



Beschrijving transmissiemodel berekening zorgbelasting

Voorlopige versie, 2 april 2021

Wat is het doel van dit rapport?

Dit is een eerste aanzet tot een "levend document" waarin het model waarmee de verwachte zorgbelasting door COVID-19 in Nederland wordt berekend, op inzichtelijke wijze wordt beschreven. Dit model wordt gebruikt bij het adviseren van de bestrijding van de COVID-19 pandemie. De uitkomsten van dit model worden regelmatig gepresenteerd in de technische briefing aan de Tweede Kamer. De kwaliteit van de gemaakte prognoses is door een ieder na te gaan. Dat kan door deze publiekelijk beschikbare prognoses van aantal opnames op de intensive care (IC) per dag te vergelijken met de later gerapporteerde waargenomen aantal opnames per dag in Nederland.

Wat is het doelpubliek van dit rapport?

Dit rapport is specifiek gericht op onderzoekers (niet-modelleurs). We komen hiermee tegemoet aan opmerkingen dat onze communicatie voornamelijk gericht is op of het algemeen publiek of op vakinhoudelijke experts (modelleurs). Maar niet op andere onderzoekers die geen modelleur zijn. Dit rapport beoogt ook dit publiek aan te spreken. We gaan er in dit rapport van uit dat dit publiek toegang heeft tot relevante wetenschappelijke artikelen en handboeken.

Waarom dit rapport?

De onderwerpen hieronder komen overeen met de belangrijkste punten die benoemd moeten worden om een model transparant te beschrijven, conform een publicatie over transparante beschrijving van modellen (Eddy et al, 2012).

Op welke vragen moet dit model antwoord geven?

- Het primaire doel is de berekening van het aantal verwachte IC-opnames en ziekenhuisopnames met COVID-19 in Nederland in de komende één tot drie weken, en berekening van de verwachte bezetting van IC en ziekenhuisbedden met COVID-19 patiënten in die periode.
- Het secundaire doel is het doorrekenen van scenario's voor alternatieve maatregelenpakketten. Ook bij dit secundaire doel worden ziekenhuis- en IC-opnames en bezette bedden berekend, maar dan om uitkomsten onderling te kunnen vergelijken.
- Een tertiair doel is het berekenen van tegenfeitelijke uitkomsten ("counterfactuals"). Die laten zien wat er gebeurd zou zijn als er geen verandering in een maatregelenpakket zou zijn geweest, ondanks het feit dat er wel verandering is geweest.

Waar wordt dit model voor gebruikt?

Het model wordt gebruikt voor het analyseren en vertalen van de binnenkomende gegevens over de epidemie naar een prognose voor wat komen gaat, met inachtneming van onzekerheden. Specifiek gaat het bij het primaire doel om korte

termijn prognoses van het verwachte aantal IC-opnames per dag bij gegeven maatregelen en gedrag. Bij het secundaire en tertiaire doel gaat het om de middellange termijn prognoses van het aantal IC-opnames per dag.

Omdat we hier een nieuwe ziekte modelleren, moet het model regelmatig aangepast kunnen worden. Bijvoorbeeld voor nieuwe varianten en voor het vaccinatieprogramma. Het model moet daarom flexibel zijn, en deze beschrijving van het model is een "levend document" dat als dat nodig is steeds aangepast wordt.

Wat is de relatie van dit model tot andere rekenmodellen die het RIVM gebruikt?

Soms wordt gesproken van "het" RIVM model, waarbij de website deelaspecten geeft van dit model, terwijl een openbaar rapport waarin "het" model in zijn geheel wordt beschreven zou ontbreken. Dat ligt genuanceerder. Het RIVM gebruikt verschillende rekenmodellen en modelleringstudies voor verschillende aspecten van de COVID-19 epidemie, die elk een verschillend doel hebben. Deze modellen en studies worden beschreven op de website. In dit rapport geven we aan op welke punten het specifieke model voor berekening van verwachte aantallen IC opnames per dag gebruik maakt van informatie uit die andere modellen en studies.

Hoe is het werk gefinancierd?

Het werk is ontwikkeld binnen de afdeling modellering van infectieziekten bij het RIVM. Het RIVM is een onafhankelijk agentschap van het ministerie van Volksgezondheid, Welzijn en Sport. De afdeling modellering van infectieziekten ontvangt geen financiering van onderzoeksprojecten door de farmaceutische industrie.

Wat zijn de belangrijkste uitkomstvariabelen van het model

- Aantal IC opnames per dag met COVID-19
- Aantal ziekenhuisopnames per dag met COVID-19
- IC bezetting per dag met COVID-19 patiënten
- Ziekenhuisbezetting per dag met COVID-19 patiënten

Wat is de structuur van het model?

Het transmissiemodel is een zogenoemd compartimentenmodel, de dynamiek kan worden weergegeven door een stelsel gewone differentiaalvergelijkingen.

De bevolking is voor dit model ingedeeld in leeftijdsgroepen. Een model met een dergelijke structuur heet leeftijdsgestructureerd.

Belangrijkste toestandsvariabelen voor elk van deze leeftijdsgroepen zijn het aantal bevattelijke mensen (*S*, susceptibles), het aantal mensen dat besmet is maar nog niet besmettelijk (*E*, exposed), het aantal mensen dat besmettelijk is (*I*, infectious), en het aantal mensen dat immuun is (*R*, removed). Een model waarbij de compartimenten deze indeling hebben is van het SEIR-type.

Met dit transmissiemodel kan worden bepaald hoeveel nieuwe besmettingen per dag plaatsvinden in elke leeftijdsgroep. Voor elke leeftijdsgroep weten we de kans dat besmetting leidt tot ziekenhuisopname en tot IC-opname, en welke verdeling van tijdsduur er is tussen besmetting en ziekenhuisopname, en van

ziekenhuisopname tot IC-opname. Zo kunnen we de besmettingen per dag omrekenen naar aantallen ziekenhuis- en IC-opnames per dag.

Waarom leeftijdsstructuur?

