**Time Series Clustering of COVID Symptoms - The KCL-ZOE analysis team**

**Explanation**

Subject selection

From the data on over 2.2 million people recorded via the Covid19Tracker app we selected a subsample of participants that tested positive for Covid19 and/or indicated a visit to hospital if they met the following criteria:

- At least 3 inputs of symptoms over a minimum of 4 days between start of symptoms and symptom peak - to ensure time series are meaningful
- An existing initial record of minimum symptom - to ensure that the whole symptom development (from the beginning) is captured
- For the people that tested positive but did not visit hospital, a sign of recovery (last recorded input decreased compared to symptom peak - to ensure that the whole symptom development (up to the peak) is captured

Clustering method

Overall, 479 individuals reporting hospital visit and 294 tested positive without reported hospital visit were selected. Recorded symptom series were automatically aggregated using an unsupervised time series clustering technique into 5 different groups.

The cluster number was chosen to provide the best balance between data fit (distance to the group centre) and model complexity (number of clusters).

For each group, the time series were then averaged across all belonging participants and visualised for comparison.

Visualisation and interpretation

The 5 main clusters presented different patterns as can be seen below with for each of the different clusters, each row representing a symptom. Color intensity represents the proportion of people reporting a given symptom. Note that the duration of symptom course varies from one cluster to another.
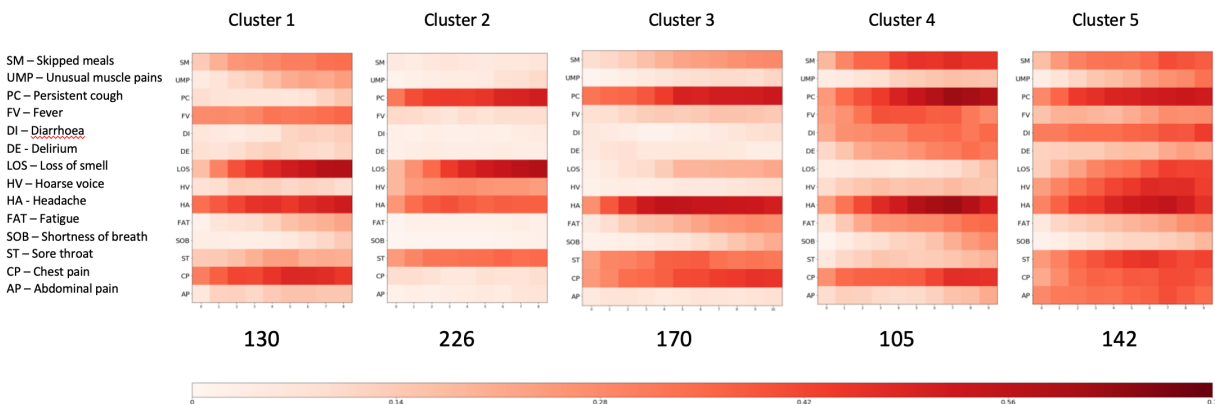


These 5 clusters from left to right can be described as follows:

1. Consistent occurrence of chest pain, headache and fever to a lesser degree with an increased reporting of loss of smell over time.

2.  Headache, sore throat and hoarse voice consistently reported from day 2 onwards, increased occurrence of persistent cough and loss of smell over time.
3.  Increased occurrence of chest pain, sore throat, headache and persistent cough over time
4.  Consistent report of chest pain, oscillating fever and diarrhoea reporting,  increasing headache, meals skipped, fatigue and persistent cough reported over time
5.  Consistent reporting of abdominal, chest and throat pain over time. Increasing report of headache and persistent cough. Larger variety of symptoms consistently reported with increased unusual muscle pains.

Conclusion- our preliminary analysis suggests specific symptom clusters that have different trajectories and outcomes. It also underlines the importance of reporting non-classic symptoms like loss of smell.