

To: [redacted] [redacted]@rivm.nl
From: [redacted]
Sent: Thur 9/10/2020 6:24:52 PM
Subject: Antw: Vraag verschillen datasets data.rivm.nl
Received: Thur 9/10/2020 6:24:52 PM

Ja inderdaad. Volgende week verder. Fijn weekend!

Van: [redacted] <[redacted]@rivm.nl>
Datum: 10 september 2020 om 20:13:35 CEST
Aan: [redacted] <[redacted]@rivm.nl>
Onderwerp: Doorst: Vraag verschillen datasets data.rivm.nl

Zie onder.. volgens mij is ze teruggefloten.
Maar goed, laten we het volgende week maar officieel oppakken!

[redacted]

Van: [redacted] [redacted]@rivm.nl
Datum: 10 september 2020 om 16:45:03 CEST
Aan: [redacted] <[redacted]@rivm.nl>
Onderwerp: RE: Vraag verschillen datasets data.rivm.nl

Beste [redacted]

Ik heb nagevraagd over het delen van de data.

Het hele RIVM het gebruiksrecht van de OSIRIS data, dus jullie mogen deze gebruiken.

Uiteraard is de wel data vertrouwelijk.

Het is echter wel een vrij groot verzoek, dus het is het beste als degene die u aanstuurt (afdelingshoofd?) hierover in overleg gaat met [redacted], [redacted] of [redacted]. Er kan dan afgestemd worden hoe deze data het handigst bij jullie kan komen.

Vriendelijke groet,

[redacted]

From: [redacted] [redacted]@rivm.nl
Sent: woensdag 9 september 2020 15:55
To: [redacted] <[redacted]@rivm.nl>
Subject: RE: Vraag verschillen datasets data.rivm.nl

Beste [redacted]

Dank voor de uitleg! Wat betreft welke variabelen verder beschikbaar zijn, voor jullie zullen de volgende variabelen het meest relevant zijn:

Datum:

- MELGGDontvDt, datum waarop de melding door de GGD is gedaan
- NCOVdat1eposncov, datum eerste positieve uitslag
- ZIE1eZiekteDt, datum eerste ziekte dag
- NCOVdat1ezkhopn, datum ziekenhuisopname
- ZIEDtOverlijden, datum overlijden
- MELRIVMONTvDt, datum waarop de melding bij het RIVM is ontvangen
- MELRIVMONTdt_Ziekenhuis, datum waarop de melding van ziekenhuisopname bij het RIVM is ontvangen
- MELRIVMONTdt_Overleden, datum waarop de melding van overlijden bij het RIVM is ontvangen

De MELRIVM/MELGGD variabelen zijn altijd volledig, de andere datavelden ontbreken soms. Daarom is er ook een variabele "date used for statistics" gemaakt, waar de datum eerste ziekte dag in staat indien bekend, anders datum uitslag, en anders datum melding GGD.

Locatie:

- NCOVpostcode, op basis van woonadres
- Gemeente, op basis van woonadres

Verder zullen wij, gezien de mogelijke herleidbaarheid van de data, intern moeten bespreken of en hoe wij deze data kunnen delen. Zodra hierover gesproken is, zal ik dit doorgeven.

Vriendelijke groet,

5.1.2e

From: 5.1.2e <5.1.2e@rivm.nl>
Sent: dinsdag 8 september 2020 16:30
To: 5.1.5; 5.1.2e@rivm.nl
Subject: RE: Vraag verschillen datasets data.rivm.nl

Beste 5.1.2e

Het gebruik van de datum van ziekenhuisopname en datum van positieve test ipv de datums van melding zou ons

model echt nauwkeuriger maken aangezien we een parameter ertussenuit halen die gemodelleerd is (de lags tussen plaatsvinden en melding). Het early warning model dat we aan het ontwikkelen zijn voor het monitoren van het virus in rioolwater zal elke dag gedraaid worden voor de meest actuele stand van zaken. Dit betekent dat we deze variabelen dan inderdaad elke dag nodig.

Wat we in elk geval **per gemeente per datum** nodig hebben zijn: aantal ziekenhuisopnames, aantal positieve testen, aantal doden. De datum is dan dus de echte datum van opname, testresultaat, overlijden en niet van melding. Is het mogelijk om een dump van de totale dataset te zien zodat we zien welke variabelen nog meer beschikbaar zijn? Zo kunnen we nu in 1 keer de goede dataset extraheren zodat we niet over een tijdje dit weer moeten veranderen omdat er andere variabelen zijn die we ook graag zouden willen hebben. Wij hebben ook contact met [redacted 5.1.2e] van C1b en zij kunnen volgens mij ook in de database en hier een **eigen selectie in maken**. Of is het handig om contact op te nemen met diegene die dit heeft ingericht? Is dat ook [redacted 5.1.2e]? Hij heeft de infrastructuur hiervoor ook ingericht is mij verteld.

