

To: 5.1.2e [5.1.2e @rivm.nl]; 5.1.2e [5.1.2e @rivm.nl]; 5.1.2e
Cc: 5.1.2e [5.1.2e @rivm.nl]; 5.1.2e [5.1.2e @rivm.nl]
From: 5.1.2e
Sent: Mon 9/25/2023 1:52:12 PM
Subject: RE: symptoom combinaties en clustering in long covid
Received: Mon 9/25/2023 1:52:12 PM

Hoi 5.1.2e

Interessante vraag waar we niet meteen een antwoord op hebben. Aangezien het neigt naar een data-science achtige vraag (clustering en visualisatie), kun je dit beter vragen aan een van onze I-collega's (de I van data science in SIM). @ 5.1.2e zou hier graag een keer met je over willen sparren (in English).

Groeten,

5.1.2e

5.1.2e 5.1.2e 5.1.2e
 Afdeling Statistiek, Data Science en Modelling - SIM
 Rijksinstituut voor Volksgezondheid en Milieu - RIVM
 Postbus 1 | 3720 BA | Bilthoven | NL
 +31 5.1.2e | 5.1.2e @rivm.nl
 5.1.2e

From: 5.1.2e [5.1.2e @rivm.nl]
Sent: maandag 25 september 2023 14:49
To: 5.1.2e <5.1.2e @rivm.nl>; 5.1.2e <5.1.2e @rivm.nl>
Cc: 5.1.2e <5.1.2e @rivm.nl>
Subject: symptoom combinaties en clustering in long covid

Hoi 5.1.2e

Voor mijn onderzoek naar de prevalentie van symptomen 12 maanden na een covid infectie ben ik bezig om te kijken of er verschillende symptomen (uit een lijst van 13) zich vaker samen voordoen of juist alleen voorkomen en hoe dit over tijd verandert. Hiervoor ben ik al bezig geweest om pairwise correlatie heatmaps te maken, bijvoorbeeld voor mensen 3 maanden na een delta infectie. Daarnaast heb ik ook een aantal dendrogrammen gemaakt via de Ward methode van hiërarchisch clusteren. Echter wordt er bij pairwise correlatie niet rekening gehouden met het gegeven dat meer dan 2 symptomen tegelijk kunnen voorkomen en bij hiërarchisch clusteren dat er niet echt te spreken valt van een hiërarchie in symptomen. Wel is het zo dat sommige symptomen natuurlijk vaker voorkomen dan andere.

Om te kijken wat de mogelijke combinaties van symptomen zijn en hoe vaak ze voorkomen (dus niet relatief maar absolute zin) heb ik nog een extra analyse gedaan. In sunburstR heb ik de mogelijke combinaties van symptomen geteld, waarbij ik dus wel een volgorde moest aanhouden van de symptomen bij het tellen van de patronen. Hiervan heb ik vervolgens een sunburst plot gemaakt (enkel interactief via de bijgevoegde html). Dat is leuk en aardig, maar misschien niet ideaal omdat dus er een bepaalde volgorde inzit die ik heb aangebracht. Je kan bijvoorbeeld de route fatigue > dyspnea > loss of smell afleggen maar niet loss of smell > dyspnea > fatigue.

Nu is het zo dat ik voor mijn huidige paper niet nog extra ruimte heb om die clustering te onderzoeken, maar wil ik me meer focussen op hoe vaak de symptoomcombinaties voorkomen om te kijken of er zo fenotypes te onderscheiden zijn en dan later in een ander paper nog de clustering goed te onderzoeken.

Kunnen jullie een methode aanbevelen waarmee ik de frequenties van de verschillende patronen goed inzichtelijk kan maken (en het liefst dus ook visualiseren)?

En denken jullie dat hiërarchische clustering met Ward methode (voor het volgende paper dus) een goed idee is of kan je beter terugvallen op methodes als principal component analysis of exploratory factor analysis om structuren te ontdekken?

Bedankt alvast!

Groeten,

5.1.2e