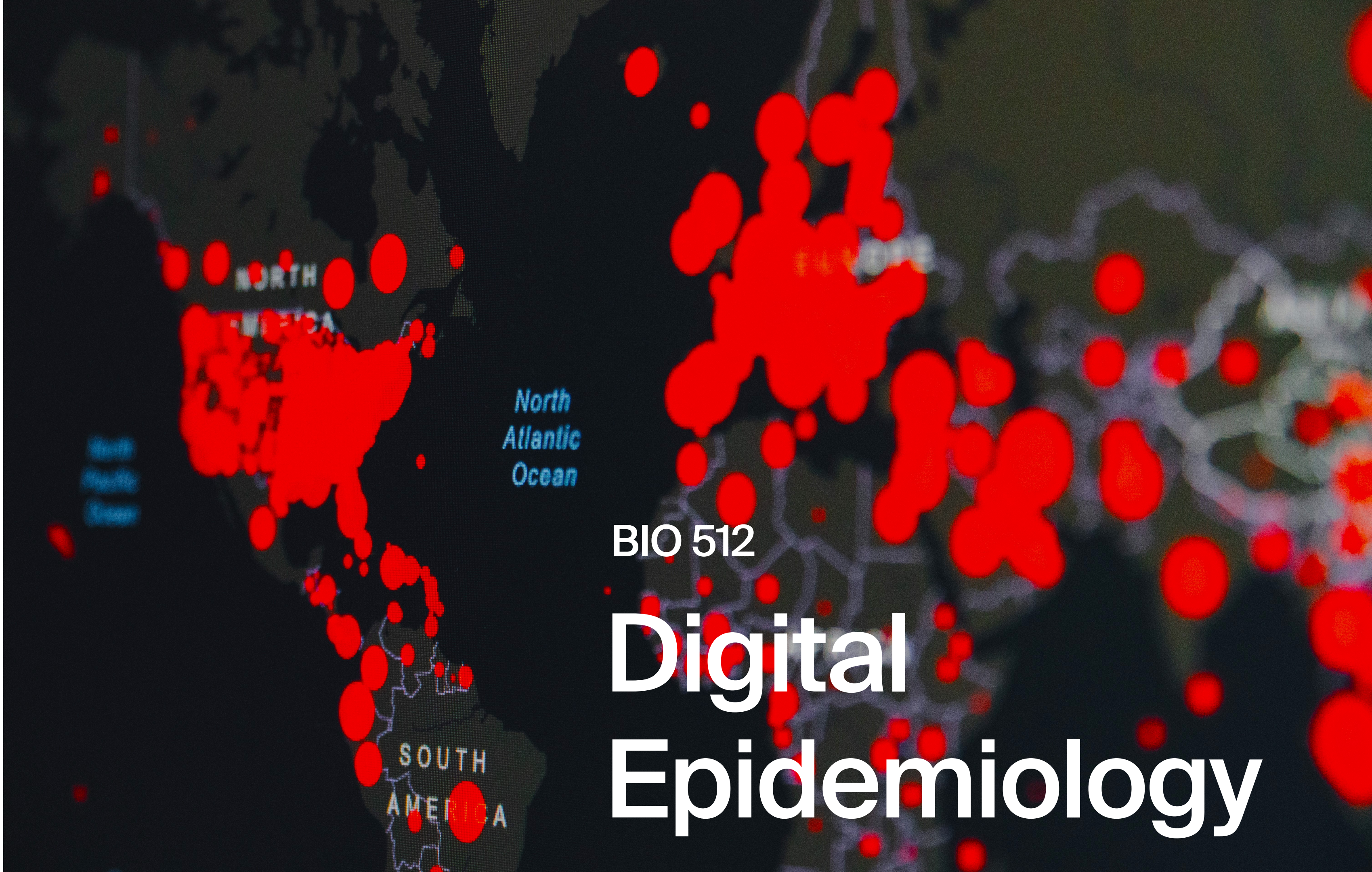
Digital Epidemiology Lab

BIO-521

Spring semester 2025



## BIO 512

# Digital Epidemiology

# Digital Public Health Surveillance

## Learning Objectives

- Learn different types of digital data sources / methods
- Read some of the original work
- Understand the challenges of digital public health surveillance
- Discuss the future of digital public health surveillance

# Epidemiology

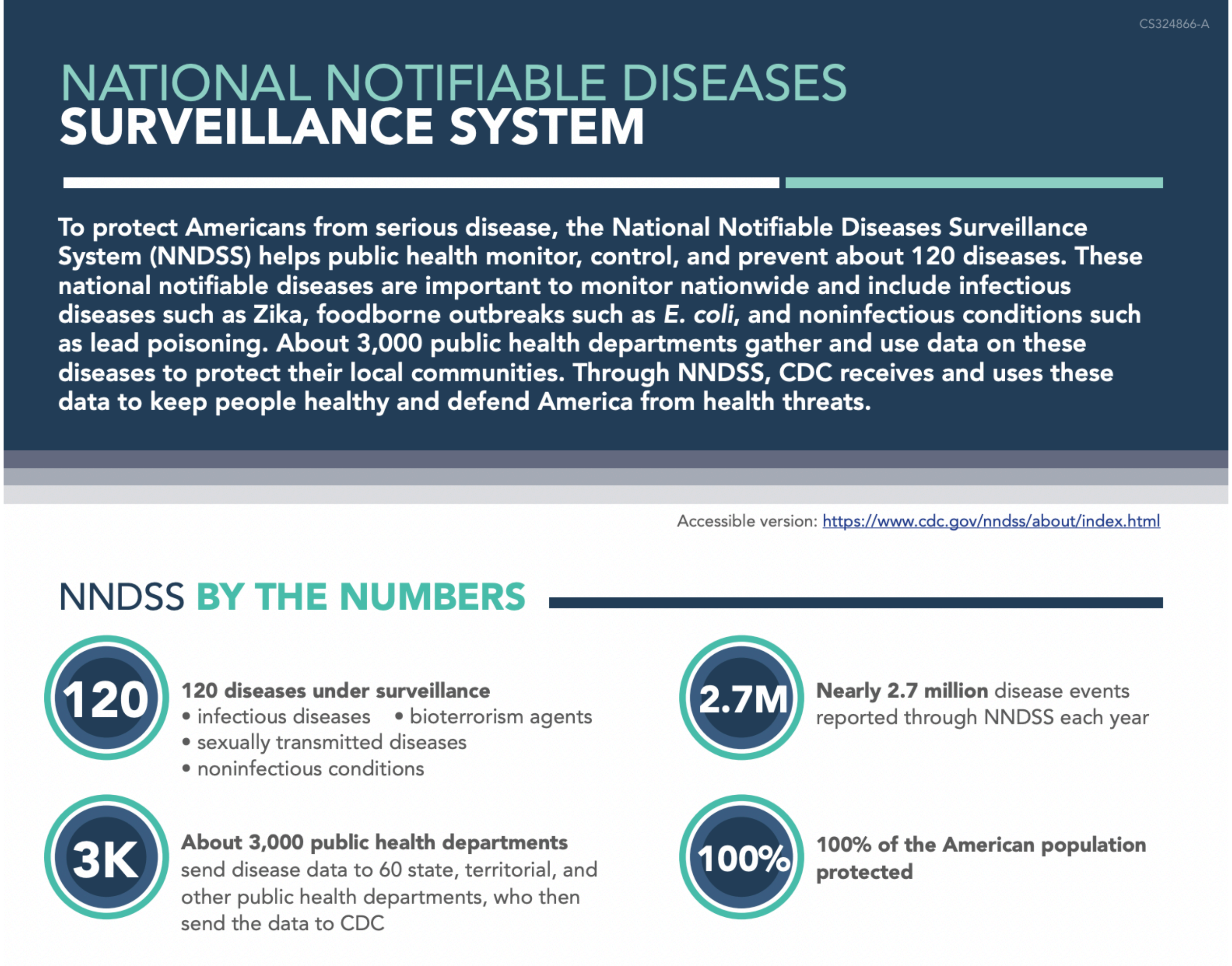
## Surveillance

- Public health surveillance - many different definitions of what counts as surveillance.
- WHO: “the continuous, systematic collection, analysis and interpretation of health-related data”

# Epidemiology

## Surveillance

- Passive vs active surveillance



CS324866-A

# NATIONAL NOTIFIABLE DISEASES SURVEILLANCE SYSTEM

To protect Americans from serious disease, the National Notifiable Diseases Surveillance System (NNDSS) helps public health monitor, control, and prevent about 120 diseases. These national notifiable diseases are important to monitor nationwide and include infectious diseases such as Zika, foodborne outbreaks such as *E. coli*, and noninfectious conditions such as lead poisoning. About 3,000 public health departments gather and use data on these diseases to protect their local communities. Through NNDSS, CDC receives and uses these data to keep people healthy and defend America from health threats.

Accessible version: https://www.cdc.gov/nndss/about/index.html

## NNDSS **BY THE NUMBERS**

**120**

**120 diseases under surveillance**
- infectious diseases
- bioterrorism agents
- sexually transmitted diseases
- noninfectious conditions

**2.7M**

**Nearly 2.7 million** disease events reported through NNDSS each year

**3K**

**About 3,000 public health departments** send disease data to 60 state, territorial, and other public health departments, who then send the data to CDC

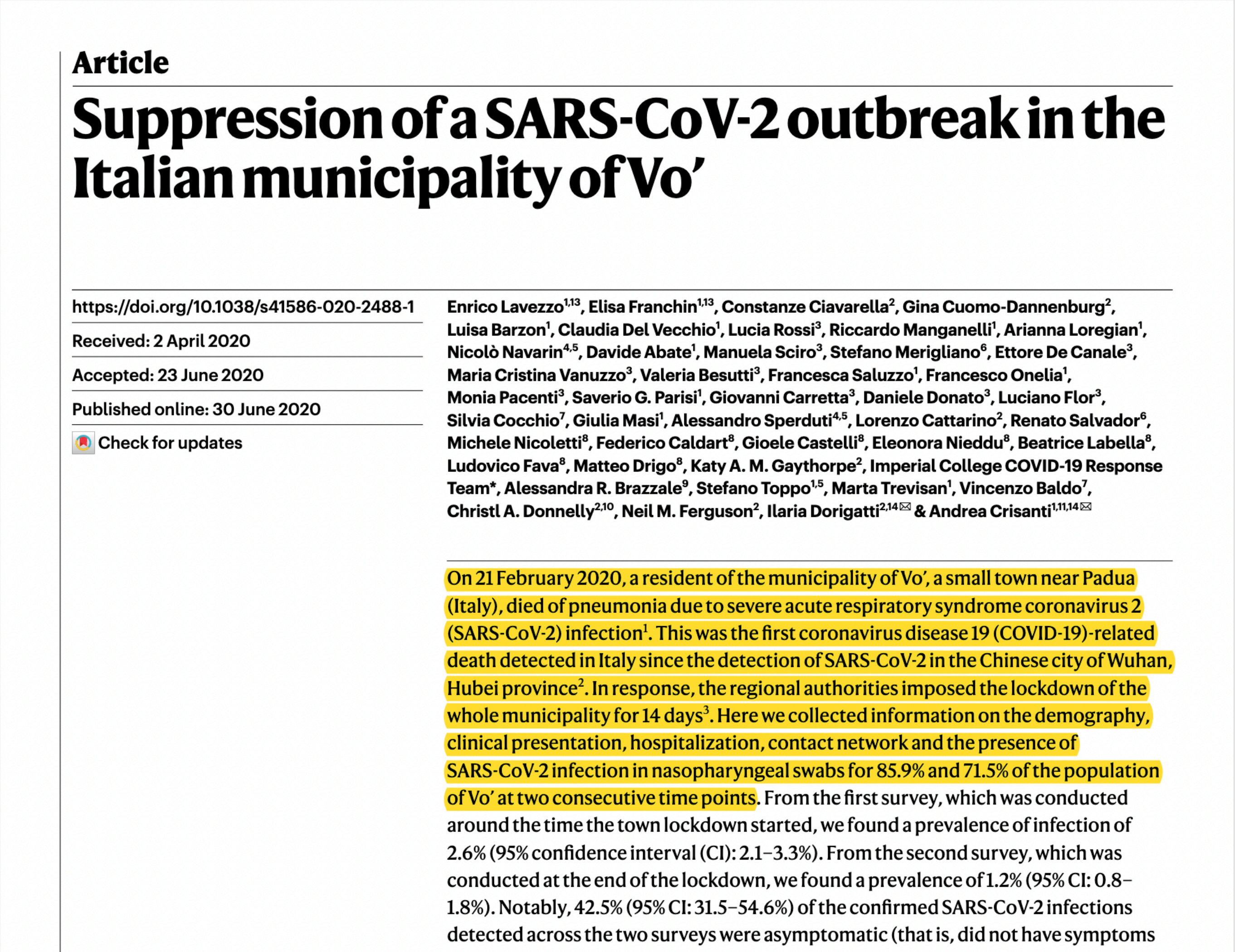
**100%**

**100% of the American population protected**

# Epidemiology

## Surveillance

- Passive vs active surveillance



Article

# Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'

https://doi.org/10.1038/s41586-020-2488-1
Received: 2 April 2020
Accepted: 23 June 2020

Enrico Lavezzo[1,13], Elisa Franchin[1,13], Constanze Ciavarella[2], Gina Cuomo-Dannenburg[2], Luisa Barzon[1], Claudia Del Vecchio[1], Lucia Rossi[3], Riccardo Manganelli[1], Arianna Loregian[1], Nicolò Navarin[4,5], Davide Abate[1], Manuela Sciro[3], Stefano Merigliano[6], Ettore De Canale[3], Maria Cristina Vanuzzo[3], Valeria Besutti[3], Francesca Saluzzo[1], Francesco Onelia[1], Monia Pacenti[3], Saverio G. Parisi[1], Giovanni Carretta[3], Daniele Donato[3], Luciano Flor[3], Silvia Cocchio[7], Giulia Masi[1], Alessandro Sperduti[4,5], Lorenzo Cattarino[2], Renato Salvador[6], Michele Nicoletti[8], Federico Caldart[8], Gioele Castelli[8], Eleonora Nieddu[8], Beatrice Labella[8], Ludovico Fava[8], Matteo Drigo[8], Katy A. M. Gaythorpe[2], Imperial College COVID-19 Response Team*, Alessandra R. Brazzale[9], Stefano Toppo[1,5], Marta Trevisan[1], Vincenzo Baldo[7], Christl A. Donnelly[2,10], Neil M. Ferguson[2], Ilaria Dorigatti[2,14] ✉ & Andrea Crisanti[1,11,14] ✉

On 21 February 2020, a resident of the municipality of Vo', a small town near Padua (Italy), died of pneumonia due to severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infection[1]. This was the first coronavirus disease 19 (COVID-19)-related death detected in Italy since the detection of SARS-CoV-2 in the Chinese city of Wuhan, Hubei province[2]. In response, the regional authorities imposed the lockdown of the whole municipality for 14 days[3]. Here we collected information on the demography, clinical presentation, hospitalization, contact network and the presence of SARS-CoV-2 infection in nasopharyngeal swabs for 85.9% and 71.5% of the population of Vo' at two consecutive time points. From the first survey, which was conducted around the time the town lockdown started, we found a prevalence of infection of 2.6% (95% confidence interval (CI): 2.1–3.3%). From the second survey, which was conducted at the end of the lockdown, we found a prevalence of 1.2% (95% CI: 0.8–1.8%). Notably, 42.5% (95% CI: 31.5–54.6%) of the confirmed SARS-CoV-2 infections detected across the two surveys were asymptomatic (that is, did not have symptoms