Ik hoor graag van je!
Groeten,

[redacted 5.1.2e]

From: [redacted 5.1.5] <[redacted 5.1.2e]@rivm.nl>
Sent: vrijdag 4 september 2020 16:34
To: [redacted 5.1.2e] <[redacted 5.1.2e]@rivm.nl>
Subject: RE: Vraag verschillen datasets data.rivm.nl

Beste [redacted 5.1.2e]

Er bestaat een variabele in OSIRIS met datum ziekenhuisopname, dus deze gegevens zijn er wel.

Hebben jullie deze data eenmalig nodig of willen jullie deze standaard geleverd krijgen? En wat precies willen jullie hebben qua data/variabelen?

Het is namelijk niet zo 1-2-3 te regelen dat deze data dan bij jullie komt, maar als jullie deze echt nodig hebben, dan zullen we dat natuurlijk gewoon gaan regelen!

Wat betreft de delay in positieve testen en de melding daarvan, hier verwacht ik nog een antwoord op dus daar kom ik nog op terug!

Ik hoor het graag,

[redacted 5.1.2e]

From: [redacted 5.1.2e] <[redacted 5.1.2e]@rivm.nl>
Sent: donderdag 3 september 2020 11:35
To: [redacted 5.1.5] <[redacted 5.1.2e]@rivm.nl>
Subject: RE: Vraag verschillen datasets data.rivm.nl

Beste 5.1.2e

Hartelijk dank voor je antwoord! Even voor de zekerheid: er is geen dataset beschikbaar waarbij de data zoals de ziekenhuisopnames beschikbaar is PER DAG VAN OPNAME ipv PER DAG VAN MELDING AAN RIVM toch? De informatie over de delay in opname en melding is dan erg waardevol voor ons!

Nog een laatste vraag voor jou. Hebben jullie ook data over de delay in positieve testen en melding daarvan? Dus eigenlijk wat je me hieronder laat zien maar dan voor positieve testen ipv ziekenhuisopnames?

Excuus dat ik nu weer met een vraag kom. Ik begrijp dat al deze vragen achter elkaar voor jullie ook extra werkdruk geven. De informatie die je geeft is wel echt heel waardevol voor ons!

Alvast heel erg bedankt!

Vriendelijke groet,

5.1.2e

From: 5.1.5 5.1.2e <@rivm.nl>
Sent: woensdag 2 september 2020 13:37
To: 5.1.2e <5.1.2e@rivm.nl>
Subject: RE: Vraag verschillen datasets data.rivm.nl

Beste 5.1.2e

Ik heb even geïnformeerd naar de vertraging tussen ziekenhuisopname en het moment dat de opname in de RIVM data gemeld wordt.

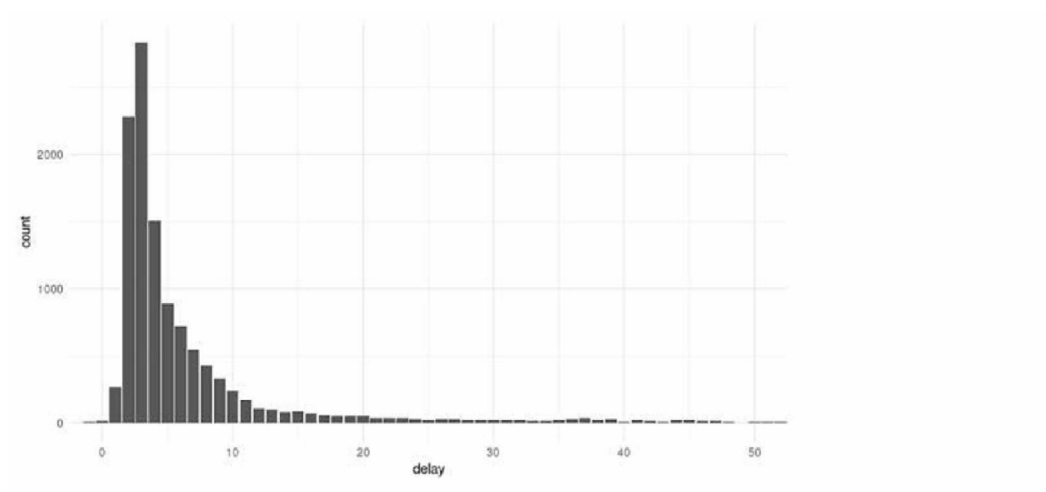
Dit is het volgende:

Hieronder de verdeling voor patiënten opgenomen vanaf 1 maart 2020. De meest voorkomende delay is 3 dagen, maar het gemiddelde is 6.6 dagen (IQR 3 – 7). Die gemiddelde delay neemt wel iets af: voor de data vanaf 1 juni is het 5.5 dagen (IQR 2 – 6).

