdeling van studenten met een niet-Europese migratieachtergrond per combinatie leeftijds- en afstands categorie in de uitwonendenbeurspopulatie-2014 (n=248.650)

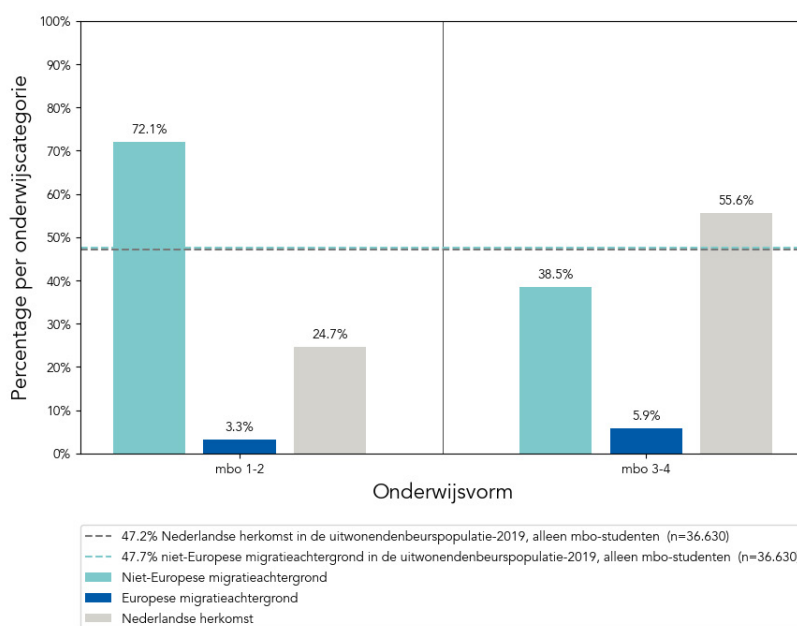


Figuur 19 – Verdeling van studenten met een niet-Europese migratieachtergrond per leeftijds- en afstands categorie in de uitwonendenbeurspopulatie-2014 (n=248.650)

### Diagrammen en cijfers 2019

#### Onderwijsvorm

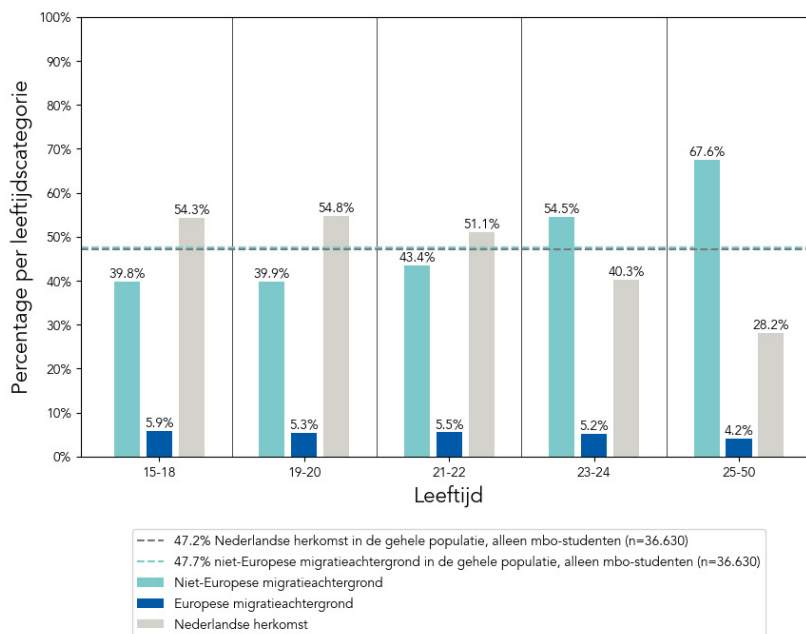
Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per onderwijsvorm in de uitwonendenbeurspopulatie-2019 (n=36.630)



Figuur 20 – Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per onderwijsvorm in de uitwonendenbeurspopulatie-2019 (n=36.630)

## Leeftijd

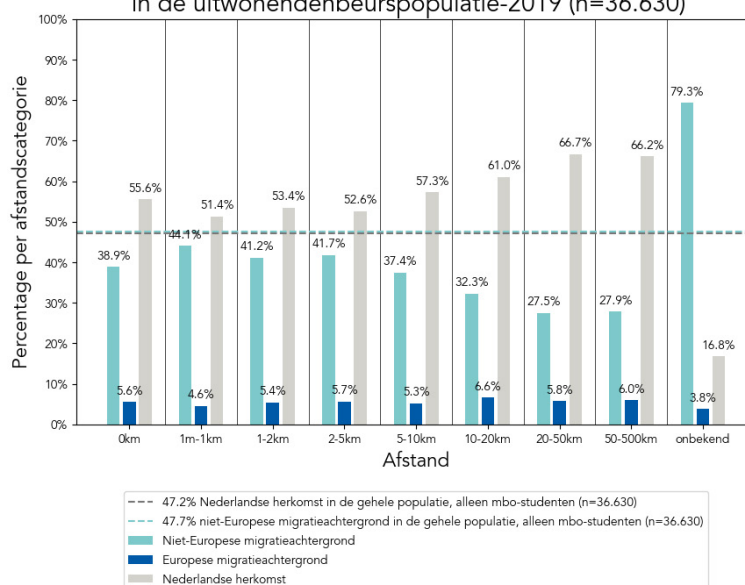
Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per leeftijdscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)



Figuur 21 – Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per leeftijdscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)

## Afstand tot ouder(s)

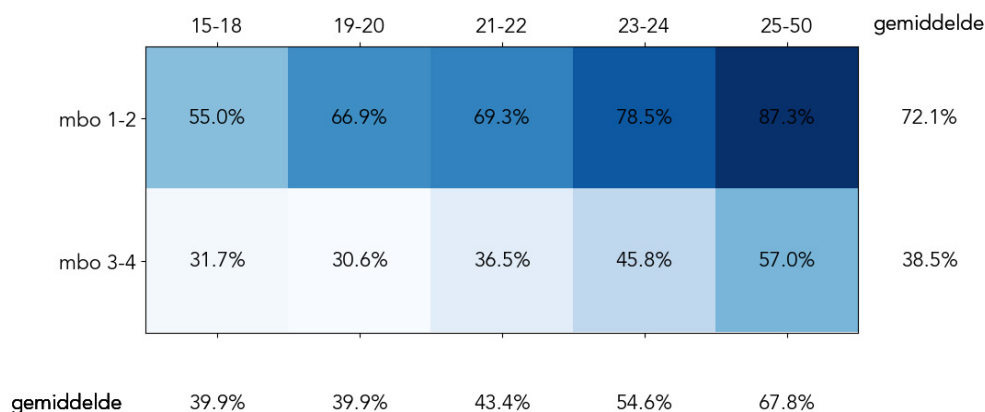
Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per afstandscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)



Figuur 22 – Verdeling van mbo-studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per afstandscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)

### Onderwijsvorm in combinatie met leeftijd

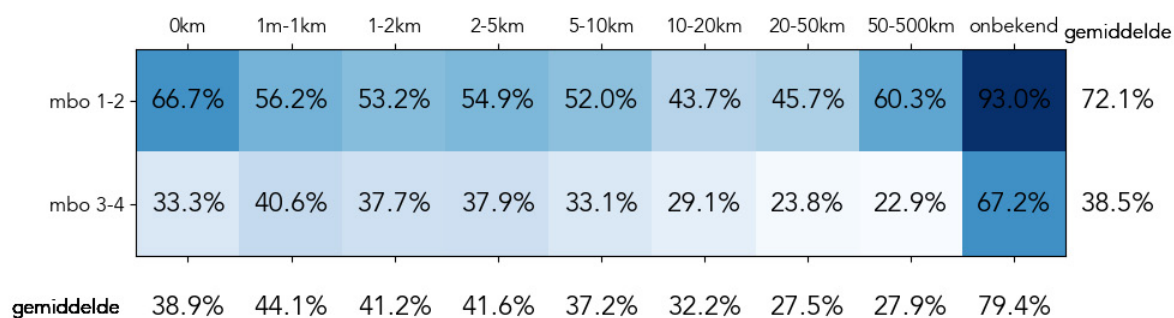
Verdeling van mbo-studenten met een niet-Europese migratieachtergrond per combinatie onderwijsvorm en leeftijdscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)



Figuur 23 – Verdeling van mbo-studenten met een niet-Europese migratieachtergrond per combinatie onderwijsvorm en leeftijdscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)

### Onderwijsvorm in combinatie met afstand tot ouder(s)

Verdeling van mbo-studenten met een niet-Europese migratieachtergrond per combinatie onderwijsvorm en afstandscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)



Figuur 24 – Verdeling van mbo-studenten met een niet-Europese migratieachtergrond per combinatie onderwijsvorm en afstandscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)

### Leeftijd in combinatie met afstand tot ouder(s)

Verdeling van mbo-studenten met een niet-Europese migratieachtergrond per combinatie leeftijds- en afstandscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)

	0km	1m-1km	1-2km	2-5km	5-10km	10-20km	20-50km	50-500km	onbekend	gemiddelde
15-18	33.3%	50.0%	41.7%	38.3%	38.9%	27.8%	22.0%	20.0%	66.3%	39.9%
19-20	28.6%	45.2%	38.9%	38.6%	31.4%	27.7%	21.2%	21.7%	70.5%	39.9%
21-22	50.0%	37.9%	33.3%	33.3%	29.0%	24.2%	21.0%	16.9%	66.0%	43.4%
23-24	25.0%	35.5%	31.9%	32.8%	31.2%	22.1%	17.4%	16.7%	65.8%	54.6%
25-50	25.0%	42.2%	39.7%	39.3%	37.5%	34.2%	27.1%	27.7%	79.8%	67.8%
gemiddelde	38.9%	44.1%	41.2%	41.6%	37.2%	32.2%	27.5%	27.9%	79.4%	

Figuur 25 – Verdeling van mbo-studenten met een niet-Europese migratieachtergrond per combinatie leeftijds- en afstandscategorie in de uitwonendenbeurspopulatie-2019 (n=36.630)

## 4.4 Resultaten onderzoeksvraag 3 (handmatige selectie)

### Onderzoeksvraag 3

Zijn studenten met een migratieachtergrond in 2014 en 2019 vaker geselecteerd voor controle tijdens handmatige selectie dan studenten met een Nederlandse herkomst?

### Antwoord op onderzoeksvraag 3

Ja. Studenten met een niet-Europese migratieachtergrond zijn in 2014 en 2019 vaker geselecteerd voor controle tijdens handmatige selectie dan studenten met een Nederlandse herkomst. In 2014 hadden studenten met een niet-Europese migratieachtergrond 6.2x meer kans om geselecteerd te worden voor een controle tijdens bij handmatige selectie (stap 3 van het CUB-proces) dan studenten met een Nederlandse herkomst. Voor 2019 was deze kans 3.6x hoger.

