

Health Data Science in the COVID-19 Era

HDR UK 'ONE INSTITUTE' CONFERENCE 2020

Title of session

- Chairs:
 - Helen Parkinson, HDR UK Cambridge
 - Richard Dobson, HDR UK London
- Panellists:
 - Tim Spector, Kings College London
 - Paul Clarke, Ocado
 - Indra Joshi, NHSX

This session will start at 10:45 BST.
Please use the Q&A function to ask questions to speakers. You are welcome to comment using the chat function, but we cannot guarantee this will be monitored.



PREDICTING COVID WITH 3.5 MILLION CITIZEN SCIENTISTS

COVID SYMPTOM STUDY APP

Tim Spector

@timspector



Endorsed and amplified by



The Scottish Government
Riaghaltas na h-Alba



Llywodraeth Cymru
Welsh Government



British Heart Foundation



twinsUK



GIG Cymru NHS Wales



Royal College of Surgeons of England

ADVANCING SURGICAL CARE



American Cancer Society



Nurses' Health Study



Royal College of Physicians

Tommy's

Together, for every baby

KING'S College LONDON



IMPROVING LIVES. CURING TYPE 1 DIABETES.

DiABETES UK
KNOW DIABETES. FIGHT DIABETES.



Alzheimer's Research UK

Make breakthroughs possible



CANCER RESEARCH UK

VERSUS ARTHRITIS

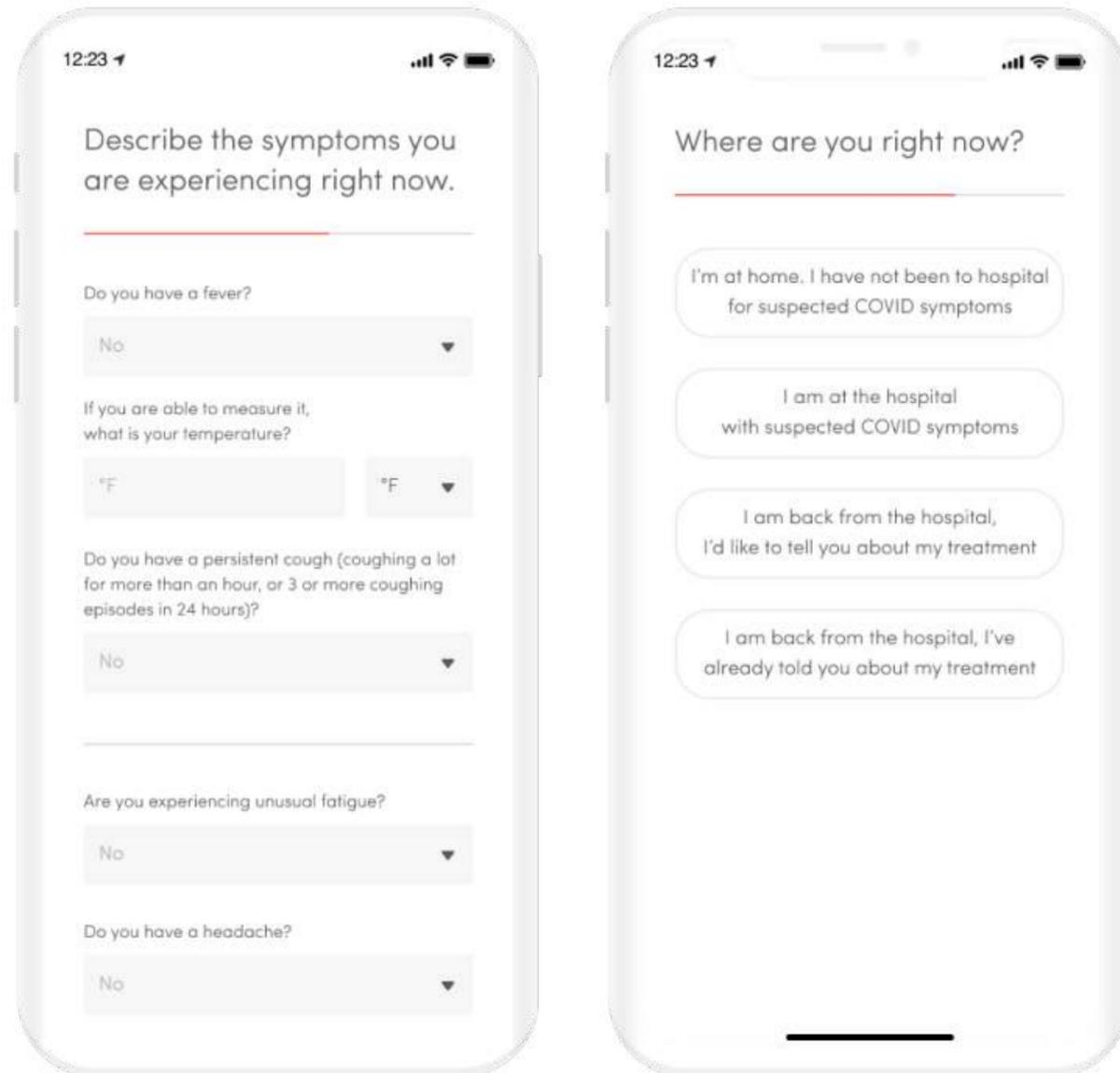


Kidney Care UK

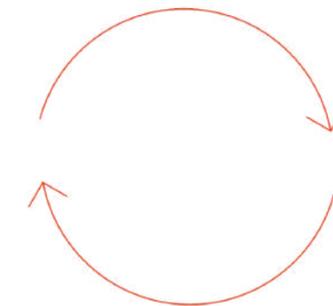


chronic disease research foundation

We can predict "Symptomatic COVID" using Machine Learning



Symptom clusters



Extrapolate to population

Days with symptoms



COVID Symptom Study

Model validation with routine testing from the DHSC

- 6-10,000 tests per week

Infectious users

Identified by the algorithm

Government testing

PCA/Swab testing at testing centres and home kits across the country –all those reporting ANY symptom

Model is tested and improved

When a user reports their results in the app
150,000 tests carried out-
Approx 13,000 positive

New prospective test of diagnostic algorithm planned compared to swab testing

Subset of 600 Twins having antibody tests as part of BRC study



Published scientific papers: 12+ more in pipeline

**nature
medicine**

Brief Communication |
Published: 11 May 2020

Real-time tracking of self-reported symptoms to predict potential COVID-19

Cristina Menni , Ana M. Valdes, Maxim B. Freidin, Carole H. Sudre, Long H. Nguyen, David A. Drew, Sajaysurya Ganesh, Thomas Varsavsky, M. Jorge Cardoso, Julia S. El-Sayed Moustafa, Alessia Visconti, Pirro Hysi, Ruth C. E. Bowyer, Massimo Mangino, Mario Falchi, Jonathan Wolf, Sebastien Ourselin, Andrew T. Chan, Claire J. Steves & Tim D. Spector  - Show fewer authors