Leeftijd is de belangrijkste variabele die de ernst van de ziekte bepaalt. Het aantal contacten tussen personen waarbij SARS-CoV-2 kan worden overgedragen is ook sterk afhankelijk van leeftijd. Maatregelen, zoals het sluiten van scholen, zorgen voor minder contacten in bepaalde leeftijdsgroepen, zoals kinderen. Daarnaast zijn de prioriteiten van het vaccinatieschema in belangrijke mate gericht op specifieke leeftijdsgroepen.

Waarom SEIR-type?

Dit is een standaard modelstructuur voor besmettelijke infectieziekten die van mens op mens overdraagbaar zijn. Vele internationale publicaties van transmissiemodellen nemen deze structuur aan bij het beschrijven van SARS-CoV-2 transmissie. Er zijn veel eigenschappen van dit model type bekend en beschreven in handboeken.

Wat zijn de gegevensstromen die het model informeren?

- Ziekenhuisgegevens : leeftijd, dag van opname in ziekenhuis, dag van opname op de intensive care, sterfte, ligduur op de intensive care, ligduur in het ziekenhuis. Dit wordt gehaald uit gegevens van de stichting NICE. Deze ruwe gegevens zijn geen eigendom van het RIVM, en bevatten privacygevoelige informatie zoals leeftijd, ziektehistorie, woonplaats, zodat ze niet gedeeld kunnen worden. De geaggregeerde gegevens worden dagelijks gepubliceerd op de RIVM website (<https://data.rivm.nl/covid-19/>).
- Meldingen naar leeftijd, eerste ziektedag en ziekenhuisopname. Infecties met SARS-CoV-2 zijn meldingsplichtig. Ze worden vastgelegd in het Osiris registratiesysteem. De ruwe gegevens uit het Osiris registratiesysteem bevatten privacygevoelige informatie zodat ze niet gedeeld kunnen worden. Geaggregeerde en gepseudonimiseerde gegevens worden dagelijks gepubliceerd op de RIVM website (<https://data.rivm.nl/covid-19/>).
- Aandeel van mensen met aantoonbare antistoffen tegen het SARS-CoV-2 virus, naar leeftijd, en tijdsperiode van bloedafname. Deze informatie wordt verkregen uit de Pienter-Corona studie (<https://www.rivm.nl/pienter-corona-studie>).
- Het aantal personen per leeftijdsgroep in 2020, dit zijn door het CBS gepubliceerde gegevens.
- Het aantal contacten per dag, naar leeftijdsgroep zoals zelf gerapporteerd door deelnemers die een dwarsdoorsnede vertegenwoordigen van de Nederlandse bevolking. In de studie zijn de contacten uitgesplitst naar verschillende situaties, zoals werk, school, thuis en vrije tijd. Deze studie is beschreven in Backer et al., 2021, en te vinden onder het kopje "De impact van COVID-19 maatregelen op aantal contacten per persoon in Nederland" op de website <https://www.rivm.nl/coronavirus-covid-19/hoe-berekeningen-bijdragen-aan-bestrijding-van-virus/rekenmodellen>
- Het serieel interval. Informatie hierover was in eerste instantie gebaseerd op analyse van gegevens uit het buitenland (Tindale et al., 2020, Ganyani et al., 2020), maar leunt nu op analyse van data uit het Osiris registratiesysteem.

- De incubatieperiode. Informatie hierover komt van analyses uit het buitenland (Lauer et al., 2020), bevestigd door data uit het Osiris registratiesysteem.
- Voor het bepalen van een seizoenseffect gebruiken we dagelijkse klimatologische gegevens (temperatuur en luchtvochtigheid) in de Bilt, <https://www.knmi.nl/nederland-nu/klimatologie-metingen-en-waarnemingen>
- Voor het bepalen van een seizoenseffect gebruiken we ook de schatting van reproductiegetal R naar dag, zoals we die in de open data van het RIVM publiceren <https://data.rivm.nl/covid-19/>
- Het aandeel van varianten met dag van monstername in een representatieve steekproef van bemonsterde patiënten. Dit wordt verkregen uit de kiemsurveillance, <https://www.rivm.nl/coronavirus-covid-19/onderzoek/kiemsurveillance> .
- Het geleverde (verleden) en geplande (toekomst) aantal vaccinaties per week voor elke leeftijdsgroep, uitgesplitst naar vaccin. Dit wordt samengesteld op basis van de vaccinatieplanning gemaakt door RIVM, Logistiek Coördinatie Centrum COVID-vaccinatie, voor het ministerie van VWS.

Hoe zijn de gegevensbronnen geselecteerd?

De belangrijkste gegevensbron is aantal opnames per dag op de intensive care in Nederland. Deze gegevens worden verzameld en gerapporteerd door de stichting NICE, vanaf het begin van de epidemie. Vrijwel iedereen die opgenomen wordt op de IC wordt getest op COVID-19. Deze gegevens zijn gebaseerd op patiënten. Andere beschikbare gegevensstromen zijn minder geschikt omdat ze gebaseerd zijn op aantal bezette bedden, of omdat ze niet alle Nederlandse ziekenhuizen beslaan, of omdat ze maar voor een gedeelte van de epidemie zijn verzameld en gerapporteerd.

Hoe wordt het model gefit aan de gerapporteerde IC-opnames per dag?

De berekende aantallen IC opnames per dag worden regelmatig (minimaal wekelijks) gefit aan de meest recente rapportages van aantal IC opnames per dag. De te fitten parameter is een maat voor besmettelijkheid en naleving van de maatregelen, de waarde is constant tussen omslagpunten. De omslagpunten vallen samen met invoering van nieuwe maatregelen of invoering van versoepelingen. Dit zijn de momenten waarop ook contactfrequenties veranderen zoals hier beschreven.

De gebruikte likelihoodfunctie veronderstelt een Poisson verdeling voor het aantal opnames per dag. Het aantal omslagpunten p wordt bepaald door middel van minimalisatie van het zogeheten Akaike Informatie Criterium (AIC), waarbij een keuze wordt gemaakt voor het aantal omslagpunten waarbij de AIC de kleinste waarde heeft.