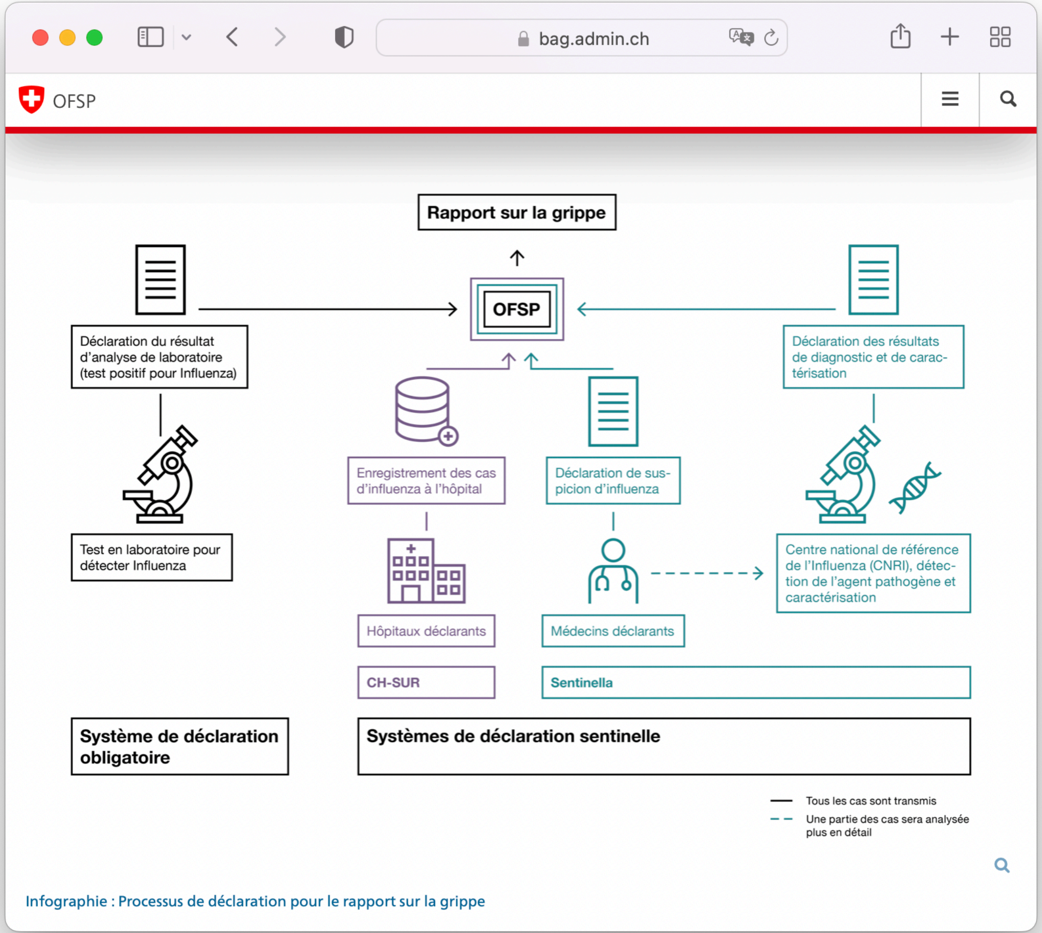
# Epidemiology

## Surveillance

- Population-based vs sentinel surveillance

# Epidemiology
## Surveillance

- Influenza:
  1-5% of physicians voluntarily reporting



Infographie : Processus de déclaration pour le rapport sur la grippe

# Epidemiology

## Surveillance

- Case-based vs aggregate surveillance

often a question of resources

# Epidemiology

## Surveillance

- Syndromic vs laboratory surveillance

Digital public health surveillance always syndromic, but may change soon.

# Epidemiology

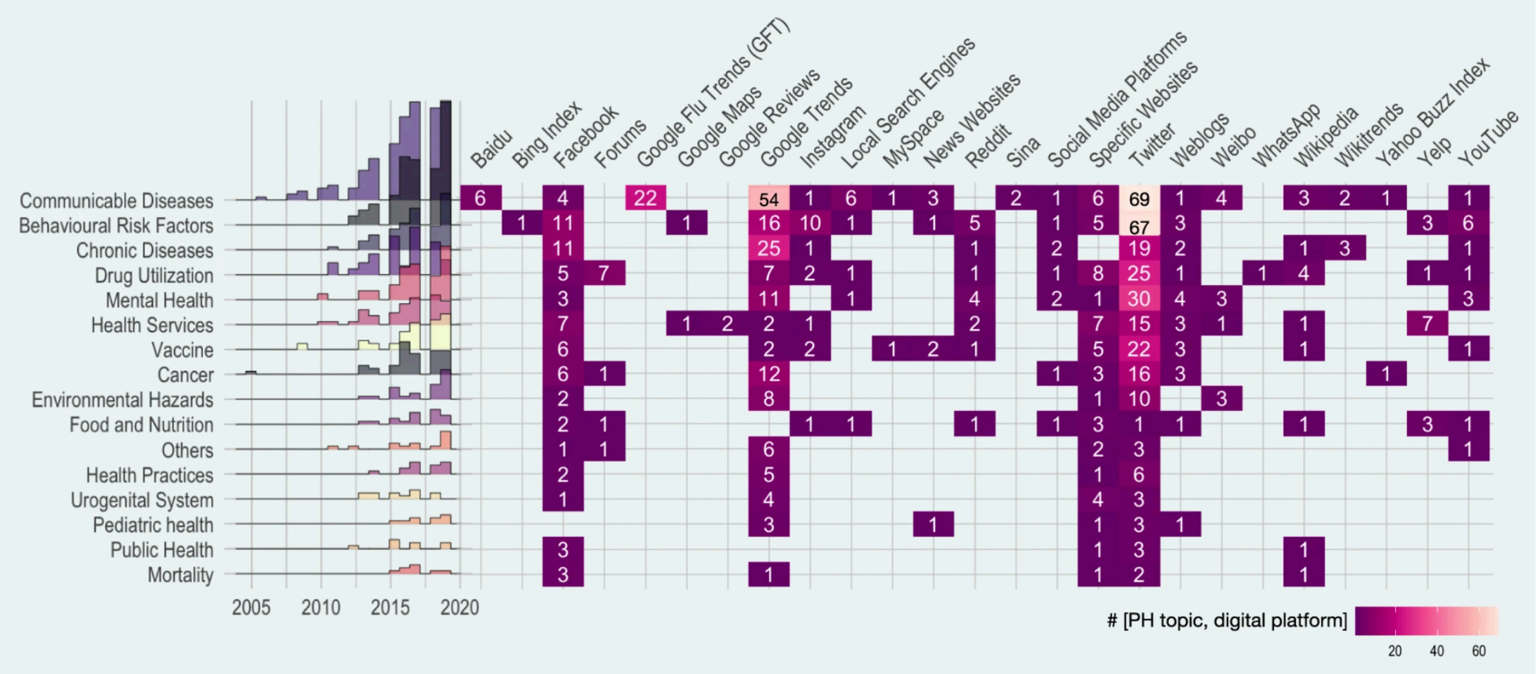
## Surveillance

- Event-based vs indicator-based surveillance

Most of digital public health surveillance is event-based

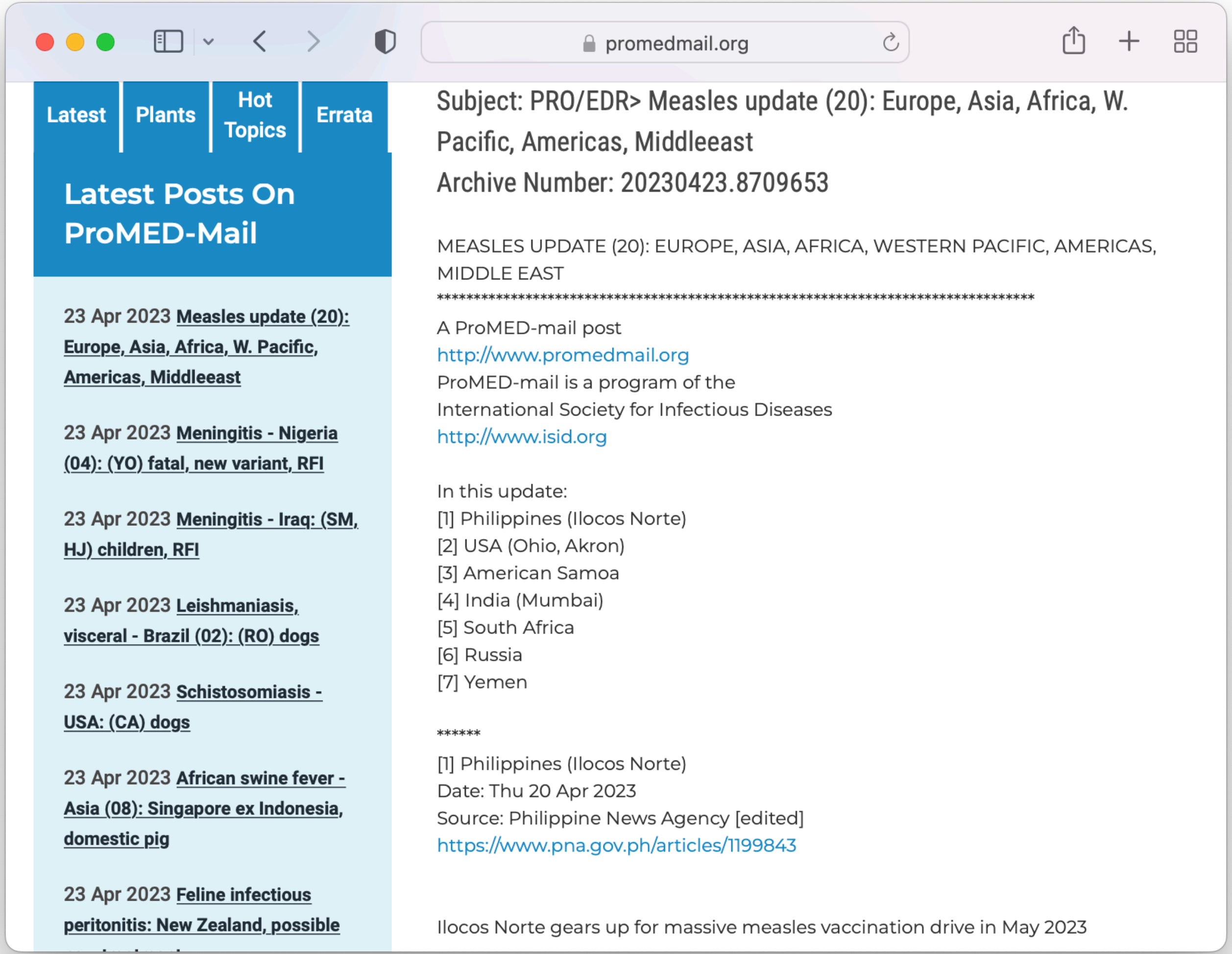
# Digital Public Health Surveillance

## Overview



| | Baidu | Bing Index | Facebook | Forums | Google Flu Trends (GFT) | Google Maps | Google Reviews | Google Trends | Instagram | Local Search Engines | MySpace | News Websites | Reddit | Sina | Social Media Platforms | Specific Websites | Twitter | Weblogs | Weibo | WhatsApp | Wikipedia | Wikitrends | Yahoo Buzz Index | Yelp | YouTube |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Communicable Diseases | 6 | | 4 | | 22 | | | 54 | 1 | 6 | 1 | 3 | | 2 | 1 | 6 | 69 | 1 | 4 | | 3 | 2 | 1 | | 1 |
| Behavioural Risk Factors | | 1 | 11 | | | 1 | | 16 | 10 | 1 | | 1 | 5 | | 1 | 5 | 67 | 3 | | | | | | 3 | 6 |
| Chronic Diseases | | | 11 | | | | | 25 | 1 | | | | 1 | | 2 | | 19 | 2 | | | 1 | 3 | | | 1 |
| Drug Utilization | | | 5 | 7 | | | | 7 | 2 | 1 | | | 1 | | 1 | 8 | 25 | 1 | | 1 | 4 | | | 1 | 1 |
| Mental Health | | | 3 | | | | | 11 | | 1 | | | 4 | | 2 | 1 | 30 | 4 | 3 | | | | | | 3 |
| Health Services | | | 7 | | | 1 | 2 | 2 | 1 | | | | 2 | | | 7 | 15 | 3 | 1 | | 1 | | | 7 | |
| Vaccine | | | 6 | | | | | 2 | 2 | | 1 | 2 | 1 | | | 5 | 22 | 3 | | | 1 | | | | 1 |
| Cancer | | | 6 | 1 | | | | 12 | | | | | | | 1 | 3 | 16 | 3 | | | | | 1 | | |
| Environmental Hazards | | | 2 | | | | | 8 | | | | | | | | 1 | 10 | | 3 | | | | | | |
| Food and Nutrition | | | 2 | 1 | | | | | 1 | 1 | | | 1 | | 1 | 3 | 1 | 1 | | | 1 | | | 3 | 1 |
| Others | | | 1 | 1 | | | | 6 | | | | | | | | 2 | 3 | | | | | | | | 1 |
| Health Practices | | | 2 | | | | | 5 | | | | | | | | 1 | 6 | | | | | | | | |
| Urogenital System | | | 1 | | | | | 4 | | | | | | | | 4 | 3 | | | | | | | | |
| Pediatric health | | | | | | | | 3 | | | | 1 | | | | 1 | 3 | 1 | | | | | | | |
| Public Health | | | 3 | | | | | | | | | | | | | 1 | 3 | | | | 1 | | | | |
| Mortality | | | 3 | | | | | 1 | | | | | | | | 1 | 2 | | | | 1 | | | | |

Source: Shakeri Hossein Abad et al. et al, 2021



## Latest Posts On ProMED-Mail

23 Apr 2023 **Measles update (20): Europe, Asia, Africa, W. Pacific, Americas, Middleeast**

23 Apr 2023 **Meningitis - Nigeria (04): (YO) fatal, new variant, RFI**

23 Apr 2023 **Meningitis - Iraq: (SM, HJ) children, RFI**

23 Apr 2023 **Leishmaniasis, visceral - Brazil (02): (RO) dogs**

23 Apr 2023 **Schistosomiasis - USA: (CA) dogs**

23 Apr 2023 **African swine fever - Asia (08): Singapore ex Indonesia, domestic pig**

23 Apr 2023 **Feline infectious peritonitis: New Zealand, possible**

## Subject: PRO/EDR> Measles update (20): Europe, Asia, Africa, W. Pacific, Americas, Middleeast
## Archive Number: 20230423.8709653

MEASLES UPDATE (20): EUROPE, ASIA, AFRICA, WESTERN PACIFIC, AMERICAS, MIDDLE EAST
********************************************************************************

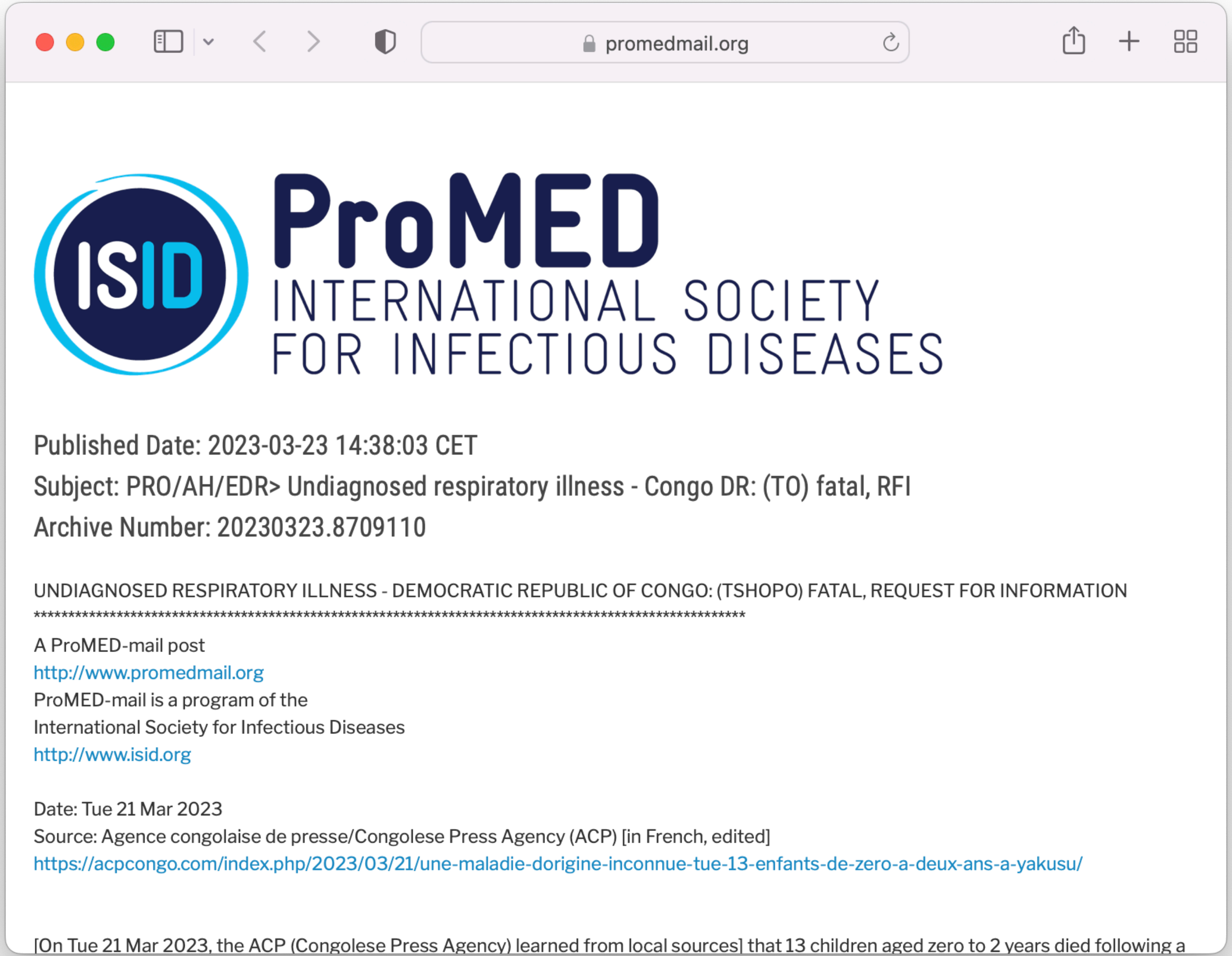
A ProMED-mail post
http://www.promedmail.org
ProMED-mail is a program of the
International Society for Infectious Diseases
http://www.isid.org