Nature Medicine (2020) | Cite this article

Science

REPORT

Rapid implementation of mobile technology for real-time epidemiology of COVID-19

David A. Drew^{1,*}, Long H. Nguyen^{1,*}, Claire J. Steves^{2,3}, Cristina Menni², Maxim Freydin², Thomas Varsavsky⁴, Carole H. Sudre⁴, M. Jorge Cardoso⁴, Sebastien Ourselin⁴, Jonathan Wolf⁵, Tim D. Spector^{2,5,†}, Andrew T. Chan^{1,6,†,‡}, COPE Consortium§

*Corresponding author. Email: achan@mgh.harvard.edu

¹Clinical & Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, 55 Fruit St., Boston, MA 02114, USA.

²Department of Twin Research and Genetic Epidemiology, King's College London, Westminster Bridge Road, London SE17EH, UK.

³Department of Ageing and Health, Guys and St. Thomas's NHS Foundation Trust, Lambeth Palace Road, London SE17EHIL, UK.

⁴School of Biomedical Engineering & Imaging Sciences, King's College London, 1 Lambeth Palace Road, London SE1 7EU, UK.

⁵Zoe Global Limited, 164 Westminster Bridge Road, London SE17RW, UK.

⁶Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, 665 Huntington Avenue, Boston, MA 02114, USA.

* These authors contributed equally to this work.

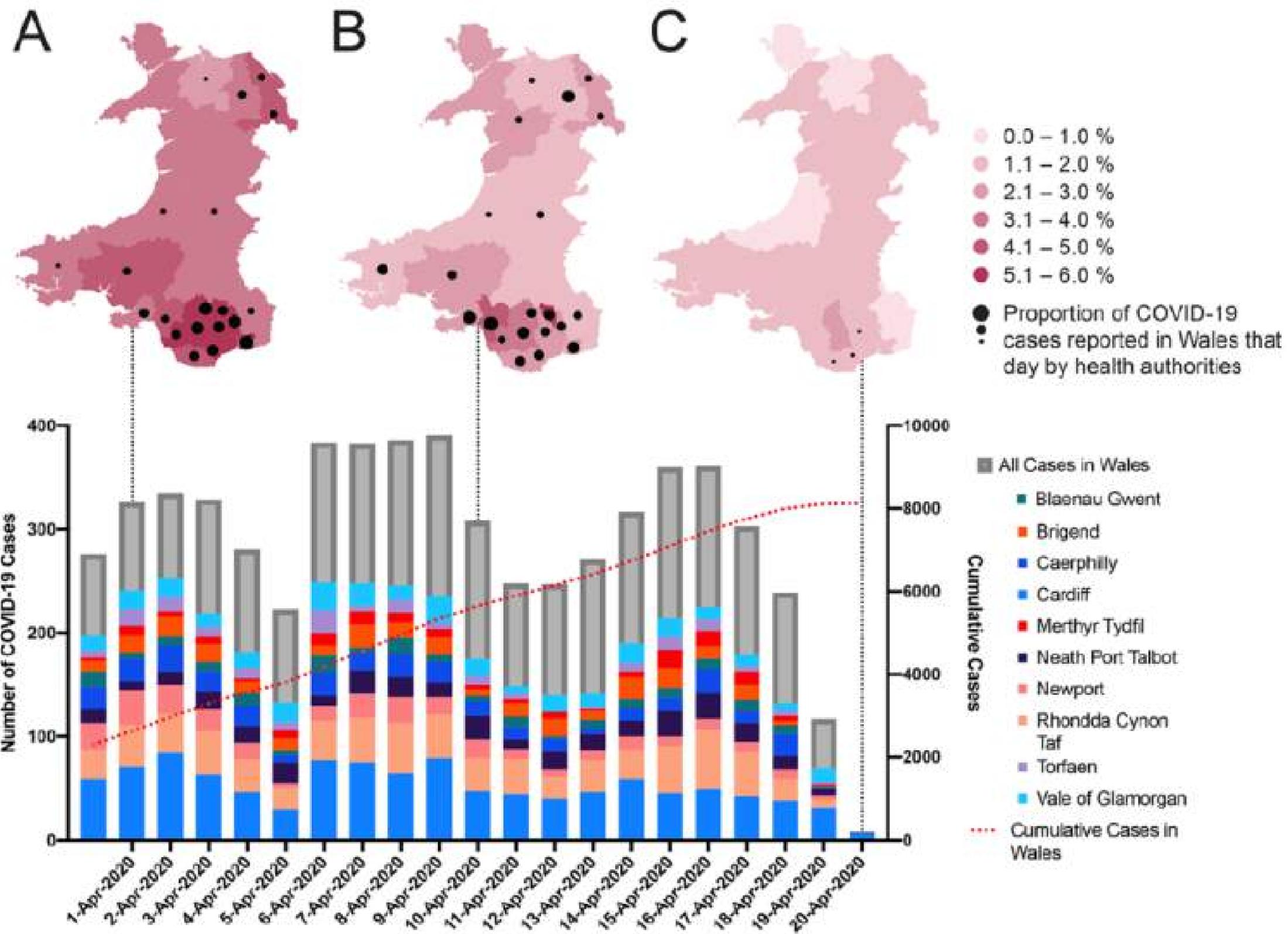
† These authors contributed equally to this work.

§ COPE Consortium members and affiliations are listed in the supplementary materials.