Door het optimaliseren van de likelihoodfunctie worden de aantallen infecties, via gemodelleerde IC opnames gefit aan het waargenomen aantal IC-opnames.

Hoe worden waarden van overige parameters en variabelen bepaald?

De meeste gegevens worden geanalyseerd om de parameter waarden te bepalen. Dit gebeurt regelmatig, voordat het model aan de IC-opnames per dag gefit wordt.

- We fitten de overgangskansen tussen compartimenten E,I,R zodanig dat de gemiddelde generatietijd, de tijdsduur tussen moment van infectie van de besmetter en de besmette persoon, overeenkomt met de waargenomen serieel interval, de tijdsduur tussen begin van symptomen van een primair geval en een secundair geval.
- De begincondities worden zo gekozen dat ze overeenkomen met een geheel vatbare populatie. De aantallen personen in elke leeftijdsgroep is zoals gerapporteerd door het CBS.
- De contactfrequenties waarmee de leeftijdsgroepen onderling contact maken komen uit de contactstudies (Backer et al., 2021). Bij elke verandering van maatregelenpakket worden andere frequenties gebruikt, die worden aangepast door in te schatten in welke mate contacten in elk van de situaties uit de contactstudie (zoals school, werk, thuis) worden gereduceerd. Voor dit inschatten wordt gebruik gemaakt van openbare gegevens over aantallen mensen werkzaam in verschillende beroepsgroepen, en vrijetijdsbesteding. Zo mogelijk worden de ingeschatte frequenties gevalideerd aan nieuwe contactstudiegegevens verzameld tijdens de COVID-19-epidemie.
- Een leeftijdsafhankelijke factor voor vatbaarheid en besmettelijkheid. We fitten het model aan serologische gegevens die aangeven hoeveel mensen in elke leeftijdsgroep eerder zijn besmet geraakt met het virus (<https://www.rivm.nl/pienter-corona-studie/resultaten>).
- Een seizoensafhankelijke vermenigvuldigingsfactor voor de transmissiekans, berekend uit de data van het KNMI en de geschatte waarde van het reproductiegetal R_t .
- Een tijdsafhankelijke vermenigvuldigingsfactor voor de transmissiekans, voor verschillende varianten, een berekend via data van de kiemsurveillance.
- Voor het vaccineren een leeftijds- en tijdsafhankelijke snelheid waarmee mensen direct van vatbaar (S) naar immuun (R) gaan, zonder besmet te worden. Deze komt uit de vaccinatieplanning
- De kans om na besmetting naar ziekenhuis en IC gaan, per leeftijdsgroep. Deze zijn berekend uit de serologische gegevens (<https://www.rivm.nl/pienter-corona-studie/resultaten>), en de ziekenhuis- en IC-opnames (stichting NICE)
- De tijdsintervallen tussen infectie, ziekenhuisopname- en ontslag, IC-opname en ontslag, per leeftijdsgroep. Deze gegevens komen deels uit het registratiesysteem Osiris, en deels uit de registratie van ziekenhuis- en IC-patiënten van stichting NICE

Hoe weten we of het de samenhang tussen de waarden van de variabelen klopt?

Hier maken we gebruik van eigenschappen die voor SEIR-type modellen al eerder zijn uitgezocht. Door het fitten van het model aan aantal IC-opnames per dag krijgen we de juiste groeisnelheid van de epidemie. Omdat we zorgen dat de gemiddelde generatietijd in het model overeenkomt met waarnemingen klopt de tijdschaal van de infectiecyclus. Tezamen met het fitten aan IC opnames zorgt dit

ervoor dat de waarde van de groeisnelheid wordt vertaald in de juiste waarde voor groeisnelheid per generatie van infectie; dit bepaalt dat het reproductiegetal en daarmee de benodigde bestrijdingsintensiteit klopt (Wallinga & Lipsitch, 2007). Door de schatting van de leeftijdsafhankelijke factor voor vatbaarheid en besmettelijkheid kloppen de leeftijdsspecifieke aantallen nieuwe infecties en de infectiedruk in elke leeftijdsgroep. Dit garandeert dat de sensitiviteit van het model voor leeftijdsspecifieke interventies klopt (Wallinga et al., 2010).

Hoe wordt rekening gehouden met vaccinatie?

We bepalen de vaccineffectiviteit na de eerste en tweede dosis voor de verschillende typen vaccins op basis van gepubliceerde gegevens in de wetenschappelijke literatuur (zie <https://www.rivm.nl/documenten/expected-outcome-of-covid-19-vaccination-strategies-voor-een-samenvatting>). Op basis van de vaccinatieplanning bepalen we hoeveel vaccins van welk type per dag naar welke leeftijdsgroep gaat als eerste of als tweede dosis. In het model berekenen we dan wanneer en hoeveel vatbare mensen in elke leeftijdsgroep door vaccinatie immuun worden. De leeftijdsspecifieke toestandsvariabele S neemt af en de leeftijdsspecifieke toestandsvariabele R neemt toe met dit aantal per dag. Een onderliggende aanname is dat het vaccin volgens een alles-of-niets manier werkt: na vaccinatie is de gevaccineerde of volledig vatbaar of volledig immuun.

Hoe wordt rekening gehouden met varianten?

Een virus verandert voortdurend. Er circuleren vele varianten van het SARS-CoV-2 virus. We onderscheiden de oude variant, de Britse variant, de Zuid-Afrikaanse variant en de Braziliaanse variant. In het model berekenen we de verhouding in het aantal infecties tussen deze varianten door op de gegevens uit de kiemsurveillance te fitten. Op basis van de relatieve toenamesnelheid van een variant berekenen we de relatieve besmettelijkheid van deze variant. Door de toename te extrapoleren berekenen we de verandering in verwachte besmettelijkheid van infecties.

Hoe wordt de effectiviteit van maatregelen ingeschat?