In this update:
[1] Philippines (Ilocos Norte)
[2] USA (Ohio, Akron)
[3] American Samoa
[4] India (Mumbai)
[5] South Africa
[6] Russia
[7] Yemen

******

[1] Philippines (Ilocos Norte)
Date: Thu 20 Apr 2023
Source: Philippine News Agency [edited]
https://www.pna.gov.ph/articles/1199843

Ilocos Norte gears up for massive measles vaccination drive in May 2023



Published Date: 2023-03-23 14:38:03 CET

Subject: PRO/AH/EDR> Undiagnosed respiratory illness - Congo DR: (TO) fatal, RFI

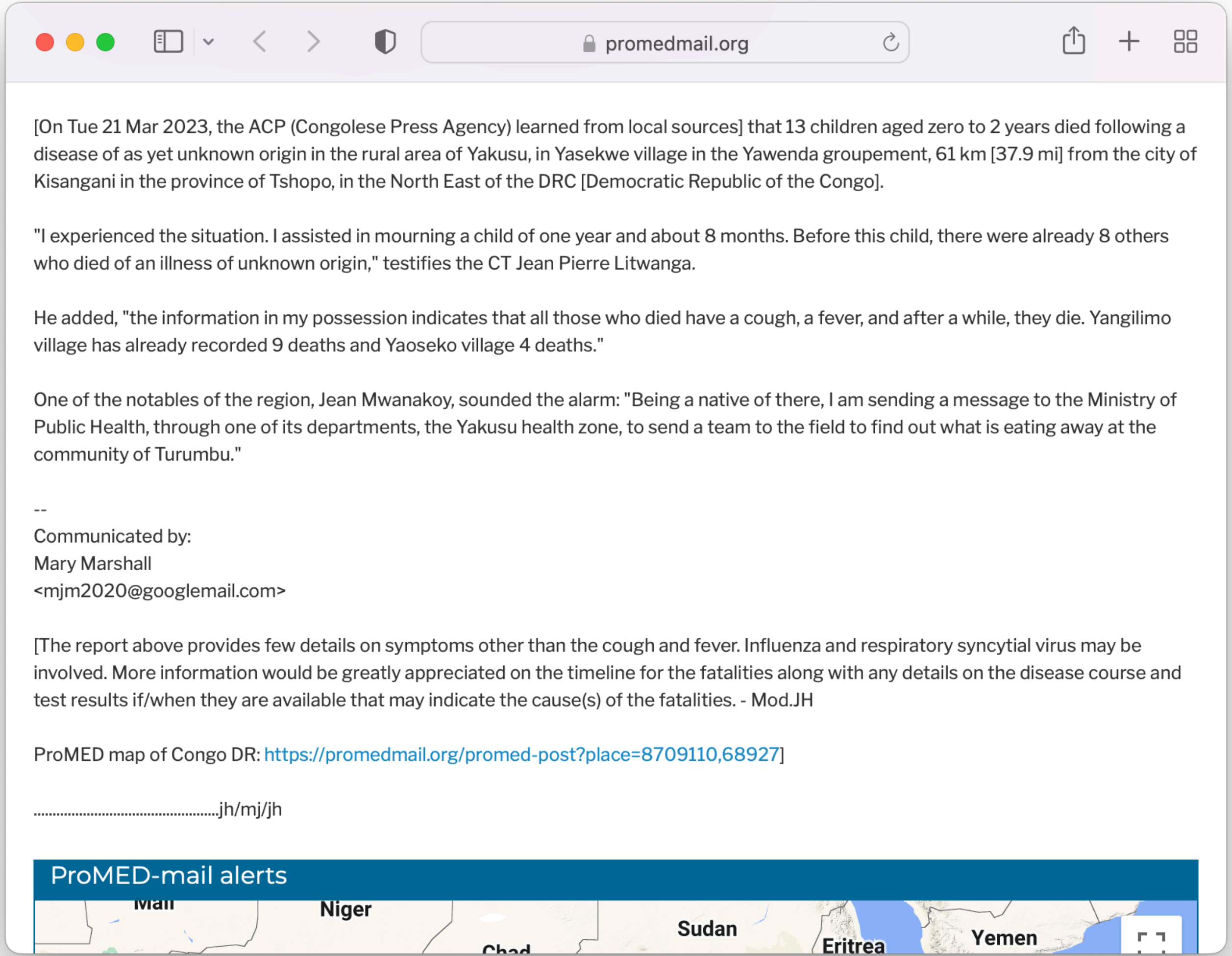
Archive Number: 20230323.8709110

UNDIAGNOSED RESPIRATORY ILLNESS - DEMOCRATIC REPUBLIC OF CONGO: (TSHOPO) FATAL, REQUEST FOR INFORMATION
******************************************************************************************

A ProMED-mail post
http://www.promedmail.org
ProMED-mail is a program of the
International Society for Infectious Diseases
http://www.isid.org

Date: Tue 21 Mar 2023
Source: Agence congolaise de presse/Congolese Press Agency (ACP) [in French, edited]
https://acpcongo.com/index.php/2023/03/21/une-maladie-dorigine-inconnue-tue-13-enfants-de-zero-a-deux-ans-a-yakusu/

[On Tue 21 Mar 2023, the ACP (Congolese Press Agency) learned from local sources] that 13 children aged zero to 2 years died following a



[On Tue 21 Mar 2023, the ACP (Congolese Press Agency) learned from local sources] that 13 children aged zero to 2 years died following a disease of as yet unknown origin in the rural area of Yakusu, in Yasekwe village in the Yawenda groupement, 61 km [37.9 mi] from the city of Kisangani in the province of Tshopo, in the North East of the DRC [Democratic Republic of the Congo].

"I experienced the situation. I assisted in mourning a child of one year and about 8 months. Before this child, there were already 8 others who died of an illness of unknown origin," testifies the CT Jean Pierre Litwanga.

He added, "the information in my possession indicates that all those who died have a cough, a fever, and after a while, they die. Yangilimo village has already recorded 9 deaths and Yaoseko village 4 deaths."

One of the notables of the region, Jean Mwanakoy, sounded the alarm: "Being a native of there, I am sending a message to the Ministry of Public Health, through one of its departments, the Yakusu health zone, to send a team to the field to find out what is eating away at the community of Turumbu."

--
Communicated by:
Mary Marshall
<mjm2020@googlemail.com>

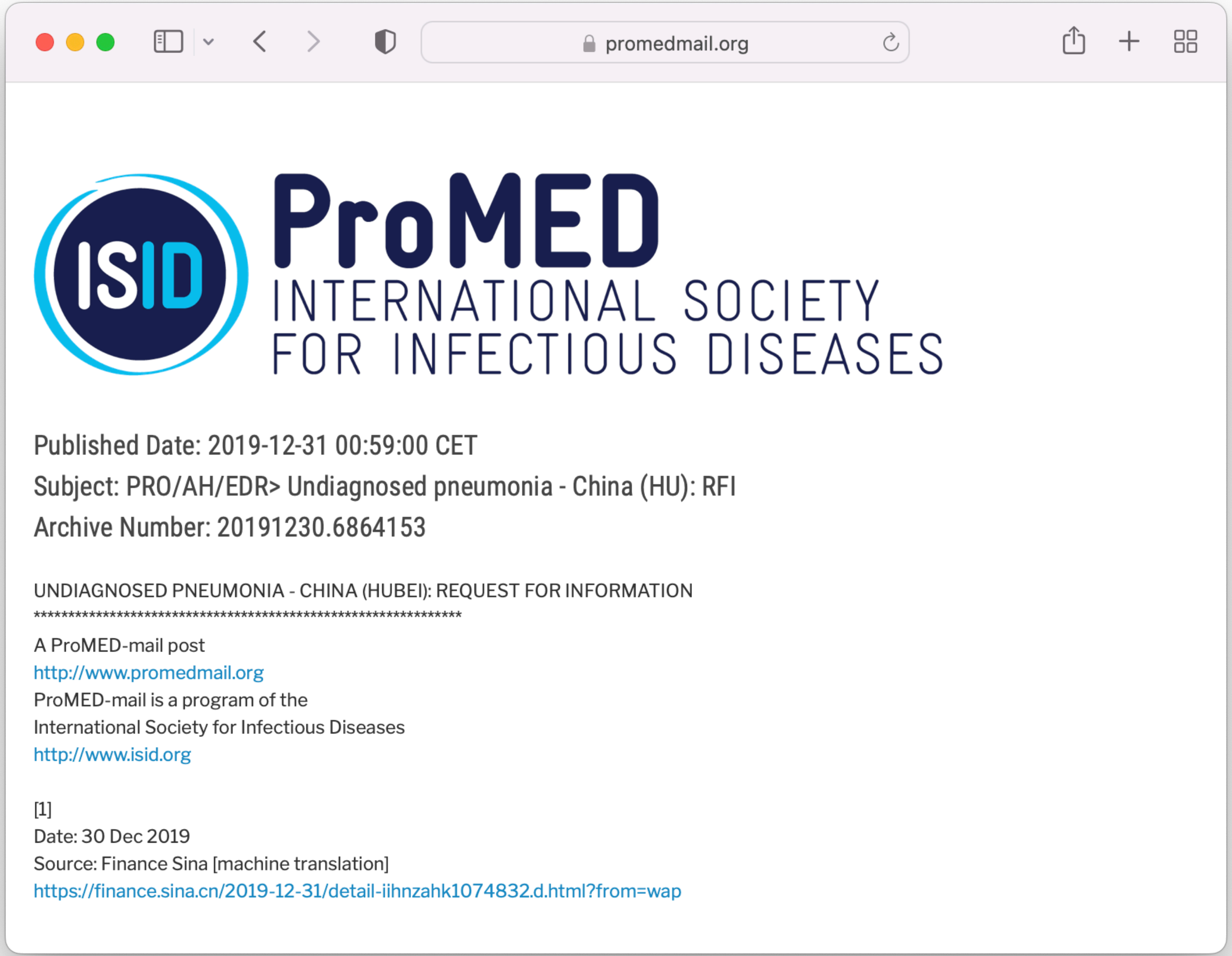
[The report above provides few details on symptoms other than the cough and fever. Influenza and respiratory syncytial virus may be involved. More information would be greatly appreciated on the timeline for the fatalities along with any details on the disease course and test results if/when they are available that may indicate the cause(s) of the fatalities. - Mod.JH

ProMED map of Congo DR: https://promedmail.org/promed-post?place=8709110,68927]

...........................................jh/mj/jh

ProMED-mail alerts

# December 31, 2019



Published Date: 2019-12-31 00:59:00 CET

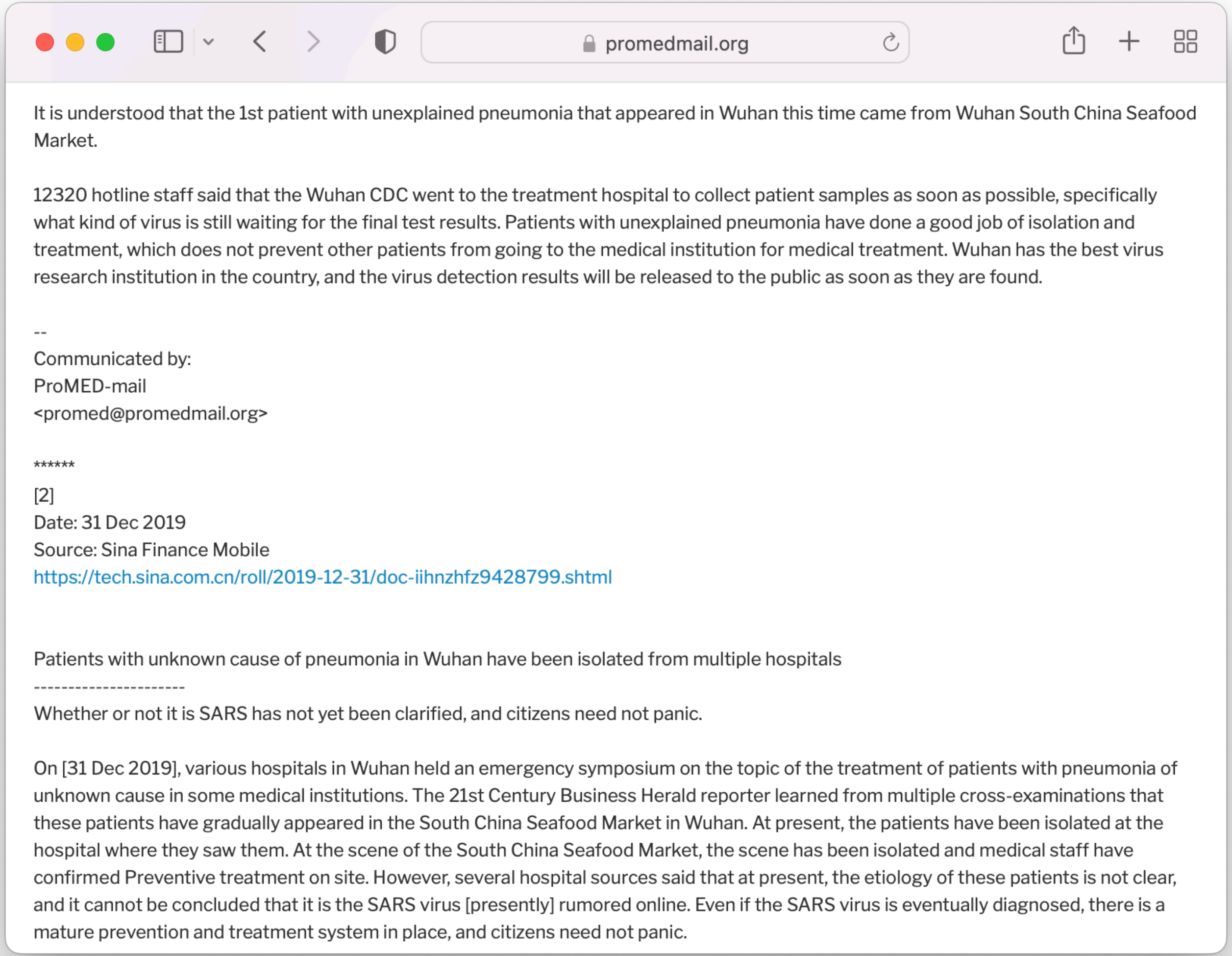
Subject: PRO/AH/EDR> Undiagnosed pneumonia - China (HU): RFI

Archive Number: 20191230.6864153

UNDIAGNOSED PNEUMONIA - CHINA (HUBEI): REQUEST FOR INFORMATION
*****************************************************************

A ProMED-mail post
http://www.promedmail.org
ProMED-mail is a program of the
International Society for Infectious Diseases
http://www.isid.org

[1]
Date: 30 Dec 2019
Source: Finance Sina [machine translation]
https://finance.sina.cn/2019-12-31/detail-iihnzahk1074832.d.html?from=wap



It is understood that the 1st patient with unexplained pneumonia that appeared in Wuhan this time came from Wuhan South China Seafood Market.