 Hide authors and affiliations

Science 05 May 2020:
eabc0473
DOI: 10.1126/science.abc0473

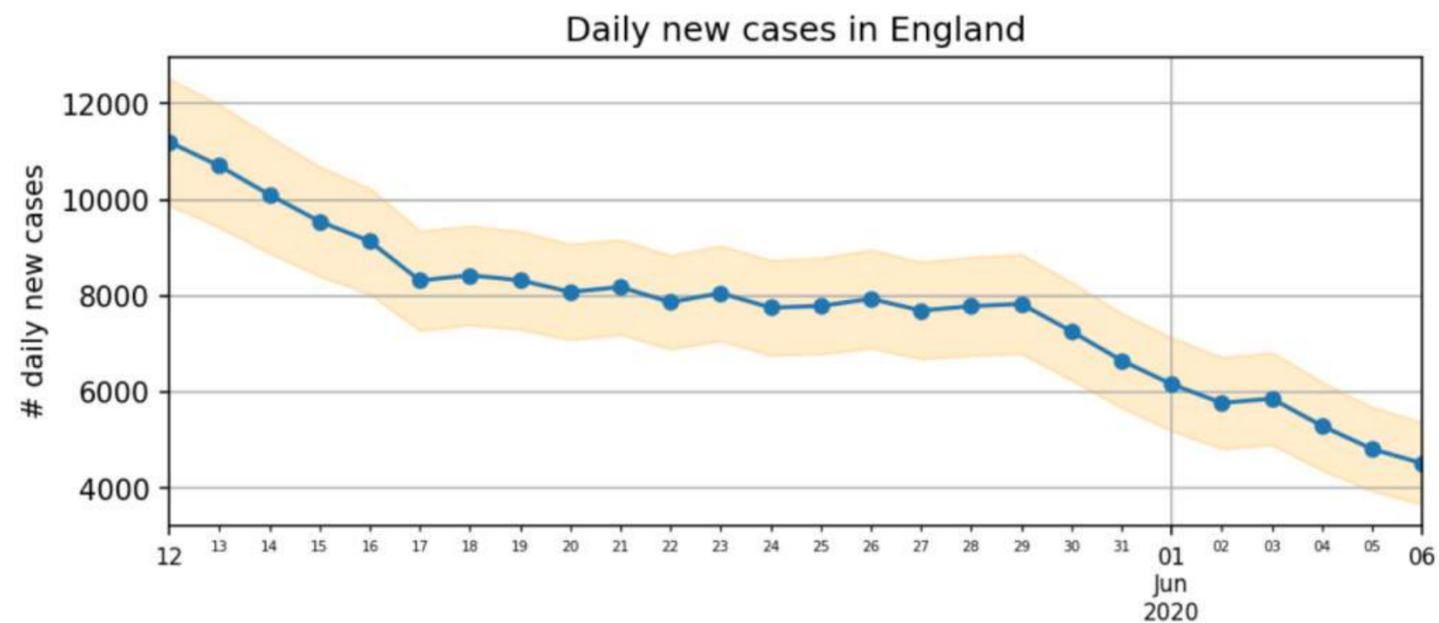
- Deprivation
- Genetics
- Obesity
- Smoking
- Health care workers
- BAME
- Prediction of admissions
- HRT
- Vitamins



Incidence rates England over the last 14 days based on DHSC swab tests

nhser19nm	reg_healthy	newly_sick	invited	tested	tested_positive	% newly sick	% invited	% tested	% tested positive	% +ve 95% lolim	% +ve 95% uplim
East of England	136305.0	11455.0	10411.0	2384.0	32.0	7.752436	90.886076	22.898857	1.342282	0.880345	1.804218
London	185214.0	14432.0	13076.0	2978.0	33.0	7.228795	90.604213	22.774549	1.108126	0.732150	1.484103
Midlands	138105.0	12009.0	10779.0	2388.0	44.0	7.999920	89.757682	22.154189	1.842546	1.303157	2.381935
North East and Yorkshire	115985.0	10301.0	9312.0	2125.0	37.0	8.156882	90.398990	22.820017	1.741176	1.185047	2.297306
North West	94747.0	8282.0	7487.0	1654.0	39.0	8.038513	90.400869	22.091625	2.357920	1.626674	3.089166
South East	233042.0	18626.0	16964.0	4329.0	51.0	7.401020	91.076989	25.518746	1.178101	0.856682	1.499521
South West	132133.0	10904.0	9899.0	2191.0	22.0	7.623202	90.783199	22.133549	1.004108	0.586637	1.421578
England	1035531.0	86009.0	77928.0	18049.0	258.0	7.668830	90.604472	23.161123	1.429442	1.256270	1.602614

Changes in daily new cases of COVID in UK (based on 2 weeks up to each date)



Cite as: D. A. Drew *et al.*, *Science*
10.1126/science.abc0473 (2020).

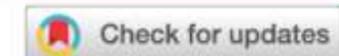
Rapid implementation of mobile technology for real-time epidemiology of COVID-19

David A. Drew^{1*}, Long H. Nguyen^{1*}, Claire J. Steves^{2,3}, Cristina Menni², Maxim Freydin², Thomas Varsavsky⁴, Carole H. Sudre⁴, M. Jorge Cardoso⁴, Sebastien Ourselin⁴, Jonathan Wolf⁵, Tim D. Spector^{2,5†}, Andrew T. Chan^{1,6†‡}, COPE Consortium§

