



INCODA

Initiative COVID-19 Data

Eindrapportage INCODA-project

(1 september 2020-31 mei 2022)

ZonMw projectnummer 10430022010019

Table of Contents

1.	Management samenvatting	2
2.	Werkpakket 1: Combineren van datasets	4
2.1	Doel	4
2.2	Uitvoering en resultaten	4
2.2.1	CBS-microdata omgeving	4
2.2.2	Gegevens rondom ziekenhuisopnames	4
2.2.3	Covid Predict vs LBZ	6
2.3	Test data	7
2.4	Dataset gebruik voor verdere analyses in dit rapport	8
3.	Werkpakket 2: Duiding infectierisico van verschillende groepen	9
3.1	Doel	9
3.2	Uitvoering en Resultaten	9
3.2.1	Beschrijving van de data en definities	9
3.2.2	Infectierisico voor verschillende groepen I	9
3.2.3	Infectierisico voor verschillende groepen II – aangescherpte risicoprofielen	11
3.2.4	Verschillen VRs en G4 met Nederland	13
4.	Werkpakket 3: Visualisatie verloop infectie in verschillende groepen	16
4.1	Doel	16
4.2	Uitvoering en resultaten	16
4.2.1	Beschrijving van de data en definities	16
4.2.2	Resultaat	16
4.2.3	Online dashboard 'Het verloop van de COVID-19 pandemie over tijd en onder verschillende Nederlandse bevolkingsgroepen'	18
5.	Werkpakket 4: Duiden van groepen met een hoog risico voor COVID-19 met een ernstig beloop	19
5.1	Doel	19
5.2	Uitvoering en resultaten	19
5.2.1	Beschrijving van de data en definities	19
5.2.2	Verschillen tussen groepen m.b.t. ziekenhuisopname, IC-opname en overlijden	19
5.2.3	Ziekenhuisopname: inzicht in risico's en absolute aantallen	21
5.2.4	Intensive care-opname: inzicht in risico's en absolute aantallen	23
5.2.5	Overlijden: inzicht in risico's en absolute aantallen	24
5.2.6	Uitwerking cascade van zorg	25
6.	Conclusies INCODA	27
7.	Referenties	28
	Bijlagen	29





1. Management samenvatting

In juli 2020 is het project *Initiative COVID data (INCODA)* gehonoreerd met een subsidie van ZonMW vanuit het COVID-19 onderzoeksprogramma. Het consortium, bestaande uit Amsterdam health & technology institute, Amsterdam UMC – afdeling Global Health en Amsterdam UMC – afdeling Intensive care, heeft de afgelopen periode gewerkt aan de volgende doelstellingen:

- Duiden van groepen met een afwijkend risico voor een COVID-19 infectie; in termen van demografische, sociaaleconomische en medische kenmerken, zowel univariabel als multivariabel en over de verschillende domeinen heen (werkpakket 2).
- Onderzoeken hoe het week-by-week infectie verloop is geweest voor verschillende groepen (werkpakket 3).
- Visualiseren hoe het verloop al dan niet samenhangt met getroffen maatregelen en onderzoeken hoe dit verschilt tussen verschillende groepen (werkpakket 3).
- Duiden van groepen met een hoog risico voor COVID-19 met ernstig beloop; in termen van demografische, sociaaleconomische en medische kenmerken, zowel univariabel als multivariabel en over de verschillende domeinen heen (werkpakket 4).
- In kaart brengen van de 'cascade of care' rondom COVID-19, van testen tot infectie en ziekenhuisopname (werkpakket 4).

Belangrijke basis voor het beantwoorden van de onderzoeksvragen is een gecombineerde dataset van 17.4 miljoen inwoners van Nederland, met daarin 16 miljoen GGD test uitslagen en gegevens over ruim 30 duizend COVID-19 gerelateerde ziekenhuis- en IC-opnames (werkpakket 1). Met deze gecombineerde dataset is vervolgens gekeken naar zowel infectierisico als het risico op een ernstig beloop. In alle resultaten komt naar voren dat sociaal demografische factoren, zoals inkomen, leeftijd en migratieachtergrond zowel het risico op een SARS-Cov-2 infectie als het risico op een ernstig beloop van COVID-19 sterk beïnvloeden. Specifiek kijkend naar infectierisico's, wordt duidelijk dat huishoudsamenstelling van grote invloed is; voor risico op ernstig beloop wordt duidelijk dat het gebruik van zorg een bepalende factor is. Daarnaast blijkt dat het inzoomen binnen groepen inzicht geeft in de invloed van de verschillende variabelen (zowel risico verhogend als juist beschermend).

Ook komt naar voren dat het per stad of regio kan verschillen welke groep het hoogste infectierisico heeft. Bij verdere verdieping in deze verschillen is het mogelijk om op geografisch (gemeente, veiligheidsregio, etc.) niveau de risicoprofielen in kaart te brengen. Voor toekomstige epidemieën, zou deze verzameling van inzichten een specifiek en op bepaalde groepen of regio gericht beleid kunnen ondersteunen. Hierbij kan er retrospectief gekeken worden naar de effectiviteit van maatregelen op verschillende groepen en regio's, kijkend naar de daling of stijging van het aantal positieve testen. Ook kunnen deze inzichten ondersteunen in preventief beleid om mensen met hoog risico op ernstig beloop te beschermen.

Aangezien zorgdruk een belangrijke sturende factor is geweest in het Corona-beleid van het VWS én veelbesproken onderwerp in de media, presenteert dit rapport ook cijfers over *aantallen* gevallen met een COVID-19 infectie en mensen met ernstig beloop waarbij ook weer onderscheid gemaakt wordt tussen verschillende groepen. Door het naast elkaar presenteren van de resultaten van deze verschillende benaderingen van de data, wordt inzichtelijk dat verschillende vragen of vraagstellingen leidt tot verschillende inzichten.

Dit rapport geeft slechts een selectie weer van inzichten die uit de data zijn verkregen. Er is een online dashboard ontwikkeld dat het voor iedereen mogelijk maakt om een gedetailleerder beeld te krijgen van welke groepen, waar en op welk moment gingen testen, positief testten voor het coronavirus, wie er in het ziekenhuis lagen en zijn overleden aan de gevolgen van het virus.

Naast inzichten op het gebied van infectierisico en risico op ernstig beloop van COVID-19 heeft dit project ook veel inzichten opgeleverd over de processen rondom het uitwisselen van medische gegevens. Ondanks een extern juridisch advies bleek dat de kennis, ervaring en interpretatie van wet- en regelgeving rondom (CBS-) data erg verschilt per organisatie/ziekenhuis.



2. Werkpakket 1: Combineren van datasets

2.1 Doel

Op persoonsniveau de relevante data samenbrengen binnen de bestaande onderzoeksdataset in de beveiligde CBS Remote Access (CBS RA-) omgeving van ahti.

2.2 Uitvoering en resultaten

2.2.1 CBS-microdata omgeving

CBS-Microdata

De onderzoeksdataset voor dit project betreft een data-infrastructuur bestaande uit een gecombineerde set van microdatabestanden van het Centraal Bureau voor de Statistiek (CBS). Zie bijlage 1 voor de omschrijving van de CBS-microdata omgeving.

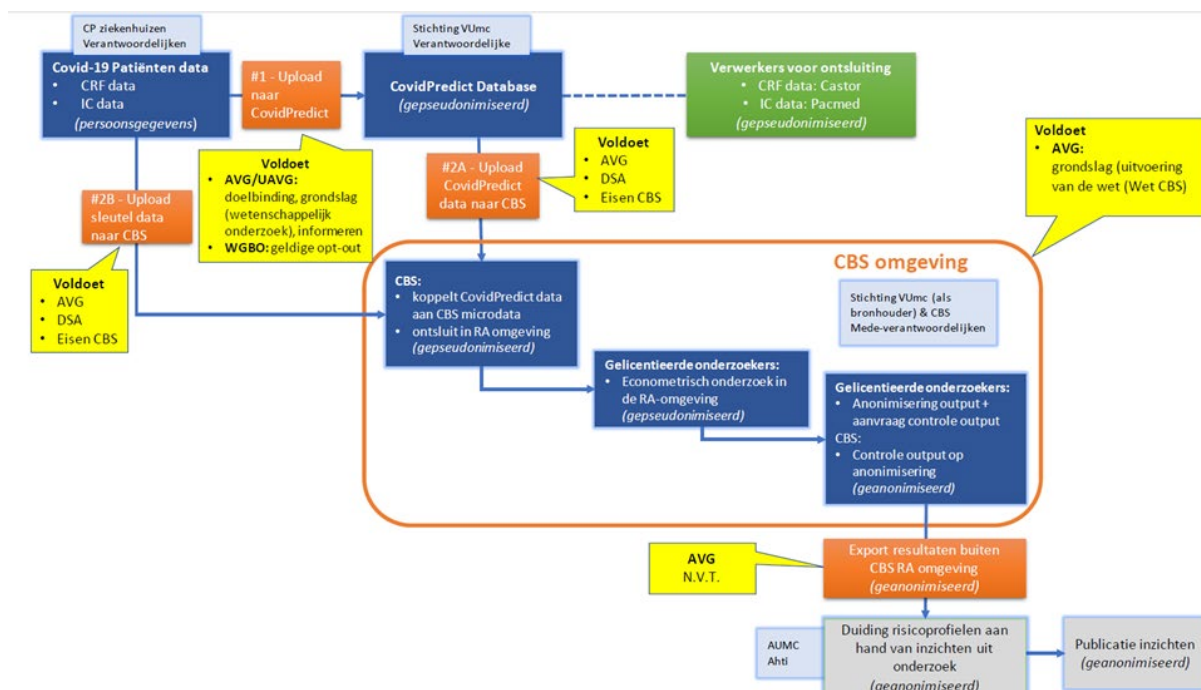
Rondom de medische voorgeschiedenis en zorggebruik zijn de gegevens binnen de CBS-omgeving vaak completer dan binnen de ziekenhuisomgeving. Daarbij geldt wel dat deze gegevens een vertraging kennen voordat ze bij CBS bekend zijn. Verder geven ziekenhuisdata slechts een beperkt beeld van de COVID-19 uitbraak. Voor de uitvoering van het INCODA-project is de microdata omgeving daarom verrijkt met een aantal specifieke bestanden. De benodigde activiteiten om dit te bewerkstelligen zijn uitgevoerd binnen werkpakket 1. Zoals ook al in eerder voortgangsrapportages omschreven is dit een langdurig proces geweest met wisselend resultaat, hieronder kort omschreven.

2.2.2 Gegevens rondom ziekenhuisopnames

Databron 1: CovidPredict data

Met steun van de Nederlandse Vereniging voor Intensive Care en de Stichting Nationale Intensive Care Evaluatie is er vanaf het begin van de uitbraak in Nederland een zeer brede samenwerking tussen ziekenhuizen tot stand gekomen om data te verzamelen van patiënten met COVID-19 onder het CovidPredict project (www.covidpredict.org). Deze dataset is uniek in de wereld qua omvang en datadichtheid en kan daarmee potentieel een zeer belangrijk bijdrage leveren aan COVID-19 onderzoek. Het gaat hierbij enerzijds om data die door onderzoekers wordt ingevoerd in elektronische caserecord forms (eCRF) en anderzijds om ruwe data rondom intensive care-opnames die rechtstreeks en geautomatiseerd uit het elektronisch patiëntendossier wordt verkregen. Voor de beantwoording van de onderzoeksvragen in dit project zijn de gegevens rondom de ziekenhuisopnames (de eCRFs) het meest relevant, daar is in dit werkpakket dan ook de focus op gelegd.

Zoals te verwachten roept het bij elkaar brengen van persoonsgegevens vragen op rondom privacy en gebruik van de data. Daarom is gekozen om de voorgestelde dataverwerkingen juridisch te laten beoordelen. De uitkomst van deze juridische beoordeling was dat de beoogde dataverwerkingen (met name de upload van de data naar het CBS), voldoen aan alle relevante wet- en regelgevingen (zie figuur 1).