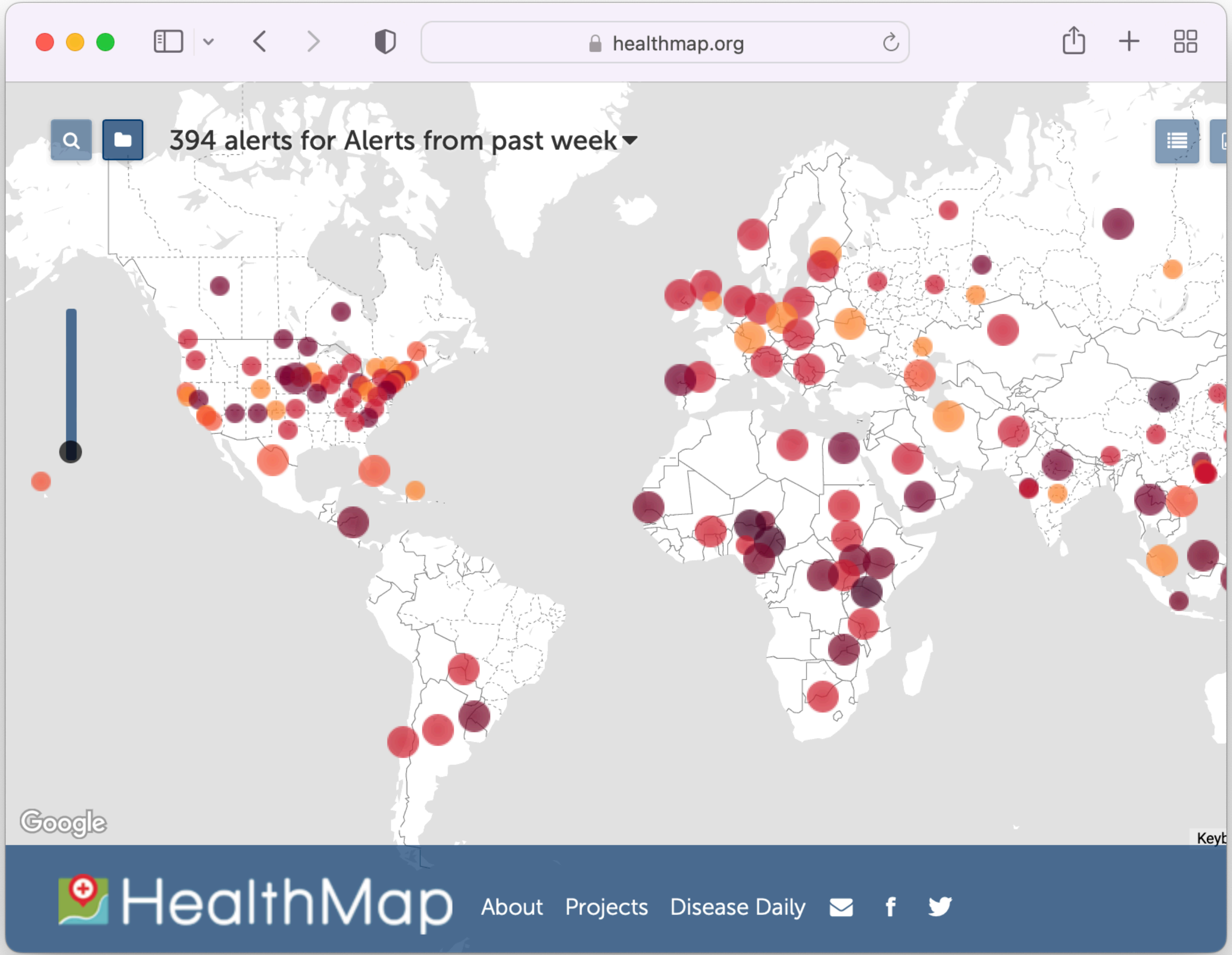
12320 hotline staff said that the Wuhan CDC went to the treatment hospital to collect patient samples as soon as possible, specifically what kind of virus is still waiting for the final test results. Patients with unexplained pneumonia have done a good job of isolation and treatment, which does not prevent other patients from going to the medical institution for medical treatment. Wuhan has the best virus research institution in the country, and the virus detection results will be released to the public as soon as they are found.

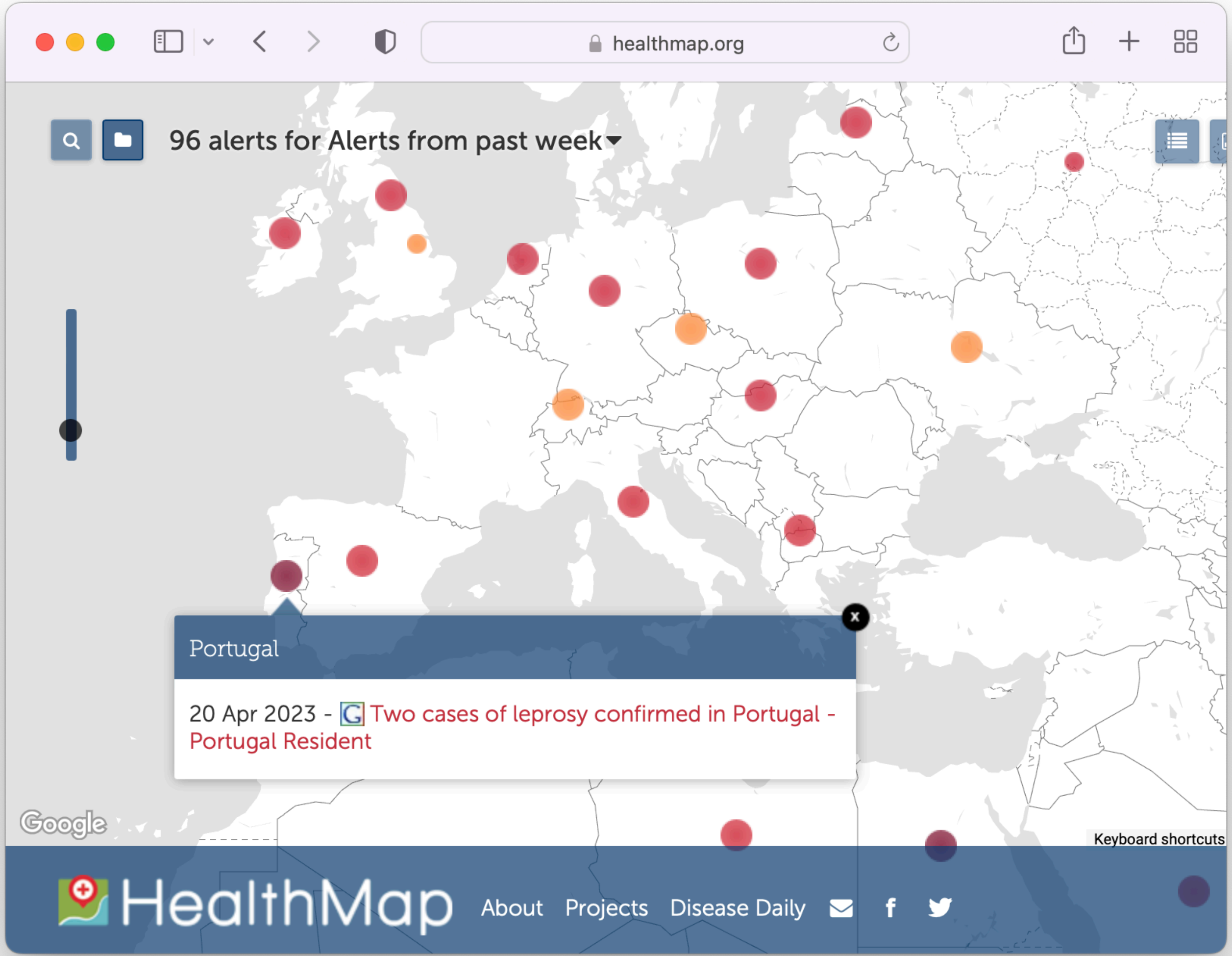
--
Communicated by:
ProMED-mail
<promed@promedmail.org>

******
[2]
Date: 31 Dec 2019
Source: Sina Finance Mobile
https://tech.sina.com.cn/roll/2019-12-31/doc-iihnzhfz9428799.shtml

Patients with unknown cause of pneumonia in Wuhan have been isolated from multiple hospitals
----------------------
Whether or not it is SARS has not yet been clarified, and citizens need not panic.

On [31 Dec 2019], various hospitals in Wuhan held an emergency symposium on the topic of the treatment of patients with pneumonia of unknown cause in some medical institutions. The 21st Century Business Herald reporter learned from multiple cross-examinations that these patients have gradually appeared in the South China Seafood Market in Wuhan. At present, the patients have been isolated at the hospital where they saw them. At the scene of the South China Seafood Market, the scene has been isolated and medical staff have confirmed Preventive treatment on site. However, several hospital sources said that at present, the etiology of these patients is not clear, and it cannot be concluded that it is the SARS virus [presently] rumored online. Even if the SARS virus is eventually diagnosed, there is a mature prevention and treatment system in place, and citizens need not panic.


healthmap.org
394 alerts for Alerts from past week
Google
Keyb
HealthMap
About
Projects
Disease Daily



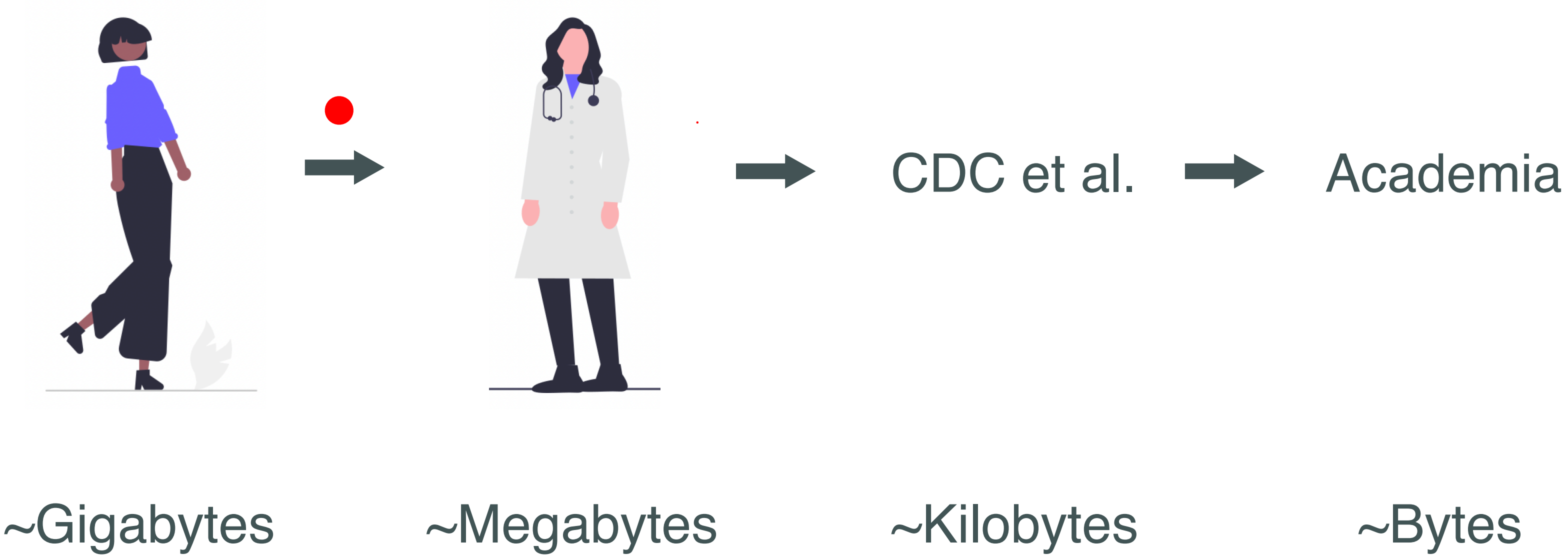
96 alerts for Alerts from past week

Portugal

20 Apr 2023 - Two cases of leprosy confirmed in Portugal - Portugal Resident

# Digital Public Health Surveillance

## Overview



CDC et al.

Academia

~Gigabytes

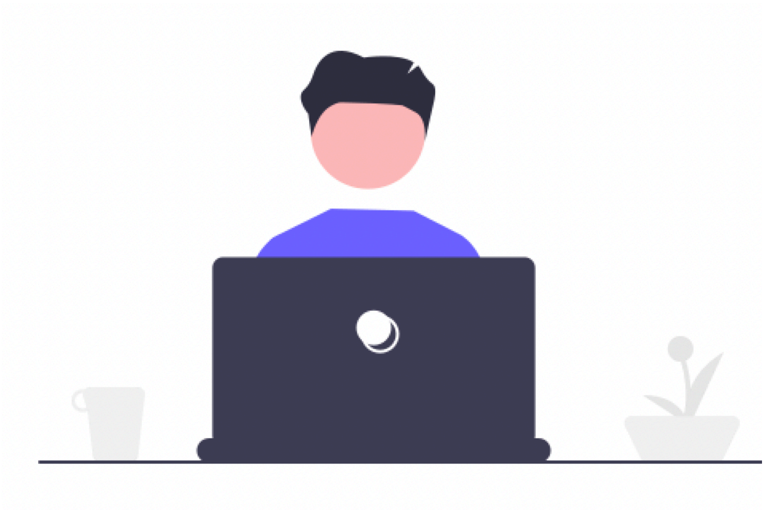
~Megabytes

~Kilobytes

~Bytes

# Digital Public Health Surveillance

## Overview

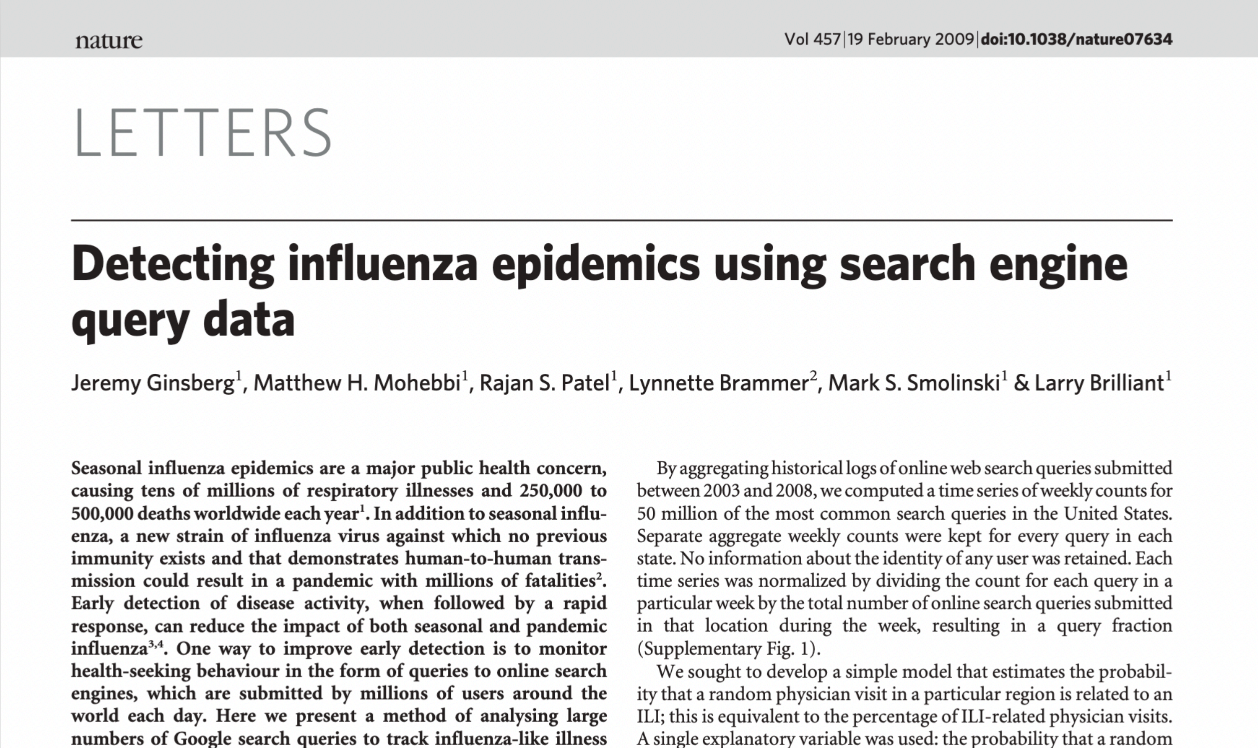


Search: “Fever, shivering, cough, what disease”

Social media: “Really not feeling great since two days 🤒 Hope I don’t have the flu…”

# Digital Public Health Surveillance

## Search Query Data



nature Vol 457|19 February 2009|doi:10.1038/nature07634

LETTERS

### Detecting influenza epidemics using search engine query data

Jeremy Ginsberg[1], Matthew H. Mohebbi[1], Rajan S. Patel[1], Lynnette Brammer[2], Mark S. Smolinski[1] & Larry Brilliant[1]

**Seasonal influenza epidemics are a major public health concern, causing tens of millions of respiratory illnesses and 250,000 to 500,000 deaths worldwide each year[1]. In addition to seasonal influenza, a new strain of influenza virus against which no previous immunity exists and that demonstrates human-to-human transmission could result in a pandemic with millions of fatalities[2]. Early detection of disease activity, when followed by a rapid response, can reduce the impact of both seasonal and pandemic influenza[3,4]. One way to improve early detection is to monitor health-seeking behaviour in the form of queries to online search engines, which are submitted by millions of users around the world each day. Here we present a method of analysing large numbers of Google search queries to track influenza-like illness**

By aggregating historical logs of online web search queries submitted between 2003 and 2008, we computed a time series of weekly counts for 50 million of the most common search queries in the United States. Separate aggregate weekly counts were kept for every query in each state. No information about the identity of any user was retained. Each time series was normalized by dividing the count for each query in a particular week by the total number of online search queries submitted in that location during the week, resulting in a query fraction (Supplementary Fig. 1).