nature
medicine

BRIEF COMMUNICATION

<https://doi.org/10.1038/s41591-020-0916-2>

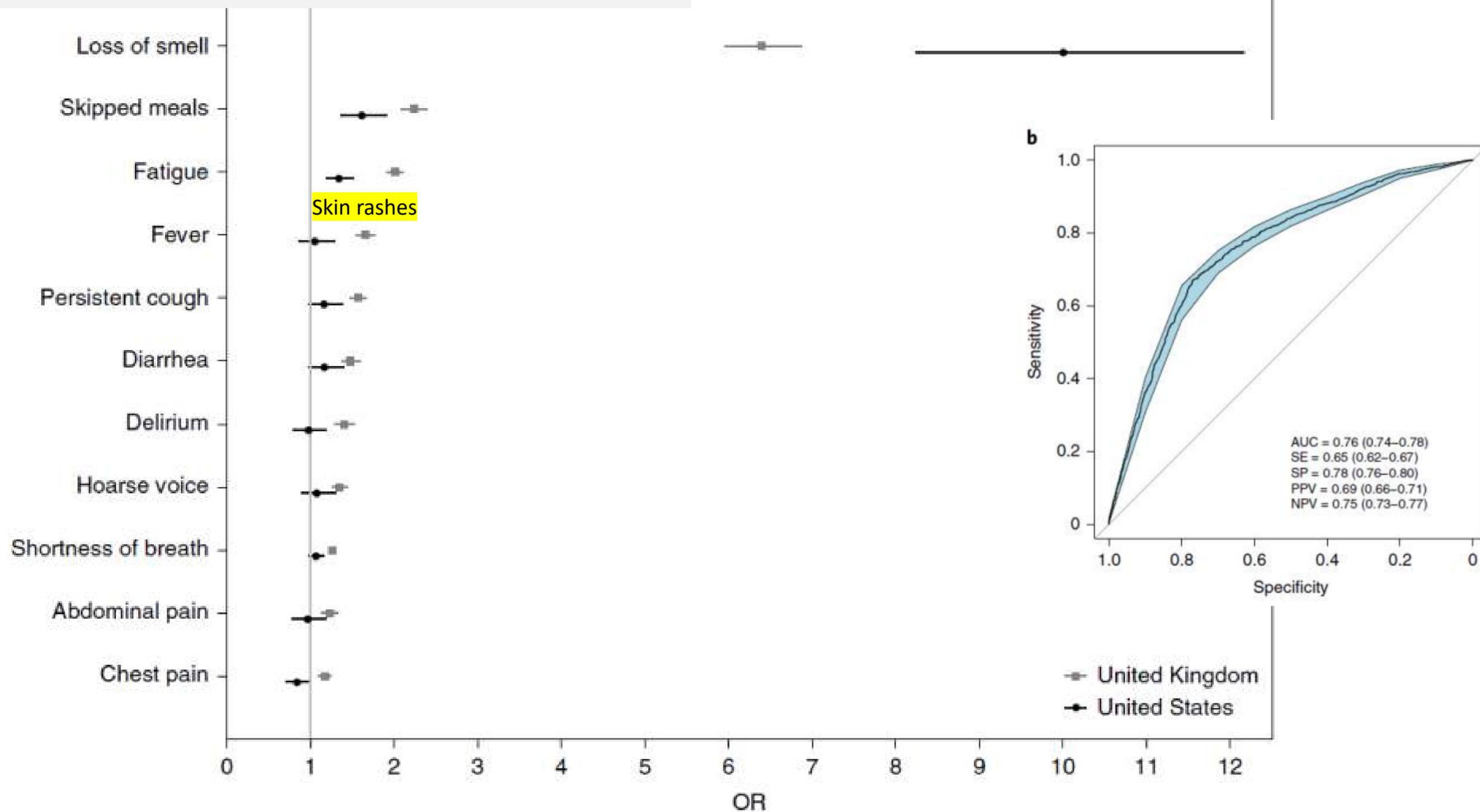


Real-time tracking of self-reported symptoms to predict potential COVID-19

Cristina Menni ^{1,7}✉, Ana M. Valdes ^{1,2,7}, Maxim B. Freidin ¹, Carole H. Sudre³, Long H. Nguyen ⁴, David A. Drew ⁴, Sajaysurya Ganesh ⁵, Thomas Varsavsky³, M. Jorge Cardoso³, Julia S. El-Sayed Moustafa ¹, Alessia Visconti¹, Pirro Hysi¹, Ruth C. E. Bowyer ¹, Massimo Mangino ^{1,6}, Mario Falchi¹, Jonathan Wolf⁵, Sebastien Ourselin³, Andrew T. Chan⁴, Claire J. Steves^{1,8}

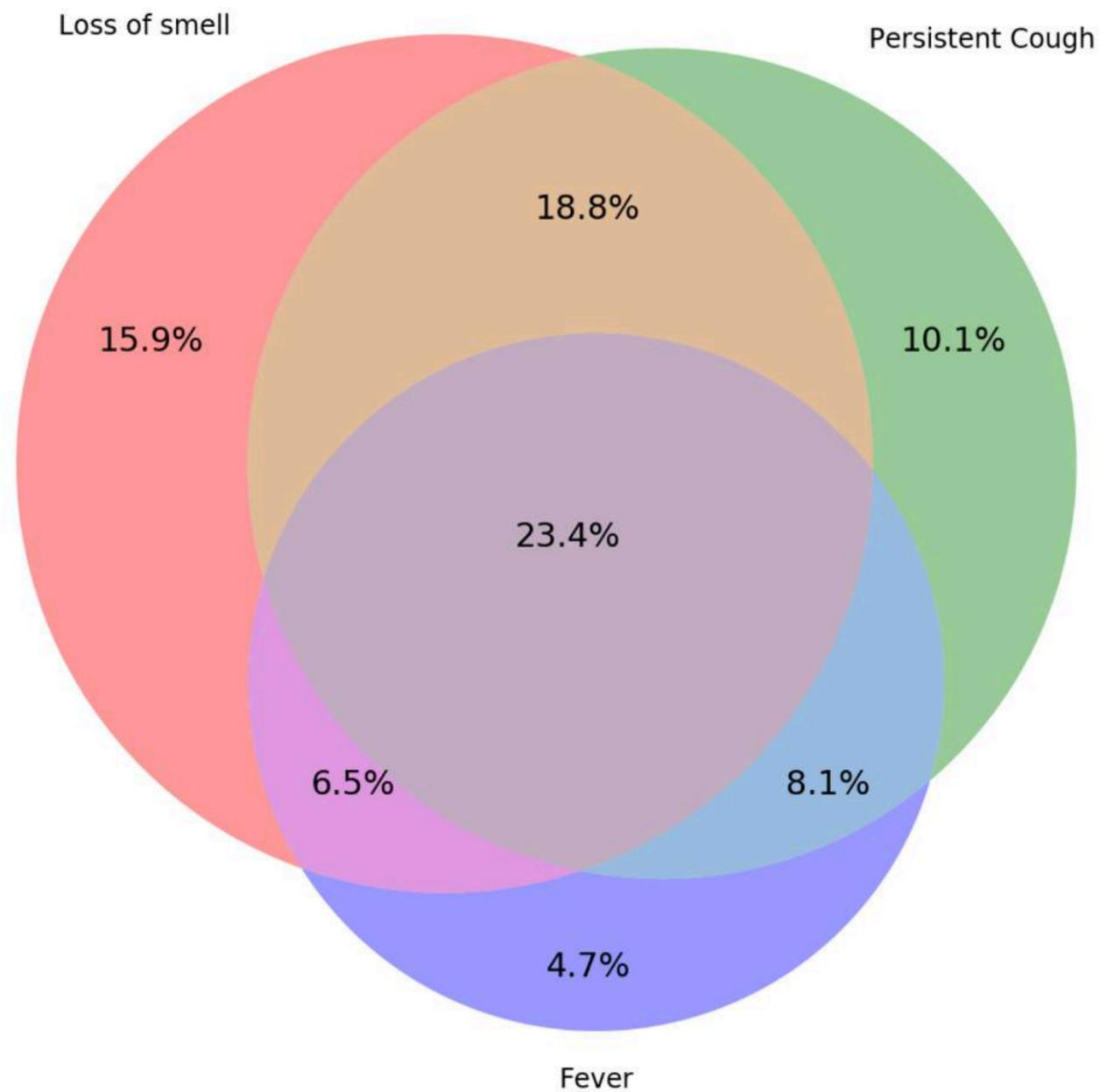
Menni C et al Nat Med May 11 2020

Symptoms predictive of a positive Covid-19 PCR test in 2.816 Million respondents in UK and USA