### Toelichting onderzoeksvraag 3

De resultaten worden toegelicht gebaseerd op de beschikbare CUB-data voor 2014 en 2019.

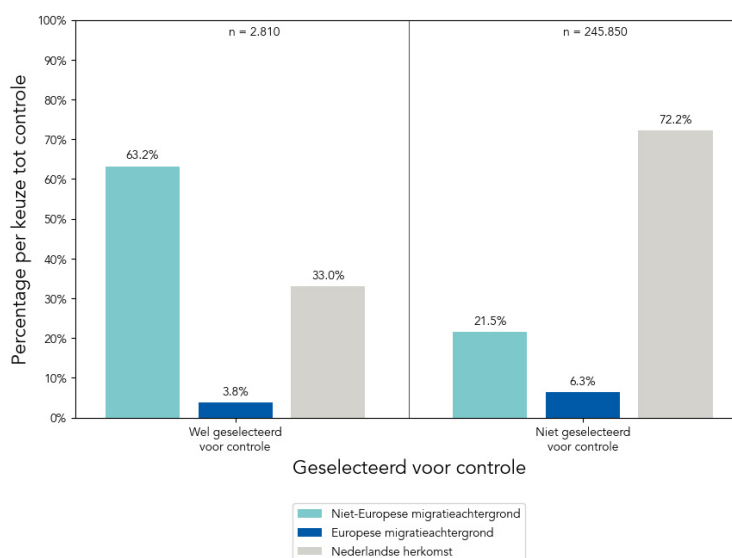
In [Figuur 26](#) wordt voor de gehele uitwonendenbeurspopulatie-2014 weergegeven welke groepen studenten bij handmatige selectie in stap 3 van het CUB-proces geselecteerd zijn voor een controle. Van de 248.650 studenten die in 2014 een uitwonendenbeurs hebben aangevraagd zijn 2.810 studenten geselecteerd voor een controle (1.1%). Van de 2.810 studenten die zijn geselecteerd voor een controle had 63.2% een niet-Europese migratieachtergrond. Dat is een sterke oververtegenwoordiging, zowel ten aanzien van de startpopulatie (zie onderzoeksvraag 1) als ten aanzien van de scheve verhoudingen die uit toepassingen van het risicoprofiel zijn voortgekomen (zie onderzoeksvraag 2). Van de 248.650 studenten die niet zijn geselecteerd voor een controle heeft 21.5% een niet-Europese migratieachtergrond.

In 2019 is de verdeling vergelijkbaar.<sup>24</sup> Van de 740 studenten die zijn geselecteerd voor een controle had 69.1% een niet-Europese migratieachtergrond ([Figuur 27](#)). Van de 49.500 studenten die niet zijn geselecteerd voor een controle had 38.4% een niet-Europese migratieachtergrond. Deze cijfers liggen dicht bij elkaar dan in 2014 aangezien een groter deel van de gehele uitwonendenbeurspopulatie een niet-Europese migratieachtergrond had als gevolg van het grotere aandeel mbo-studenten.

Voor zowel 2014 als 2019 zijn de verschuivingen van het aandeel studenten met een niet-Europese migratieachtergrond na handmatige selectie voor controle in stap 3 van het CUB-proces sterk. Gebaseerd op de CUB-data uit 2014 verschuift het aandeel studenten met een niet-Europese migratieachtergrond na stap 1 van het CUB-proces (risicoprofiel) van 36.0% naar 63.2%. Voor 2019 geldt een toename van 37.7% naar 69.1%.

## 2014

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst geselecteerd voor controle in de uitwonendenbeurspopulatie-2014 (n=248.650)

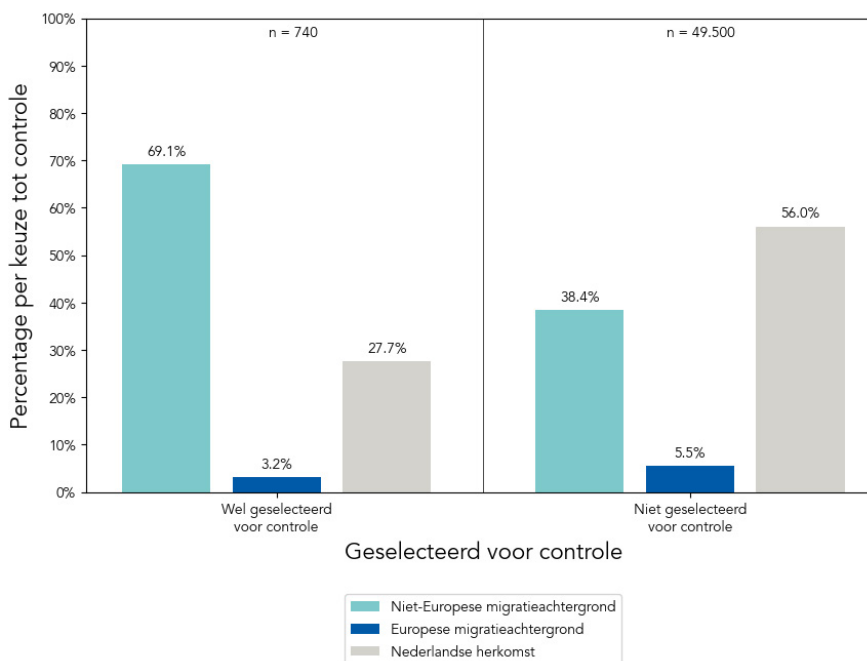


**Figuur 26** – Verdeling van studenten met een (niet-)Europese migratieachtergrond geselecteerd voor controle in de uitwonendenbeurspopulatie-2014 (n=248.650)

<sup>24</sup> Voor de volledigheid wordt opgemerkt dat dit de cijfers zijn voor alle studenten die in 2019 een uitwonendenbeurs ontvingen; niet alleen de mbo-studenten zoals in enkele andere delen van dit addendum, maar ook uitfaserende wo- en hbo-studenten die voor introductie van het leenstelsel in 2015 al recht hadden op een uitwonendenbeurs. Mbo-studenten konden voor deze stap technisch niet gemakkelijk gescheiden worden van de gehele populatie.

2019

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst geselecteerd voor controle uit de uitwonendenbeurspopulatie-2019 (n=50.230)



Figuur 27 – Verdeling van studenten met een (niet-)Europese migratieachtergrond geselecteerd voor controle in de uitwonendenbeurspopulatie-2019 (n=50.230)

## 4.5 Resultaten onderzoeksvraag 3a (mate waarin voorspelling algoritme wordt gevolgd)

### Onderzoeksvraag 3a

Wat is de mate waarin de uitkomst van het risicoprofiel en de handmatige selectie voor huisbezoek overeenkomt, in het bijzonder voor studenten met een migratieachtergrond?

### Antwoord op onderzoeksvraag 3a

Binnen alle risicocategorieën worden studenten met een migratieachtergrond in verhouding vaker geselecteerd voor een huisbezoek dan studenten met een Nederlandse herkomst. In 2014 hadden studenten met een niet-Europese migratieachtergrond die als hoog risico zijn geclassificeerd door het risicoprofiel 5.5x meer kans om geselecteerd te worden voor controle tijdens handmatige selectie dan studenten met Nederlandse herkomst en dezelfde risicocategorie. Studenten met een niet-Europese migratieachtergrond die als laag risico zijn geclassificeerd door het risicoprofiel hadden 1.8x meer kans om alsnog geselecteerd te worden voor controle dan studenten met Nederlandse herkomst en dezelfde risicocategorie. Studenten met een niet-Europese migratieachtergrond die als onbekend risico zijn geclassificeerd door het risicoprofiel hadden 2.3x meer kans om alsnog geselecteerd te worden voor controle dan studenten met Nederlandse herkomst en dezelfde risicocategorie. Voor 2019 zijn deze aantallen respectievelijk 3.6x (hoog risico), 4.2x (laag risico) en 1.0x (onbekend risico) hoger dan studenten met Nederlandse herkomst en dezelfde risicocategorie.

### Toelichting onderzoeksvraag 3a

In 2014 zijn 42.800 studenten door het risicoprofiel ingedeeld in de groep hoog risico. 36.0% van die studenten had een niet-Europese migratieachtergrond. Niet alle studenten die zijn ingedeeld in de groep hoog risico zijn geselecteerd voor een controle; 2.400 studenten zijn dat wel, waarvan 73.9% een niet-Europese migratieachtergrond heeft. 40.400 studenten die als hoog risico zijn geclassificeerd zijn niet geselecteerd voor een controle, waarvan 33.7% behoorde tot de groep niet-Europese migratieachtergrond (Figuur 28). Van de 171.810 studenten die door het risicoprofiel ingedeeld zijn in de groep laag risico zijn 640 studenten wel en 171.180 studenten niet handmatig geselecteerd voor controle. Van deze groepen behoorde respectievelijk 25.0% en 15.3% tot de groep niet-Europese migratieachtergrond (Figuur 29). Dit is een aanzienlijk mildere afwijking dan voor de hoog risico categorie. Van de 34.050 studenten die door het risicoprofiel ingedeeld zijn in de groep onbekend risico zijn 140 studenten wel en 33.910 studenten niet handmatig geselecteerd voor controle. Van deze groepen behoorde respectievelijk 60.0% en 38.2% tot de groep niet-Europese migratieachtergrond (Figuur 30). Dit is een sterkere afwijking dan waargenomen voor de lage risico categorie, maar niet een dermate sterke nadelige afwijking voor studenten met een niet-Europese migratieachtergrond die door het risicoprofiel als hoog risico zijn geclassificeerd.

Voor 2019 volgt voor zowel studenten die als hoog en laag risico zijn geclassificeerd door het risicoprofiel een afwijking ten nadele van studenten met een niet-Europese migratieachtergrond tijdens handmatige selectie. Van de 14.850 studenten die door het risicoprofiel ingedeeld zijn in de groep hoog risico zijn 840 studenten wel en 14.010 studenten niet handmatig geselecteerd voor controle. Van deze groepen behoorde respectievelijk 67.9% en 35.9% tot de groep niet-Europese migratieachtergrond (Figuur 31). Van de 24.130 studenten die door het risicoprofiel ingedeeld zijn in de groep laag risico zijn 60 studenten wel en 24.070 studenten niet handmatig geselecteerd voor controle. Van deze groepen behoorde respectievelijk 60.0% en 24.8% tot de groep niet-Europese migratieachtergrond (Figuur 32). De afwijking voor deze groepen zit in dezelfde orde grootte. Voor de studenten die door het risicoprofiel als onbekend risico zijn geclassificeerd volgt een ander beeld. Van de 11.260 studenten die door het risicoprofiel ingedeeld zijn in de groep onbekend risico zijn 40 studenten wel en 11.220 studenten niet handmatig geselecteerd voor controle. Van deze groepen behoorde respectievelijk 75.0% en 70.7% tot de groep niet-Europese migratieachtergrond (Figuur 33).