Figuur 1: Gegevensstroom CovidPredict data t.b.v. INCODA. Bron: Juridische beoordeling door Prof. Dr. Lokke Moerel en mr. Dr. Theo Hooghiemstra d.d. 1 oktober 2020.

Stichting VUmc is – middels afspraken vastgelegd in een Data Sharing Agreement – verwerkingsverantwoordelijke over de CovidPredict data, welke bestaat uit de patiëntgegevens met een patiënt specifiek ID. De afzonderlijke ziekenhuizen hebben de sleutel tussen persoonsgegevens en het CovidPredict ID, en zijn verwerkingsverantwoordelijke over deze 'sleuteldata'. Er is dus sprake van twee soorten verwerkingen: de upload van de data naar het VUmc (1) en de upload van de CovidPredict data en de sleuteldata naar het CBS (2A en 2B).

Met de verschillende ziekenhuizen is na de juridische beoordeling het gesprek aangegaan om de juiste koppelgegevens (zijnde de sleuteldata) te verstrekken aan het CBS via de beveiligde upload faciliteit. Ondanks de juridische beoordeling en de uitgebreide gesprekken is het uiteindelijke resultaat dat er gedurende de looptijd van het project van slechts 6 (van de 11) instellingen sleuteldata is geüpload in de CBS-microdata omgeving (in totaal 1595 patiënten). Doordat de coronacrisis gedurende het project doorliep en de dataverzameling per ziekenhuis in een verschillend tempo gebeurde, zijn er van sommige ziekenhuizen alleen gegevens geüpload van de eerste golf (Amsterdam UMC-beide locaties, NoordWest Ziekenhuisgroep, OLVG). Andersom stonden 2 ziekenhuizen alleen inclusie van patiënten toe na aanpassing van de patiënten informatie, waardoor er alleen data van de tweede golf is geüpload (Antonius Ziekenhuis, Flevoziekenhuis).

Na upload van de bestanden zijn de CovidPredict bestanden met behulp van de sleutelbestanden door CBS gekoppeld aan de betreffende persoonsgegevens in de beveiligde RA-omgeving (het zogenaamde verrinnen).

Data bron 2: Dutch Hospital Data 2020

In parallel is er gezocht naar aanvullende gegevens over ziekenhuisopnames in Nederland. De meest geschikte bron bleek uiteindelijk het microdatabestand van de Landelijke Basisregistratie Ziekenhuiszorg (LBZ) van Dutch Hospital Data (DHD). Dit bestand, waar middels een speciale procedure toegang voor moest worden aangevraagd bij DHD, bevat gegevens over ziekenhuisopnames in Nederland van personen die staan ingeschreven in de Basisregistratie Personen (BRP). In april 2022 kwam de data van 2020

beschikbaar, wat de mogelijkheid heeft geboden om voor de eerste 6-9 maanden van de coronacrisis inzichten te verkrijgen over ziekenhuisopnames door (verdenking op) COVID-19 (ICD-10 codes U07.1 (Virus geïdentificeerd) en U07.2 (Verdenking van COVID-19, virus niet geïdentificeerd)). Deze data zijn dan ook aangevraagd en voor dit project beschikbaar gekomen.

2.2.3 Covid Predict vs LBZ

De CovidPredict en LBZ-databestanden zijn na verrinning door CBS gekoppeld op persoonsniveau in de microdata omgeving aan andere microdatabestanden. In tabel 1 zijn de eigenschappen van de beide ziekenhuisdatabronnen te zien ten opzichte van de bevolkingssamenstelling van Nederland. Het gaat hierbij om de situatie van deze personen in het jaar 2020. In de microdata omgeving zijn dus 1.595 mensen opgenomen die in de CovidPredict database staan. Interessant genoeg zijn deze mensen niet allemaal (1.340, 84.01%) opgenomen in de Landelijke Basisregistratie Ziekenhuiszorg. De oorzaak hierachter moet nog worden onderzocht.

Opvallende bevindingen uit de descriptieve vergelijking tussen de CovidPredict database en LBZ-data:

- Er zijn geen grote verschillen tussen de beide datasets voor wat betreft de variabelen rondom zorggebruik zoals medicatie voor CRVM, Diabetes, COPD en gebruik van Wlz en Wmo.
- Er zijn relatief grote verschillen in het percentage mensen met een migratieachtergrond tussen beide datasets, in CovidPredict zijn bijvoorbeeld relatief meer mensen geïncludeerd met een niet-westerse migratieachtergrond.
- De LBZ-dataset bevat daarentegen relatief meer ouderen (75+).
- Er zijn verschillen tussen beide datasets voor wat betreft gezinssamenstelling. In de LBZ lijken relatief meer mensen geregistreerd die wonen in een huishouden zonder kinderen.
- Zoals te verwachten is de populatie die in het ziekenhuis is opgenomen (zowel in LBZ als CovidPredict) met (verdenking op) COVID-19 vaker ouder, een man, heeft een niet-westerse migratieachtergrond en een laag inkomen en maakt meer gebruik van zorg. Deze risicoprofielen worden verder uitgewerkt in werkpakket 4.

Gezien de overall dekking, de hogere aantallen én de verschillen tussen de datasets in demografische gegevens zijn vervolganalyses uitgevoerd met de LBZ-data van 2020.

Tabel 1: Eigenschappen van de CovidPredict data en Landelijke Basisregistratie Ziekenhuishouding (LBZ) uitgezet tegen de bevolkingssamenstelling (aantallen en proportie (%)) van Nederland.

Variabele [§]	Totaal in NL	CovidPredict	LBZ*
Geslacht			
Man	8.648.030 (49,68%)	990 (62,07%)	18.075 (60,06%)
Vrouw	8.759.555 (50,32%)	605 (37,93%)	12.020 (39,94%)
Leeftijd			
0-11	2.154.675 (12,38%)	n/a [#]	110 (0,37%)
12-18	1.396.090 (8,02%)	n/a [#]	80 (0,27%)
19-27	1.96.225 (11,47%)	30 (1,88%)	310 (1,03%)
28-65	8.675.270 (49,84%)	855 (53,61%)	12.690 (42,17%)
66-75	1.877.645 (10,79%)	390 (24,45%)	7.730 (25,69%)
75+	1.307.680 (7,51%)	320 (20,06%)	9.175 (30,49%)
Migratieachtergrond			
Geen migratieachtergrond	13.186.880 (75,75%)	805 (50,47%)	20.770 (69,01%)
Niet-westers	2.392.060 (13,74%)	630 (39,50%)	6.455 (21,45%)
Westers	1.828.645 (10,50%)	160 (10,03%)	2.865 (9,52%)
Inkomen			
<140%	3.285.075 (18,87%)	515 (32,29%)	9.795 (32,55%)
140 – 280%	4.441.735 (25,52%)	440 (27,59%)	9.730 (32,33%)
>280%	9.051.140 (52,00%)	605 (37,93%)	9.685 (32,18%)
onbekend / institutioneel	629.635 (3,62%)	30 (1,88%)	885 (2,94%)
Huishoudsamenstelling			
Tweepersoonshuishouden met kinderen	8.913.950 (51,21%)	525 (32,92%)	7.820 (25,98%)
Tweepersoonshuishouden zonder kinderen	4.305.470 (24,73%)	575 (36,05%)	13.045 (43,35%)
Eenpersoonshuishouden	2.991.645 (17,19%)	395 (24,76%)	7.270 (24,16%)
Institutioneel huishouden	210.030 (1,21%)	20 (1,25%)	770 (2,56%)
Overig huishouden	986.490 (5,67%)	80 (5,02%)	1.190 (3,95%)
Zorg- en sociale voorzieningengebruik			
Gebruik zorg uit de Wlz [£]	294.145 (1,69%)	50 (3,13%)	1.505 (5,00%)
Gebruik zorg uit de Wmo [£]	1.027.795 (5,90%)	365 (22,88%)	7.305 (24,27%)
Gebruik Diabetes [£] medicatie	800.115 (4,60%)	410 (25,71%)	6.985 (23,21%)
Gebruik COPD [£] medicatie	1.501.300 (8,62%)	360 (22,57%)	7.525 (25,00%)
Gebruik CVRM [£] medicatie	3.585.910 (20,60%)	930 (58,31%)	18.545 (61,62%)
[§] Voor een beschrijving van de variabelen zie bijlage 2			
[*] LBZ = alle ziekenhuisopnames met indicatie (verdenking op) COVID (ICD-10 U07.01 en U07.2)			
[#] N/a: aantallen kleiner dan 10, niet door CBS outputcontrole.			
[£] Wlz: wet langdurige zorg; Wmo: wet maatschappelijke ondersteuning; CVRM: cardiovasculair risicomanagement; COPD: Chronic obstructive pulmonary disorder			

2.3 Test data

Eind 2021 zijn de gegevens over GGD testafspraken en testuitslagen (CoronIT) beschikbaar gekomen voor CBS-microdata instellingen. Opgenomen in deze dataset zijn de datum en het tijdstip van de testafname en de testuitslag van personen die zich hebben laten testen op het SARS-CoV-2 virus bij de GGD in de periode van 13 mei 2020 tot en met 6 september 2021.

Bij het interpreteren van de resultaten is het van belang om rekening te houden met de volgende beperkingen van de data:

- Het betreft naar schatting informatie over 85% van de positieve testen in Nederland over deze periode en geeft dus geen compleet beeld.
- Het testgebruik van verschillende bevolkingsgroepen kan verschillend zijn.
- Om de resultaten in deze rapportage goed te interpreteren, is het naast bovenstaande punten belangrijk in gedachten te houden dat de teststrategie gedurende de pandemie is veranderd, hieronder op hoofdlijnen de belangrijkste aanpassingen:
 - Tot en met november 2020: Volwassenen alleen testen bij corona-gerelateerde klachten, zoals neusverkoudheid, hoesten, benauwdheid of kortademigheid, (lichte) koorts en niet meer kunnen ruiken of proeven.
 - Vanaf 1 december 2020: Ook zonder klachten kunnen volwassenen zich laten testen als zij in contact zijn geweest met een persoon die positief is getest op SARS-CoV-2 of 5 dagen na terugkeer uit een land met een oranje reisadvies.
 - Vanaf 1 januari 2021: Er geldt voor kinderen hetzelfde testbeleid als voor volwassenen.
 - Er zijn enkele pilots uitgevoerd rondom grootschalig testen die op specifieke momenten afwijkende patronen in de data als gevolg kunnen hebben. Meer informatie:
<http://www.rijksoverheid.nl/onderwerpen/coronavirus-covid-19/testen/testbeleid/grootschalig-en-risicogericht-testen>
 - De onderliggende testdata is niet bewerkt of opgeschoond. Personen die bijvoorbeeld meerdere keren positief zijn getest in korte tijd worden als individuele gevallen gezien.

Bij het beschikbaar komen van de CORONIT data is bij VWS een data-aanvraag ingediend met daarbij omschreven de doelstellingen van werkpakket 2, 3 en 4. Een commissie met daarin leden van o.a. het CBS en RIVM heeft uiteindelijk toegestemd in het gebruik van de CORONIT testdata voor dit project. Waarna de testdata eind 2021 zijn gekoppeld aan de microdata van de INCODA dataset.

2.4 Dataset gebruik voor verdere analyses in dit rapport

Resultaat van werkpakket 1 is een gecombineerde dataset van 17.4 miljoen inwoners van Nederland, met daarin 16 miljoen GGD testuitslagen en van ruim 30 duizend COVID-19 gerelateerde ziekenhuis en IC-opnames de gegevens. Deze gecombineerde dataset is gebruikt voor werkpakket 2 tot en met 4.

3. Werkpakket 2: Duiding infectierisico van verschillende groepen

3.1 Doel

Het doel van WP2 is het creëren van multivariabele risicogroepen, gestratificeerd naar de kans op een positieve SARS-CoV-2 test. Voor de duiding van de risicogroepen is er een multivariabel profiel ontwikkeld van de personen die positief getest zijn op SARS-CoV-2, gestratificeerd naar kans op infectie in termen van demografische en sociaaleconomische factoren.