We sought to develop a simple model that estimates the probability that a random physician visit in a particular region is related to an ILI; this is equivalent to the percentage of ILI-related physician visits. A single explanatory variable was used: the probability that a random

# Digital Public Health Surveillance

## Search Query Data

- Idea wasn’t new: Gunther Eysenbach published on the same idea - but not being Google, he had to resort to a trick.

# Digital Public Health Surveillance

## Search Query Data

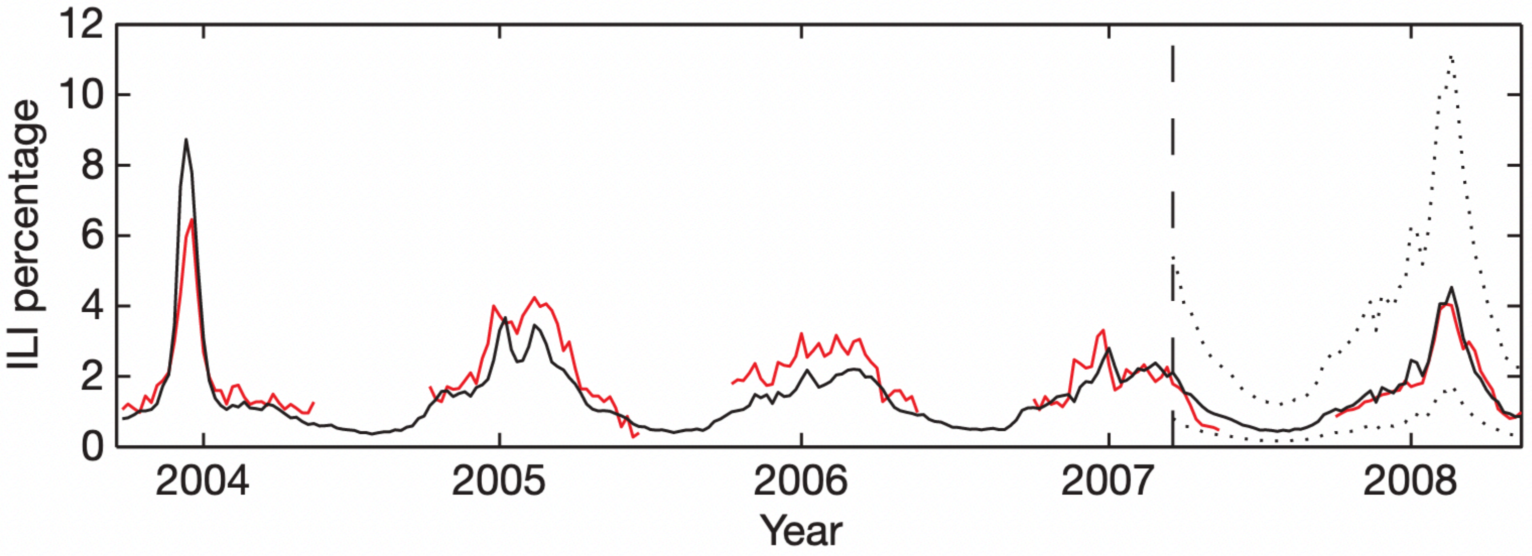


**Figure 2 | A comparison of model estimates for the mid-Atlantic region (black) against CDC-reported ILI percentages (red), including points over which the model was fit and validated.** A correlation of 0.85 was obtained

# Digital Public Health Surveillance

## Search Query Data

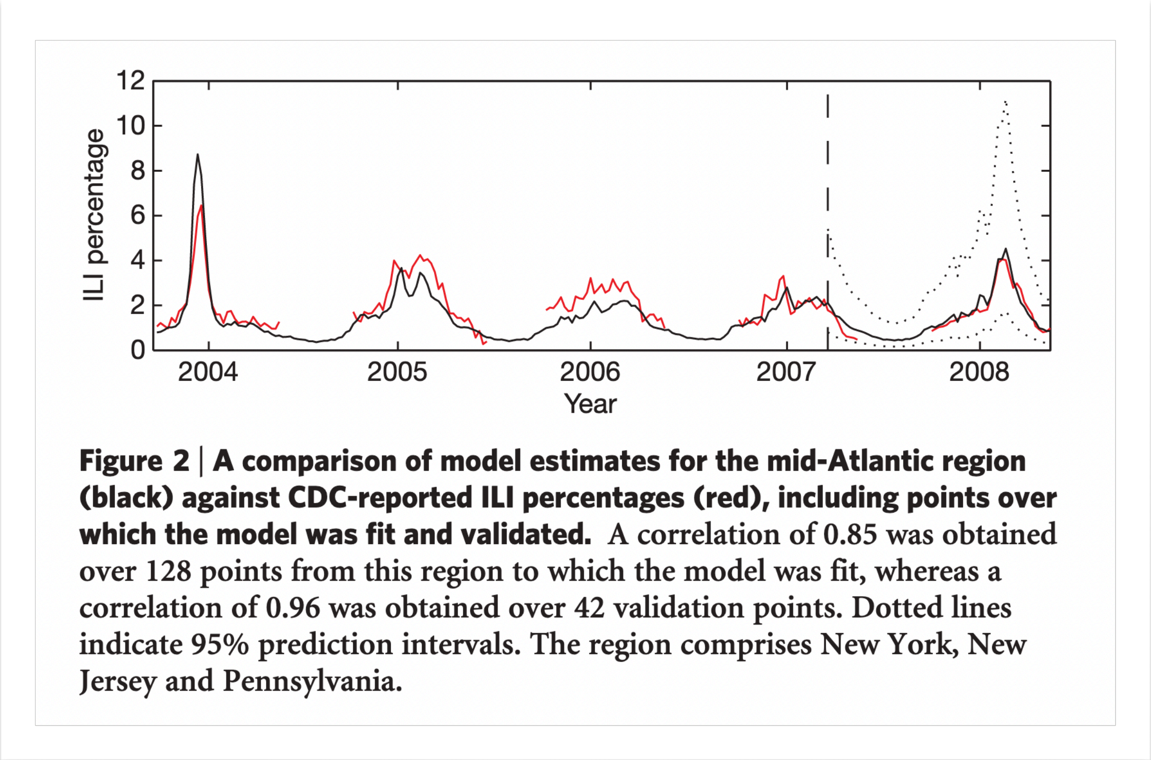


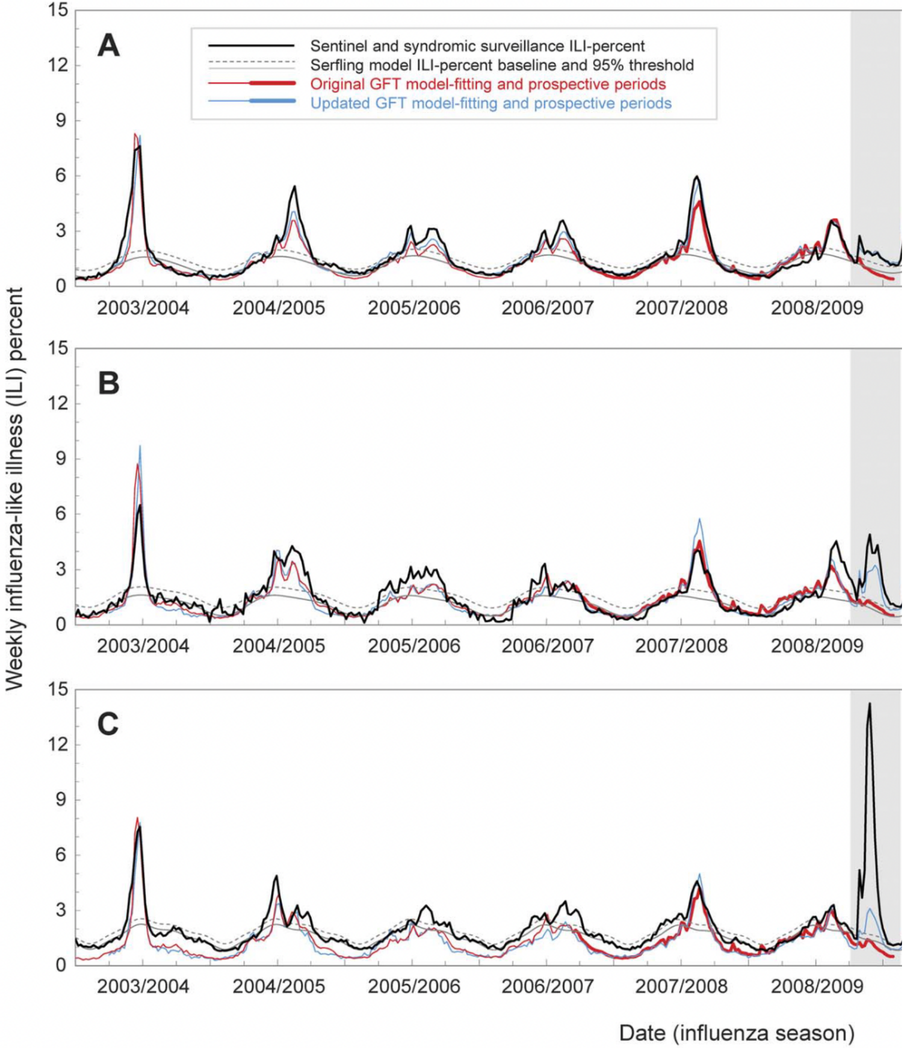
**Figure 2 | A comparison of model estimates for the mid-Atlantic region (black) against CDC-reported ILI percentages (red), including points over which the model was fit and validated.** A correlation of 0.85 was obtained over 128 points from this region to which the model was fit, whereas a correlation of 0.96 was obtained over 42 validation points. Dotted lines indicate 95% prediction intervals. The region comprises New York, New Jersey and Pennsylvania.

# Digital Public Health Surveillance

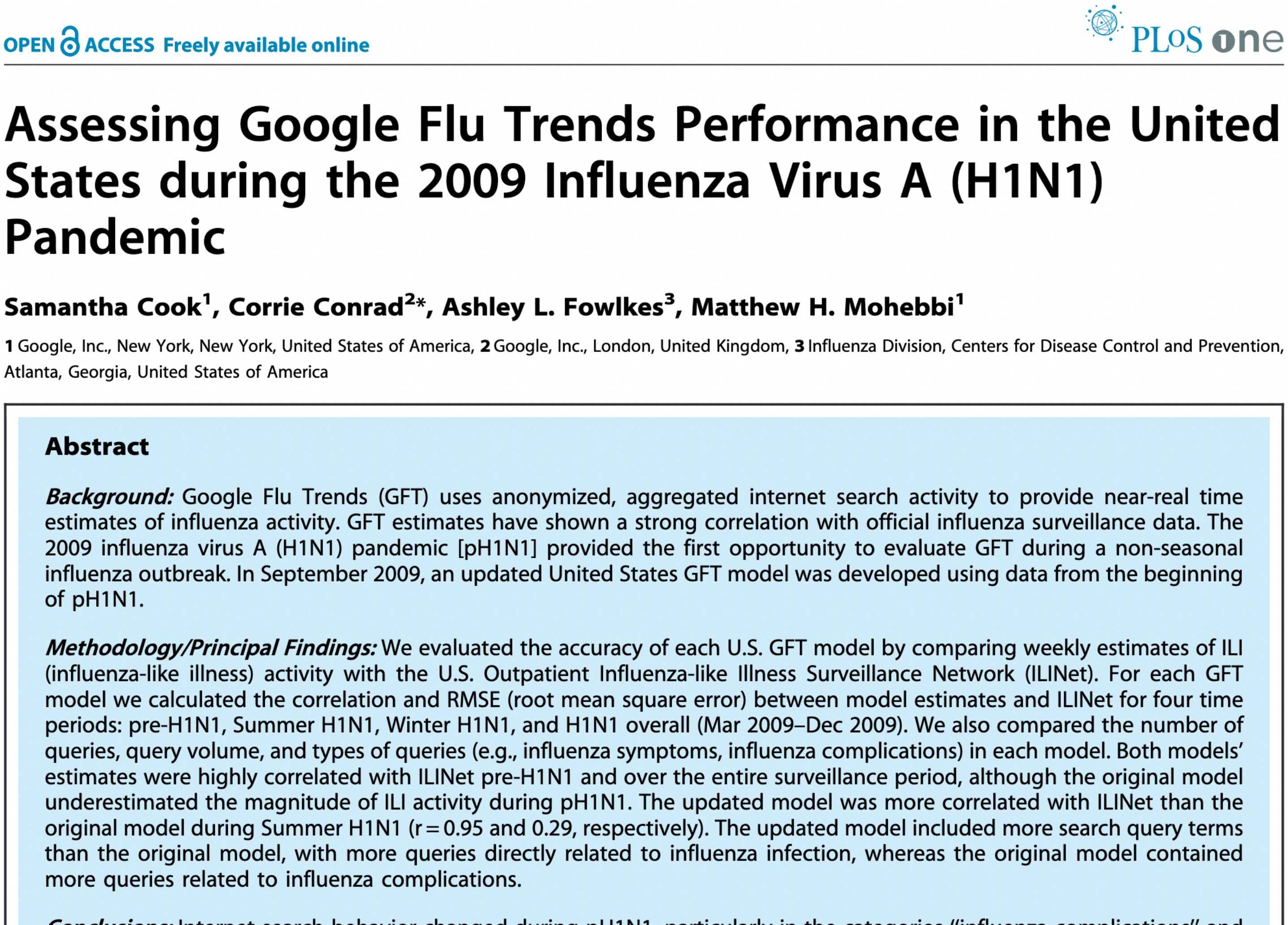
## Search Query Data

- “This system is not designed to be a replacement for traditional surveillance networks or supplant the need for laboratory-based diagnoses and surveillance.”

- “Despite strong historical correlations, our system remains susceptible to false alerts caused by a sudden increase in ILI-related queries.”