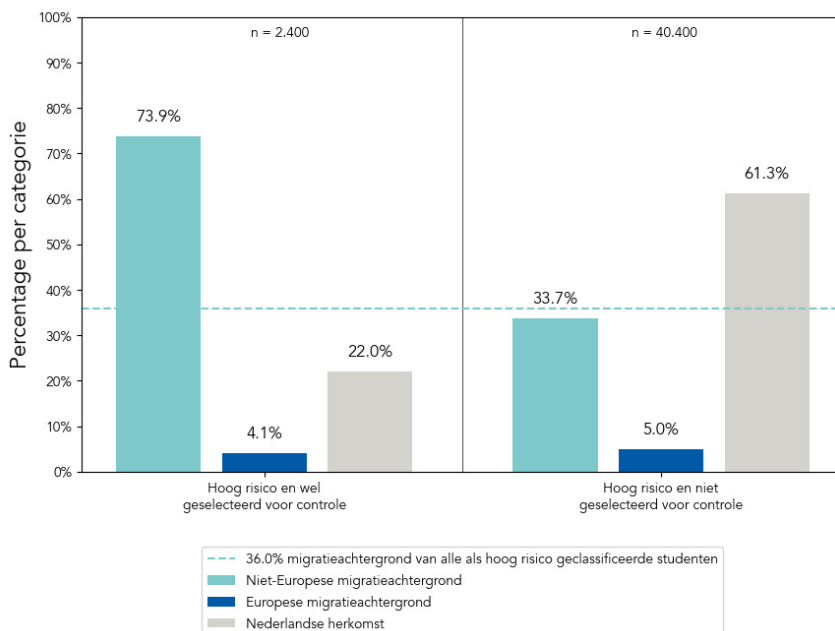
Door bovenstaande percentages te relateren aan het gemiddelde aandeel studenten met een niet-Europese migratieachtergrond per risicocategorie kunnen de verhoogde kansen worden bepaald voor deze groep om handmatig te worden geselecteerd voor controle. Aangezien al deze kansen (zoals gegeven hierboven in het antwoord op de onderzoeksvraag) vaak aanzienlijk groter dan 1.0x zijn kan worden geconcludeerd dat vooringenomenheid ten aanzien van studenten met een niet-Europese migratieachtergrond zich voor heeft gedaan tijdens handmatige selectie voor controle.

Er zijn uit het onderzoek geen aanwijzingen naar voren gekomen dat deze vooringenomenheid door menselijke vooringenomenheid van individuele ambtenaren komt. Vooringenomenheid kan ook voortkomen uit de aard van de werkinstructies, zoals de uitsluiting van studentenhuizen van huisbezoek, of andere institutionele oorzaken. De vraag hoe de vooringenomenheid in de handmatige selectie ontstond, valt buiten de reikwijdte van dit onderzoek.



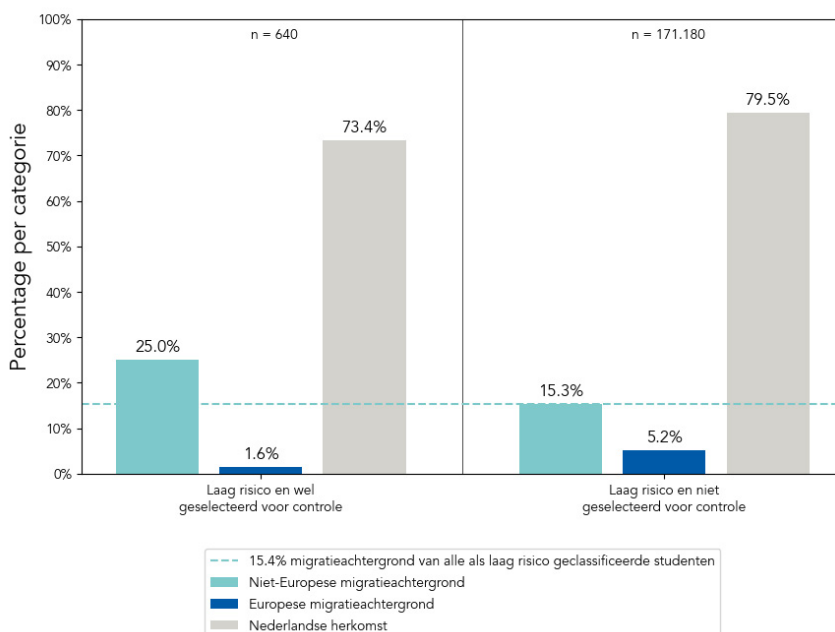
2014

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst geassocieerd als hoog risico door profiel en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2014 (n=42.800)



Figuur 28 – Verdeling van studenten met een (niet-)Europese migratieachtergrond geassocieerd als hoog risico en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2014 (n=42.800)

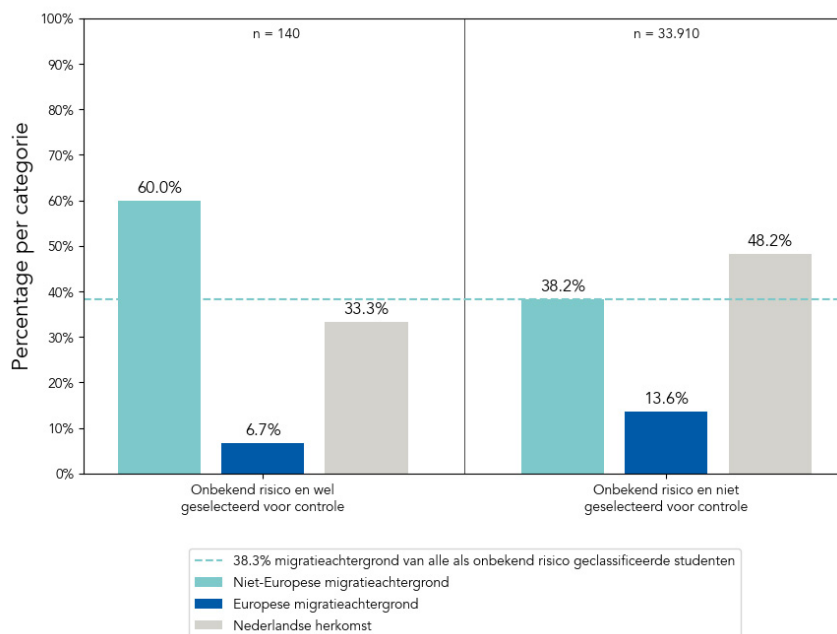
Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst geassocieerd als laag risico door profiel en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2014 (n=171.810)



Figuur 29 – Verdeling van studenten met een (niet-)Europese migratieachtergrond geassocieerd als laag risico en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2014 (n=171.810)



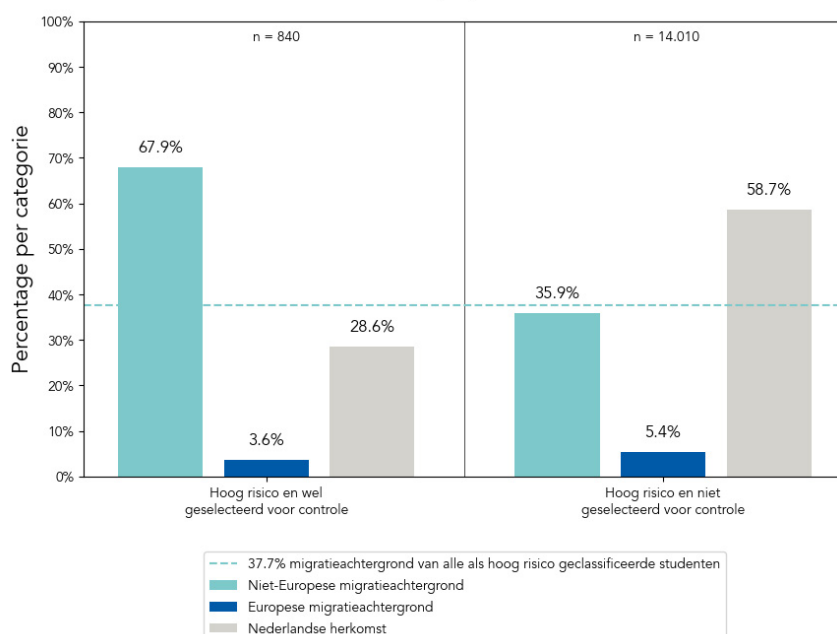
Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst geassocieerd als onbekend risico door profiel en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2014 (n=34.050)



Figuur 30 – Verdeling van studenten met een (niet-)Europese migratieachtergrond geassocieerd als onbekend risico en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2014 (n=34.050)

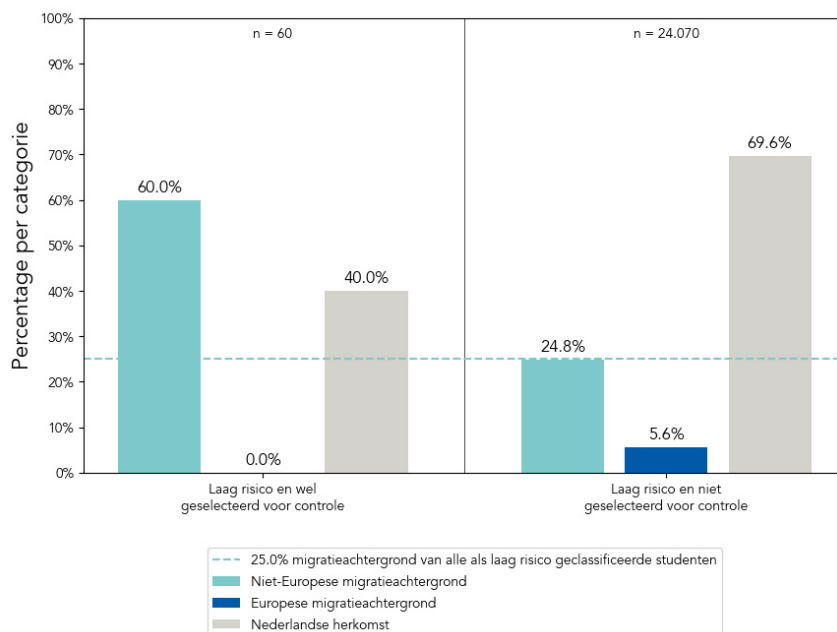
2019

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst geassocieerd als hoog risico door profiel en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2019 (n=14.850)