3.2 Uitvoering en Resultaten

3.2.1 Beschrijving van de data en definities

Om de risicogroepen in kaart te brengen is er gebruik gemaakt van de GGD testdata (CORONIT) uit de periode 1 juni 2020 tot 1 september 2021 en de ahti zorg-data infrastructuur in de CBS RA-omgeving (zie bijlage 1).

Het infectierisico wordt in dit rapport gedefinieerd als het percentage positieve testuitslagen binnen een specifieke groep in de periode 06 juni 2020 t/m 30 augustus 2021 gedeeld door het totaal aantal personen waar deze groep uit bestaat. Merk op dat een aantal factoren, zoals het veranderde testbeleid en het ontbreken van testdata van ziekenhuizen en commerciële teststraten hierin niet is meegenomen (zie ook WP1).

3.2.2 Infectierisico voor verschillende groepen I

Voor 5 variabelen (zie bijlage 2 voor een omschrijving van de variabelen en de bijbehorende categorieën) is gekeken naar de invloed ervan op het infectierisico.

Invloed van sociaal demografische variabelen op infectie risico

Gemiddeld was de kans 9,37% dat iemand in de periode van 1 juni 2020 tot 1 september 2021 in NL een positieve test had (1.630.955 positieve testen op 17.408.280 Nederlanders, waarbij herbesmetting in deze periode niet uitgesloten kan worden). Uit tabel 2 blijkt dat sociaal demografische factoren (migratieachtergrond, inkomen, leeftijd en huishoudsamenstelling) een invloed hebben op de kans op een positieve test. De variabele "geslacht" lijkt weinig invloed te hebben op de kans op een positieve test.

Tabel 2: Infectierisico (% positieve testen voor SARS-CoV-2) op basis van sociaal demografische variabelen

Variabele [§]	Aantal in NL	Positieve test
Geslacht		
Man	8.648.425	9.25%
Vrouw	8.759.855	9.49%
Leeftijd		
0-11	2.154.305	5.52%
12-18	1.396.075	16.44%
19-27	1.996.410	15.56%
28-65	8.675.765	9.79%
66-75	1.877.775	4.56%
75+	1.307.950	2.85%
Migratieachtergrond		
Geen migratieachtergrond	13187110	9.02%
Niet-westers	2392095	12.75%
Westers	1829075	7.44%
Inkomen		
<140%	2.782.480	7.06%
140 – 280%	4.442.080	8.19%
>280%	9.051.575	10.92%
Onbekend/institutioneel	1.132.145	7.27%
Huishoudsamenstelling		
Huishouden met kinderen	8.939.765	11.89%
Tweepersoonshuishouden zonder kinderen	4.347.570	6.73%
Eénpersoonshuishouden	2.922.565	6.45%
Institutioneel huishouden	210.110	3.50%
Ander type huishouden	988.270	8.05%

[§] Voor een beschrijving van de variabelen zie bijlage 2

Aangescherpte risicoprofielen d.m.v. Classification and Regression Tree (CART) methode

Het combineren van de variabelen leidt tot specifiekere risicoprofielen voor infectie. Uit de analyses blijkt dat wanneer er bijvoorbeeld drie van de geselecteerde sociaal demografische variabelen worden meegenomen er risicoprofielen naar voren komen met meer dan 20% kans op een positieve test. Analyses met de vooraf gekozen variabelen en categorieën uit tabel 2 leidde tot een groot aantal, zeer variabele risicoprofielen en geven een indicatie dat binnen de gekozen variabelen een andere onderverdeling nodig is. Voor het kiezen van passendere categorieën binnen bovenstaande variabelen is gekozen voor toepassing van een 'Classification and Regression Tree' algoritme (hierna CART methode, zie technische omschrijving in bijlage 3). Daarnaast zijn voor de analyses op de test data, op basis van andere studies en verschenen artikelen, 2 variabelen toegevoegd, namelijk "personen per huishouden" (Marmot, 2020; Reukers, 2021) en "stedelijkheid niveau" (Duca, 2020; Smit, 2020). Verder is "geslacht" uit de analyse gehaald omdat er minimaal verschil te zien was in infectierisico op basis van geslacht. Tot slot is ook het "institutioneel huishouden" en "institutioneel inkomen" uit de analyses gehaald aangezien personen die in een institutioneel huishouden wonen niet of nauwelijks zijn getest door de GGD. Het includeren van institutioneel huishouden zou daarom een vertekend beeld kunnen geven. De CART heeft geleid tot een nieuwe categorisering van variabelen (omschreven in bijlage 2) welke voor verdere analyses zijn gebruikt.

3.2.3 Infectierisico voor verschillende groepen II – aangescherpte risicoprofielen

Tabel 3 geeft de infectierisico's weer voor de nieuw gekozen categorieën van variabelen. Zoals eerder beschreven leidt het combineren van de variabelen tot specifiekere risicoprofielen, met relatief hogere of juist lagere infectierisico's. Zie ter illustratie figuur 2, waarbij er is gekeken naar de kans op een positieve COVID-19 test onder jongeren van de leeftijd 14-18 (infectierisico totale groep: 18.54%) die nog thuis wonen (huishouden met kinderen).

Tabel 3: Infectierisico (% positieve testen op SARS-CoV-2) op basis van sociaal demografische variabelen (nieuwe indeling op basis van CART)

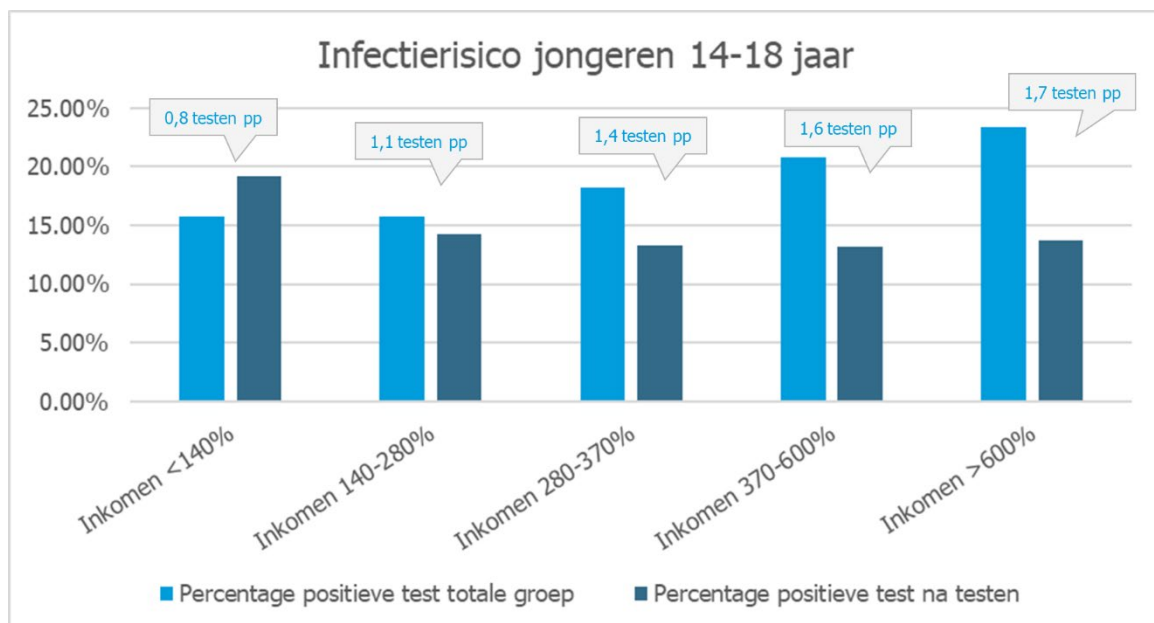
Variable [§]	Aantal in NL	Positieve test
Leeftijd		
0-8	1.387.820	4.58%
9-13	934.575	10.01%
14-18	995.250	18.54%
19-26	1.674.830	16.58%
27-59	7.378.660	10.51%
60-70	2.279.765	6.01%
>70	2.115.040	3.46%
Migratieachtergrond		
Geen migratieachtergrond	12.871.550	9.17%
Surinaams	343.780	12.85%
Turks	400.910	17.37%
Marokkaans	394.800	16.08%
Anders	2.754.895	8.98%
Inkomen		
<140%	2.159.775	7.00%
140-280%	5.064.190	8.08%
280-370%	3.595.730	10.21%
370-600%	4.360.845	11.39%
>600%	1.585.395	11.41%
Huishoudsamenstelling		
Huishouden met kinderen	7.529.470	12.04%
Tweepersoonshuishouden zonder kinderen	4.305.470	6.77%
Eénpersoonshuishouden	2.991.645	6.41%
Eénouderhuishouden	1.384.480	11.13%
Ander type huishouden	554.875	10.94%
Aantal personen per huishouden		
Minder dan 4 personen per huishouden	10.883.260	7.98%
4 of meer personen per huishouden	5.882.680	12.52%
Stedelijkheid		
Stedelijkheid niveau 1	4.048.610	10.55%
Stedelijkheid niveau 2-5	12.717.330	9.26%

[§] Voor een beschrijving van de variabelen zie bijlage 2

Binnen deze groep blijkt dat jongeren uit huishoudens met een inkomen van meer dan 600% van het sociaal minimum (n=89.080), bijna 8% meer kans hebben op een positieve COVID-19 test (23.41%) dan jongeren uit een huishouden met een inkomen onder de 140% van het sociaal minimum (15.72%). Daarnaast is er gekeken naar het percentage positieve testuitslagen van de totale testuitslagen (zowel positief als negatief) van deze groep. Hieruit blijkt dat jongeren uit rijke gezinnen *relatief* minder vaak positief testen:

- Gemiddeld werden er 1,7 testen uitgevoerd per jongere uit een huishouden met inkomen boven de 600% van het sociaal minimum.
- Gemiddeld werden er 0,8 testen uitgevoerd per jongere uit een huishouden met inkomen onder de 140% van het sociaal minimum.
- 14-18-jarigen uit een huishouden met inkomen boven de 600% van het sociaal minimum: 13,77% van de 15.470 testen positief.
- 14-18-jarigen uit een huishouden met inkomen onder de 140% van het sociaal minimum 19,19% van de 36.765 testen positief.

Met andere woorden: jongeren uit relatief rijke gezinnen hebben zich vaker laten testen terwijl ze geen infectie hadden. En jongeren uit armere gezinnen hebben zich minder vaak bij de GGD laten testen, maar als zij zich testten was er vaker een positieve uitslag.



Figuur 2: Infectierisico binnen de groep thuiswonende jongeren (huishouden met kinderen) van 14-18 jaar, uitgesplitst naar inkomen.

De hierop volgende resultaten zijn gericht op risicoprofielen met een hoog risico op een positieve uitslag, onafhankelijk van het aantal testen wat is uitgevoerd.

In tabel 4 zijn de 10 profielen (op basis van 2 variabelen) weergegeven met het hoogste risico op een positieve test. Hieruit blijkt dat een combinatie van twee van de volgende variabelen: leeftijd (14-18 en 19-26 jaar), een Turkse of Marokkaanse migratieachtergrond en inkomen (>600%, 370-600% en 280-370%), een duidelijke relatie hebben met een hoog percentage positieve testen. Binnen de groep met de leeftijd van 19-26 jaar hebben personen met een Turkse migratieachtergrond gemiddeld een ongeveer

7% hogere kans op een positieve test (23,4% kans) ten opzichte van het gemiddelde van de 19-26-jarigen (16,6%). In tabel 5 is verder ingezoomd op deze groep. Het is te zien dat binnen deze groep de middeninkomens (280-600% van sociaal minimum) en de huishoudens met kinderen een hoger risico hebben op een positieve test.