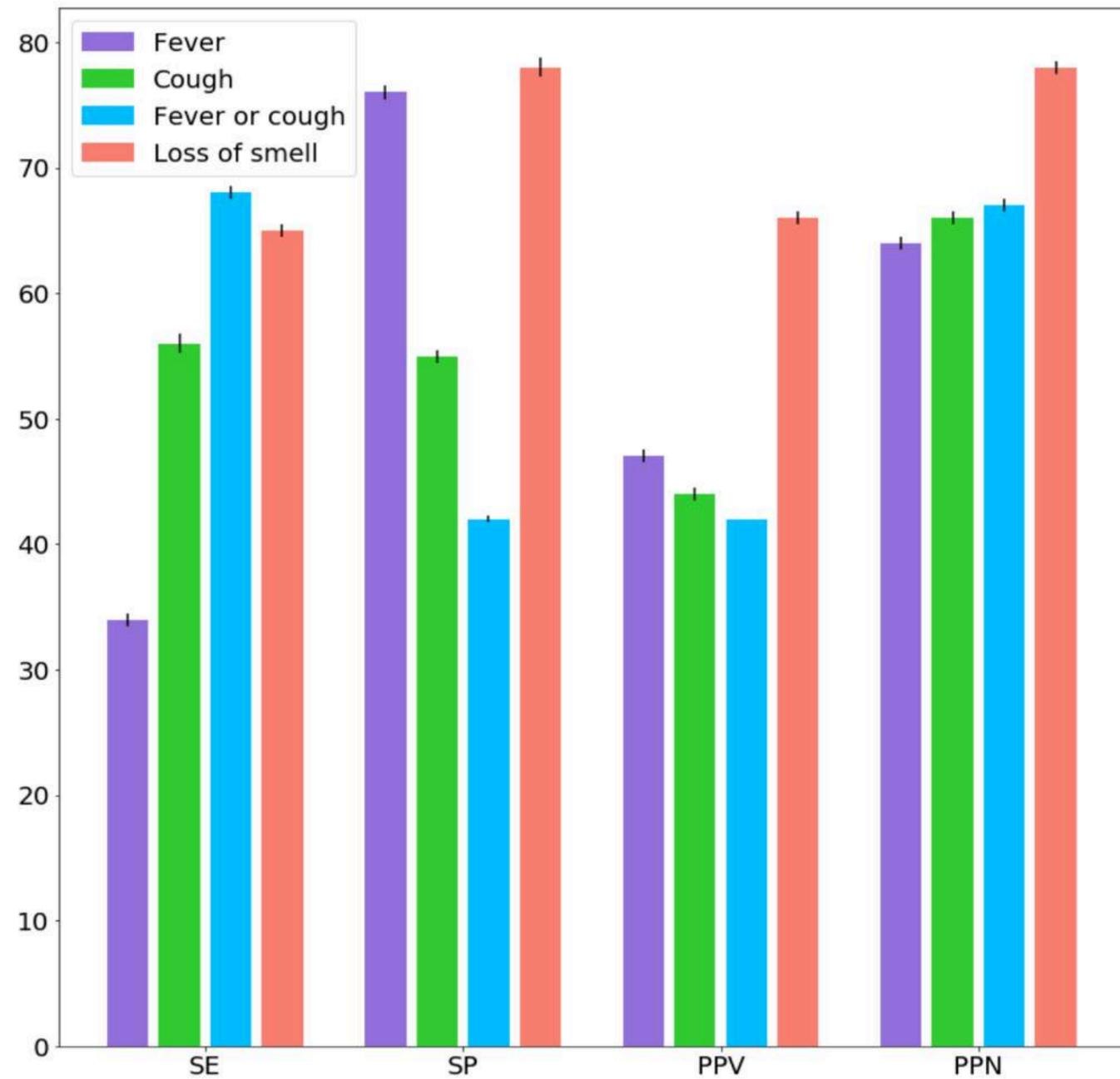


IS sense of smell important ?

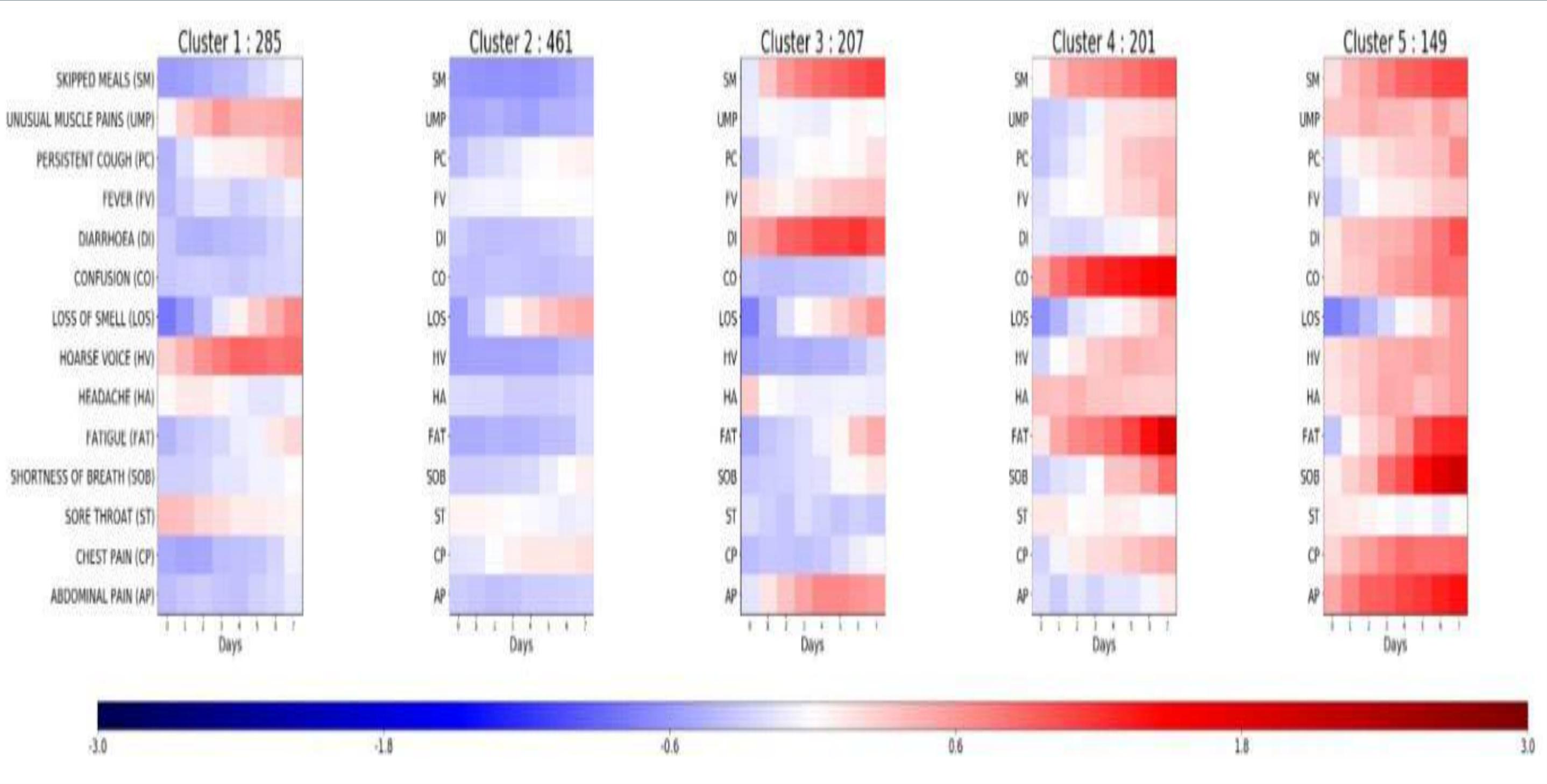
a)