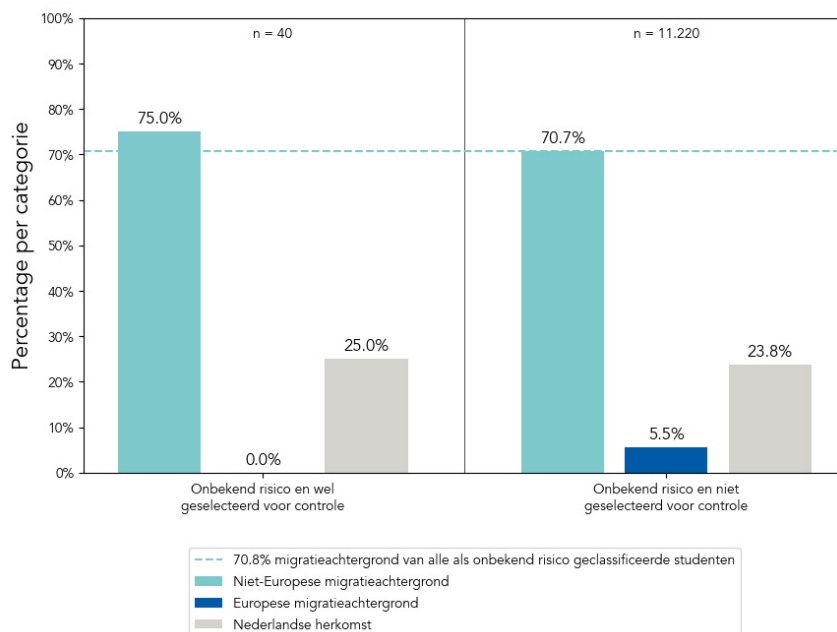
Figuur 31 – Verdeling van studenten met een (niet-)Europese migratieachtergrond geassocieerd als hoog risico en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2019 (n=14.850)

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst geassocieerd als laag risico door profiel en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2019 (n=24.130)



Figuur 32 – Verdeling van studenten met een (niet-)Europese migratieachtergrond geassocieerd als laag risico en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2019 (n=24.130)

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst geassocieerd als onbekend risico door profiel en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2019 (n=11.260)



Figuur 33 – Verdeling van studenten met een (niet-)Europese migratieachtergrond geassocieerd als onbekend risico en wel/niet geselecteerd voor controle in de uitwonendenbeurspopulatie-2019 (n=11.260)

## 4.6 Resultaten onderzoeksvraag 4 (onrechtmatig gebruik)

### Onderzoeksvraag 4

*Wat was in 2014 en 2019 het aandeel studenten met een migratieachtergrond in de onrechtmatig gebruik-populatie?*

### Antwoord op onderzoeksvraag 4

Gebaseerd op de uitwonendenbeurspopulatie-2014 had 86.3% van de studenten waarvan door DUO onrechtmatig gebruik is vastgesteld een niet-Europese migratieachtergrond. Gebaseerd op de uitwonendenbeurspopulatie-2019 was dit 75.8%.

Gebaseerd op de uitwonendenbeurspopulatie-2014 had van de groep studenten waarvan na een huisbezoek is vastgesteld dat zij rechtmatig gebruik maakte van de uitwonendenbeurs 40.5% een migratieachtergrond. Gebaseerd op de uitwonendenbeurspopulatie-2019 was dit aandeel 56.1%.

### Toelichting onderzoeksvraag 4

Gebaseerd op de beschikbare CUB-data uit 2014 en 2019 (Figuur 34-35) kan worden geconcludeerd dat van meer studenten met een niet-Europese migratieachtergrond na een huisbezoek door DUO is vastgesteld dat zij onrechtmatig gebruik maakten van de uitwonendenbeurs. Of dit ook betekent dat studenten met een migratieachtergrond vaker onrechtmatig gebruik maken van de uitwonendenbeurs kan niet worden vastgesteld op basis van de beschikbare data. Er ontstaat door de stappen van het CUB-proces heen een vergrootglaseffect met betrekking tot de groep met een niet-Europese migratieachtergrond. Het mogelijke aandeel van een hoger onrechtmatig gebruik-percentage van studenten met een niet-Europese migratieachtergrond kan niet worden geïsoleerd van de vooringenomenheid van het CUB-proces in stap 1 (risicoprofiel) en stap 3 (handmatige selectie). Daarnaast is de aselechte steekproef te klein om dit onafhankelijk van het CUB-proces te meten. Of er naast het vergrootglaseffect verdere vooringenomenheid schuilt in het proces van de huisbezoeken zelf kon eveneens niet worden onderzocht.

In aanvulling op bovenstaande bevinding wordt voor 2014 en 2019 in [Figuur 36](#) en [Figuur 37](#) de verdeling van de uitkomst van huisbezoeken per migrantengroepen gegeven. In 2014 is bij 2.810 studenten een huisbezoek afgelegd. Op basis daarvan is van 1.570 studenten vastgesteld dat zij toch rechtmatig gebruik maakten van de uitwonendenbeurs. Van die groep heeft 45.5% een niet-Europese migratieachtergrond. Hieruit blijkt niet dat studenten met een niet-Europese migratieachtergrond en studenten met Nederlandse herkomst gelijke kans hebben op een zogenaamd 'onterecht' huisbezoek, hiervoor zouden immers de verhoudingen uit de basispopulatie aangehouden moeten worden (22% vs. 71.7%).<sup>25</sup> Van 1.240 studenten is vastgesteld dat zij onrechtmatig gebruik maakten van de uitwonendenbeurs. Daarvan heeft 86.3% een niet-Europese migratieachtergrond.

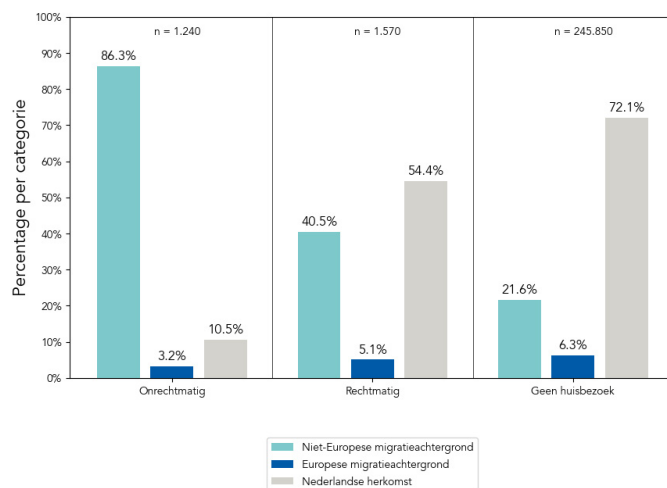
In 2019 is bij 740 studenten een huisbezoek afgelegd. Van de 410 studenten is aan de hand van dit bezoek vastgesteld dat zij rechtmatig gebruik hebben gemaakt van de uitwonendenbeurs. Van deze groep had 59.1% een niet-Europese migratieachtergrond. Van de 330 studenten waarvan DUO onrechtmatig gebruik heeft vastgesteld had 75.8% een migratieachtergrond. Ook voor 2019 geldt dat studenten met een niet-

<sup>25</sup> 'Onterecht huisbezoek' is een werknaam voor een huisbezoek waaruit later geen onrechtmatig gebruik bleek

Europese migratieachtergrond en studenten met Nederlandse herkomst geen gelijke kans hadden op een onterecht huisbezoek, hiervoor zouden de verhoudingen uit de basispopulatie aangehouden moeten worden (22% vs. 71.7%).

## 2014

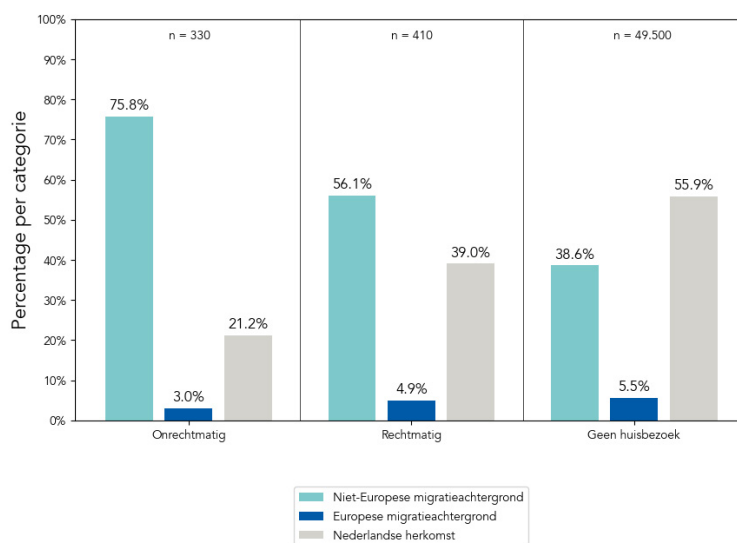
Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst per uitkomst huisbezoek in de uitwonendenbeurspopulatie-2014 (n=248.650)



Figuur 34 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst per uitkomst huisbezoek in de uitwonendenbeurspopulatie-2014 (n=248.650)

## 2019

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst per uitkomst huisbezoek in de uitwonendenbeurspopulatie-2019 (n=50.230)

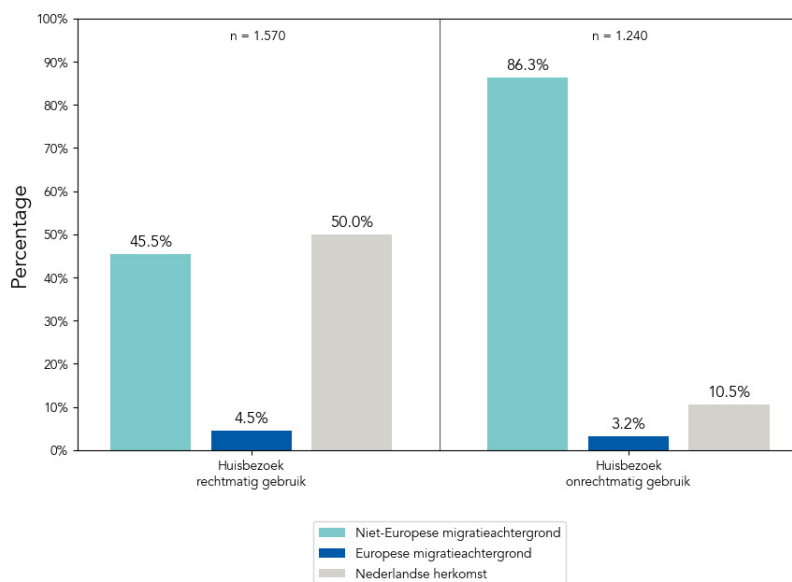


Figuur 35 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst per uitkomst huisbezoek in de uitwonendenbeurspopulatie-2019 (n=50.230)