Het contactpatroon tussen leeftijdsgroepen wordt regelmatig gemeten in Nederland in twee verschillende studies (Backer et al., 2021; Jarvis et al., 2020). De verzamelde gegevens worden online beschikbaar gesteld. Voor nieuwe maatregelen kan vooraf worden ingeschat voor welke leeftijdsgroepen dit een verandering geeft in het aantal contacten, in welke categorie van contacten ze een verandering betekenen (thuis, werk, school, transport, vrije tijd, overig), en hoe groot deze veranderingen kunnen zijn. Het effect van een maatregelenpakket wordt dus uitgedrukt in een verandering van het aantal contacten, per leeftijdsgroep, per contactcategorie. Op basis hiervan wordt een nieuw contactpatroon berekend. Omdat deze inschatting van verandering onzeker is, houden we hier expliciet rekening met onzekerheid. Dat doen we door de ingeschatte percentages verandering in contactfrequentie per contactcategorie aan te passen met een willekeurig gekozen percentage, van 5 procent eraf tot 5 procent erbij. Zo verkrijgen we bij elke inschatting 200 nieuwe contactpatronen, en elk van die nieuwe contactpatronen wordt gebruikt als het nieuwe maatregelenpakket in gaat.

Evaluatie achteraf gebeurt enerzijds door het fitten van het model aan de IC-opnames. Hier wordt via de te fitten parameter (een maat voor besmettelijkheid en naleving van de maatregelen) zo gecorrigeerd dat de uiteindelijke stijging of daling overeenkomt met de waargenomen aantallen IC-opnames per dag. Evaluatie achteraf gebeurt anderzijds door de gebruikte contactpatronen te vergelijken met

de binnenkomende gegevens voor zelf-gerapporteerde contactpatronen in de studies. Discrepancies kunnen hier worden gesignaleerd en gecorrigeerd.

Hoe wordt rekening gehouden met rapportagevertraging?

Het aantal IC opnames per dag dat wordt gerapporteerd door de stichting NICE wordt voor de meest recente dagen regelmatig aangevuld. Dat houdt in dat het aantal opnames dat voor vandaag en gisteren is gemeld altijd te laag is. Hier houden we rekening mee door te corrigeren voor deze rapportagevertraging. Die correctie is direct berekend uit de historie van IC-opnames tijdens de COVID-19-epidemie en de dagen waarop elke opname in de gegevens terecht kwam. De statistische achtergrond van dergelijke algoritmes voor correctie voor rapportagevertraging is te vinden in bijvoorbeeld Van de Kastele et al. 2017.

Hoe wordt rekening gehouden met een seizoenseffect?

We berekenen het reproductiegetal R per dag in de loop van de tijd (<https://www.rivm.nl/coronavirus-covid-19/rekenmodellen#hoe-bereken-je-de-r-454461-more>). We hebben ook tijdreeksen met diverse factoren die mogelijk dit reproductiegetal beïnvloeden. Dit zijn de genomen maatregelen, en interne factoren zoals de opbouw van immuniteit in de bevolking, maar ook externe factoren zoals temperatuur. Met een statistische analyse (te Beest et al) kunnen we de bepalen welke mate van effect aan externe factoren zoals temperatuur zou kunnen worden toegeschreven. Dit gebruiken we als schatting van een seizoenseffect.

Hoe wordt de onzekerheid in de uitkomst berekend?

Er worden 200 sets met parameterwaarden getrokken, en voor elke set wordt het model doorgerekend. Van deze 200 uitkomsten voor elke dag berekenen we de 2,5e percentiel, het 50e percentiel en het 97,5e percentiel. Deze geven we weer als ondergrens, mediaan, en bovengrens van de voorspellingen.

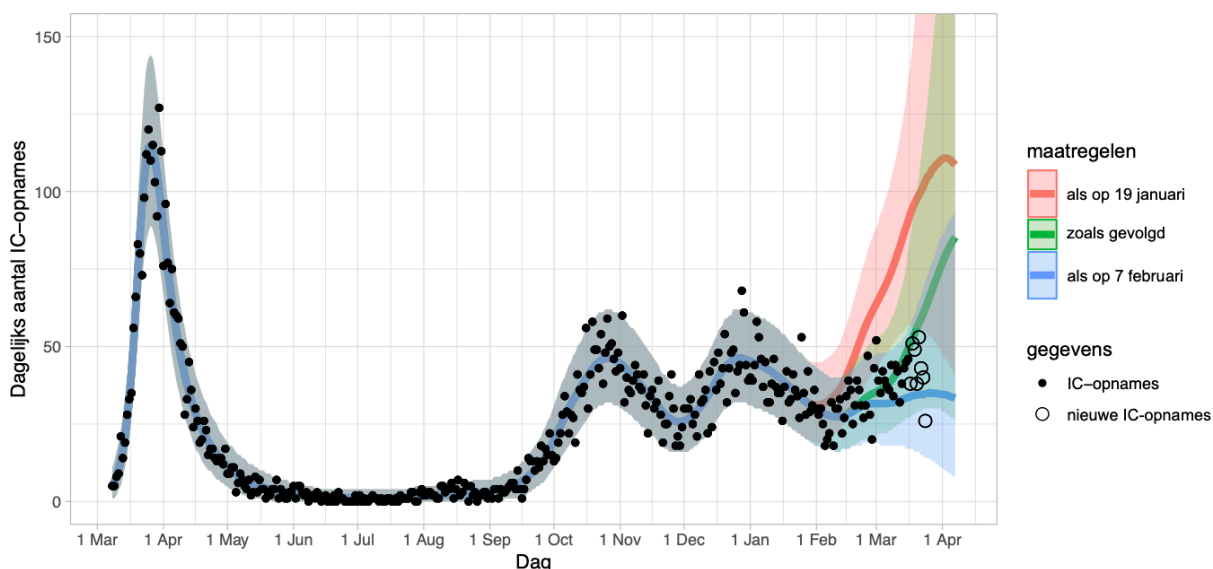
De verdeling van parameterwaarden waaruit getrokken wordt is bepaald door de intrinsieke onzekerheid rondom de parameterwaarden, het relatieve gewicht van een parametercombinatie in de trekking wordt bepaald door de likelihoodfunctie. Parametercombinaties die resulteren in een goede fit op de waargenomen IC-opnames worden zo zwaarder gewogen dan parametercombinaties die niet passen bij de waarnemingen.

Hoe is het model gevalideerd?