Tabel 4: Top 10 risicogroepen met het hoogste risico op een positieve testuitslag (%) op basis van 2 variabelen

Profiel		Aantal in NL	Positieve test
Categorie variabele 1	Categorie variabele 2		
19-26	Turks	56.135	23,36%
14-18	>600%	108.500	22,21%
19-26	Marokkaans	51.155	21,86%
14-18	370-600%	313.005	20,72%
Turks	370-600%	55.175	20,58%
14-18	Turks	30.345	20,38%
Marokkaans	370-600%	37.735	20,35%
Turks	280-370%	73.550	20,20%
Marokkaans	280-370%	55.630	19,65%
27-59	Marokkaans	181.955	19,39%

Tabel 5: Invloed van andere variabelen op het infectierisico binnen de groep 19-26-jarigen met een Turkse migratieachtergrond

Variabele	Categorie	Aantal in NL	Positieve test
Inkomen	370-600%	9.275	26,20%
Inkomen	280-370%	11.140	25,72%
Huishoudsamenstelling	Huishouden met kinderen	30.780	24,97%
Stedelijkheidsniveau	1	28.100	24,04%
Inkomen	140-280%	21.235	23,99%

3.2.4 Verschillen VRs en G4 met Nederland

De grote hoeveelheid data maakt het ook mogelijk om risicogroepen op een kleiner geografisch niveau te bekijken. Ter illustratie is er gekeken naar de G4 (Den Haag, Utrecht, Rotterdam en Amsterdam), zie tabel 6, en naar de veiligheidsregio's, zie tabel 7. Bij deze vergelijking is specifiek gekeken naar risicogroepen met twee variabelen aangezien de groeps grootte te klein is wanneer er naar drie variabelen wordt gekeken.

Deze resultaten illustreren dat het per stad of regio kan verschillen welk profiel het hoogste infectierisico heeft. Verdere verdieping in regionale verschillen tussen de steden kan van belang zijn in het ontwerp en toepassen van toekomstige maatregelen en beleid.

Tabel 6: Top 5 groepen met het hoogste risico op een positieve testuitslag op basis van 2 variabelen, weergegeven per G4 gemeente.

vergeleken per G4 gemeente.

Top 5 risicoprofielen in G4 gemeente		Aantal in G4 gemeente	Positieve test
Categorie variabele 1	Categorie variabele 2		
Utrecht			
14-18	>600%	2.360	27,54%
14-18	370-600%	5.310	22,98%
14-18	Geen migratieachtergrond	9.525	22,47%
19-26	Huishouden met kinderen	13.620	21,92%
19-26	Turks	1.930	21,76%
Rotterdam			
14-18	>600%	3.245	27,12%
14-18	Huishouden zonder kinderen	60*	25,00%
14-18	370-600%	5.905	24,72%
Marokkaans	370-600%	3.360	24,26%
19-26	Turks	6.730	24,22%
Amsterdam			
14-18	>600%	4.780	27,09%
14-18	Geen migratieachtergrond	12.935	25,67%
14-18	370-600%	6.625	25,21%
19-26	Turks	5.515	24,48%
19-26	Huishouden met kinderen	26.355	22,69%
Den Haag			
19-26	Turks	5.430	26,24%
19-26	Marokkaans	3.905	24,20%
19-26	4 of meer personen per huishouden	15.600	21,86%
14-18	Huishouden zonder kinderen	70*	21,43%
Turkse	370-600%	4.365	21,42%
Gemiddeld Nederland			
19-26	Turks	56.135	23,36%
14-18	>600%	108.500	22,21%
19-26	Marokkaans	51.155	21,86%
14-18	370-600%	313.005	20,72%
Turks	370-600%	55.175	20,58%

*waarschijnlijk is dit een (kleine) groep 18-jarigen die alleen wonen

*waarschijnlijk is dit een (kleine) groep 18-jarigen die alleen wonen

Tabel 7: Groepen met het hoogste risico op een positieve testuitslag op basis van 2 variabelen, weergegeven voor 10 veiligheidsregio's.

Risicoprofielen in 10 veiligheidsregio's		Aantal in veiligheidsregio	Positieve test
Categorie variabele 1	Categorie variabele 2		
Gooi en Vechtstreek			
14-18	>600%	3.765	30,15%
14-18	Stedelijkheidniveau 1	10.455	24,20%
Zuid-Holland-Zuid			
Turks	370-600%	1.765	28,33%
19-26	Turks	1.710	26,61%
Limburg Noord			
19-26	Turks	1.165	28,33%
Turkse	370-600%	820	26,83%
Amsterdam-Amstelland			
14-18	>600%	6.935	26,82%
14-18	Geen migratieachtergrond	20.010	24,86%
Noord-Holland-Noord			
14-18	Turks	560	26,79%
19-26	Marokkaanse	620	24,19%
Brabant-Noord			
19-26	Turks	1.815	26,72%
19-26	Marokkaans	965	24,87%
Rotterdam - Rijnmond			
19-26	Turks	9.795	24,66%
14-18	>600%	7.235	23,91%
Haaglanden			
19-26	Turks	6.490	25,35%
19-26	Marokkaans	5.075	23,35%
Utrecht			
14-18	>600%	11.980	25,00%
19-26	Turks	4.185	22,22%
Twente			
14-18	370-600%	11.820	24,15%
19-26	Turks	3.035	23,06%

4. Werkpakket 3: Visualisatie verloop infectie in verschillende groepen

4.1 Doel

Het doel is om voor geïdentificeerde groepen het week-by-week infectie verloop laten zien. Deze tijdlijn zal worden gevisualiseerd, waarbij ook gekeken wordt naar de getroffen maatregelen en het effect op deze groepen.

4.2 Uitvoering en resultaten

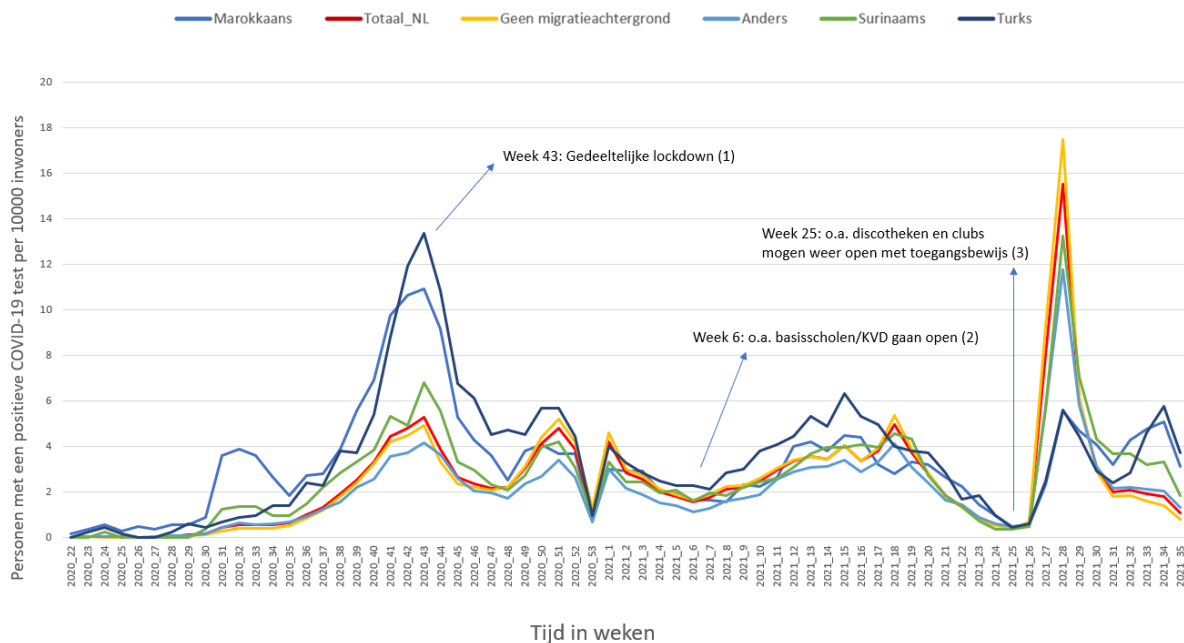
4.2.1 Beschrijving van de data en definities

De gebruikte data in WP3 komt overeen met de data gebruikt in WP2: GGD-testdata uit de periode 06 juni 2020 tot en met 30 augustus 2021 en de ahti zorg-data infrastructuur in de CBS RA-omgeving (zie bijlage 1). Er is gekozen om in de beantwoording van de vragen van dit werkpakket te kijken naar risicogroepen waarbij twee variabelen worden meegenomen gezien de noodzakelijke minimale groepsomvang per week ($n \geq 10$ per groep). Om dezelfde reden is er ook voor gekozen om in deze rapportage naar heel Nederland te kijken.

4.2.2 Resultaat

Kijkende naar de resultaten van WP2 is ervoor gekozen om het week-by-week verloop voor de groep 19-26-jarigen te visualiseren, waarbij verschillende migratieachtergronden worden vergeleken (zie figuur 3). Uit figuur 3 komen duidelijke verschillen in het aantal positieve testuitslagen per week naar voren tussen verschillende groepen binnen deze leeftijdscategorie. Wanneer er bijvoorbeeld wordt ingezoomd op week 34 tot en met 50 van 2020, blijkt dat 19-26-jarigen met een Turkse migratieachtergrond het hoogste aantal positieve testen hebben, gevolgd door 19-26-jarigen met een Marokkaanse migratieachtergrond. In de piek vanaf week 26 tot week 29 van 2021, zijn juist de 19-26-jarigen zonder migratieachtergrond verantwoordelijk voor het hoogste aantal positieve testen per 1000.

De oorzaak voor deze verschillen zou kunnen zijn dat mensen met verschillende migratieachtergronden gemiddeld anders reageren op nieuwe maatregelen (gedeeltelijke lockdown) en versoepelingen ('Dansen met Janssen'), dit pleit voor meer onderzoek naar het effect van maatregelen op risicogroepen met verschillende migratieachtergronden.



Figuur 3a: Het wekelijks verloop van het gemiddeld aantal positieve testen voor de groep 19-26-jarigen met verschillende migratieachtergronden. Belangrijk om te vermelden is dat de dip in week 53 wordt veroorzaakt door een afwijking in de registratie.

14 oktober Gedeeltelijke lockdown

- Mensen mogen max. 3 bezoekers ontvangen per dag (kinderen tot 13 jaar tellen niet mee), met 1,5 m afstand.
- Sluiting horeca.
- Verbod op evenementen.
- In het voorgetzet onderwijs (VO), MBO en hoger onderwijs (HO) draagt iedereen een mondkapje buiten de les.
- Sport max. 4 personen, op 1,5 m. Wedstrijden zijn niet toegestaan.

(1) Maatregelen 14 oktober 2020

8 februari

- Basisscholen en KDV weer open.
- Voor niet-essentiële winkels click and collect toegestaan.

(2) Maatregelen 8 februari 2021

26 juni

- Er zijn geen regels meer voor het aantal bezoekers thuis.
- Er zijn geen regels meer voor het aantal mensen in groepen buiten. 1,5 m afstand houden geldt nog steeds.
- Mondkapje alleen nog verplicht op plekken waar 1,5 m afstand houden niet kan. Wel nog verplicht in ov, op vliegvelden en op middelbare scholen.
- Minder thuiswerken. Advies is om max. de helft van de werktijd op kantoor te werken.
- Winkels, restaurants, cafés en theaters mogen langer open. Er gelden nog wel regels om 1,5 m afstand te kunnen houden.
- Discotheken en clubs mogen weer open. Met coronatoegangsbewijs.
- Sommige evenementen kunnen weer met een coronatoegangsbewijs.
- Amateurwedstrijden mogen weer. Ook met publiek. Publiek houdt 1,5 m afstand.

(3) Maatregelen 26 juni 2021

4.2.3 Online dashboard 'Het verloop van de COVID-19 pandemie over tijd en onder verschillende Nederlandse bevolkingsgroepen'

Om de beschikbare data rondom het testen voor een COVID-19 infectie inzichtelijk te maken voor een breed publiek, is er een online dashboard gebouwd via deze link: <https://healthinsights.ahti.nl/ahticovid/>. Op dit dashboard kun je testdata per veiligheidsregio zien, evenals de ontwikkeling van het aantal COVID-19 infecties over een periode van ruim een jaar. Hierbij kan een keuze gemaakt worden voor welke categorieën binnen de demografische kenmerken testdata wordt vergeleken. Zie figuur 4 voor een screenshot van het online dashboard.



Figuur 4: Screenshot van online dashboard "Het verloop van de COVID-19 pandemie over tijd en onder verschillende Nederlandse bevolkingsgroepen".