## Welke groepen hebben terecht/onterecht een huisbezoek gekregen?

2014

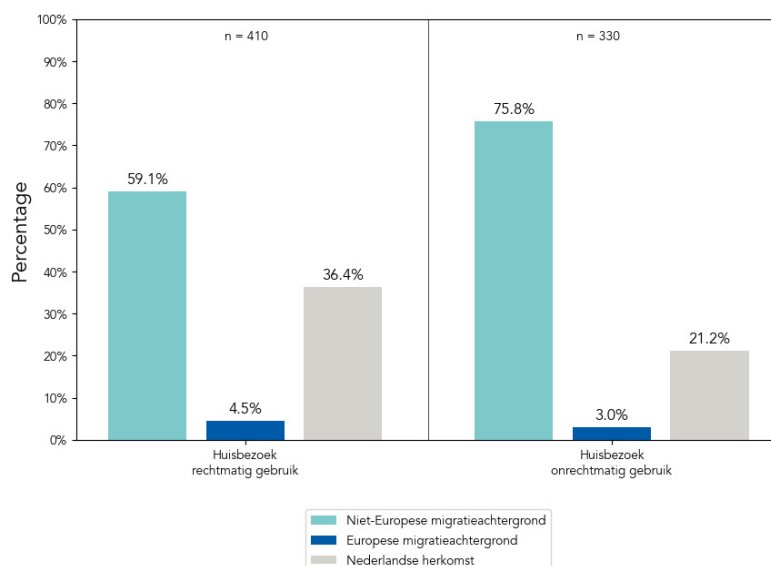
Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst die terecht/onterecht een huisbezoek hebben gehad in de huisbezoekpopulatie-2014 (n=2.810)



Figuur 36 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst die terecht/onterecht een huisbezoek hebben gehad in de uitwonendenbeurspopulatie-2014 (n=2.810)

2019

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst die terecht/onterecht een huisbezoek hebben gehad in de huisbezoekpopulatie-2019 (n=740)



Figuur 37 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met een Nederlandse herkomst die terecht/onterecht een huisbezoek hebben gehad in de uitwonendenbeurspopulatie-2019 (n=740)

## 4.7 Resultaten onderzoeksvraag 5 (bezwaarpopulaties)

### Onderzoeksvraag 5

Wat is in 2014, 2019, 2021 en 2022 de verdeling van studenten met een migratieachtergrond geweest voor selectie voor bezwaarpopulatie?

### Antwoord op onderzoeksvraag 5

Voor de jaren 2014, 2019, 2021 en 2022 behoort respectievelijk 85.2%, 79.3%, 80.0% en 82.6% van de studenten uit de bezwaarpopulatie tot de groep studenten met een niet-Europese migratieachtergrond.

### Toelichting onderzoeksvraag 5

De resultaten zijn gebaseerd op de CBS-data met gegevens over de bezwaarpopulatie-2014, -2019, -2021 en -2022. De resultaten worden per jaar toegelicht.

Uit [Figuur 38](#) volgt dat 85.2% van de studenten uit de bezwaarpopulatie-2014 behoort tot de groep studenten met een niet-Europese migratieachtergrond. Uit [Figuur 39](#) volgt dat deze 85.2% voor 70.3% bestaat uit kinderen van een niet-Europese migrant en voor 14.8% uit niet-Europese migranten.

Uit [Figuur 40](#) volgt dat 79.3% van de studenten uit de bezwaarpopulatie-2019 behoort tot de groep studenten met een niet-Europese migratieachtergrond. Uit [Figuur 41](#) volgt dat deze 79.3% voor 62.1% bestaat uit kinderen van een niet-Europese migrant en voor 17.2% uit niet-Europese migranten.

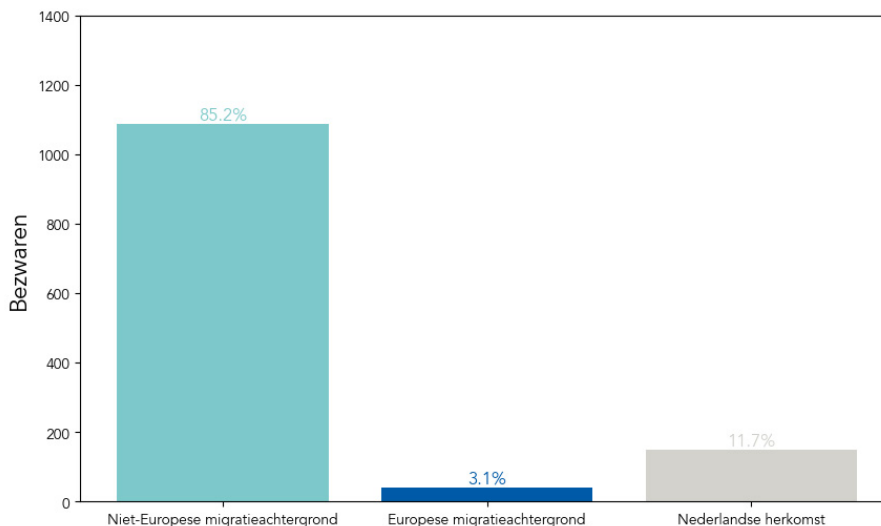
Uit [Figuur 42](#) volgt dat 80.0% van de studenten uit de bezwaarpopulatie-2021 behoort tot de groep studenten met een niet-Europese migratieachtergrond. Uit [Figuur 43](#) volgt dat deze 80.0% voor 65.0% bestaat uit kinderen van een niet-Europese migrant en voor 15.0% uit niet-Europese migranten.

Uit [Figuur 44](#) volgt dat 82.6% van de studenten uit de bezwaarpopulatie-2022 behoort tot de groep studenten met een niet-Europese migratieachtergrond. Uit [Figuur 45](#) volgt dat deze 82.6% voor 60.9% bestaat uit kinderen van een niet-Europese migrant en voor 21.7% uit niet-Europese migranten.

De bezwaarpopulaties bevat dus een disproportioneel aandeel kinderen van een niet-Europese migrant vergeleken met de startpopulatie (22.0% studenten met een niet-Europese migratieachtergrond) wat verklaard kan worden door vooringenomenheid ten aanzien van deze groep in verschillende stappen van het CUB-proces.

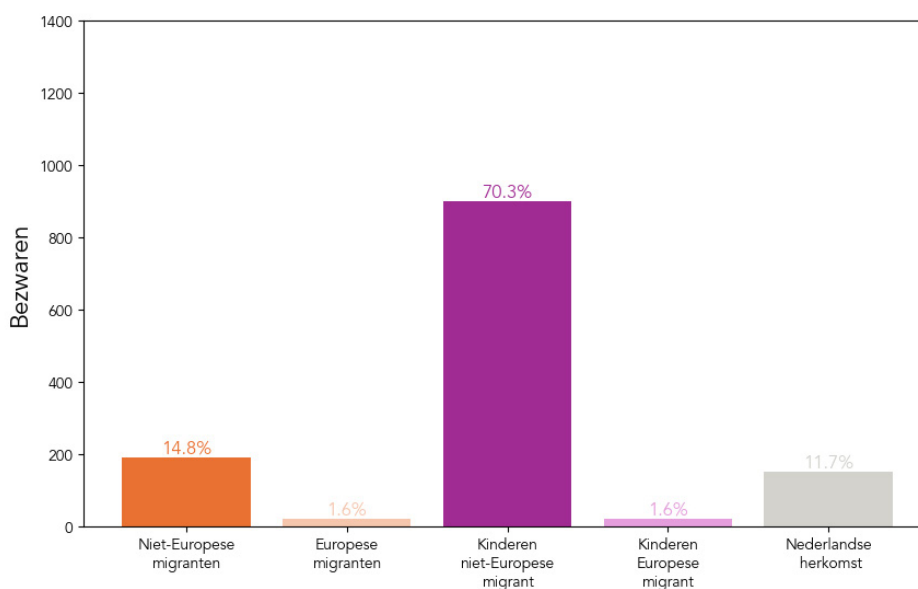
2014

Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de bezwaarpopulatie-2014 (n=1.290)



Figuur 38 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de bezwaarpopulatie-2014 (n=1.290)

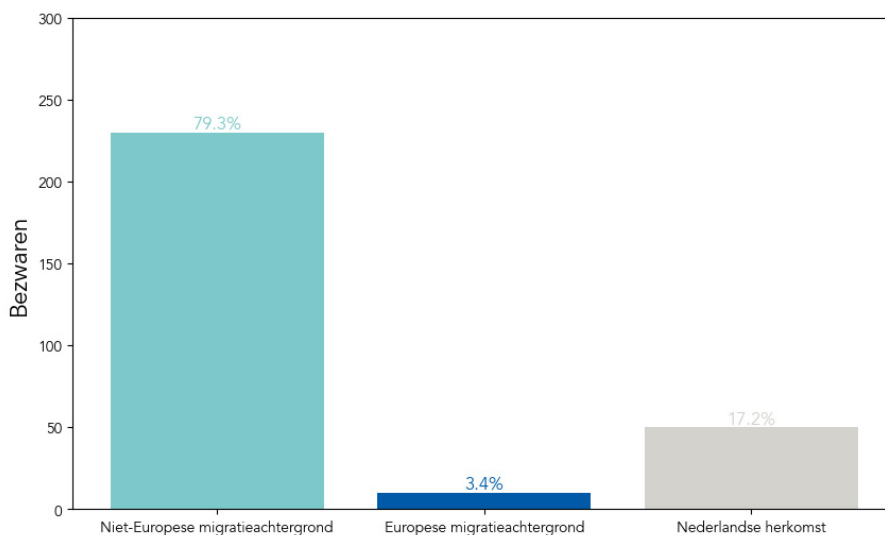
Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst in de bezwaarpopulatie-2014 (n=1.290)



Figuur 39 – Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst in de bezwaarpopulatie-2014 (n=1.290)

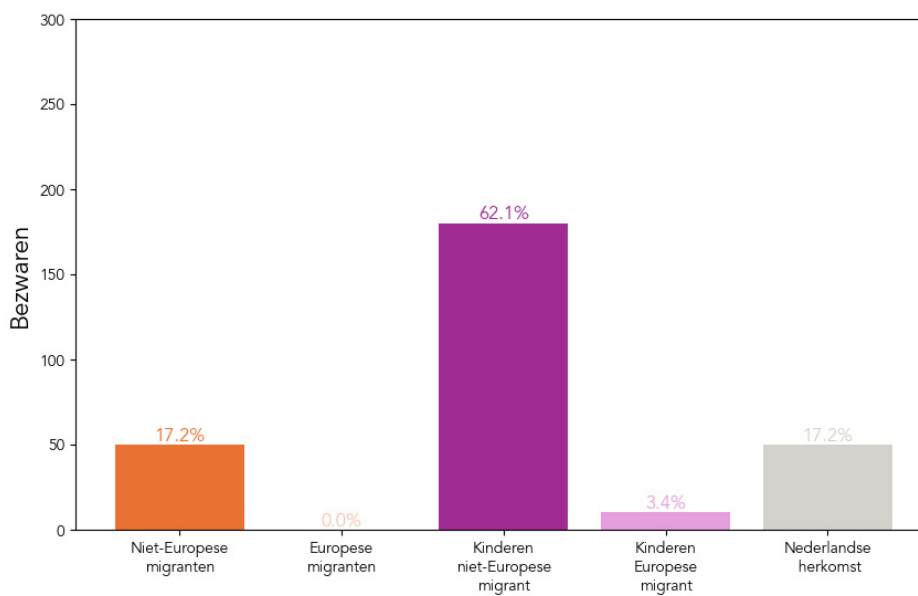
2019

Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de bezwaarpopulatie-2019 (n=280)