Source: Olson et al, 2013



OPEN ACCESS Freely available online
PLoS one

# Assessing Google Flu Trends Performance in the United States during the 2009 Influenza Virus A (H1N1) Pandemic

**Samantha Cook**[1], **Corrie Conrad**[2]*, **Ashley L. Fowlkes**[3], **Matthew H. Mohebbi**[1]

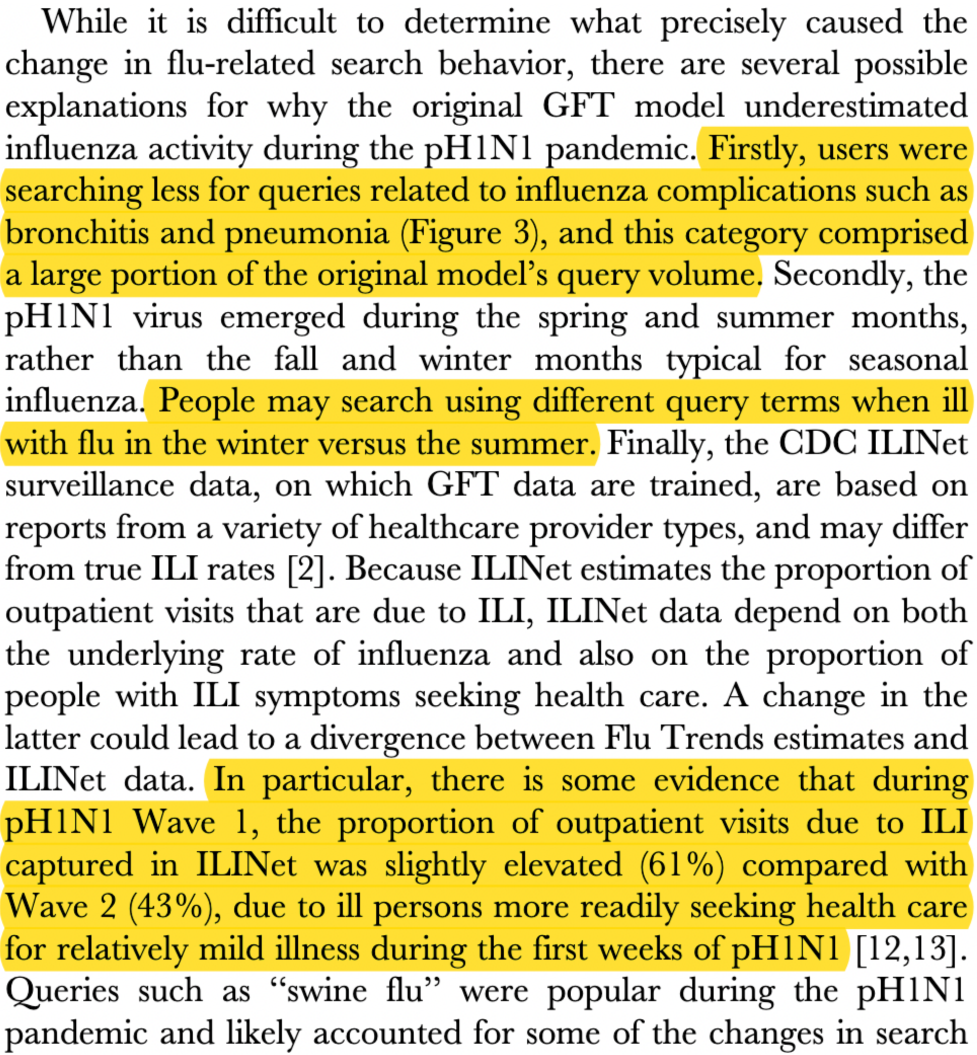
**1** Google, Inc., New York, New York, United States of America, **2** Google, Inc., London, United Kingdom, **3** Influenza Division, Centers for Disease Control and Prevention, Atlanta, Georgia, United States of America

## Abstract

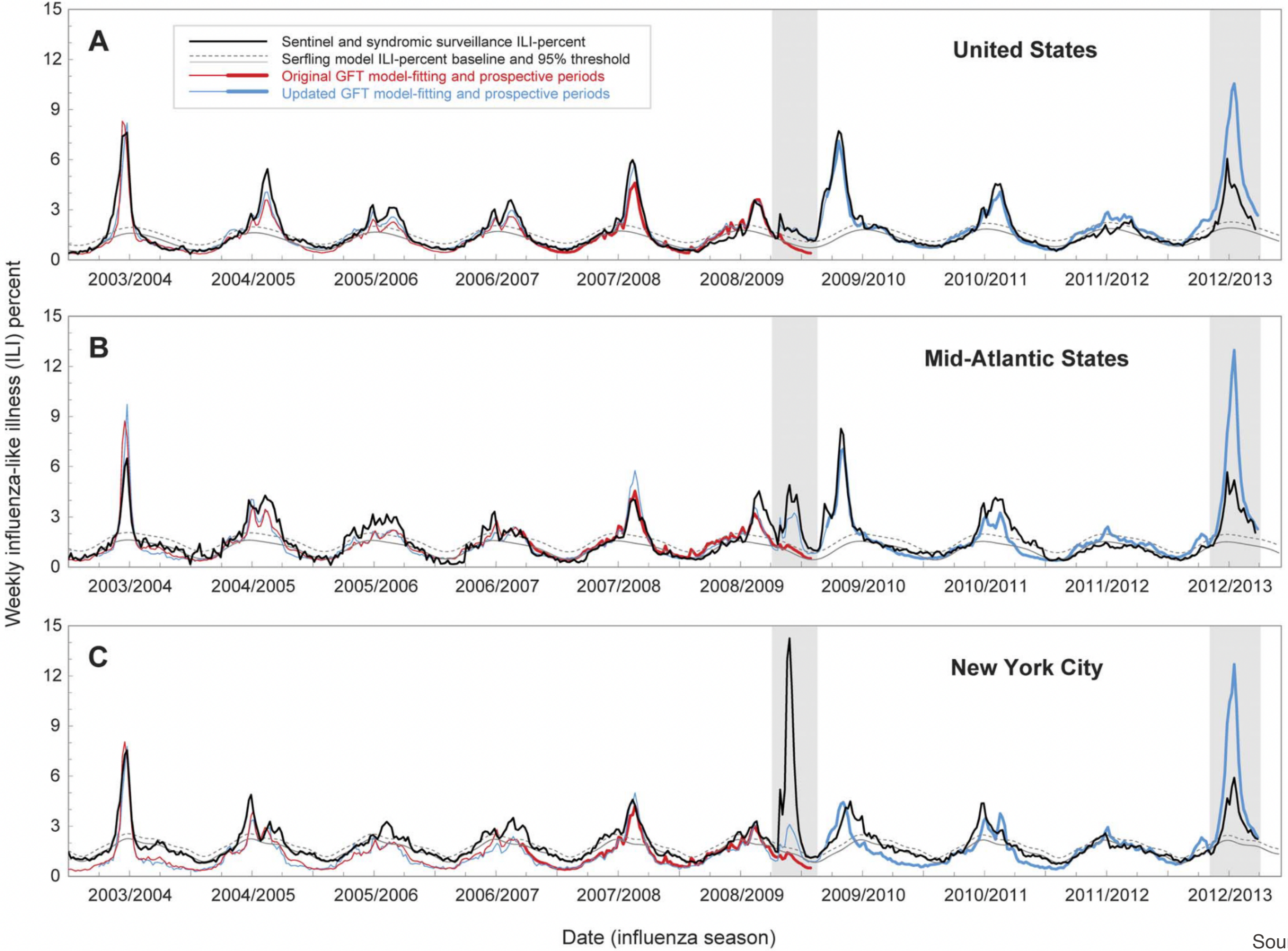
***Background:*** Google Flu Trends (GFT) uses anonymized, aggregated internet search activity to provide near-real time estimates of influenza activity. GFT estimates have shown a strong correlation with official influenza surveillance data. The 2009 influenza virus A (H1N1) pandemic [pH1N1] provided the first opportunity to evaluate GFT during a non-seasonal influenza outbreak. In September 2009, an updated United States GFT model was developed using data from the beginning of pH1N1.

***Methodology/Principal Findings:*** We evaluated the accuracy of each U.S. GFT model by comparing weekly estimates of ILI (influenza-like illness) activity with the U.S. Outpatient Influenza-like Illness Surveillance Network (ILINet). For each GFT model we calculated the correlation and RMSE (root mean square error) between model estimates and ILINet for four time periods: pre-H1N1, Summer H1N1, Winter H1N1, and H1N1 overall (Mar 2009–Dec 2009). We also compared the number of queries, query volume, and types of queries (e.g., influenza symptoms, influenza complications) in each model. Both models' estimates were highly correlated with ILINet pre-H1N1 and over the entire surveillance period, although the original model underestimated the magnitude of ILI activity during pH1N1. The updated model was more correlated with ILINet than the original model during Summer H1N1 ($r = 0.95$ and $0.29$, respectively). The updated model included more search query terms than the original model, with more queries directly related to influenza infection, whereas the original model contained more queries related to influenza complications.

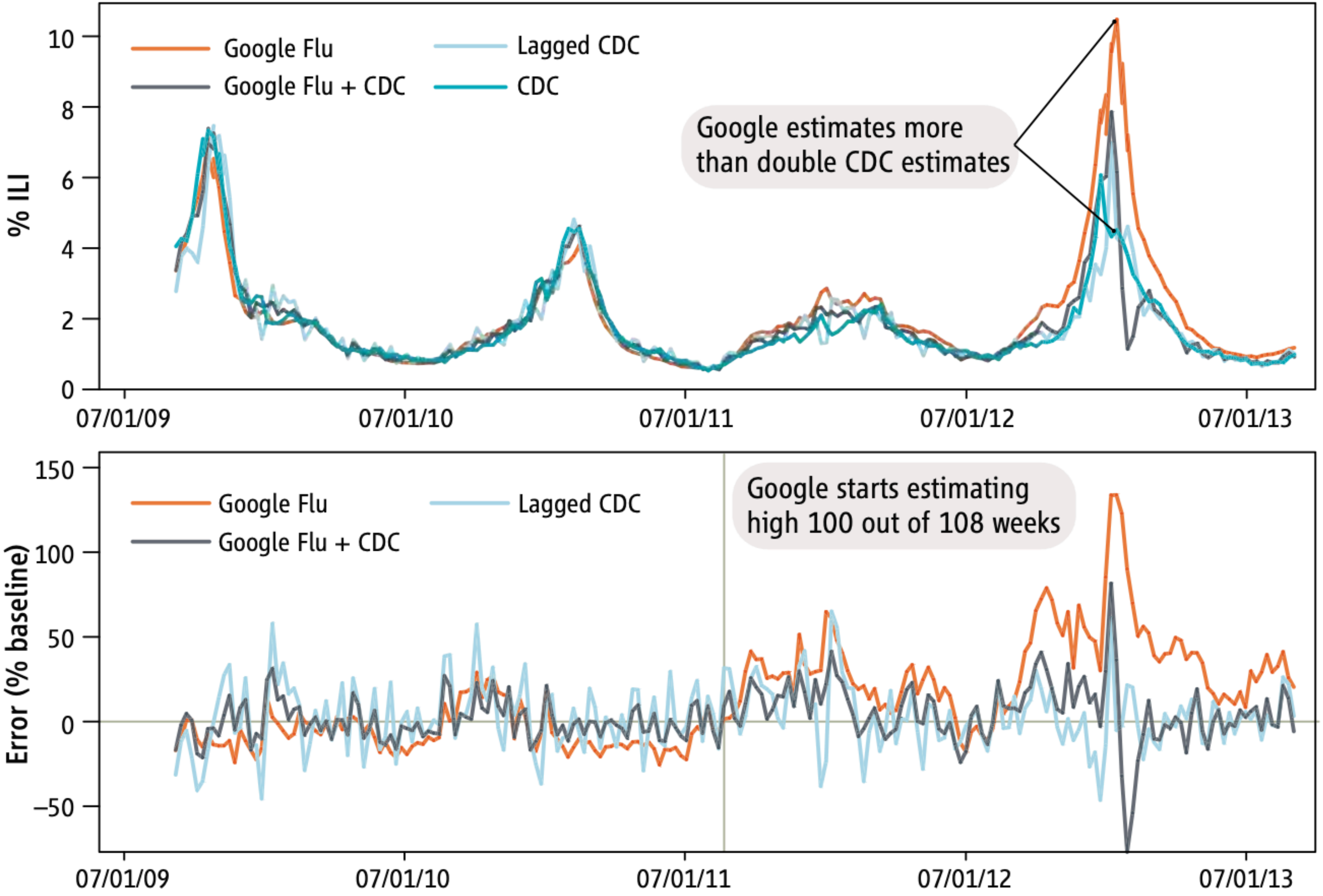
***Conclusions:*** Internet search behavior changed during pH1N1, particularly in the categories "influenza complications" and



While it is difficult to determine what precisely caused the change in flu-related search behavior, there are several possible explanations for why the original GFT model underestimated influenza activity during the pH1N1 pandemic. Firstly, users were searching less for queries related to influenza complications such as bronchitis and pneumonia (Figure 3), and this category comprised a large portion of the original model's query volume. Secondly, the pH1N1 virus emerged during the spring and summer months, rather than the fall and winter months typical for seasonal influenza. People may search using different query terms when ill with flu in the winter versus the summer. Finally, the CDC ILINet surveillance data, on which GFT data are trained, are based on reports from a variety of healthcare provider types, and may differ from true ILI rates [2]. Because ILINet estimates the proportion of outpatient visits that are due to ILI, ILINet data depend on both the underlying rate of influenza and also on the proportion of people with ILI symptoms seeking health care. A change in the latter could lead to a divergence between Flu Trends estimates and ILINet data. In particular, there is some evidence that during pH1N1 Wave 1, the proportion of outpatient visits due to ILI captured in ILINet was slightly elevated (61%) compared with Wave 2 (43%), due to ill persons more readily seeking health care for relatively mild illness during the first weeks of pH1N1 [12,13]. Queries such as "swine flu" were popular during the pH1N1 pandemic and likely accounted for some of the changes in search



Source: Olson et al, 2013



Source: Lazer et al, 2014

# Digital Public Health Surveillance

## Search Query Data

- Challenges
  1. Queries not related to flu incidence
  2. Reproducibility, transparency: Google never provided the 45 search terms used in GFT.
  3. Blue team / red team issues

- In 2015, GFT stopped being accessible to the public.

# Digital Public Health Surveillance

## Search Query Data

# Digital Public Health Surveillance

## Search Query Data

- Algorithms can be continuously improved / adjusted
- Is comparison to ILI from CDC fair? It turns out that the data are reported with quite some delay (2-3 weeks until finalized). So the fair comparison is with the data available at the time.

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

# Reappraising the utility of Google Flu Trends

**Sasikiran Kandula*, Jeffrey Shaman**

Department of Environmental Health Sciences, Columbia University, New York, New York, United States of America

* sk3542@cumc.columbia.edu

Check for updates

OPEN ACCESS

**Citation:** Kandula S, Shaman J (2019) Reappraising the utility of Google Flu Trends. PLoS Comput Biol 15(8): e1007258. https://doi.org/10.1371/journal.pcbi.1007258

**Editor:** Nicola Segata, University of Trento, ITALY

**Received:** October 29, 2018

**Accepted:** July 9, 2019

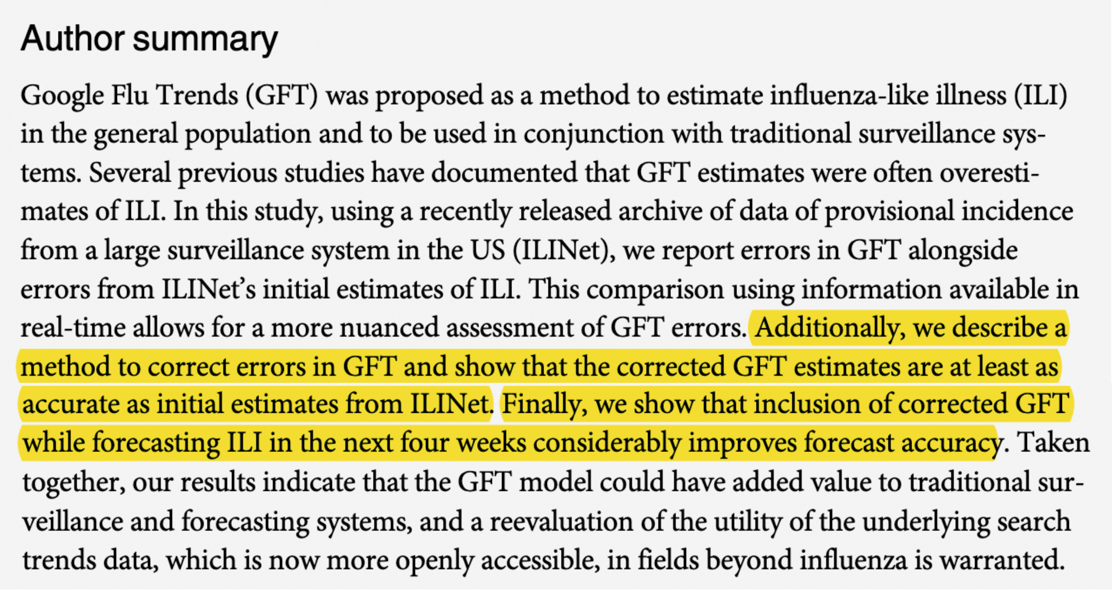
**Published:** August 2, 2019

## Abstract

Estimation of influenza-like illness (ILI) using search trends activity was intended to supplement traditional surveillance systems, and was a motivation behind the development of Google Flu Trends (GFT). However, several studies have previously reported large errors in GFT estimates of ILI in the US. Following recent release of time-stamped surveillance data, which better reflects real-time operational scenarios, we reanalyzed GFT errors. Using three data sources—GFT: an archive of weekly ILI estimates from Google Flu Trends; ILIf: fully-observed ILI rates from ILINet; and, ILIp: ILI rates available in real-time based on partial reporting—five influenza seasons were analyzed and mean square errors (MSE) of GFT and ILIp as estimates of ILIf were computed. To correct GFT errors, a random forest regression model was built with ILI and GFT rates from the previous three weeks as predictors. An overall reduction in error of 44% was observed and the errors of the corrected GFT are lower than those of ILIp. An 80% reduction in error during 2012/13, when GFT had large errors, shows that extreme failures of GFT could have been avoided. Using autoregressive integrated moving average (ARIMA) models, one- to four-week ahead forecasts were generated with two separate data streams: ILIp alone, and with both ILIp and corrected GFT. At all forecast targets and seasons, and for all but two regions, inclusion of GFT lowered MSE. Results from two alternative error measures, mean absolute error and mean absolute proportional error, were largely consistent with results from MSE. Taken together these findings