In de eerste golf van de pandemie vergeleken we de modeluitkomsten met de uitkomsten van vergelijkbare modellen voor zorgbelasting door COVID-19 in Nederland van de onafhankelijke teams Gupta in opdracht van het LCPS (niet gepubliceerd) en TNO (van Wees et al., 2020). Dit is een externe validatie.

We vergelijken modelprognoses voor IC-opnames met de daadwerkelijke gerapporteerde aantallen. Dit is een check op het gebruik voor het maken van korte termijn voorspellingen. Omdat de prognoses voor IC-opnames per dag en IC-bezetting regelmatig worden gepresenteerd in de technische briefing en omdat de gerapporteerde aantallen IC opnames per dag en IC bezetting te raadplegen zijn op de website van de stichting NICE kan iedereen achteraf de betrouwbaarheid van de voorspellingen vaststellen.

Een voorbeeld is de prognose van 18 maart 2021, waarbij het model (groene lijn) is gefit op beschikbare gegevens op 18 maart (zwarte punten), en de figuur ook de meer recente waarnemingen tot 25 maart 2021 laat zien (open cirkels). De aangegeven trend van de voorspelling, een stijging in aantal IC-opnames per dag, is inderdaad uitgekomen. De waarnemingen liggen op en onder de mediaan, hetgeen aangeeft dat de voorspelling voor die dag iets te somber was. Hierbij moet worden aangetekend dat de meest recente waarnemingen niet gecorrigeerd zijn voor rapportagevertraging, de uiteindelijke aantallen opnames zullen dus hoger liggen dan de open cirkels nu aangeven, en beduidend hoger voor de meest recente waarneming, de open cirkel rechtsonder.



Figuur 1. Prognose van IC opnames gedaan op 18 maart op basis van beschikbare gegevens bij Stichting NICE. Het model met maatregelen zoals gevolgd is gefit aan deze gegevens (groene lijn geeft de mediaan van model uitkomsten, de groene band geeft het 95% interval). De dagen waarvoor op 18 maart gegevens beschikbaar waren (zwarte punten) en waarvoor op 18 maart geen gegevens beschikbaar waren maar wel op 26 maart (open cirkels), met waarden tot en met 25 maart zoals gerapporteerd door NICE op 26 maart, met informatie voor meest recente dagen niet gecorrigeerd voor rapportagevertraging. De rode lijn geeft de mediane uitkomsten van het model in de tegenfeitelijke (counterfactual) situatie waarbij vanaf 19 januari geen verandering in de maatregelen zouden zijn geweest; de rode band geeft het 95% interval. De blauwe lijn geeft de mediane uitkomsten van het model in de tegenfeitelijke (counterfactual) situatie waarbij vanaf 7 februari geen verandering in de maatregelen zouden zijn geweest; de blauwe band geeft het 95% interval.

Waar zijn de wetenschappelijke rapporten waarin validiteit en betrouwbaarheid van scenario studies beoordeeld kunnen worden?

We krijgen van onderzoekers (niet-modelleurs) soms de opmerkingen dat wetenschappelijke rapporten ontbreken waarin de validiteit en betrouwbaarheid van scenariostudies beoordeeld kunnen worden.

Dit klopt, dergelijke rapporten zijn er niet. Het is immers niet mogelijk de validiteit en betrouwbaarheid van verschillende hypothetische scenario's te beoordelen, aangezien deze scenario's nooit tegelijkertijd werkelijkheid kunnen worden. En vaak zelfs niet één van de scenario's de werkelijkheid wordt, bijvoorbeeld doordat maatregelen in de tussentijd worden aangepast. En dus kunnen de uitkomsten van deze hypothetische scenario's in de praktijk niet achteraf met elkaar vergeleken worden. Door onderzoekers die deze opmerking maken, worden dan ook geen voorbeelden aangehaald van wetenschappelijke rapporten door andere onderzoeksgroepen waarin de validiteit en betrouwbaarheid van scenariostudies wel beoordeeld zou kunnen worden.

Zoals aan het begin van dit document beschreven, is het primaire doel van het model voor IC opnames voor de toepassing van korte-termijn-prognoses. Hiervoor is het model ontwikkeld en wordt het op getest. Een tertiair doel was laten zien wat er gebeurd zou zijn als een maatregelenpakket voortgezet zou zijn, hoewel de werkelijkheid anders was, een tegenfeitelijke (counterfactual) uitkomst. Deze zijn weergegeven als rode en blauwe lijn en band in figuur 1. Voor de eerste paar weken zijn deze uitkomsten even valide als korte-termijn-prognoses.

Het is ook mogelijk om hetzelfde model in te zetten voor het verkennen van toekomstige situaties waarin verschillende alternatieven voor interventies worden verkend. Dat is een scenario analyse, en het secundaire doel van dit model. Zoals hierboven uitgelegd, zijn er geen waarnemingen om te testen of de modelvoorspellingen in deze situaties uit zouden zijn gekomen. Impliciete criteria voor valide scenario analyses bestaan uit een combinatie van de volgende factoren: het bereik van de uitkomstvariabelen is beperkt tot realistische waarden (het berekende aantal IC-opnames kan niet negatief zijn, of groter dan de bevolkingsomvang); de causale structuur van het model beschrijft het proces (leidt vaccinatie tot de juiste veranderingen in aantal vatbaren en in aantal IC-opnames per dag); en voorspellend vermogen in een veranderende omgeving (komend de voorspelde waarden voor IC-opnames per dag overeen met latere waarnemingen). Deze criteria zijn afgedekt met het gebruik van een standaard model zoals het SEIR-type waarin transmissie van infectie expliciet wordt meegenomen, en met het in vele eerdere infectieziektmodelleringsstudies aantoonbaar voorspellend vermogen van zo'n SEIR-model.

Wat is de afbakening van het toepassingsgebied?