5. Werkpakket 4: Duiden van groepen met een hoog risico voor COVID-19 met een ernstig beloop

5.1 Doel

Het doel van WP4 is het duiden van groepen met een hoog risico op COVID-19 met ernstig beloop en in kaart brengen van gehele 'cascade of care' rondom COVID-19, van testen tot infectie tot (IC-)opname in het ziekenhuis tot overlijden.

5.2 Uitvoering en resultaten

5.2.1 Beschrijving van de data en definities

Om mensen met COVID-19 met ernstig beloop in kaart te brengen is er gebruik gemaakt van de CORONIT-data, data van de Landelijke Basisregistratie Ziekenhuiszorg van het jaar 2020, periode 1 juni 2020 tot en met 31 december 2020 (LBZ, Dutch Hospital data via CBS-microdata), voor de sociaal demografische factoren en overige informatie omtrent aandoeningen en zorggebruik (Wmo, Wlz) is gewerkt met de relevante CBS microdatabestanden binnen de ahti zorg-data infrastructuur in de CBS RA-omgeving (zie bijlage 1).

Er is voor de volgende categorieën van beloop gekozen:

1. Alleen positieve test, geen ziekenhuisopname (CORONIT)
2. Klinische opname, ICD10 code bevestigde (U07.1) of vermoedelijke (U07.2) SARS-Cov-2-infectie (LBZ)
3. IC-opname, ICD10 code bevestigde (U07.1) of vermoedelijke (U07.2) SARS-Cov-2-infectie (LBZ)
4. Doodsoorzaak bevestigde (U07.1) of vermoedelijke (U07.2) SARS-Cov-2-infectie (DOODSOORZAKENTAB)

Er is gewerkt met data uit de periode 1 juni 2020 tot 31 december 2020, omdat voor die periode zowel testdata als LBZ-data beschikbaar is. Net als in werkpakket 2 is er gekozen om gebruik te maken van een classification and regression tree (CART) algoritme om binnen de beschikbare data de relevante variabelen en bijbehorende categorieën te kiezen die het meeste invloed hebben op het risico voor ernstig beloop (zie bijlage 3). In bijlage 2 zijn de variabelen en de categorieën weergegeven.

Ook bij deze analyses is ervoor gekozen om de institutionele huishoudens en inkomens niet mee te nemen.

5.2.2 Verschillen tussen groepen m.b.t. ziekenhuisopname, IC-opname en overlijden

Voor alle inzichten die te genereren zijn uit de data geldt dat de vraagstelling bepaalt op welke manier er naar de data gekeken wordt. Is de vraag welke mensen een hoog risico hebben op ernstig beloop? Het antwoord op deze vraag kan waardevol zijn om gericht preventieve maatregelen te kunnen nemen om mensen te beschermen. Of is de vraagstelling gericht op het krijgen van inzicht over de groepen die zorgden voor de meeste ziekenhuisopnames? Met het beantwoorden van deze vraag wordt duidelijk welke groepen het meeste bijdragen aan de druk op de zorg. Aangezien zorgdruk een belangrijke sturende factor is geweest in het Corona-beleid van het VWS én veelbesproken onderwerp in de media, gaat dit hoofdstuk ook in op de groepen mensen die zorgden voor de grootste *aantallen* met ernstig beloop.

In de onderzochte periode (1 juni - 31 december 2020) zijn in de onderzoekspopulatie (n=16.765.940) 19.350 ziekenhuisopnames (0,12%), 1.115 IC-opnames (0,007%) en 7.780 overlijdens (0,046%) geregistreerd. Tabel 8 geeft de risico's en aantallen weer voor ziekenhuisopname, IC-opname en overlijden, uitgaande van één variabele.

Tabel 8: Aantallen en risico's (%) voor ziekenhuisopname, intensive care-opname en overlijden opgesplitst voor verschillende variabelen.

Variabele	Aantal in NL	In ziekenhuis [§]	IC-opname [§]	Overlijden [§]
Leeftijd				
80+	753.165	3.830 (0,51%)	15 (0,002%)	4.640 (0,616%)
70-79	1.552.000	5.300 (0,34%)	255 (0,016%)	2.225 (0,143%)
60-69	2.089.640	4.285 (0,21%)	380 (0,018%)	680 (0,033%)
50-59	2.496.070	3.340 (0,13%)	280 (0,011%)	175 (0,007%)
<50	9.875.065	2.595 (0,03%)	185 (0,002%)	65 (0,001%)
Migratieachtergrond				
Marokkaans	394.800	1.515 (0,38%)	100 (0,025%)	220 (0,056%)
Turks	400.910	1.360 (0,34%)	110 (0,027%)	180 (0,045%)
Surinaams	343.780	1.085 (0,32%)	80 (0,023%)	185 (0,054%)
Anders	2.754.895	3.050 (0,11%)	230 (0,008%)	945 (0,034%)
Geen migratieachtergrond	12.871.550	12.340 (0,10%)	590 (0,005%)	6.255 (0,049%)
Inkomen[#]				
<140%	2.159.775	5.575 (0,26%)	275 (0,013%)	2.860 (0,132%)
140-280%	5.064.190	7.820 (0,15%)	410 (0,008%)	3.875 (0,077%)
280-370%	3.595.730	2.735 (0,08%)	175 (0,005%)	670 (0,019%)
370-600%	4.360.845	2.550 (0,06%)	205 (0,005%)	310 (0,007%)
>600%	1.585.395	675 (0,04%)	50 (0,003%)	65 (0,004%)
Aandoeningen[#]				
Meer dan 1 aandoening	1.622.480	8.660 (0,534%)	405 (0,025%)	4.195 (0,259%)
1 aandoening	3.032.500	5.455 (0,180%)	345 (0,011%)	2.070 (0,068%)
Geen aandoening	12.110.955	5.235 (0,043%)	365 (0,003%)	1.515 (0,013%)
Wmo gebruik				
Gebruik zorg uit de Wmo	845.125	4.415 (0,522%)	120 (0,014%)	3.615 (0,428%)
Gebruik geen zorg uit de Wmo	15.920.815	14.940 (0,094%)	995 (0,006%)	4.165 (0,026%)
Wlz gebruik				
Gebruik zorg uit de Wlz	123.585	510 (0,413%)	n/a*	2.085 (1,687%)
Gebruik geen zorg uit de Wlz	16.642.355	18.845 (0,113%)	1.110 (0,007%)	5.695 (0,034%)
Geslacht				
Man	8.321.985	11.505 (0,138%)	775 (0,009%)	4.620 (0,056%)
Vrouw	8.443.955	7.850 (0,093%)	340 (0,004%)	3.160 (0,037%)
*n/a: Aantallen te klein				
#: Voor een beschrijving van de variabelen, zie bijlage 2.				
§: Risico op ziekenhuisopnames, IC-opnames en overlijdens is het percentage ziekenhuisopnames, IC-opnames of aantal overlijdens binnen het aantal mensen in een bepaalde categorie. Daarbij moet opgemerkt worden dat uit de beschikbare data niet is uit te sluiten dat dezelfde persoon meer dan één keer is opgenomen geweest.				

Duidelijk is dat alle variabelen impact hebben op het beloop van COVID-19. In de volgende paragrafen wordt voor ziekenhuisopname (5.2.3), intensive care-opname (5.2.4) en overlijden (5.2.5) een selectie van inzichten uit de data gepresenteerd, waarbij zowel wordt ingezoomd op risicogroepen alsook de

groepen met de hoogste aantallen. Voornaamste doel is om inzichtelijk te maken wat de mogelijkheden zijn met de beschikbare data. Middels het dashboard zijn alle inzichten online beschikbaar.

5.2.3 Ziekenhuisopname: inzicht in risico's en absolute aantallen

Tabel 8 geeft een eerste inzicht in de impact van verschillende variabelen op het risico op een ernstig beloop van COVID-19. In de onderzochte periode zijn in de onderzoekspopulatie (n=16.765.940) 19.350 ziekenhuisopnames (0,12%) geregistreerd. Om meer inzicht te krijgen in de groepen met hoogste risico en aantallen is het waardevol om naar groepen te kijken ingedeeld op basis van meer dan één variabele. Wanneer meerdere variabelen worden meegenomen, zijn er specifieke groepen te onderscheiden met een afwijkende kans op ziekenhuisopname ten opzichte van de hele populatie. Werken met minder variabelen leidt tot werkbaardere inzichten en groepen met hogere aantallen. Bij het toepassen van de inzichten moet hier rekening mee worden gehouden.

Allereerst wordt maximaal ingezoomd: welk profiel heeft het hoogste risico op ziekenhuisopname? En welke zorgt voor de meeste ziekenhuisopnames? Bij het combineren van 7 variabelen, (maximaal aantal wat de data toelaat) is het profiel met de hoogste kans op een ziekenhuisopname: *vrouwen van >80 jaar met een Turkse migratieachtergrond en een inkomen van ≤140% van het sociaal minimum, met meer dan één aandoening en die zowel zorg uit de Wlz als uit de Wmo gebruikt*. Deze groep heeft 8.7% kans op een ziekenhuisopname. Echter zijn er in Nederland maar 115 vrouwen die binnen deze groep vallen. Het profiel dat in absolute aantallen voor de meeste ziekenhuisopnames zorgt zijn *mannen zonder migratieachtergrond, in de leeftijd 70-79 met een inkomen van 140-280% van het sociaal minimum en meer dan één aandoening, die geen zorg uit de Wmo en geen zorg uit de Wlz gebruiken*, namelijk 665 (3.44%). Dit zijn 2 groepen die weinig overeenkomsten laten zien en illustreert dat een andere manier van kijken naar de beschikbare data geheel verschillende inzichten geeft.

Om middels deze 2 verschillende aanvieligrotes (gericht op inzicht in risico versus aantallen) werkbare inzichten te verkrijgen, wordt er ingezoomd op profielen op basis van 2 en 3 variabelen, in plaats van het maximum van 7. Voor ziekenhuisopname, IC-opname en overlijden wordt er telkens eerst een profiel uitgelicht met een hoog risico en vervolgens een profiel dat verantwoordelijk is voor hoge aantallen. Een voorbeeld van een profiel met een hoog *risico* op ziekenhuisopname zijn mensen met een Marokkaanse migratieachtergrond en die meer dan 1 aandoening hebben (2.25%). Als er vervolgens binnen dit profiel een 3^e variabele wordt meegenomen, komen profielen naar voren met een nog hogere kans op ziekenhuisopname, zie tabel 9. Deze top 5 laat zien dat mensen die gebruik maken van zorg uit de Wlz binnen dit profiel meer dan 2 keer zo veel kans hebben op ziekenhuisopname (namelijk 5.88%) dan de gemiddelde kans op ziekenhuisopname van mensen uit dit profiel (2.25%).

Tabel 9: Top 5 risicogroepen voor ziekenhuisopname, binnen de groep 'Marokkaanse' en 'Meer dan 1 aandoening'.