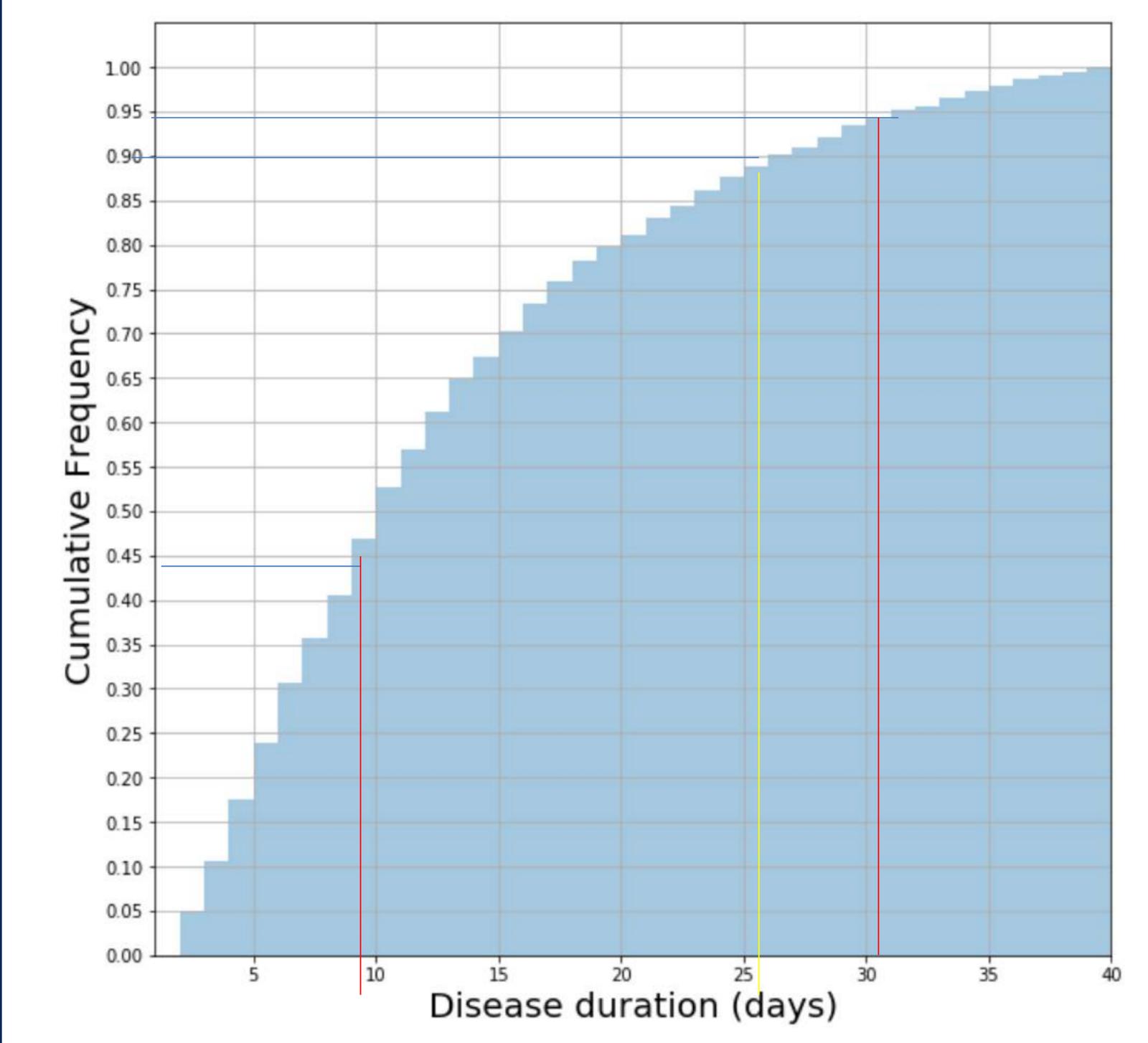
b)



COVID subtypes – 5 clusters of increasing respiratory severity



Symptom duration on 2000 swab positives on whom we had daily data for 40 days



The COVID Symptom app: Predicting COVID-19

- 3.5 million users across the United Kingdom
- Predicts "Symptomatic COVID" via algorithm
- Can predict COVID and hospitalisations
- Can predict cases days earlier than conventional test and trace
- Aiding NHS and the Department of Health
- Perform real time research on risk factors

Not-for-profit initiative by health science company ZOE and King's College London
Helped by HDR-UK to disseminate data



Why does it matter?

With 3.5 million users we can accelerate testing, support contact tracing, expedite a safe lockdown release and will be able to detect second wave quickly by region.