Figuur 40 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de bezwaarpopulatie-2019 (n=280)

Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst in de bezwaarpopulatie-2019 (n=280)

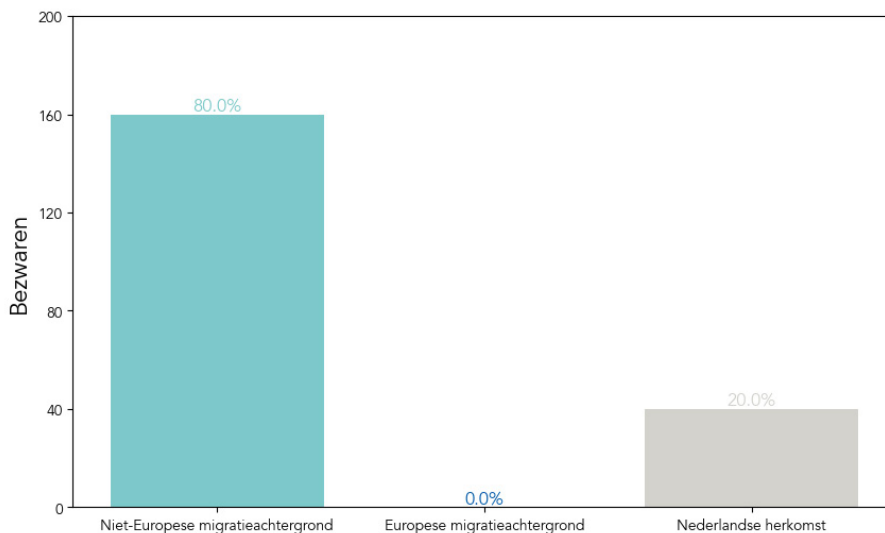


Figuur 41 – Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst in de bezwaarpopulatie-2019 (n=280)



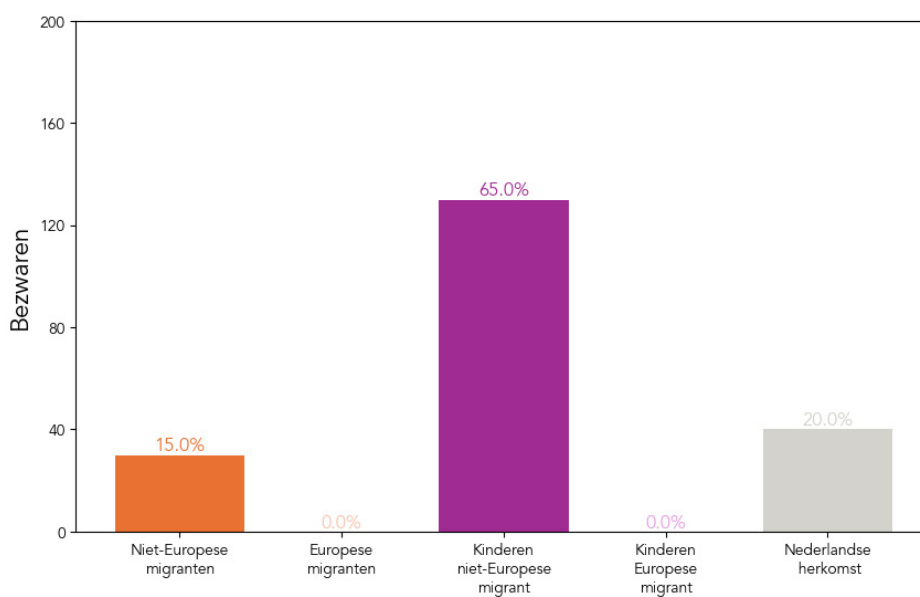
2021

Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de bezwaarpopulatie-2021 (n=200)



Figuur 42 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de bezwaarpopulatie-2021 (n=200)

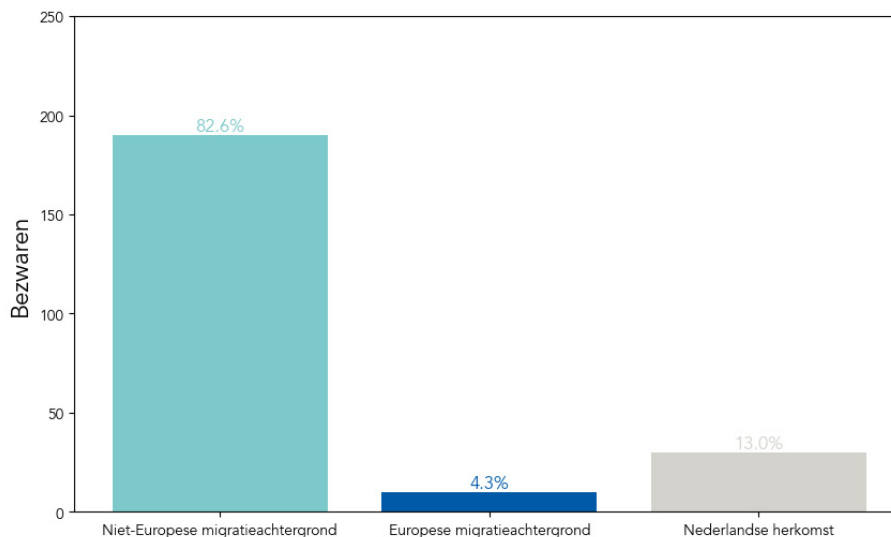
Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst in de bezwaarpopulatie-2021 (n=200)



Figuur 43 – Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst in de bezwaarpopulatie-2021 (n=200)

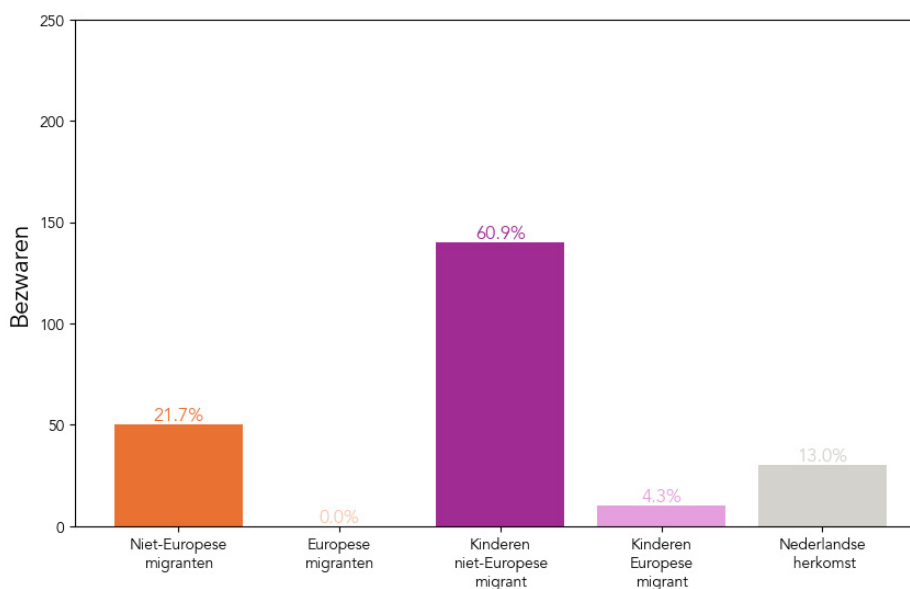
2022

Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de bezwaarpopulatie-2022 (n=230)



Figuur 44 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en met Nederlandse herkomst in de bezwaarpopulatie-2022 (n=230)

Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst in de bezwaarpopulatie-2022 (n=230)



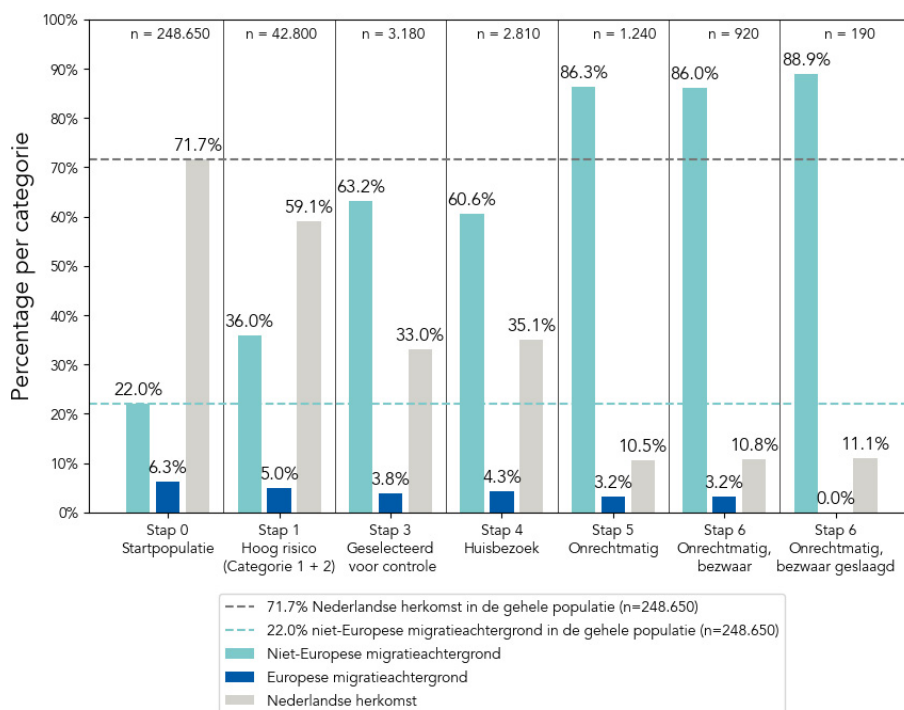
Figuur 45 – Verdeling van (niet-)Europese migranten, kinderen van een (niet-)Europese migrant en studenten met een Nederlandse herkomst in de bezwaarpopulatie-2022 (n=230)

## 4.8 Overzicht van het gehele CUB-proces

In [Figuur 46](#) en [Figuur 47](#) worden bovenstaande resultaten voor 2014 en 2019 samengevat in één figuur (funnel-visualizatie). Deze funnel geeft voor het gehele CUB-proces de verhoudingen studenten met een migratieachtergrond (Europees en niet-Europees) weer ten opzichte van de studenten met een Nederlandse herkomst. De concrete cijfers zijn in bovenstaande secties besproken. Bevindingen over het CUB-proces als geheel worden gedeeld in [6. Conclusie](#).

