

## Supplementary Materials for

### **Symptom clusters in COVID-19: A potential clinical prediction tool from the COVID Symptom Study app**

Carole H. Sudre\*, Karla A. Lee, Mary Ni Lochlainn, Thomas Varsavsky, Benjamin Murray, Mark S. Graham, Cristina Menni, Marc Modat, Ruth C. E. Bowyer, Long H. Nguyen, David A. Drew, Amit D. Joshi, Wenjie Ma, Chuan-Guo Guo, Chun-Han Lo, Sajaysurya Ganesh, Abubakar Buwe, Joan Capdevila Pujol, Julien Lavigne du Cadet, Alessia Visconti, Maxim B. Freidin, Julia S. El-Sayed Moustafa, Mario Falchi, Richard Davies, Maria F. Gomez, Tove Fall, M. Jorge Cardoso, Jonathan Wolf, Paul W. Franks, Andrew T. Chan, Tim D. Spector, Claire J. Steves, Sébastien Ourselin\*

\*Corresponding author. Email: [carole.sudre@kcl.ac.uk](mailto:carole.sudre@kcl.ac.uk) (C.H.S.); [sebastien.ourselin@kcl.ac.uk](mailto:sebastien.ourselin@kcl.ac.uk) (S.O.)

Published 19 March 2021, *Sci. Adv.* 7, eabd4177 (2021)  
DOI: 10.1126/sciadv.abd4177

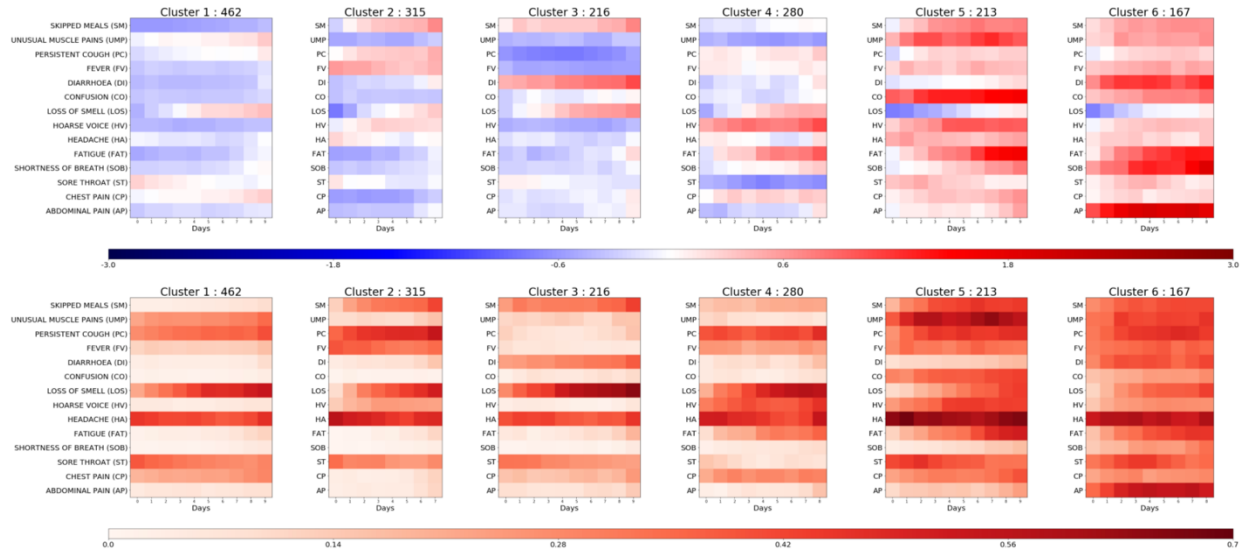
#### **This PDF file includes:**

Figs. S1 to S4  
Tables S1 and S2

**Fig S1 Bayesian information criterion for the choice of number of clusters**



**Fig S2 Top:** associated Z-Score of presentation of symptoms over overall symptom distribution (red = reported more than average; blue = reported less than average) and **Bottom:** Frequency of positive answers per symptoms across days for each cluster (darker = reported more frequently)



**Fig S3 Confusion matrix for cluster assignment on the independent replication set using the projections learnt with limited number of days from the training set**