Hopelijk is dit voldoende informatie zo.

Vriendelijke groet,

5.1.2e



From: [redacted] <[redacted]@rivm.nl>
Sent: dinsdag 1 september 2020 13:21
To: [redacted] <[redacted]@rivm.nl>
Cc: [redacted] <[redacted]@rivm.nl>
Subject: RE: Vraag verschillen datasets data.rivm.nl

Beste [redacted]

Hartelijk dank voor je antwoord! Helemaal duidelijk zo.

Nog een kleine aanvullende vraag naar aanleiding hiervan. De datum in het bestand aantallen_per_dag is de dag waarop de melding is gedaan. Kan dit dan enkele dagen later zijn dan de datum waarop iemand is opgenomen in het ziekenhuis? Bijvoorbeeld omdat door drukte het administratieve gedeelte even is uitgesteld en er niet op dezelfde dag een melding is gemaakt als een patiënt is opgenomen? Zo ja, hebben jullie een idee wat dit aantal dagen ongeveer is, bijv. gemiddeld of maximaal? Ik snap ook dat je hier geen hard getal aan kunt/wilt hangen, maar misschien een indicatie?

Alvast hartelijk dank voor jullie tijd!

Vriendelijke groet,

[redacted]

[redacted]
[redacted]

.....

Centrum Veiligheid | Model & Scenario Onderzoek**Rijksinstituut voor Volksgezondheid en Milieu**
Antonie van Leeuwenhoeklaan 9 | 3721 MA Bilthoven

.....
T 5.1.2e
M 5.1.2e
5.1.2e @rivm.nl
<http://www.rivm.nl>
.....

RIVM *De zorg voor morgen* begint vandaag

From: covid-19 surveillance <5.1.2e @rivm.nl>
Sent: dinsdag 1 september 2020 11:54
To: 5.1.2e <5.1.2e @rivm.nl>
Cc: 5.1.2e <5.1.2e @rivm.nl>
Subject: FW: Vraag verschillen datasets data.rivm.nl

Beste 5.1.2e

Jullie kunnen het beste het bestand COVID-19_aantallen_per_dag gebruiken.

Wat de verschillen in datasets betreft:

Onderstaand is een figuur met aantallen, rood is cumulatief en blauw is per dag.

Deze kleine verschillen komen doordat het bestand aantallen_per_dag met terugwerkende kracht wordt aangepast. Gewiste cases worden hieruit verwijderd, waardoor deze aantallen altijd iets lager zijn dan het cumulatief aantal.

Daarnaast is het zo dat de datum in het bestand aantallen_per_dag op basis is van de dag waarop de melding is gedaan.