The Crowdsourcing and Sharing of Digital Assets for Societal Good

Paul Clarke

Chief Technology Officer, Ocado







COVID-19
TEST

05:26:48 AM





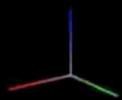
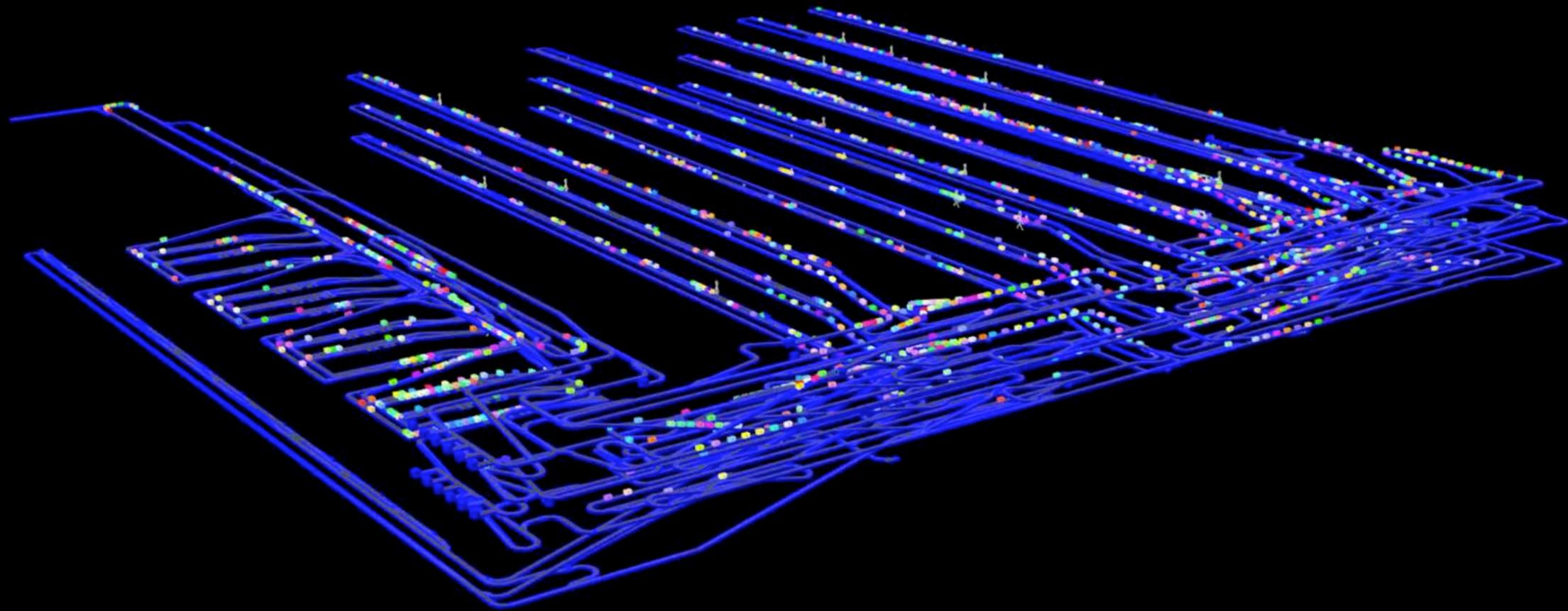
An Comhthéacht Eorpach
European Community

ÉIRE
IRELAND

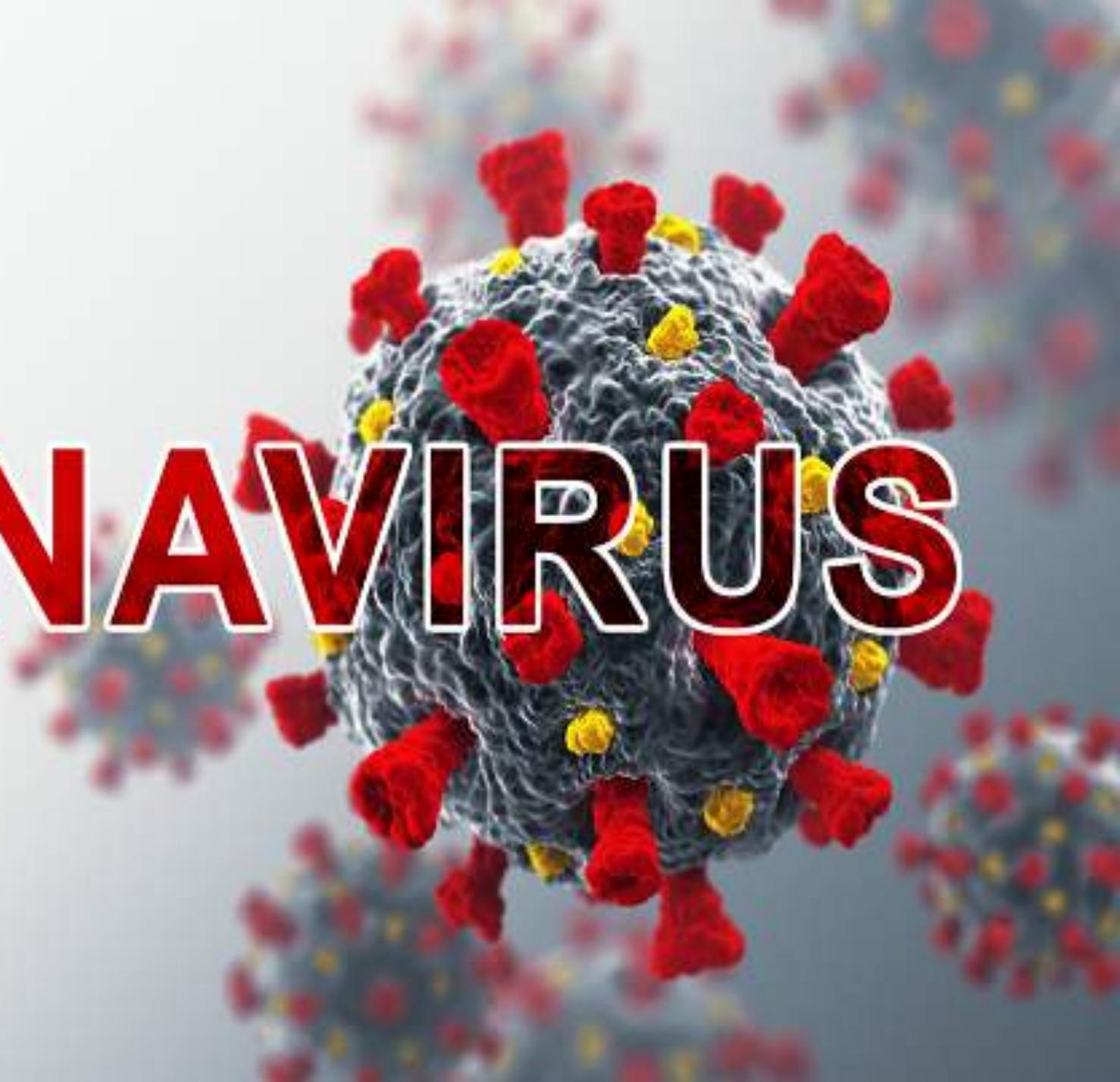


Pas
Passport

United States
of America



CORONAVIRUS



A vibrant teal watercolor splash is centered on a white background. The splash has a soft, irregular edge and a textured, grainy appearance. The word "CATALYST" is written in white, bold, uppercase letters across the middle of the splash.