### 2014

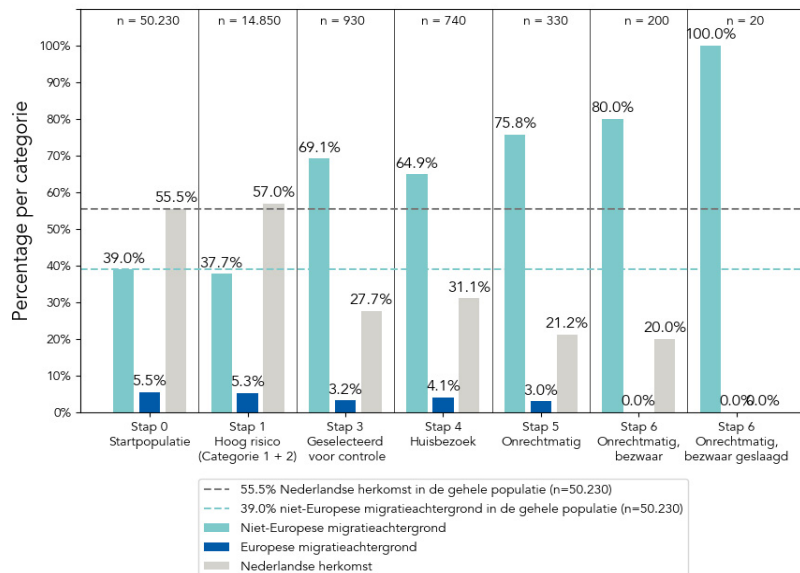
Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per stap van het CUB-proces in de uitwonendenbeurspopulatie-2014 (n=248.650)



[Figuur 46](#) – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per stap van het CUB-proces in de uitwonendenbeurspopulatie-2014 (n=248.650)

2019

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per stap van het CUB-proces in de uitwonendenbeurspopulatie-2019 (n=50.230)



Figuur 47 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per stap van het CUB-proces in de uitwonendenbeurspopulatie-2019 (n=50.230)

## 5. Voorbehouden

Verschillende voorbehouden zijn van toepassing.

De maatwerktabellen van het CBS geanalyseerd in dit rapport zijn gebaseerd op een datalevering van DUO aan het CBS. Er hebben kwaliteitscontroles plaatsgevonden op de correctheid van de data. Algorithm Audit kan echter niet garanderen dat alle bijbehorende *queries* en/of de onderliggende datastructuren bij DUO en het CBS volledig zijn en volledig vrij zijn van fouten of onvolkomenheden.

Algorithm Audit is niet verantwoordelijk voor beslissingen die worden genomen naar aanleiding van dit rapport.

Bij het opstellen van dit rapport is de bescherming van persoonsgegevens in ogenschouw genomen.

- > De interne processen van Algorithm Audit voldoen aan de AVG;
- > DUO deelde haar documenten via haar eigen platform.

Algorithm Audit heeft van DUO afkomstige documenten derhalve nooit in handen gehad. Bij het ontnemen van toegang tot de systemen van DUO heeft Algorithm Audit definitief geen toegang meer tot de documenten.

## 6. Conclusie

Uit bovenstaande biasmeting volgen vijf bevindingen.

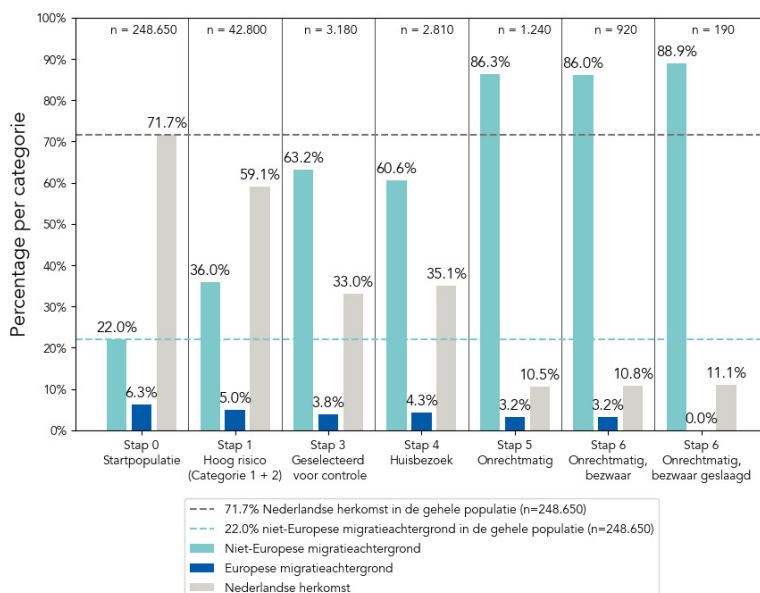
### Bevinding 1 – Het CUB-proces is als geheel vooringenomen geweest ten aanzien van studenten met een niet-Europese migratieachtergrond

In het CUB-proces bestond een sterke vooringenomenheid ten aanzien van studenten met een niet-Europese migratieachtergrond. Studenten met een niet-Europese migratieachtergrond werden 2.0x vaker bestempeld als hoog risico door het risicoprofiel dan studenten met een Nederlandse herkomst. De groep werd 6.2x vaker handmatig geselecteerd voor een huisbezoek. Uiteindelijk hadden studenten met een niet-Europese migratieachtergrond een 3.0x grotere kans om onterecht een huisbezoek te krijgen dan studenten met een Nederlandse herkomst.

Rekening houdend met de wezenlijk verschillende populaties in de verschillende jaren (in belangrijke mate als gevolg van de invoering van het leenstelsel<sup>26</sup>), is er sprake van een duidelijke structurele trend in de vooringenomenheid ten aanzien van studenten met een niet-Europese migratieachtergrond.

Figuur 48 geeft een overzicht van de verdeling studenten van verschillende herkomst per stap van het CUB-proces. Figuur 48 wordt nader toegelicht in onderstaande bevindingen en in 4. Resultaten.

Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per stap van het CUB-proces in de uitwonendenbeurspopulatie-2014 (n=248.650)



Figuur 48 – Verdeling van studenten met een (niet-)Europese migratieachtergrond en studenten met Nederlandse herkomst per stap van het CUB-proces in de uitwonendenbeurspopulatie-2014 (n=248.650)<sup>27</sup>

<sup>26</sup> In de periode 2015-2023 hadden alleen studenten die aan het mbo gingen studeren recht op een uitwonendenbeurs. Meer details kunnen worden gevonden in 3.2 Onderzoekspopulaties.

<sup>27</sup> Omdat de aggregatiestatistieken zoals samengesteld door het CBS afgerond worden op tientallen kan het voorkomen dat percentages niet exact tot 100% opsommen. Hetzelfde geldt voor de som van aantallen.

## **Bevinding 2 – Het risicoprofiel dat in het CUB-proces werd gebruikt was vooringenomen ten aanzien van studenten met een niet-Europese migratieachtergrond. Het toekennen van een hoger risico aan mbo-studenten en studenten die dicht bij het ouderlijk adres stonden ingeschreven was hier de oorzaak van.**

In het gehanteerde risicoprofiel (dat gebruik maakt van de kenmerken onderwijsvorm, leeftijd en afstand tussen het ingeschreven adres en het ouderlijk adres) ontstond een vooringenomenheid ten aanzien van migratieachtergrond. Studenten met een niet-Europese migratieachtergrond werden 2.0x vaker bestempeld als hoog risico door het risicoprofiel dan studenten met een Nederlandse herkomst. Dit is te herleiden naar het proxykarakter van met name onderwijsvorm en afstand. De kenmerken mbo 1-2 en een kleine afstand tot ouder(s) correleren sterk met de groep studenten met een niet-Europese migratieachtergrond. Door te profileren op deze kenmerken ontstaat vooringenomenheid ten aanzien van deze demografie. Een *unsupervised bias detectiemethode* bevestigt deze resultaten.

De proxy-analyse wordt nader toegelicht in [4.2 Resultaten onderzoeksvraag 2 \(risicoprofiel\)](#) en [4.3 Resultaten onderzoeksvraag 2a \(proxy-analyse\)](#). De resultaten van de *unsupervised bias detectiemethode* worden toegelicht in [Appendix A](#).

## **Bevinding 3 – Handmatige selectie versterkte de vooringenomenheid van het CUB-proces.**

Na indeling in risicocategorieën door het risicoprofiel vond handmatige selectie van studenten plaats voor controle en huisbezoek. Bij die handmatige selectie speelden schriftelijk onderzoek, werkinstructies, bepaalde uitsluitingsgronden en andere procedurele aspecten een rol. De vooringenomenheid die door het risicoprofiel ontstaat wordt in deze stap versterkt. Er zijn uit het onderzoek geen aanwijzingen naar voren gekomen dat deze vooringenomenheid door menselijke vooringenomenheid van individuele ambtenaren komt. Vooringenomenheid kan ook voortkomen uit de aard van de werkinstructies, zoals de uitsluiting van studentenhuizen van huisbezoek en zorginstellingen. Selectiegronden binnen de geselecteerde groep om wel een huisbezoek af te leggen waren onder meer de verhouding tussen oppervlakte en het aantal bewoners op het adres, inschrijving bij familie en de afstand van de inschrijving tot het adres van de ouders. De vraag hoe de vooringenomenheid in de handmatige selectie ontstond, valt buiten de reikwijdte van dit onderzoek.

Studenten met een niet-Europese migratieachtergrond die als hoog risico zijn geclassificeerd door het risicoprofiel hadden 5.5x meer kans om geselecteerd te worden voor controle dan studenten met Nederlandse herkomst en dezelfde risicocategorie. Studenten met een niet-Europese migratieachtergrond die als laag risico zijn geclassificeerd door het risicoprofiel hadden 1.8x meer kans om alsnog geselecteerd te worden voor controle dan studenten met een Nederlandse herkomst en dezelfde risicocategorie. Hierbij moet worden meegenomen dat de toegekende risicoscores zelf al vooringenomen zijn.

De analyse van de handmatige selectie van studenten wordt nader toegelicht in [4.4 Resultaten onderzoeksvraag 3 \(handmatige selectie\)](#) en [4.5 Resultaten onderzoeksvraag 3a \(mate waarin voorspelling algoritme wordt gevolgd\)](#).

**Bevinding 4 – Omdat het CUB-proces vooringenomen is geweest ten aanzien van studenten met een niet-Europese migratieachtergrond, is bij deze groep veel onrechtmatig gebruik vastgesteld. Dit is grotendeels het gevolg van overmatige controle van deze demografie.**

Door de vooringenomenheid van het gehele CUB-proces zijn studenten met een niet-Europese migratieachtergrond relatief vaker geselecteerd voor een huisbezoek dan studenten met een Nederlandse herkomst. Dit is in het geval van deze groep ook vaker een onterechte selectie: studenten met een niet-Europese migratieachtergrond hadden 3.0x meer kans op een huisbezoek waaruit later geen onrechtmatig gebruik bleek dan studenten met een Nederlandse herkomst.