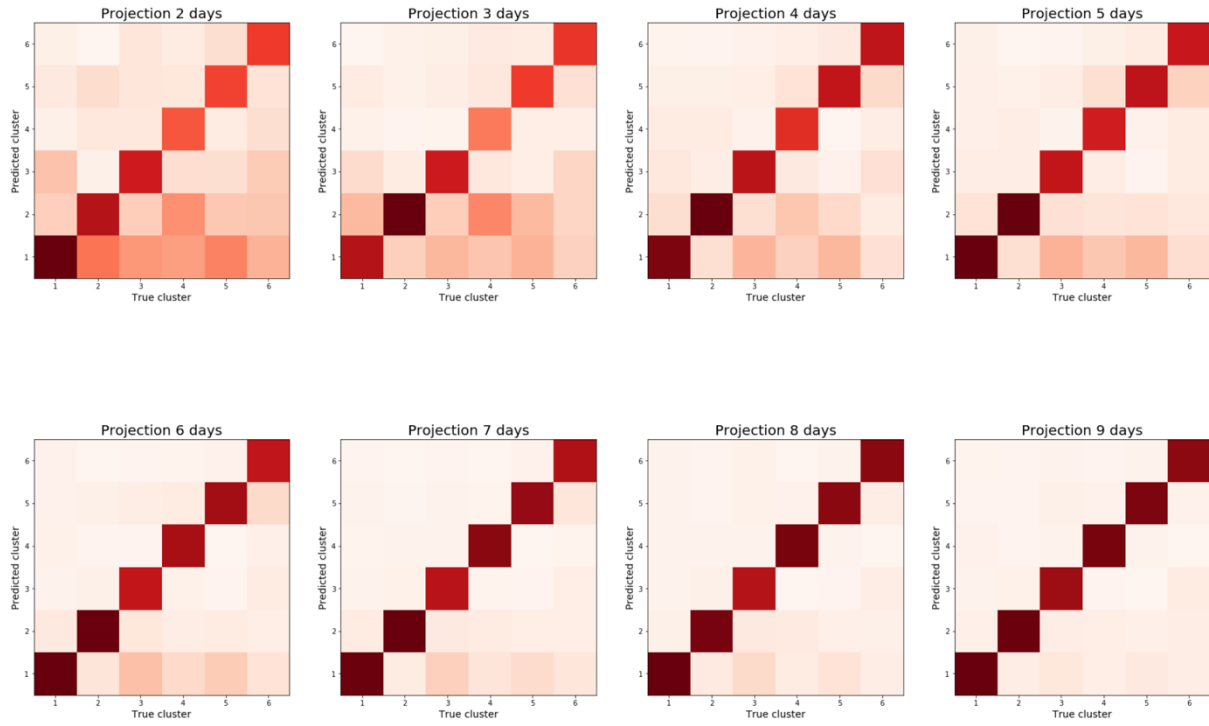


Fig S4 – Co-occurrence matrix of symptoms by cluster. Colors represent proportion of occurrence from white (0) to dark red (1)