Profiel		Aantal in NL	In ziekenhuis (risico [§])
Variabele	Categorie		
Wlz gebruik	Ja	595	35 (5,88%)
Leeftijd	80+	2.895	135 (4,66%)
Leeftijd	70-79	7.280	275 (3,78%)
Wmo gebruik	Ja	5.785	210 (3,63%)
Geslacht	Man	15.740	430 (2,73%)

[§] *Risico op ziekenhuisopnames is het percentage ziekenhuisopnames, IC-opnames of aantal overlijdens door (verdenking op) COVID-19 binnen het aantal mensen in een bepaalde categorie. Daarbij moet opgemerkt worden dat uit de beschikbare data niet is uit te sluiten dat dezelfde persoon meer dan één keer is opgenomen geweest.*

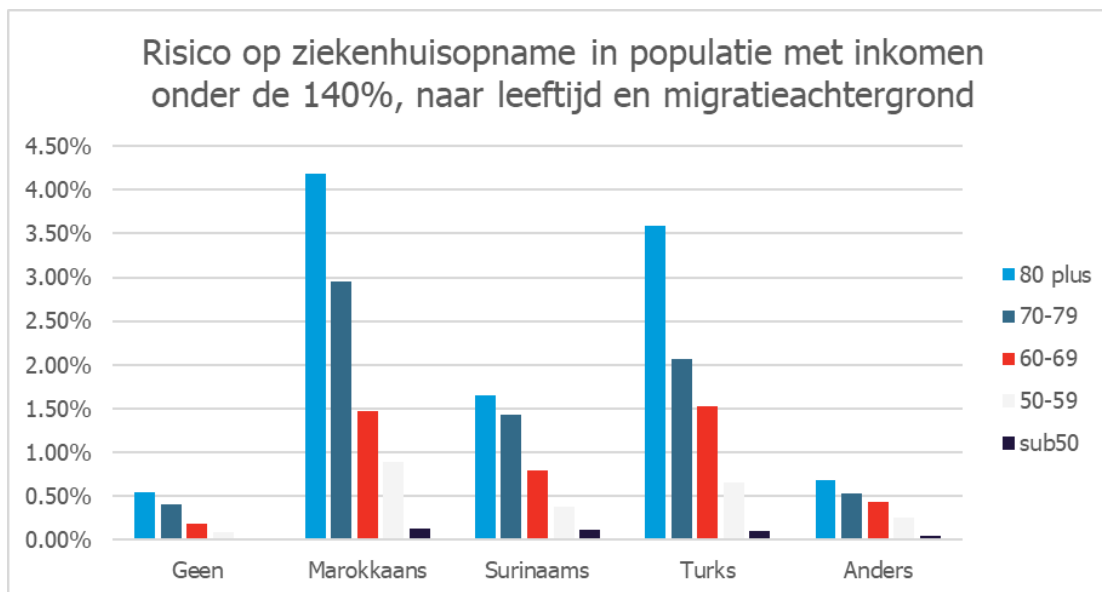
Om meer inzicht te krijgen in de profielen met grote aantallen, wordt er wederom ingezoomd vanuit één profiel. Personen zonder migratieachtergrond zorgden bijvoorbeeld voor een totaal van 12.340 (63,77%) van de ziekenhuisopnames, waarvan 7350 mannen (0.115% risico voor deze groep op opname, 38.0% van alle opnames). Uit tabel 10 wordt duidelijk dat de meeste mensen uit deze groep die in het ziekenhuis zijn opgenomen geweest 1 of meer aandoeningen hebben en geen zorg uit de Wlz of Wmo ontvangen.

Tabel 10: Top 5 profielen in aantallen ziekenhuisopnames, binnen de groep 'Geen migratieachtergrond' en 'Man'.

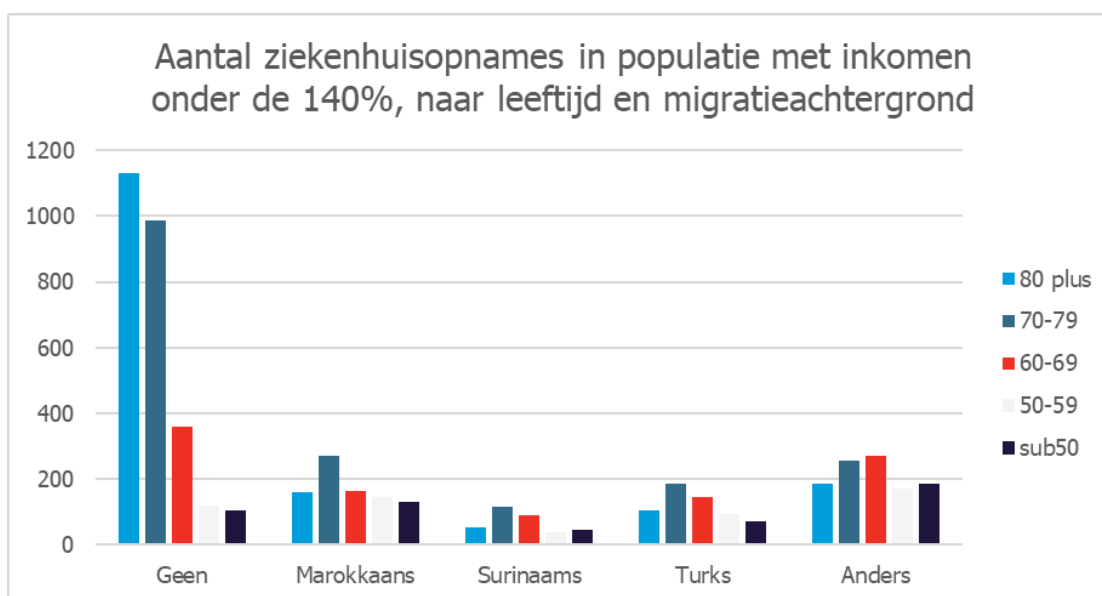
Profiel		Aantal in NL	In ziekenhuis (proportie)
Variabele	Categorie		
Wlz gebruik	Nee	6.372.665	7.200 (37,2%)
Wmo gebruik	Nee	6.173.725	6.075 (31,4%)
Aandoeningen	Meer dan 1	650.485	3.320 (17,2%)
Leeftijd	70-79	644.690	2410 (12,5%)
Aandoeningen	1	1.183.475	2220 (11,5 %)
[§] Proportie (percentage) is het aantal ziekenhuisopnames door (verdenking op) COVID-19 binnen een profiel gedeeld door het totaal aantal ziekenhuisopnames in de periode 1 juni-31 december 2020			

Om nog eens de verschillen tussen de 2 aanvliegroutes te illustreren, worden beide inzichten voor dezelfde profielen hieronder in 2 figuren weergegeven. Voor mensen met een laag inkomen (<140% van het sociaal minimum), worden in figuur 5 de risico's op ziekenhuisopname weergegeven voor profielen op basis van leeftijd en migratieachtergrond; en in figuur 6 worden de aantallen ziekenhuisopnames weergegeven voor dezelfde profielen. Uit figuur 5 wordt duidelijk dat met name de groepen met een Turkse en Marokkaanse migratieachtergrond binnen alle leeftijdsgroepen een hogere kans hebben op ziekenhuisopnames door COVID-19. Kijkende naar de absolute aantallen binnen dezelfde profielen in figuur 6, komt naar voren dat in de leeftijdsgroepen 60-69, 70-79 en 80+ mensen zonder migratieachtergrond voor de hoogste aantallen ziekenhuisopnames zorgen.

Hieruit wordt nog eens duidelijk dat de profielen met hoge risico's er anders uitzien dan de profielen die zorgen voor de meeste ziekenhuisopnames. Een voor de hand liggende verklaring voor de profielen die verantwoordelijk zijn voor de hoogste aantallen ziekenhuisopnames is de samenstelling van de bevolking. Hoewel het risico op ernstig beloop onder mensen met een migratieachtergrond hoger is voor mensen met migratieachtergrond dan zonder migratieachtergrond, zorgden zij dus niet voor de hoogste aantallen ziekenhuisopnames. Er zijn immers veel meer mensen zonder migratieachtergrond dan mensen met een migratieachtergrond in Nederland.



Figuur 5: Risico op ziekenhuisopname in populatie met inkomen onder 140% van het sociaal minimum, naar leeftijd en migratieachtergrond.



Figuur 6: Aantal ziekenhuisopnames in populatie met inkomen onder 140% van het sociaal minimum, naar leeftijd en migratieachtergrond.

5.2.4 Intensive care-opname: inzicht in risico's en absolute aantallen

In totaal zijn er 1.115 IC-opnames door (verdenking op) COVID-19 in de periode 1 juni 2020-31 december 2020 geregistreerd in de LBZ, dit is 0,007% van alle mensen in Nederland (n=16.765.940).

Net als voor ziekenhuisopnames wordt er ook voor IC-opnames ingezoomd op profielen die het hoogste *risico* hebben om op de IC te belanden en profielen die voor het hoogste *aantal* IC-opnames hebben gezorgd. Het profiel met één van de hoogste risico's op IC-opname is 'meer dan 1 aandoening', namelijk 0,025%. In tabel 11 is te zien dat mensen binnen dit profiel met een Marokkaanse migratieachtergrond 5 keer meer kans hebben (namelijk 0,130%) op IC-opname dan gemiddeld in dit profiel.

Tabel 11: Top 5 risicoprofielen voor IC-opname, binnen de groep 'meer dan 1 aandoening'.

Profiel		Aantal in NL	IC-opnames (risico [§])
Variabele	Categorie		
Migratieachtergrond	Marokkaans	30.840	40 (0,130%)
Migratieachtergrond	Turks	36.945	40 (0,108%)
Migratieachtergrond	Surinaams	44.945	35 (0,078%)
Leeftijd	60-69	430.060	170 (0,040%)
Migratieachtergrond	Anders	211.870	75 (0,035%)

[§] Risico (percentage) is het aantal IC-opnames binnen een profiel gedeeld door het totaal aantal mensen in dit profiel in Nederland (periode 1 juni-31 december 2020).

Dat migratieachtergrond impact heeft op het risico op IC opnames is ook terug te zien in de aantallen. Bijna de helft van de personen op de IC in deze periode heeft een migratieachtergrond (47.1%). Verder blijkt dat relatief veel mannen op de IC hebben gelegen in deze periode, namelijk 775. Wanneer er vervolgens verder wordt ingezoomd blijkt het dat het grootste deel van deze mensen geen migratieachtergrond heeft en dat meer dan een derde van deze groep een inkomen heeft van 140-280% van het sociaal minimum (zie tabel 12).

Tabel 12: Top 5 profielen in aantallen voor IC-opname, binnen de groep 'mannen'.

Profiel		Aantal in NL	IC-opnames (proportie [§])
Variabele	Categorie		
Migratieachtergrond	Geen	6.414.070	410 (36,8%)
Inkomen	140-280%	2.449.045	295 (26,5%)
Aandoeningen	> 1 aandoening	810.360	275 (24,7%)
Leeftijd	60-69	1.035.735	270 (24,2%)
Aandoeningen	Geen	6.070.445	255 (22,9%)

[§] Proportie (percentage) is het aantal IC-opnames binnen een profiel gedeeld door het totaal aantal IC-opnames in Nederland in de periode 1 juni 2020-31 december 2020.

5.2.5 Overlijden: inzicht in risico's en absolute aantallen

Tot slot is er gekeken naar de kans op overlijden met een COVID-19 infectie in deze periode. In Nederland zijn er in de onderzochte periode (1 juni 2020 tot 31 december 2020) 7.780 personen overleden aan (verdenking op) COVID-19. Gemiddeld in Nederland komt het risico op overlijden met een COVID-19 infectie daarmee op 0.046%.

Personen die gebruik maken van zorg uit de Wlz hebben een veel hoger risico om te overlijden aan COVID-19, namelijk 50 keer zo hoog risico als mensen die geen gebruik maken van zorg uit de Wlz (1.687%, zie ook tabel 8; en zie de CBS-publicatie 'Ontwikkelingen in sterfte in 2020 en 2021'). Niet alleen gebruik van zorg via de Wlz blijkt een risicofactor, maar ook het gebruik van zorg via de Wmo geeft een verhoogd risico op overlijden. Het risico op overlijden onder mensen die gebruik maken van zorg via de Wmo is namelijk 16 keer zo hoog als mensen die daar geen gebruik van maken (0.428%, zie ook tabel 8). Als er wordt ingezoomd op dit profiel, dus mensen die vallen binnen de categorie 'Wmo gebruik', zijn het, zoals verwacht, mensen die in de hogere leeftijdscategorie vallen een hogere kans hebben om te overlijden. Ook mensen met meer dan 1 aandoening hebben een verhoogd risico, zie tabel 13.

Tabel 13: Top 5 profielen risicogroepen voor overlijden, binnen de groep 'Wmo gebruik'.

Profiel		Aantal mensen	Overlijdens (risico)
Variabele	Categorie		
Leeftijd	80+	281.445	2.575 (0,915%)
Aandoeningen	>1	326.585	2.125 (0,651%)
Inkomen	140-280%	340.690	1.710 (0,502%)
Migratieachtergrond	Marokkaans	18.575	85 (0,458%)
Leeftijd	70-79	176.555	800 (0,453%)

^{§§} Risico (percentage) is het aantal overlijdens binnen een profiel gedeeld door het totaal aantal mensen in dit profiel in Nederland (periode 1 juni-31 december 2020).