Ook in de groep studenten waarbij onrechtmatig gebruik uiteindelijk wordt vastgesteld is het aandeel studenten met een niet-Europese migratieachtergrond groot. Er ontstaat door de stappen van het CUB-proces heen een vergrootglaseffect met betrekking tot de groep met een niet-Europese migratieachtergrond. Studenten met een Nederlandse herkomst werden relatief minder vaak gecontroleerd en onrechtmatigheid wordt in deze groep minder vaak opgespoord.

Of studenten met een migratieachtergrond ook vaker onrechtmatig gebruik maken van de uitwonendenbeurs kan niet worden vastgesteld op basis van de beschikbare data. Dat komt doordat deze verhouding niet kan worden geïsoleerd van de vooringenomenheid van het CUB-proces, en omdat de aselecte steekproef te klein is om dit onafhankelijk van het CUB-proces te meten. Of er naast het genoemde vergrootglaseffect verdere vooringenomenheid schuilt in het proces van de huisbezoeken zelf kon eveneens niet worden onderzocht.

De oververtegenwoordiging van deze groep studenten in de onrechtmatig-gebruik-populatie wordt nader toegelicht in [4.6 Resultaten onderzoeksvraag 4 \(onrechtmatig gebruik\)](#).

**Bevinding 5 – De groep studenten die in bezwaar gaat tegen vaststelling van onrechtmatig gebruik bestaat voor een overgroot deel uit studenten met een niet-Europese migratieachtergrond. Er is geen vooringenomenheid vastgesteld in de bezwaarprocedure zelf.**

Uit de verschillende peiljaren (2014, 2019, 2021 en 2022) treedt een consistent beeld naar voren dat de bezwaarpopulatie voor 79-85% uit studenten met een niet-Europese migratieachtergrond bestaat. Dit bevestigt globaal het beeld dat uit journalistiek onderzoek naar voren kwam omtrent de sterke oververtegenwoordiging van studenten met een migratieachtergrond die bezwaar maken tegen een besluit van DUO.<sup>28</sup>

Aan de andere kant wordt in de bezwaarstap in het CUB-proces geen vooringenomenheid vastgesteld. Bezwaar wordt even vaak gegrond verklaard voor alle studenten ongeacht hun herkomst.

De oververtegenwoordiging van deze groep studenten in de bezwaarpopulaties en gelijke behandeling tijdens de bezwaarprocedure wordt nader toegelicht in [4.7 Resultaten onderzoeksvraag 5 \(bezwaarpopulaties\)](#).

<sup>28</sup> B. Belleman, B. Heilbron & A. Kootstra. De discriminerende fraudecontroles van DUO. Investico Onderzoeksjournalisten, 2023



## Appendix A – Unsupervised bias detectie

De biasmeting in dit rapport is gebaseerd op aggregatiestatistiek over de migratieachtergrond van groepen studenten zoals beschikbaar gesteld door het CBS. Naar dit type biasmeting wordt gerefereerd als *supervised bias detectie*. Supervised refereert naar het feit dat er toegang is tot bijzondere persoonsgegevens (zij het op aggregatieniveau). Voor veel organisaties, zoals bedrijven, is het echter onmogelijk om een supervised bias te meten, aangezien bijzondere persoonsgegevens (ethniciteit, migratieachtergrond etc.) door deze organisaties in het kader van de Algemene Verordening Gegevens Verordening (AVG) vaak niet beschikbaar zijn. Ook bij private organisaties doet bias zich voor in algoritmes en de data, maar hoe meet je dat?

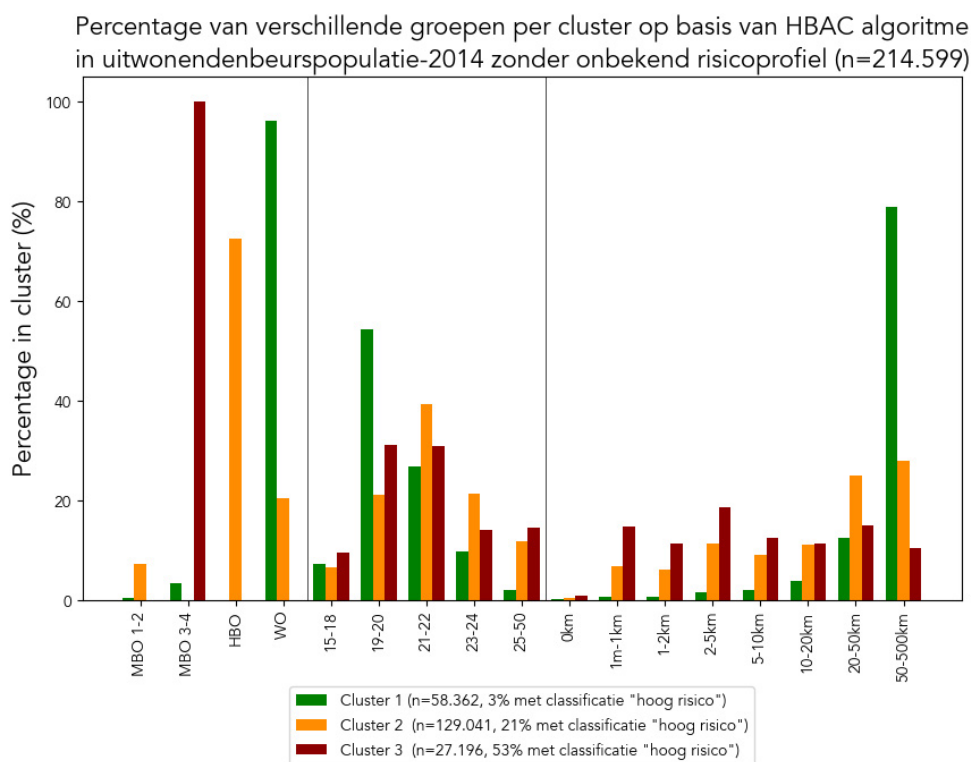
In deze appendix bestuderen we kort een alternatief voor wanneer toegang tot bijzondere persoonsgegevens niet beschikbaar zouden zijn geweest voor de biasmeting van het CUB-proces. Een dergelijke meting zonder toegang tot bijzondere persoonskenmerken wordt *unsupervised bias testing* genoemd. Een unsupervised bias detectie tool probeert zonder informatie over de migratieachtergrond van uitwonende studenten alsnog iets te zeggen over groepen studenten die in het proces (mogelijk) ongelijk behandeld worden. De groepen studenten worden aan de hand een eerlijkheidsmetriek, bijvoorbeeld 'door het risicoprofiel geclassificeerd als hoog risico' of 'geselecteerd voor controle', ingedeeld in groepen (ook wel: clusters). De afwijkende clusters kunnen handmatig geïnspecteerd worden op proxykenmerken, bijvoorbeeld of een afwijkend cluster (in termen van de eerlijkheidsmetriek) primair bestaat uit jonge wo-studenten die dicht bij hun ouders wonen, of juist bestaat uit mbo 1-2-studenten in de leeftijdscategorie 25-50 jaar.

Voor dit experiment hebben we gebruikt gemaakt van het zogenoemde Hierarchical Bias-Aware Clustering (HBAC) algoritme, zoals beschreven in wetenschappelijke literatuur.<sup>29</sup> Dit algoritme wordt toegepast op de uitwonendenbeurspopulatie-2014. Als eerlijkheidsmetriek wordt 'classificatie als hoog risico (categorie 1 of 2) door het risicoprofiel' gekozen. Alle gevallen waarbij de risicocategorie onbekend is worden verwijderd. De totale populatie bestaat zo uit 214.599 uitwonende studenten. Het HBAC-algoritme krijgt als input de onderwijsvorm, leeftijd, en afstand tot ouder(s) van een student. Voor dit experiment maakt het HBAC-algoritme gebruik van het k-modes clustering algoritme, aangezien dit clustering algoritme geschikt is voor categorische data. Het HBAC-algoritme is zo afgesteld dat het een resultaat geeft als een afwijkende cluster uit minimaal 25.000 studenten bestaat. De implementatie van het HBAC-algoritme is open-source en kan worden gevonden in Github.<sup>30</sup>

In dit experiment worden drie clusters gevonden die sterk verschillen in termen van de gekozen eerlijkheidsmetriek. In cluster 1 wordt 3% van de 58.362 studenten bestempeld als 'hoog risico', in cluster 2 is dit 21% van de 129.041 studenten en in cluster 3 is dit 53% van de 27.196 studenten. In cluster 1, het cluster met de laagste mate van vooringenomenheid, zitten met name wo-studenten (96%) en veel studenten die ver van hun ouders wonen (91% woont op meer dan 20km van hun ouder(s)). In cluster 3, het cluster met de hoogste mate van vooringenomenheid, bevat exclusief mbo 3-4- studenten en bevat veel studenten die dicht bij hun ouder(s) wonen (1m tot 5km) (44%). Zie [Figuur 49](#).

<sup>29</sup> Joanna Misztal-Radecka, Bipin Indurkha, Bias-Aware Hierarchical Clustering for detecting the discriminated groups of users in recommendation systems, Information Processing & Management, Volume 58, Issue 3, 2021.

<sup>30</sup> <https://github.com/NGO-Algorithm-Audit/unsupervised-bias-detection>



**Figuur 49** – Percentage van verschillende groepen per cluster op basis van HBAC-algoritme in uitwonendenbeurspopulatie-2014, zonder onbekend risicoprofiel (n=214.599)

Kortom, de resultaten van de unsupervised biasdetectie tool komen op hoofdlijnen overeen met de resultaten van de supervised biasmeting. Merk op dat in dit geval het HBAC-algoritme geen toegang heeft tot de gebruikte logica in het risicoprofiel, maar desondanks (sommige) vergelijkbare groepen vindt die bestempeld worden als hoog risico. Dergelijke toepassing van unsupervised bias detectie zou daarmee aan menselijke experts een signaal kunnen geven dat afwijkingen zich in een proces voordoen en dat de oorzaak hiervan verder moet worden uitgezocht.

Unsupervised bias detectie is daarmee een veelbelovende methode die in de nabije toekomst uitvoerig getest en doorontwikkeld moet worden. Het stelt publieke en private organisaties in staat om zonder toegang tot bijzondere persoonsgegevens (onwenselijke) bias in algoritmes te detecteren en te mitigeren.





[www.algorithmaudit.eu](http://www.algorithmaudit.eu)



[www.github.com/NGO-Algorithm-Audit](https://www.github.com/NGO-Algorithm-Audit)



[info@algorithmaudit.eu](mailto:info@algorithmaudit.eu)

Stichting Algorithm Audit is geregistreerd bij de Kamer  
van Koophandel onder nummer 83979212