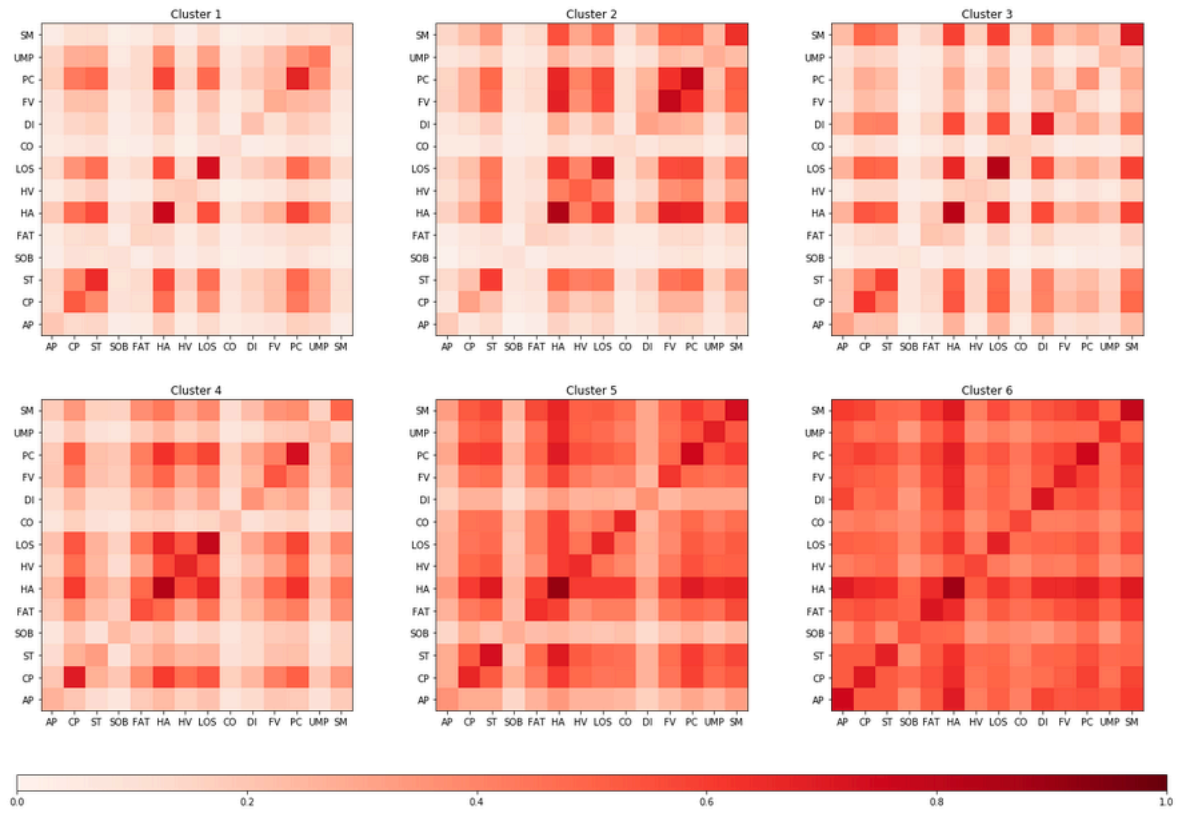


Table S1- Frequency of count until peak and associated missing logs by cluster:

Cluster	Count	Number missing	Ratio missing	% of individuals with more than 3 logs
1	8 [6;11]	2 [1;4]	0.25 [0.08;0.40]	90.3
2	8 [6;11]	2 [1;3]	0.25 [0.11;0.39]	88.3
3	8 [6;10]	2 [0;4]	0.26 [0;0.40]	85.2
4	8 [6;11]	2 [0;3]	0.20 [0;0.39]	90.0
5	8 [6;11]	2 [0;3]	0.21 [0;0.40]	86.4
6	7 [5;10]	1 [0;3]	0.20 [0;0.40]	81.4

Table S2 Frequency of occurrence of symptom according to cluster category. Abbreviations: AP – Abdominal pain; CP – chest pain; ST – sore throat; SOB – Shortness of Breath; FA – fatigue; HA – headache; HV – hoarse voice; LOS – loss of smell; CO – confusion; DI – Diarrhoea; FV- fever; PC – persistent cough; UMP – unusual muscle pains; SM – skipped meals.

Cluster	1	2	3	4	5	6
<b>AP</b>	19.48%	18.41%	31.02%	26.43%	35.68%	74.85%
<b>CP</b>	52.60%	31.43%	61.11%	69.64%	66.67%	70.66%
<b>ST</b>	64.50%	59.68%	59.26%	32.50%	73.71%	67.66%
<b>SOB</b>	10.39%	9.21%	8.80%	23.21%	27.70%	53.29%
<b>FA</b>	15.58%	16.19%	20.37%	55.36%	63.38%	71.26%
<b>HA</b>	76.41%	84.13%	81.48%	83.57%	90.61%	88.62%
<b>HV</b>	18.40%	50.48%	18.98%	67.50%	64.79%	57.49%
<b>LOS</b>	72.73%	71.43%	82.87%	78.93%	66.67%	67.66%
<b>CO</b>	12.12%	13.65%	17.13%	21.07%	66.20%	57.49%
<b>DI</b>	20.78%	30.79%	68.52%	35.36%	36.15%	71.86%
<b>FV</b>	28.57%	78.41%	28.70%	53.57%	61.97%	68.26%
<b>PC</b>	67.32%	78.73%	35.65%	73.21%	75.59%	75.45%
<b>UMP</b>	42.64%	27.94%	23.15%	25.36%	69.01%	62.28%
<b>SM</b>	14.50%	63.81%	70.83%	48.93%	74.18%	78.44%