We nemen aan dat immuniteit langdurig is. Dit is een redelijke aanname voor het afgelopen jaar en op korte termijn, maar het is onzeker of dit ook op langere termijn (komende jaren) een redelijke aanname is. Voor het primaire doel (korte-termijn-prognoses) is dat niet relevant, want dat wordt opgevangen door een andere schatting van de transmissieparameters bij fitten op de IC-incidentie. Voor het secundaire doel (scenario-analyses) zorgt dit ervoor dat de onzekerheid groter wordt, met een mogelijke onderschatting van het verloop van de epidemie.

We nemen geen geboorte en sterfte mee in het model. Op de langere termijn (komende jaren) ontstaat er discrepantie tussen aantallen mensen

per leeftijdsgroep en het aandeel immune mensen per leeftijdsgroep. Ook dit heeft geen gevolgen voor het primaire doel, maar mogelijk wel voor het secundaire doel.

We nemen aan dat eenieder in een bepaalde leeftijdsgroep eenzelfde contactpatroon heeft, en met eenzelfde kans zich laat vaccineren. Omdat er heterogeniteit is in contactpatronen en in acceptatie in vaccinatie, bijvoorbeeld in sociale groepen of geografisch, zal het model sommige aspecten van de epidemiologie missen. Dit zorgt ervoor dat het aandeel van immune mensen benodigd voor groepsimmunitet in het model een onderschatting geeft van het daadwerkelijk benodigd aandeel.

Staan de code en de data op github?

Als we broncode en datasets bij de verschillende modellen delen, lopen we tegen juridische en praktische problemen aan. Grootste probleem is de privacygevoelige informatie in veel van de databronnen die we gebruiken. Daarnaast zijn we van verschillende databronnen geen eigenaar. Om deze hindernis te omzeilen hebben we van diverse databronnen, zoals ziekenhuisopnames en IC opnames, ook geanonimiseerde en gepseudonimiseerde versies online gezet. Die staan bij de RIVM open data (<https://data.rivm.nl/covid-19/>). Die data kan je goed gebruiken als je de berekeningen zelf wilt reconstrueren, maar het zijn niet de gegevens zoals wij die gebruiken. Andere data die wij zelf verzamelen, zoals de aantal contacten in de bevolking, maken wij zelf zo snel mogelijk beschikbaar, ook via publicaties en open data.

We publiceren de broncode (zoals gebruikelijk) bij peer review publicaties in wetenschappelijke tijdschriften. Om te laten zien hoe de code werkt is het nodig de data die we niet kunnen delen te vervangen door "dummy" data. Het genereren van dummy data voor de toestand van de beschikbare datasets op elk moment, documenteren van de code, en het meegeven van instructies is tijdrovend. De code wordt tijdens een lopende epidemie met steeds nieuwe vragen doorlopend aangepast. We maken de resultaten van onze geactualiseerde berekeningen zo snel mogelijk openbaar, door middel van technische briefings en het delen van resultaten online. Het is echter onmogelijk om in eenzelfde tempo ook broncode met dummy data en instructies te ontwikkelen en beschikbaar te stellen. Op dit moment is dat onhaalbaar en we moeten ons concentreren op onze primaire taak, het adviseren ten behoeve van de infectieziekte bestrijding.

De voornaamste componenten van het model die de transmissie bepalen zijn standaard. Wij geven het type aan (SEIR-type), hier kan iedereen zelf bestaande code bij zoeken in de favoriete taal, of zelf de code schrijven. Voor degenen die het niet zelf willen opzoeken of schrijven geven we op de website ook een voorbeeld van code voor zo'n standaard model type. Een manier om zelf een vergelijkbaar model te draaien met vergelijkbare databronnen is om het zelf in elkaar te zetten.

Een snellere manier of zelf een vergelijkbaar model te draaien is om de code en data over te nemen van mensen die hetzelfde al eerder hebben gedaan. Bijvoorbeeld een model dat door TNO is ontwikkeld (van Wees et

al 2020), en een model dat door de Universiteit van Utrecht is ontwikkeld (Rozhnova et al 2021).

Is er een publicatie of technische documentatie?

De vergelijkingen die het transmissiemodel weergeven zijn standaard, vergelijkbaar met die zoals gebruikt in andere COVID-19 modellen. Op de website is daarom alleen het modeltype weergegeven, en met welk soort gegevens dit modeltype geïnformeerd wordt. De relevante informatie zit in de gegevens waarop we ons baseren.

We richten ons allereerst op het beschikbaar maken van de gegevens om dit type model te informeren. Zo worden via de opendata website het RIVM gegevens beschikbaar gemaakt over IC-opnames en ziekenhuisopnames in Nederland zonder daarbij privacy-gevoelige informatie prijs te geven. De informatie die het RIVM zelf verzamelt om dit type model te informeren wordt ook beschikbaar gesteld, via een preprint server en via een publicatie (Backer et al. Eurosurveillance 2021).

Dit "levend document" richt op onderzoekers (niet-modelleurs), en legt het model voor IC prognoses zo uit dat de betrouwbaarheid van de uitkomsten kan worden beoordeeld.

Technische documentatie volgt hieronder, dit is in ontwikkeling en zal in volgende versies van dit levende document geactualiseerd worden.

We hopen daarnaast zo snel mogelijk tijd te vinden om het model ook op te schrijven als wetenschappelijke publicatie, zodat het voor peer-review kan worden aangeboden aan andere experts. Bij het uiteindelijk publiceren worden de gebruikte code en gegevens (in dit geval, dummygegevens om geen privacy-gevoelige informatie te onthullen) openbaar.

Kan het model worden gereproduceerd?

Het model en de data zijn nauw verweven; we fitten het model wekelijks aan de meest recente gegevens, de onzekerheid in de parameters nemen we mee in de onzekerheid van de prognoses. Het exact reproduceren van het model vereist toegang tot gegevens die we vanwege o.a. privacy-aspecten niet zomaar kunnen delen.

Als er geen beperking zou zijn van menskracht zouden we dummydata aanmaken en dit met de geschoonde code op onze website verstrekken en regelmatig updaten. Dit is met beperkte menskracht simpelweg onmogelijk.

Omdat de gebruikte gegevensstromen vanwege privacyaspecten niet gedeeld kunnen worden, maakt het RIVM voor geïnteresseerden de geanonimiseerde gegevens openbaar via de open data (<https://data.rivm.nl/covid-19/>). We geven het type model aan dat we gebruiken, zodat geïnteresseerden zelf met deze open data onze berekeningen in grote lijnen kunnen volgen.