CATALYST







Thank-you

paul.clarke@ocado.com

@OcadoTechnology



Health Data Science in the COVID-19 Era

HDR UK 'ONE INSTITUTE' CONFERENCE 2020

Dr Indra Joshi, Director of AI, NHSX

HDR UK Collaboration and data with AI Lab

HDRUK
Health Data Research UK

Join the conversation: [#HDRUKcon20](https://twitter.com/HDRUKcon20)



Panel discussion and audience Q&A



Panel Q&A

#	Question	Speaker	Answer
1	Is the app predicting COVID through machine learning models? Are there any sensitivity, specificity details you can let us know?	Tim Spector	Live Answered
2	What is the data privacy on this? Is it linked to an individual or is only linked to the phone?	Tim Spector	
3	if you are changing the questions (interactively) does that not change the model features? how do you manage this when training / using the model?	Tim Spector	By using different date cut-offs - usually when we add new symptoms -such as skin rashes
4	Are any non symptom features (e.g. location, education or wealth) used as features for machine learning?	Tim Spector	We use demographics of each region - deprivation index etc as well as age, gender, postcode, BMI and other diseases
5	What are the predictions for R over the rest of the summer. Could you share this model with myself and Future NHS website.	Tim Spector	We are only doing short term predictions
6	Has this app been used in conjunction with the UK's track and trace program?	Tim Spector	No - we have proposed this to the govt but they haven't responded.
7	Since the early symptoms of COVID are similar to common flu, is it possible to differentiate between the two cases? what symptoms are discriminating?	Tim Spector	The Menni et al Nature medicine paper has the graphs that i showed - anosmia, fatigue, skin rashes, anorexia are more discriminating than others.
8	1) What machine learning method are you using? 2) Have you permissions to link this app data with routine health outcome datasets?	Tim Spector	1. logistic regression up to now but getting more complex - we have over 70 million logs now. 2. HDRUK/Swansea are facilitating linkage with many cohorts- but we don't have it upfront
9	How do you make sure that the symptoms are specific to COVID (e.g. not overlapping with seasonal allergies/hay fever)? Do you expect the classifier to perform worse in other seasons?	Tim Spector	
10	How did you decide on the set of symptoms that you tracked (eg what made you look at smell?) and what should we learn from this for other symptom trackers for other diseases?	Tim Spector	
11	Does the % of population using the app vary throughout country, and does this impact results? And what about any age / diversity measures?	Tim Spector	Live answered
12	Has Northern Ireland government been approached to endorse the app?	Tim Spector	

Panel Q&A

#	Question	Speaker	Answer
13	Are you able to tell whether those with longer-lasting disease are in higher vulnerability groups?	Tim Spector	
14	In the first talk, is the data predicting COVID available publicly ?	Time Spector	https://web.www.healthdatagateway.org/dataset/5dc5998d-efa4-490a-9a46-feb6eb27dabe
15	What software does Ocado use - are you using Tableau or Power BI?	Paul Clarke	We build our simulation, emulation, visualisation and digital twin frameworks from the ground up. When we embarked on this journey 14 years ago, no publicly available tools could have met our requirements. That is still the case
16	Is the ZOE app CE marked as a medical device?	Tim Spector	
17	Some of the speakers would be very valuable for NHS Huddles we are doing with the Future NHS website	All	
18	Absolutely agree, really interesting. We really need to find ways to bring this insight and know-how to healthcare if we are going to make a difference. How can we work together?	Paul Clarke	I have been speaking to NHSx and some clinicians about digital twins in healthcare. Those discussions got paused by the current crisis but keen to pick these up
19	Will digital twins scale to the complexity of public systems where humans, in all their individuality (and therefore spanners-in-the-works), are key players?	Paul Clarke	I/ we believe so but we will need models at differing levels of abstraction and fidelity. Rutherford said "all science in physics or stamp collecting". He was right but it's also an unhelpful comment because we cannot model social science at a molecular level. Similarly, we probably won't build a model of the NHS based on the movement of individual patients within every trust/ hospital. We will need to abstract the lower level models. This is a research topic that EPSRC are focussed on as are we
20	On the learnings from behavioural side and incentives for adherence to use of the app: - Do the healthy individuals keep using the app or lose interest after sometimes? - If symptomatic but have no confirmation/ access to test, would they keep reporting? do you give any support?	Tim Spector	

Panel Q&A

#	Question	Speaker	Answer
21	it would be good to connect with NHS X . I work for NHS England and NHS improvement. What are peoples thoughts around the Track and Trace that is currently going on?	All	
22	I like the ambition but how much storage is required to create digital twins of all our major institutions and assets? Does sufficient storage exists in the UK?	Paul Clarke	I have just answered a similar question. We will need models at multiple levels of fidelity and abstraction. These will be built to meet the needs of other specific use cases. So it won't be some massive centralised IT project. Compute power probably more significant than storage IMHO :-)
23	Loss of taste was also heavily reported in the media as predictive of a covid-19 case. Where did loss of taste factor in your analysis of symptoms predictive of a covid 19 positive test? Was it included as part of the loss of smell symptom? Thanks! (Apologies if answered in presentation)	Tim Spector	
24	Are there any plans to develop more local information than regions?	Tim Spector	
25	Is the data predicting COVID available publicly?	Tim Spector	https://web.www.healthdatagateway.org/dataset/5dc5998d-efa4-490a-9a46-feb6eb27dabe
26	How often symptom data will be collected in Zoe app??	Tim Spector	
27	How do you explain the % of answers when you send the invitation for a test, which is so important?	Tim Spector	
28	Your app has a very large user base. I wonder what strategy you used to advertise your app? and how that affected the growth of your users?		
29	Why do you not think about people who want to hide their sickness?		
30	There's an interetsing comparison with organisations like YouGov where there are large panels of people who agree to do surveys and contribute data		

Panel Q&A

#	Question	Speaker	Answer
31	<ol style="list-style-type: none">1. For an app to work in for effective contact tracing, it will need to at least reach 60-70% coverage of the population, can be hard for a voluntary app, and wonder what may be the view of the panellists.2. For effective contact tracing, real time linkage of EHR will be helpful, the question is if it may be possible for the UK current EHR (data from NHS digital and other government administrative system in the ONS) to be linked in such manner?	Tim Spector	
32	What are the current technical real time challenges in Ocado(digital twin) creation ??	Paul Clarke	Not sure there any specific challenges. We built our own simulation framework so that we could combine the simulation with the real time control system that it models. This enables us to do true emulation where we test the production software on the emulation framework. Combining codebase for the digital and physical twins avoid having to maintain them in sync with one another
33	Thoughts on data governance please?	All	