Voor wat betreft absolute aantallen zijn er een hoog aantal overlijdens binnen de groep mensen met een inkomen van minder dan 140% van het sociaal minimum, namelijk 2.860 geregistreerde gevallen (34% van alle overlijdens). Inzoomend op dit profiel, blijkt dat mensen zonder migratieachtergrond voor de hoogste aantallen overlijdens zorgen, zie tabel 14. Verder wordt uit de tabel ook duidelijk dat veel mensen met een positieve COVID-19 test die zijn overleden oudere mensen waren, meerdere aandoeningen hadden en gebruik maakten van zorg uit de Wmo.

Tabel 14: Top 5 profielen in aantallen voor overlijden, binnen de groep 'Inkomen minder dan 140% van het sociaal minimum'.

Profiel		Aantal mensen	Overlijdens (Proportie)
Variabele	Categorie		
Migratieachtergrond	Geen	1.258.390	2.115 (27,2%)
Wlz gebruik	Nee	2.114.820	2.060 (26,4%)
Leeftijd	80+	245.815	1.850 (23,8%)
Aandoeningen	>1	405.985	1.695 (21,8%)
Wmo gebruik	Ja	390.880	1.640 (21,1%)

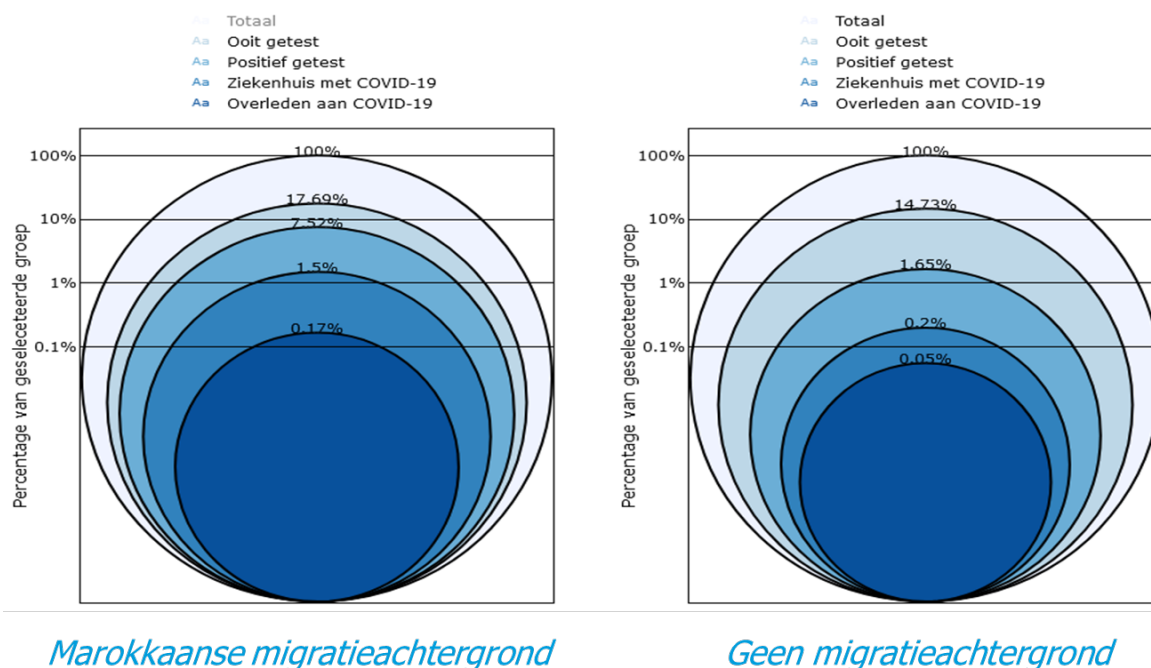
[§] Proportie (percentage) is het aantal overlijdens door (verdenking op) COVID-19 binnen een profiel gedeeld door het totaal aantal overlijdens door (verdenking op) COVID-19 in Nederland in de periode 1 juni 2020-31 december 2020.

5.2.6 Uitwerking cascade van zorg

Door het combineren van testdata, ziekenhuisdata en overlijdensdata is het mogelijk om een zogenaamde cascade van zorg in kaart te brengen. Dit geeft inzicht in waar profielen die verschillende op bepaalde kenmerken anders door de zorg bewegen.

Figuur 6 is een visuele weergave van de cascade van 2 profielen. Het gaat om de vergelijking tussen het profiel met een leeftijd van 60 – 70, inkomen <140% en een Marokkaanse migratieachtergrond en het profiel met zelfde leeftijd en inkomen, maar dan zonder migratieachtergrond, zie ook tabel 15. Het blijkt dat mensen uit een zelfde leeftijds- en inkomenscategorie een ander zorgverloop hebben wanneer ze een Marokkaanse migratieachtergrond hebben.

Zoals benoemd in paragraaf 4.2.3 is er een online dashboard ontwikkeld om de beschikbare data publiek toegankelijk te maken. Naast de testdata geeft het dashboard ook inzicht in de data rondom het ernstig beloop na een positieve test. In het tabblad 'COVID-19 Cascade van zorg' is alle beschikbare data per profiel inzichtelijk in figuur te bekijken – en te vergelijken met een ander profiel – van positieve test, tot ziekenhuisopname en overlijdens. Hier kunnen 1, 2 of 3 variabelen meegenomen worden in het profiel.



Figuur 6: Screenshot van het dashboard met de cascade of care weergegeven voor groepen met een leeftijd van 60 – 70, inkomen <140% en een Marokkaanse migratieachtergrond (links) en zonder migratieachtergrond (rechts).

Tabel 15: Absolute getallen behorende bij de cascade of care voor de groep met een leeftijd van 60 – 70, inkomen <140% en een Marokkaanse migratieachtergrond en zonder migratieachtergrond.

	Marokkaanse migratieachtergrond	Geen migratieachtergrond
Totaal	12.040	209.240
Ooit getest	2.130	30.825
Positief getest	905	3.450
Ziekenhuisopname met COVID-19	180	415
Overleden aan COVID-19	20	115

: Voor een beschrijving van de variabelen, zie bijlage 2.

6. Conclusies INCODA

Dit rapport toont de grote variabiliteit aan mogelijkheden door het koppelen van CBS-microdata aan GGD testdata en data over COVID-19 gerelateerde ziekenhuis-, IC-opnames en overlijden. De opgedane inzichten kunnen worden gebruikt om te ondersteunen in beleidskeuzes bij de aanpak tegen de COVID-19 pandemie.

Afhankelijk van de onderzoeksvraag zijn er verschillende manieren om naar de gekoppelde data te kijken. In alle resultaten komt naar voren dat demografische en sociaaleconomische factoren, zoals inkomen, leeftijd en migratieachtergrond zowel het risico op een SARS-CoV-2 infectie als het risico op een ernstig beloop van COVID-19 sterk beïnvloeden. Daarnaast blijkt dat het koppelen van meerdere sociaal demografische variabelen meer inzicht geeft in de invloed van bepaalde variabelen (zowel risico verhogend als juist beschermend). Dit zou een specifiekere en op bepaalde doelgroepen gericht beleid kunnen ondersteunen.

Ook komt naar voren dat het per stad of regio kan verschillen welke groep het hoogste infectierisico heeft. Bij verdere verdieping in deze verschillen wordt het mogelijk om op geografisch niveau (gemeente, veiligheidsregio, etc.) de risicoprofielen in kaart te brengen. Beleid kan hiermee bijvoorbeeld gericht kunnen worden ontwikkeld. Daarnaast kan er retrospectief gekeken worden naar het effectiviteit van maatregelen op verschillende groepen kijkend naar de daling of stijging van het aantal positieve testen, waardoor specifieke maatregelen in kunnen worden gezet die effectief zijn gebleken voor specifiek burgerprofielen.

Deze rapportage geeft slechts een selectie weer van inzichten die uit de data zijn verkregen. Het gecreëerde dashboard maakt het voor iedereen mogelijk om een gedetailleerder beeld te krijgen van welke groepen, waar en op welk moment gingen testen, positief testten voor het coronavirus en wie er overleden zijn aan de gevolgen van het virus. De inzichten zullen vragen en hypothesen oproepen. Het dashboard en onderliggende data maken het in sommige gevallen mogelijk om deze vragen te beantwoorden en hypothesen te toetsten, in andere gevallen is vervolgonderzoek nodig met aanvullende data.

De vele opgedane inzichten die mogelijk zijn gemaakt door het combineren van CBS microdata met 16 miljoen GGD testuitslagen en ruim 30 duizend COVID-19 gerelateerde ziekenhuis- en IC-opnames van 2020, laten slechts een deel van de mogelijkheden van deze data zien. Daarnaast heeft het werken met CBS-microdata een belangrijke beperking, aangezien CBS de microdata beschikbaar stelt met een vertraging van 2 jaar. Om nog meer uitspraken te kunnen doen over het infectierisico en de kans op een ernstig verloop met COVID-19, is het belangrijk grotere datasets beschikbaar te stellen over een langere periode. Daarnaast zou de data verrijkt kunnen worden met de testdata van de personen die niet bij de GGD getest zijn (OSIRIS) en vaccinatiedata (CIMS). Dit maakt het mogelijk om uitspraken te kunnen doen over welke profielen zich al dan niet laten vaccineren, wat het effect is van vaccinatie op verdere infectierisico's, het risico op een ernstig beloop etc.

Naast inzichten op het gebied van infectierisico en risico op ernstig beloop van COVID-19 heeft dit project ook veel inzichten opgeleverd over de processen rondom het uitwisselen van medische gegevens. Ondanks een extern juridisch advies bleek dat de kennis, ervaring en interpretatie van wet- en regelgeving rondom (CBS-) data erg verschilt per organisatie/ziekenhuis. Dit project staat hier niet alleen in. Het wordt steeds duidelijker dat in Nederland de privacywetgeving veel strikter dan de EU beoogt wordt toegepast (Zie ook: Volkskrant d.d. 23 en 30 augustus 2022).. Dit heeft een remmende werking op de uitvoering van wetenschappelijk onderzoek en vraagt wellicht om een herziening van de huidige wet- en regelgeving.

7. Referenties

De Kruif, Beudel en Vajda. (2022). *Opinie: Bescherming privacy van patiënt moet niet ten koste gaan van de zieken zelf*. Geraadpleegd op 23-08-2022, via: <https://www.volkskrant.nl/columns-opinie/opinie-bescherming-privacy-van-patient-moet-niet-ten-koste-gaan-van-de-zieken-zelf~b303f583/>

Duca, L., Coyle, J., McCabe, C., McLean, C. (2020). *COVID-19 Stats: COVID-19 Incidence, by Urban-Rural Classification – United States, January 22–October 31, 2020. Morbidity and Mortality Weekly Report*, 69(15), 465. doi: [10.15585/mmwr.mm6915e4](https://doi.org/10.15585/mmwr.mm6915e4)

Marmot M, Allen J, Goldblatt P, Herd E, Morrison J. Build Back Fairer: The COVID-19 Marmot Review. *The Health Foundation*; 2020 health.org.uk/publications/build-back-fairer-the-covid-19-marmot-review

Reukers, D., et al. (2021). High infection attack rates of SARS-CoV-2 in Dutch households revealed by dense sampling. *MedRxiv*. <https://doi.org/10.1101/2021.01.26.21250512>

Smit, M., Hospers, G. (2020). *Corona-effect 1: meer ruimte in de grote stad*. De Openbare Ruimte.NU. Geraadpleegd op 12-09-2022, via: <https://deopenbareruimte.nu/corona-effect-1-meer-ruimte-in-de-grote-stad/>

Schürmann, R., Van der Toorn, J., Stoeldraijer, L. (2022, 16 februari). *Ontwikkelingen in sterfte in 2020 en 2021*. Centraal Bureau voor de Statistiek. Geraadpleegd op 10-09-2022, via: <https://www.cbs.nl/nl-nl/longread/diversen/2022/ontwikkelingen-in-sterfte-in-2020-en-2021>

Van Rietbergen. (2022). *Opinie: Nederland handhaaft privacywetgeving veel strikter dan de EU beoogt*. Geraadpleegd op 30-08-2022, via: <https://www.volkskrant.nl/columns-opinie/opinie-nederland-handhaaft-privacywetgeving-veel-strikter-dan-de-eu-beoogt~b00a07ed/>

Bijlagen

Bijlage 1: Omschrijving van de CBS microdata omgeving

CBS heeft vanuit haar wettelijke taken inzicht in de achtergrondkenmerken van alle inwoners van Nederland. De Wet op het Centraal Bureau voor de Statistiek staat analyse toe van CBS-microdata door daartoe geautoriseerde partijen voor statistisch of wetenschappelijk onderzoek. Onderzoek met CBS-microdata kan alleen onder strikte voorwaarden door geautoriseerde onderzoekers worden uitgevoerd binnen de RA-omgeving van het CBS. Microdata mag bovendien worden "verrijkt" met persoonsgegevens mits deze rechtmatig zijn verkregen en de voorgeschreven uploadprocedures worden gebruikt. Middels deze maatregelen wordt voldaan aan de AVG.