Het is niet nodig om eerst zelf een simulatiemodel na te bouwen en te fitten aan gegevens om bepaalde uitkomsten te kunnen volgen. Men kan gebruik maken van bestaande theorie om met relatief weinig rekenwerk specifieke uitkomsten te controleren. De gebruikte modelstructuur van het SEIR-type is standaard, en het meerekenen kan op basis van de vergelijkingen en theorie zoals te vinden in wikipedia of in handboeken. Hieronder de beantwoording van een voorbeeld vraag: *Hoeveel SARS-CoV-2 infecties zijn voorkomen tijdens de eerste golf door de genomen maatregelen?*

Een eerste grove benadering:

We lezen het basis reproductiegetal R_0 af als de waarde van het reproductiegetal voordat maatregelen werden genomen. Deze informatie is te vinden in de wekelijkse rapportages van het RIVM, in de open data van het RIVM, en op het dashboard van de rijksoverheid. We nemen een waarde van R_0 ongeveer 2.

De proportie van de bevolking die besmet raakt als iedereen willekeurige contacten maakt en het gedrag niet verandert tijdens de epidemie noemen we z . De waarde van z kan worden berekend als de waarde R_0 bekend is. De berekening staat bekend als de "final size equation" (te vinden in de introductie hoofdstukken van handboeken, pagina 15 Diekmann et al, pagina 15 Bjornstad, pagina 116 Held et al, ook te vinden op wikipedia). De waarde van z is de oplossing van de vergelijking $z = 1 - e^{-R_0 z}$. Voor $R_0 = 2$ vinden we ongeveer $z = 0.8$. Dus bij een basisreproductiegetal van 2 raakt 80% van de bevolking besmet als we niet ingrijpen. We hebben waargenomen dat bij de eerste golf van COVID-19 in Nederland ongeveer 5% besmet is geraakt met SARS-CoV-2 (zie de RIVM website, Pienter Corona studie). Dat betekent dat, aan het eind van de eerste golf, ongeveer 94% van alle mogelijke infecties voorkomen zijn.

Een tweede, meer verfijnde benadering:

Bij de voorgaande berekening namen we aan dat iedereen willekeurige contacten maakt. Hier kunnen we het waargenomen leeftijdsspecifiek patroon van contacten gebruiken. Dit wordt door het RIVM regelmatig gemeten, en de resultaten zijn beschikbaar gesteld op medRxiv en bij de publicatie in Eurosurveillance. De waargenomen matrix met het aantal leeftijdsspecifieke contacten per persoon, zo geschaald dat de grootste eigenwaarde van de matrix overeenkomt met de gevonden waarde van $R_0 = 2$, noemen we \mathbf{M} . De leeftijdsspecifieke proportie van de bevolking die besmet raakt als iedereen willekeurige contacten maakt en dit gedrag niet verandert tijdens de epidemie noemen we \mathbf{z} . De waarde van \mathbf{z} kan worden berekend als de matrix \mathbf{M} bekend is. Dit doen we met de generalisatie van de "final size equation" voor een gestructureerde populatie (te vinden op pagina 220 Diekmann et al). De waarde van \mathbf{z} is de oplossing van de matrixvergelijking $\mathbf{z} = 1 - e^{-\mathbf{M}\mathbf{z}}$. De uitkomst kunnen we vergelijken met de leeftijdsspecifieke proportie dat bij de eerste golf van COVID-19 in Nederland besmet is geraakt met SARS-CoV-2 (zie de RIVM website, Pienter Corona studie). We krijgen dan een leeftijdsspecifiek percentage voorkomen infecties. Als we dit wegen met een leeftijdsspecifieke kans om in de IC opgenomen te worden na infectie, dan hebben we ook een percentage voor het voorkomen percentage IC opnames.

In ons werk verfijnen we deze aanpak verder door rekening te houden met leeftijdsspecifieke vatbaarheid, besmettelijkheid en rekening houden met de onzekerheid hieromtrent. Ook houden we rekening met de sensitiviteit en specificiteit van de serologische testen in de Pienter Corona studie, en waarbij ook de onzekerheid omtrent zowel de contactpatronen als ook de serologie volledig wordt meegenomen. Voor dit laatste is het nodig om over de brondata te beschikken. En als ook de brondata van NICE over IC opnames naar leeftijd beschikbaar is, is het mogelijk om daarmee het aantal voorkomen IC opnames te berekenen.

Technische documentatie

De bevolking wordt ingedeeld in 9 leeftijdsgroepen: 0-9 jaar, 10-19 jaar, ..., 80+ jaar oud. De aantallen per leeftijdsgroep worden overgenomen van de website van het CBS.

In de specifieke implementatie van het SEIR-type die we hier gebruiken splitsen twee van deze groepen iets verder uit: E^1 , het aantal mensen dat net is besmet, E^2 het aantal mensen dat bijna besmettelijk is, I^1 het aantal mensen dat net besmettelijk is geworden, I^2 het aantal mensen dat bijna niet meer besmettelijk is. Dit resulteert in een verdeling van de generatietijd die meer past bij de waarnemingen.