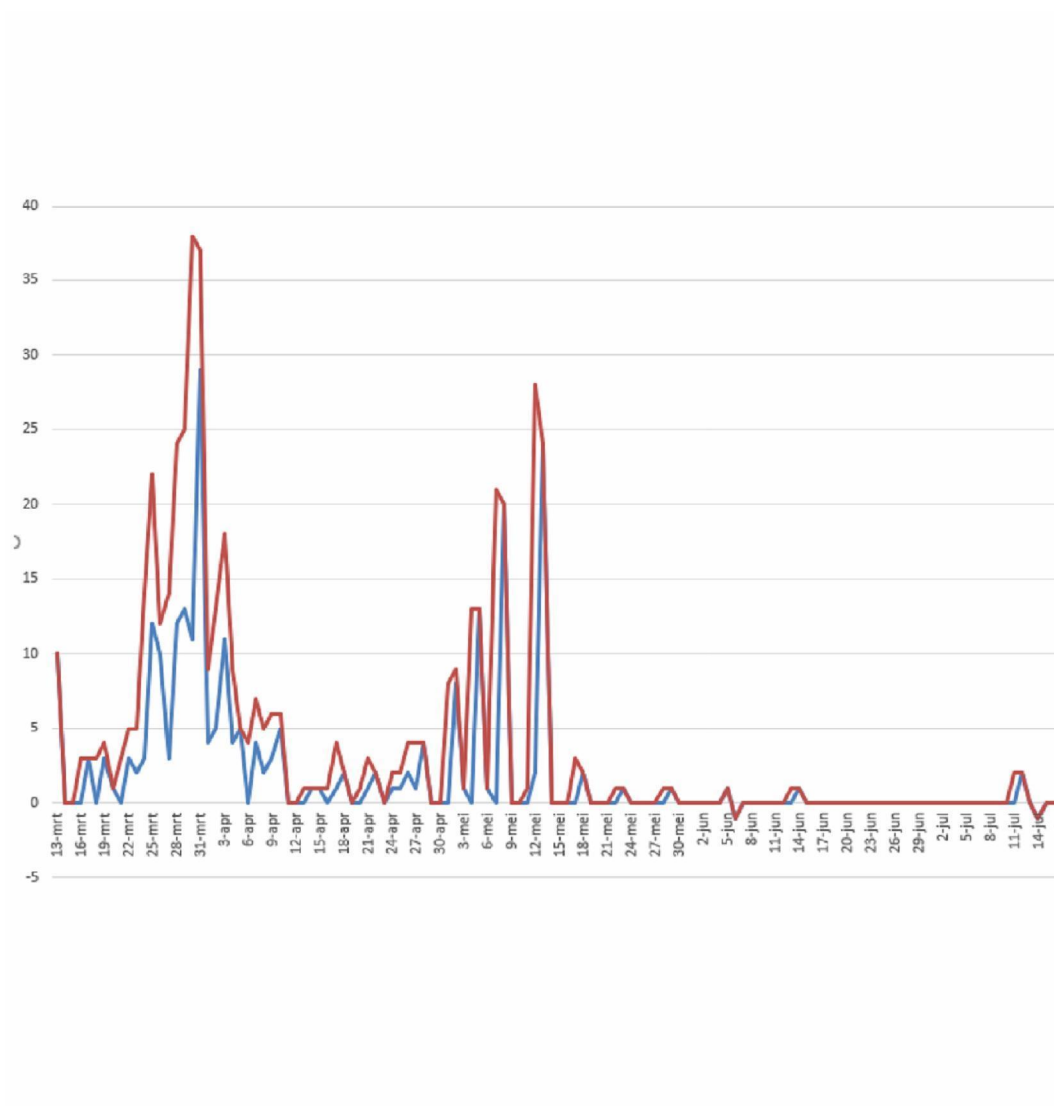
De datum in het cumulatieve bestand (COVID-19_aantallen_gemeente_cumulatief) is de dag waarop wij de rapporteren, dus hoeveel meldingen wij op dat moment binnen hadden.

Ik hoop dat het duidelijk is zo!

Mochten er nog vragen zijn, dan horen we het graag.

Vriendelijke groet,

5.1.2e namens het COVID-19 Surveillance Team



In het project over het analyseren van Sars-CoV-2 in rioolwater maken we gebruik van de data over het aantal besmette mensen dat in het ziekenhuis is opgenomen per gemeente op data.rivm.nl/covid-19. Hiernaast willen we ook de aantallen positief geteste mensen hierop per gemeente gaan gebruiken. We gebruiken hiervoor altijd het databestand COVID-19_aantallen_gemeente_cumulatief.csv en we berekenden uit de cumulatieve

ziekenhuisopnamecijfers zelf de ziekenhuisopnames per dag (door 2 opeenvolgende getallen van elkaar af te halen).

Nu zie ik ook een nieuwe dataset staan: COVID-19_aantallen_per_dag. Deze geeft net als het databestand COVID-19_aantallen_gemeente_cumulatief.csv het aantal positief getesten, het aantal ziekenhuisopnames en het aantal overledenen (plus extra info zoals de bijbehorende veiligheidsregio), maar dan per dag ipv cumulatief. Toen ik de ziekenhuisopnames van de twee datasets vergelijk zag ik wat verschillen. Ik heb ook een excel-file bijgesloten waarbij ik alleen de ziekenhuisopnames per dag van Tilburg laat zien.

In de bijgesloten excelfile is de eerste kolom met ziekenhuisopnames gedestilleerd uit de cumulatieve dataset, de tweede kolom met ziekenhuisopnames is uit de dataset die de gegevens per dag laat zien. Je ziet grofweg een verschuiving van 1 dag, maar dat is verder niet zo interessant. Je ziet ook dat de getallen iets afwijken. Het kan ook met het tijdstip van registreren te maken hebben, maar dat kan niet alle verschillen verklaren. Zou je mij kunnen uitleggen hoe de twee datasets tot stand zijn gekomen en welke dus voor ons het beste is om te gebruiken? Het fijne van de 'dataset per dag' is trouwens wel dat deze nu nooit de waarde -1 heeft, wat bij het gebruik van de cumulatieve dataset wel eens voorkomt door correcties.

Ik hoop dat mijn vraag duidelijk is en dat je mij kunt helpen. Mailen/bellen mag altijd, vrijdags ben ik altijd vrij.

Alvast dank!
Vriendelijke groet,

5.1.2e