Ahti (projectleider en penvoerder) is een CBS-microdata-geautoriseerde instelling en heeft daarmee toegang tot de CBS RA-omgeving. De uiteindelijke analyses zijn uitgevoerd in het CBS microdataproject 9096 - Aandoeningen: gebruik over de regelingen heen. Voor INCODA is de microdata (voornamelijk demografische, geografische en sociaaleconomische kenmerken en medische voorgeschiedenis en zorggebruik) verrijkt met een aantal databestanden (zie werkpakket 1). Na analyses op individueel niveau binnen de CBS RA-omgeving zijn de resultaten aangeboden ter controle op herleidbaarheid aan het CBS, waarna het toegestaan is om de resultaten te rapporteren.



Bijlage 2: Omschrijving van de variabelen

Variabelen CBS microdata

In dit onderzoek is gewerkt met verschillende (socio-demografische) variabelen, hieronder beschreven, voor alle variabelen geldt de situatie in het jaar 2020;

Migratieachtergrond

<i>Geen migratieachtergrond</i>	Een persoon zonder migratieachtergrond heeft als herkomstgroepering Nederland. Wanneer een persoon niet in Nederland geboren is, maar beide ouders zijn wel in Nederland geboren, dan wordt de herkomst-groepering van de persoon ook Nederland.
<i>Niet-westers</i>	Persoon met als geboorteland van de ouders en/of de persoon zelf een van de landen in Afrika, Latijns-Amerika en Azië (exclusief Indonesië en Japan) of Turkije.
<i>Westers</i>	Persoon met als geboorteland van de ouders en/of de persoon zelf een van de landen in Europa (exclusief Turkije), Noord-Amerika en Oceanië, en Indonesië en Japan.
<i>Turks =</i>	Persoon die in Turkije is geboren met ten minste één in Turkije geboren ouder.
<i>Marokkaans</i>	Persoon die in Marokko is geboren met ten minste één in Marokko geboren ouder.
<i>Surinaams</i>	Persoon die in Suriname is geboren met ten minste één in Suriname geboren ouder.
<i>Anders</i>	Persoon met een andere migratieachtergrond dan 'geen migratieachtergrond', 'Turkse', 'Marokkaanse' of 'Surinaamse'.

Inkomen

De inkomens zijn weergegeven in percentages van het sociaal minimum, geldend voor het jaar 2020. Het sociaal minimum is het wettelijk bestaansminimum zoals dat in de politieke besluitvorming is vastgesteld. Om te kunnen beoordelen hoe het inkomen van een huishouden zich verhoudt tot het minimum, moet aan de hand van de regelgeving worden vastgesteld welke norm voor het desbetreffende huishouden van toepassing is. De norm voor een (echt)paar met uitsluitend minderjarige kinderen is bijvoorbeeld gelijkgesteld aan de bijstandsuitkering van een echtpaar, aangevuld met de (leeftijdsafhankelijke) kinderbijslag. Bij 65-plussers is het bedrag aan AOW-pensioen als norm gekozen.

<i><140%</i>	Persoon woont in een huishouden met een inkomen tot 140% van het sociaal minimum
<i>140-280%</i>	Persoon woont in een huishouden met een inkomen van 140-280% van het sociaal minimum
<i>280-370%</i>	Persoon woont in een huishouden met een inkomen van 280-370% van het sociaal minimum
<i>370-600%</i>	Persoon woont in een huishouden met een inkomen van 370-600% van het sociaal minimum
<i>>280/>600%</i>	Persoon woont in een huishouden met een inkomen van meer dan 280%/600% van het sociaal minimum
<i>Onbekend /institutioneel:</i>	Inkomen onbekend e/o persoon woont in een instellingen zoals verpleeg-, en verzorgingshuizen, instellingen voor geestelijke gezondheidszorg, forensische centra, instellingen voor verstandelijk, lichamelijk en zintuiglijk gehandicapten, instellingen voor verslavingszorg en daklozenopvang, internaten, kloosters, gevangenissen, kazernes, en asielzoekerscentra, waarin de personen in principe voor langere tijd (zullen) verblijven.

Huishoudsamenstelling

<i>Huishouden met kinderen</i>	Persoon woont in een particulier huishouden bestaande uit twee ouders met één of meer thuiswonende kinderen.
<i>Tweepersoonshuishouden zonder kinderen</i>	Persoon woont in een tweepersoonshuishouden zonder kinderen.
<i>Eénpersoonshuishouden</i>	Persoon woont in een éénpersoonshuishouden.
<i>Eenouderhuishouden</i>	Persoon woont in een particulier huishouden bestaande uit één ouder met één of meer thuiswonende kinderen.
<i>Institutioneel huishouden</i>	Persoon woont in een huishouden met één of meer personen samen in een woonruimte en worden bedrijfsmatig voorzien in dagelijkse levensbehoeften. Ook de huisvesting vindt bedrijfsmatig plaats. Het gaat om instellingen zoals verpleeg-, en verzorgingshuizen, instellingen voor geestelijke gezondheidszorg, forensische centra, instellingen voor verstandelijk, lichamelijk en zintuiglijk gehandicapten, instellingen voor verslavingszorg en daklozenopvang, internaten, kloosters, gevangenissen, kazernes, en asielzoekerscentra, waarin de personen in principe voor langere tijd (zullen) verblijven.
<i>Ander type huishouden</i>	Overig, anders dan bovenstaande

Zorg- en sociale voorzieningen gebruik

<i>Gebruik zorg uit de Wmo</i>	Persoon heeft in 2020 gebruik gemaakt van zorg waarvan de kosten voor rekening van de Wet Maatschappelijke Ondersteuning komen.
<i>Gebruik zorg uit de Wlz</i>	Persoon heeft in 2020 gebruik gemaakt van zorg waarvan de kosten voor rekening van de Wet Langdurige Zorg komen.
<i>Gebruik Diabetes medicatie</i>	Aan deze persoon is in 2020 medicatie versterkt voor Diabetes type I of type II vergoed uit de verplichte basisverzekering voor geneeskundige zorg.
<i>Gebruik COPD medicatie</i>	Aan deze persoon is in 2020 medicatie versterkt voor COPD of astma (onderscheid tussen de ziekten is hier niet mogelijk) vergoed uit de verplichte basisverzekering voor geneeskundige zorg.
<i>Gebruik CVRM medicatie</i>	Aan deze persoon is in 2020 medicatie versterkt voor CVRM (antitrombotica, hartmiddelen, cholesterolverlagers en middelen tegen hoge bloeddruk) vergoed uit de verplichte basisverzekering voor geneeskundige zorg.
<i>Gebruik medicatie tegen hartaandoeningen</i>	Aan deze persoon is in 2020 medicatie versterkt voor hartaandoeningen vergoed uit de verplichte basisverzekering voor geneeskundige zorg.

Aandoeningen

Een persoon wordt geclassificeerd als het hebben van een aandoening wanneer aan de persoon medicatie tegen deze aandoening is verstrekt in het jaar 2020, vergoed uit de basisverzekering voor geneeskundige zorg. In dit rapport zijn de volgende aandoeningen meegenomen: astma/COPD, Diabetes type I, Diabetes type II, Hartaandoeningen, CVRM.

<i>1 aandoening</i>	Aan deze persoon is in het jaar 2020 medicatie voor 1 chronische aandoening verstrekt.
<i>Meer dan 1 aandoening</i>	Aan deze persoon is in het jaar 2020 medicatie voor meer dan 1 chronische aandoening verstrekt.
<i>Geen aandoening</i>	Aan deze persoon is in het jaar 2020 geen medicatie voor een chronische aandoening verstrekt.

Aantal personen per huishouden

<i>Minder dan 4 personen per huishouden</i>	Een persoon woont in 2020 in een huishouden bestaande uit minder dan 4 personen.
<i>4 of meer personen per huishouden</i>	Een persoon woont in 2020 in een huishouden bestaande uit 4 of meer personen.

Stedelijkheidsniveau

De indeling van gemeenten naar stedelijkheid is gebaseerd op de omgevingsadressendichtheid van de gemeente. De vijf stedelijkheidsklassen zijn gebaseerd op klassegrenzen van 2500, 1500, 1000 en 500 adressen per km².

De volgende klassen worden onderscheiden:

<i>Stedelijkheid niveau 1</i>	Zeer sterk stedelijk (omgevingsadressendichtheid van 2500 of meer)
<i>Stedelijkheid niveau 2</i>	Sterk stedelijk (omgevingsadressendichtheid van 1500 tot 2500)
<i>Stedelijkheid niveau 3</i>	Matig stedelijk (omgevingsadressendichtheid van 1000 tot 1500)
<i>Stedelijkheid niveau 4</i>	Weinig stedelijk (omgevingsadressendichtheid van 500 tot 1000)
<i>Stedelijkheid niveau 5</i>	Niet stedelijk (omgevingsadressendichtheid van minder dan 500)

Leeftijd

Afhankelijk van de onderzoeksvraag is er gewerkt met verschillende categorieën voor de variabele leeftijd:

WP1/WP2 (t/m § 3.2.2)	Na gebruik CART in WP2 (vanaf §3.2.3) en WP3	WP4
0-11	0-8	<50
12-18	9-13	50-59
19-27	14-18	60-69
26-65	19-26	70-79
65-75	27-59	80+
75+	60-70	
	>70	

Bijlage 3: Beschrijving van de classification and regression tree (CART) methode

Om de variabelen en onderliggende categorieën te identificeren die het meeste invloed hebben op de verschillende COVID-19 gerelateerde uitkomsten, hebben we een classification and regression tree (CART) algoritme toegepast op de data. Startpunt was gebaseerd op eerdere resultaten met de standaard variabelen geslacht, migratieachtergrond, inkomen, leeftijd en huishoudsamenstelling. Daarnaast zijn voor het analyseren van de SARS-CoV-2 testdata de volgende variabelen ook meegenomen in de analyse: stedelijk niveau en aantal personen in huishouden. Voor analyses rondom het beloop van COVID-19 zijn tenslotte (het aantal) chronische aandoeningen, Wmo gebruik en Wlz gebruik meegenomen. Om het CART algoritme te implementeren maakten we gebruik van de rpart toolbox binnen de R software (versie 4.2.1).

In deze statische analyse, schatten we een 'regression tree' op de data. Hierbij labelden we de uitkomstmaat (bijvoorbeeld 'test positief' of 'in ziekenhuis opgenomen') met '100' wanneer deze is gerealiseerd en '0' wanneer de uitkomstmaat *niet* is gerealiseerd.

In eerste instantie kozen we ervoor om een grote boom te genereren, door de drempelwaarde van verbetering in root-mean-square-error (RMSE) op het lage niveau van $10e-7$ in te stellen. Vervolgens hebben we de boom verkleind door middel van 10-voudige 'cross validation'. Door visuele inspectie van de out-of-sample error hebben we de originele boom teruggebracht tot een punt waar de verdere verbetering in RMSE bij hogere diepte minimaal werd. Dit resulteerde in de identificatie van bepalende categorieën met de meeste invloed op de COVID-19 gerelateerde uitkomsten (zie bijlage 1 voor een uitleg van de variabelen en de gekozen categorieën).