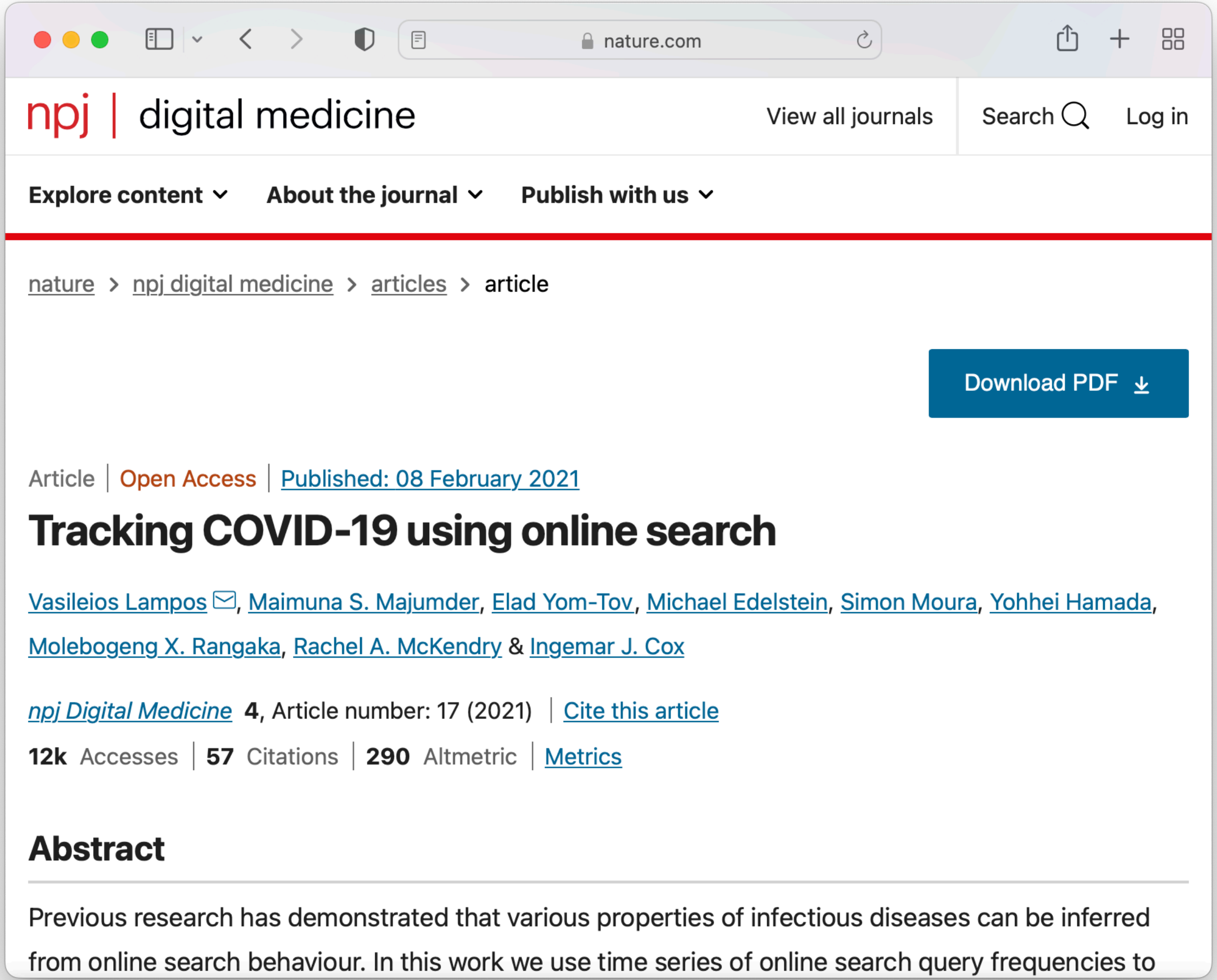
# Digital Public Health Surveillance

## Search Query Data



### Author summary

Google Flu Trends (GFT) was proposed as a method to estimate influenza-like illness (ILI) in the general population and to be used in conjunction with traditional surveillance systems. Several previous studies have documented that GFT estimates were often overestimates of ILI. In this study, using a recently released archive of data of provisional incidence from a large surveillance system in the US (ILINet), we report errors in GFT alongside errors from ILINet's initial estimates of ILI. This comparison using information available in real-time allows for a more nuanced assessment of GFT errors. Additionally, we describe a method to correct errors in GFT and show that the corrected GFT estimates are at least as accurate as initial estimates from ILINet. Finally, we show that inclusion of corrected GFT while forecasting ILI in the next four weeks considerably improves forecast accuracy. Taken together, our results indicate that the GFT model could have added value to traditional surveillance and forecasting systems, and a reevaluation of the utility of the underlying search trends data, which is now more openly accessible, in fields beyond influenza is warranted.



Article | Open Access | Published: 08 February 2021

# Tracking COVID-19 using online search

Vasileios Lampos, Maimuna S. Majumder, Elad Yom-Tov, Michael Edelstein, Simon Moura, Yohhei Hamada, Molebogeng X. Rangaka, Rachel A. McKendry & Ingemar J. Cox

*npj Digital Medicine* **4**, Article number: 17 (2021) | Cite this article

**12k** Accesses | **57** Citations | **290** Altmetric | Metrics

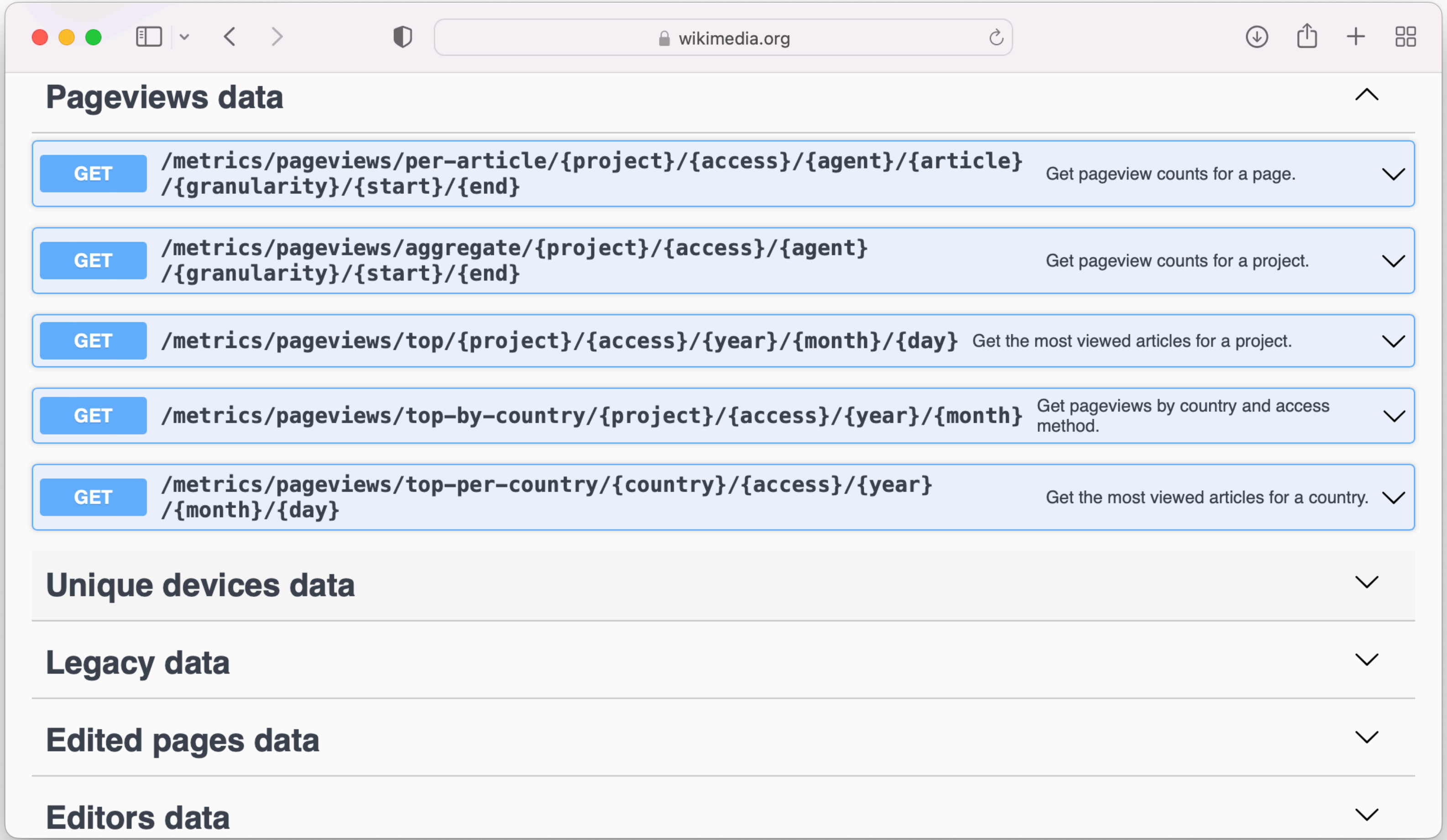
## Abstract

Previous research has demonstrated that various properties of infectious diseases can be inferred from online search behaviour. In this work we use time series of online search query frequencies to

# Digital Public Health Surveillance

## Log Data

- Wikimedia has an API: https://wikimedia.org/api/rest_v1/



## Pageviews data

**GET** `/metrics/pageviews/per-article/{project}/{access}/{agent}/{article}/{granularity}/{start}/{end}` Get pageview counts for a page.

**GET** `/metrics/pageviews/aggregate/{project}/{access}/{agent}/{granularity}/{start}/{end}` Get pageview counts for a project.

**GET** `/metrics/pageviews/top/{project}/{access}/{year}/{month}/{day}` Get the most viewed articles for a project.

**GET** `/metrics/pageviews/top-by-country/{project}/{access}/{year}/{month}` Get pageviews by country and access method.

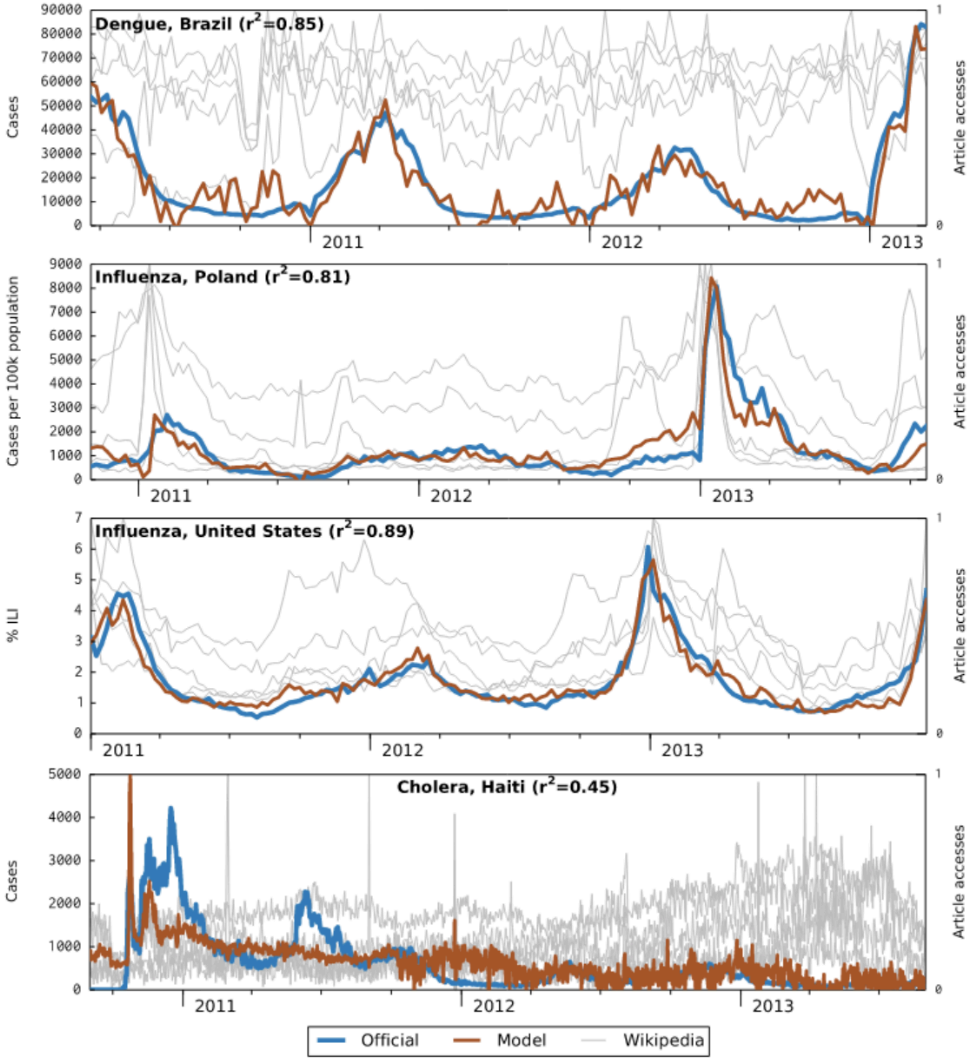
**GET** `/metrics/pageviews/top-per-country/{country}/{access}/{year}/{month}/{day}` Get the most viewed articles for a country.