De modelvergelijkingen zijn voor elke leeftijdsgroep i :

$$\dot{S}_i = -\lambda_i S_i$$

$$\dot{E}_i^1 = \lambda_i S_i - \gamma E_i^1$$

$$\dot{E}_i^2 = \gamma E_i^1 - \gamma E_i^2$$

$$\dot{I}_i^1 = \gamma E_i^2 - \gamma I_i^1$$

$$\dot{I}_i^2 = \gamma I_i^1 - \gamma I_i^2$$

$$\dot{R}_i = \gamma I_i^2$$

De parameter γ bepaald hoe snel de infectiestadia doorlopen worden. De waarde van deze parameter wordt zo gekozen dat het gemiddelde generatie interval in het model overeenkomt met wat we uit de Osiris gegevens kunnen bepalen. De parameter λ is de kans per tijd om besmet te raken (infectiedruk) in leeftijdsgroep i . Deze kans hangt af van het aantal besmettelijke personen in de leeftijdsgroepen, en hun kans om contact te maken met een persoon in leeftijdsgroep i .

$$\lambda_i = \beta_i \sigma_i \sum_j c_{ij} \rho_j (I_j^1 + I_j^2) / N$$

De parameter ρ geeft de relatieve besmettelijkheid van iemand in leeftijdsgroep j , de parameter σ geeft de relatieve vatbaarheid van iemand in leeftijdsgroep i . De parameter c_{ij} geeft een maat voor de intensiteit van contacten tussen leeftijdsgroepen i en j . De parameter β geeft dan een maat voor de besmettelijkheid van het virus. De waarde van deze parameter wordt geschat door het model te fitten aan waarnemingen. N is hier de grootte van de bevolking.

De parameter c_{ij} wordt als volgt berekend. We geven het totaal aantal contacten van personen in leeftijdsgroep i met personen in leeftijdsgroep j weer als q_{ij} . De parameter c_{ij} is dan $c_{ij} = q_{ij} N / (N_i \times N_j)$, waarbij N_i het aantal mensen in leeftijdsgroep i is, N_j het aantal mensen in leeftijdsgroep j is, N het totaal aantal mensen in de bevolking is. De aantallen slaan hier op het moment van meting van het aantal contacten.

Met dit transmissiemodel kan worden bepaald hoeveel nieuwe besmettingen per dag plaatsvinden in elke leeftijdsgroep. Voor elke leeftijdsgroep weten we de kans dat besmetting leidt tot ziekenhuisopname en tot IC-opname, en welke verdeling van tijdsduur er is tussen besmetting en ziekenhuisopname, en van ziekenhuisopname tot IC-opname. Zo kunnen we uitrekenen welke aantallen IC-opnames er per dag verwacht kunnen worden.

$$\mu_{i,t} = \int g(\lambda_{i,\tau} S_{i,\tau}) d\tau$$

De berekende aantallen IC opnames per dag worden gefit aan het waargenomen aantal IC opnames per dag. De gefitte parameter is een maat voor besmettelijkheid en naleving van de maatregelen, de waarde is constant tussen omslagpunten. De omslagpunten vallen samen met invoering van nieuwe maatregelen of invoering van versoepelingen.

De gebruikte likelihoodfunctie veronderstelt een Poisson verdeling voor het aantal opnames per dag, y_t .

$$L(\beta | y_t) = \text{Poisson}(\sum_i \mu_{i,t}; \beta)$$

Referenties

- Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic?. *The lancet*, 395(10228), 931-934.
- Backer, J. A., Mollema, L., Vos, E. R., Klinkenberg, D., Van Der Klis, F. R., De Melker, H. E., van den Hof, S. & Wallinga, J. (2021). Impact of physical distancing measures against COVID-19 on contacts and mixing patterns: repeated cross-sectional surveys, the Netherlands, 2016–17, April 2020 and June 2020. *Eurosurveillance*, 26(8), 2000994.
- Bjørnstad, O. N. (2018). *Epidemics. Models and data using R*: Springer International Publishing, 318
- Diekmann, O., Heesterbeek, H., & Britton, T. (2012). *Mathematical tools for understanding infectious disease dynamics (Vol. 7)*. Princeton University Press.
- Eddy, D. M., Hollingworth, W., Caro, J. J., Tsevat, J., McDonald, K. M., & Wong, J. B. (2012). Model transparency and validation: a report of the ISPOR-SMDM Modeling Good Research Practices Task Force-7. *Medical Decision Making*, 32(5), 733-743.
- Ganyani, T., Kremer, C., Chen, D., Torneri, A., Faes, C., Wallinga, J., & Hens, N. (2020). Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance*, 25(17), 2000257.
- Jarvis, C. I., Van Zandvoort, K., Gimma, A., Prem, K., Klepac, P., Rubin, G. J., & Edmunds, W. J. (2020). Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC medicine*, 18, 1-10.
- Held, L., Hens, N., D O'Neill, P., & Wallinga, J. (Eds.). (2019). *Handbook of infectious disease data analysis*. CRC Press.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., Azman, A.S., Reich, N.G. & Lessler, J. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9), 577-582.
- Van Wees, J. D., Osinga, S., van der Kuip, M., Tanck, M., Hanegraaf, M., Pluymaekers, M., Leeuwenburgh, O., van Bijsterveldt, L., Zindler, J. & van Furth, M. T. (2020). Forecasting hospitalization and ICU rates of the COVID-19 outbreak: An efficient SEIR model. *Bull World Health Organ*.
- te Beest, D. E., van Boven, M., Hooiveld, M., van den Dool, C., & Wallinga, J. (2013). Driving factors of influenza transmission in the Netherlands. *American journal of epidemiology*, 178(9), 1469-1477.
- Rozhnova, G., van Dorp, C. H., Bruijning-Verhagen, P., Bootsma, M. C., van de Wiggert, J. H., Bonten, M. J., & Kretzschmar, M. E. (2021). Model-based evaluation of school-and non-school-related measures to control the COVID-19 pandemic. *Nature communications*, 12(1), 1-11.
- Tindale, L. C., Stockdale, J. E., Coombe, M., Garlock, E. S., Lau, W. Y. V., Saraswat, M., Lee, Y-H.B., Chen, D., Wallinga, J., & Colijn, C. (2020). Evidence for transmission of COVID-19 prior to symptom onset. *Elife*, 9, e57149.

- van de Kastele, J., Eilers, P. H., & Wallinga, J. (2019). Nowcasting the number of new symptomatic cases during infectious disease outbreaks using constrained P-spline smoothing. *Epidemiology* (Cambridge, Mass.), 30(5), 737.
- Wallinga, J., van Boven, M., & Lipsitch, M. (2010). Optimizing infectious disease interventions during an emerging epidemic. *Proceedings of the National Academy of Sciences*, 107(2), 923-928.
- Wallinga, J., & Lipsitch, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proceedings of the Royal Society B: Biological Sciences*, 274(1609), 599-604.