## Unique devices data

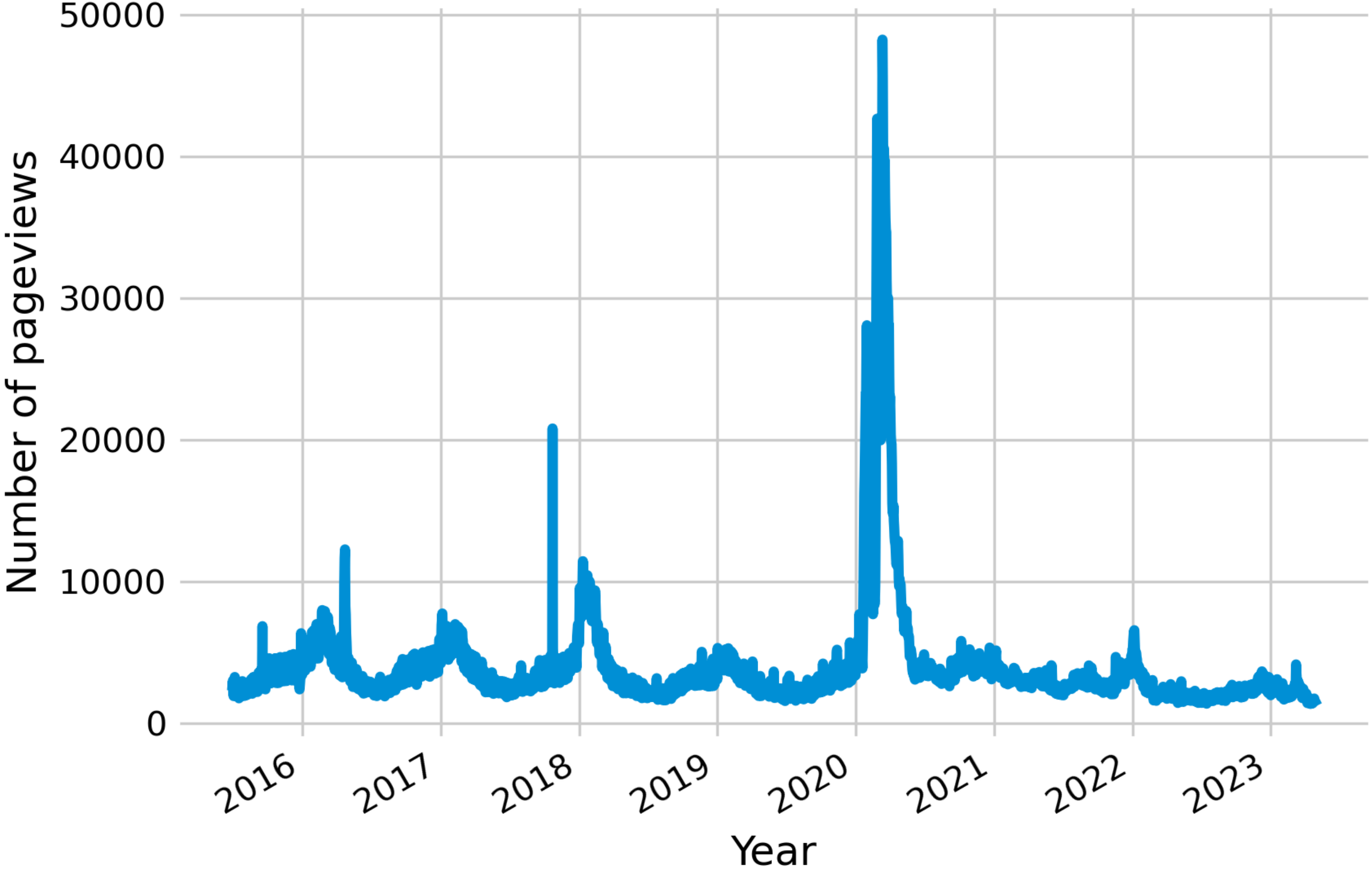
## Legacy data

## Edited pages data

## Editors data



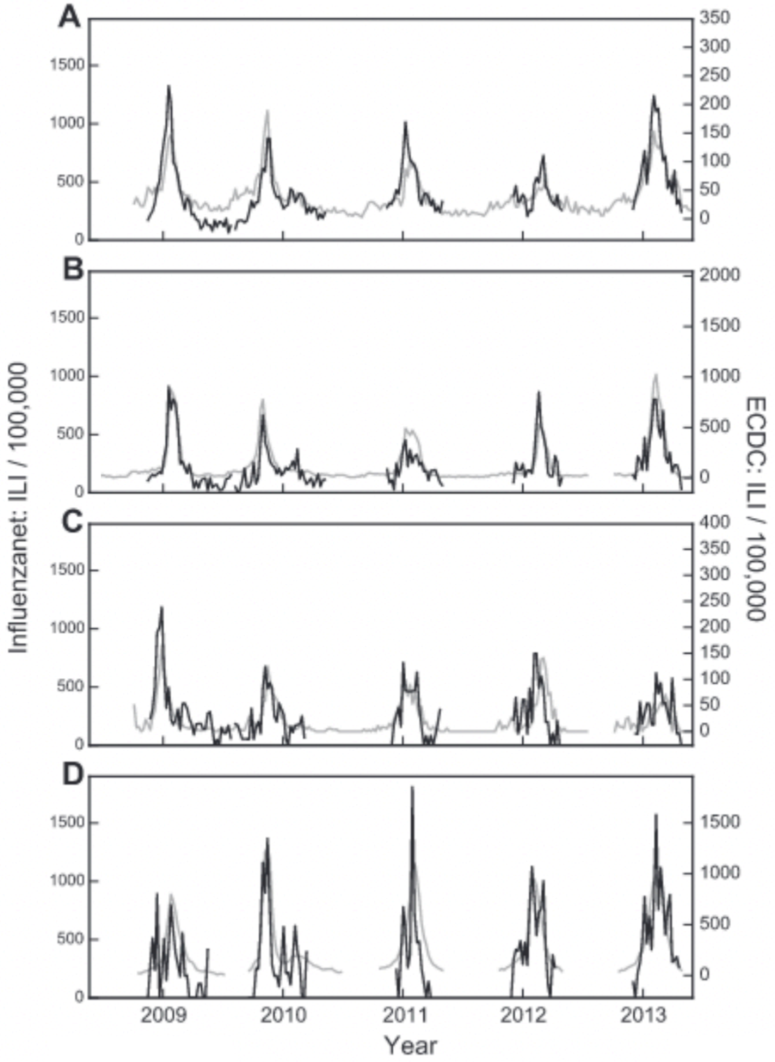
Source: Generous et al, 2014


Number of pageviews
50000
40000
30000
20000
10000
0
2016
2017
2018
2019
2020
2021
2022
2023
Year

# Digital Public Health Surveillance

## Participatory Surveillance

- Invite the population to share health data with public health
- Example: https://outbreaksnearme.org/
- Example: https://grippenet.ch/



- A)Netherlands,
- B) Belgium
- C) Portugal
- D) Italy

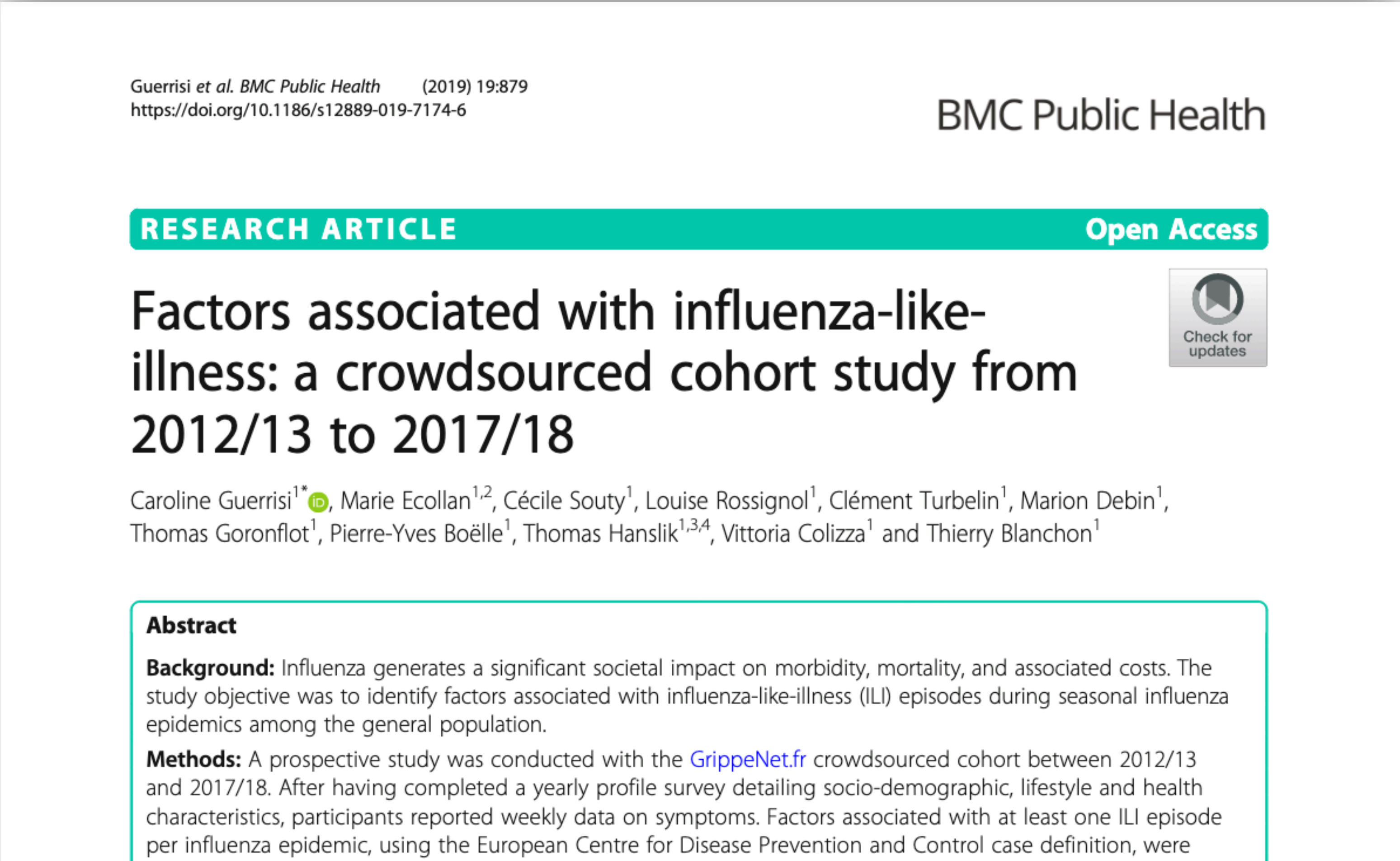
Source: Noort et al, 2015

# Digital Public Health Surveillance

## Participatory Surveillance

- Can ask about disease symptoms, but much more

- Not limited to humans



Guerrisi *et al. BMC Public Health* (2019) 19:879
https://doi.org/10.1186/s12889-019-7174-6

BMC Public Health

**RESEARCH ARTICLE** **Open Access**

# Factors associated with influenza-like-illness: a crowdsourced cohort study from 2012/13 to 2017/18

Caroline Guerrisi[1*], Marie Ecollan[1,2], Cécile Souty[1], Louise Rossignol[1], Clément Turbelin[1], Marion Debin[1], Thomas Goronflot[1], Pierre-Yves Boëlle[1], Thomas Hanslik[1,3,4], Vittoria Colizza[1] and Thierry Blanchon[1]

## Abstract

**Background:** Influenza generates a significant societal impact on morbidity, mortality, and associated costs. The study objective was to identify factors associated with influenza-like-illness (ILI) episodes during seasonal influenza epidemics among the general population.

**Methods:** A prospective study was conducted with the GrippeNet.fr crowdsourced cohort between 2012/13 and 2017/18. After having completed a yearly profile survey detailing socio-demographic, lifestyle and health characteristics, participants reported weekly data on symptoms. Factors associated with at least one ILI episode per influenza epidemic, using the European Centre for Disease Prevention and Control case definition, were

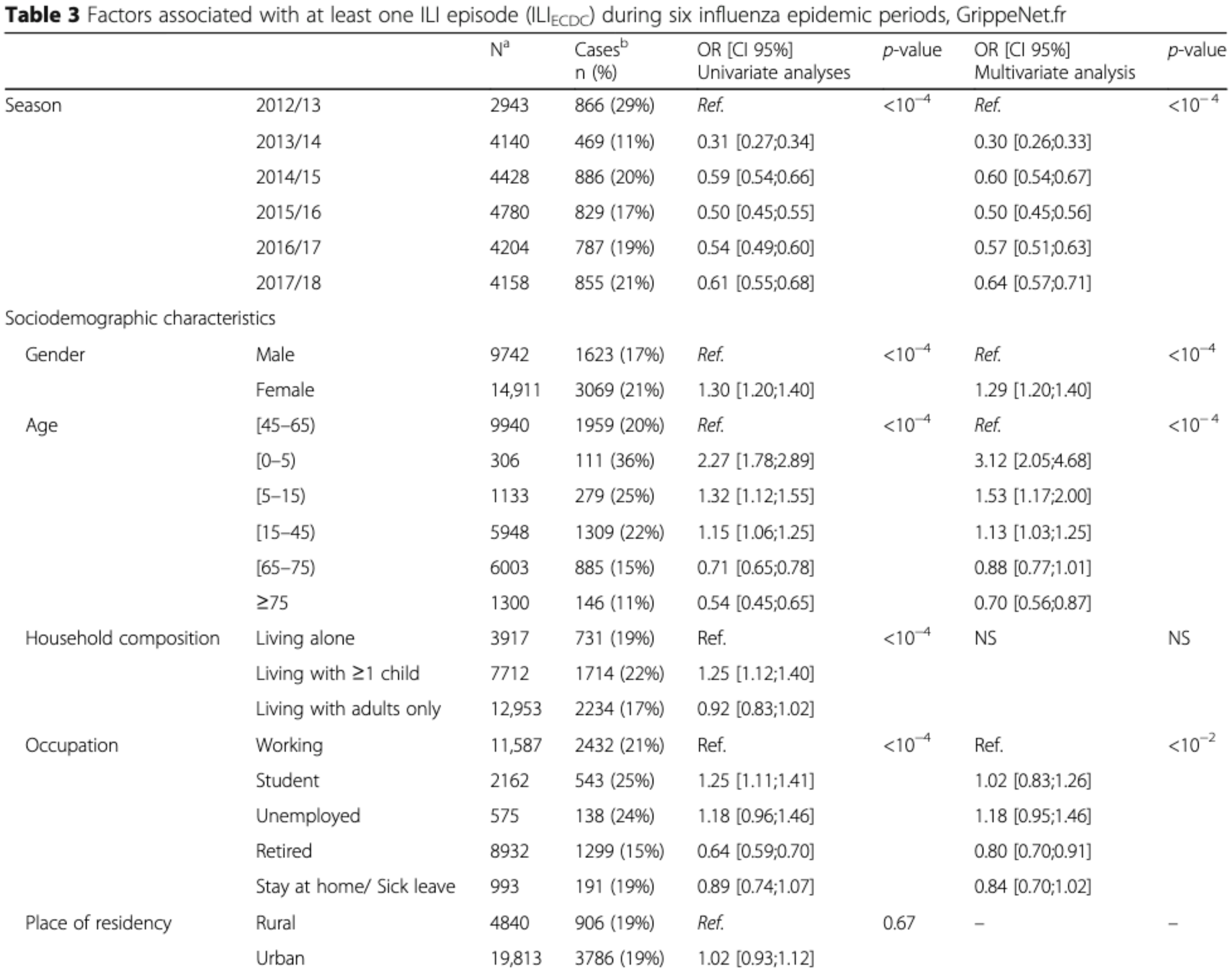
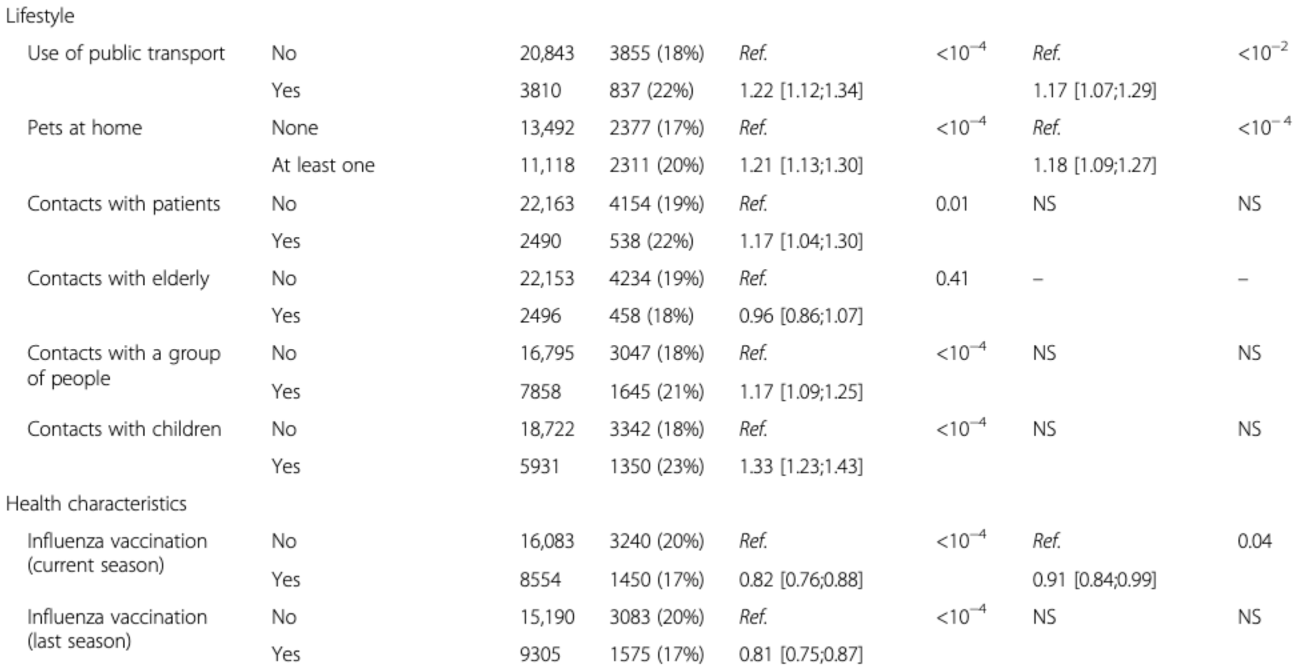


**Table 3** Factors associated with at least one ILI episode ($ILI_{ECDC}$) during six influenza epidemic periods, GrippeNet.fr

| | | N[a] | Cases[b] n (%) | OR [CI 95%] Univariate analyses | *p*-value | OR [CI 95%] Multivariate analysis | *p*-value |
|---|---|---|---|---|---|---|---|
| Season | 2012/13 | 2943 | 866 (29%) | *Ref.* | $<10^{-4}$ | *Ref.* | $<10^{-4}$ |
| | 2013/14 | 4140 | 469 (11%) | 0.31 [0.27;0.34] | | 0.30 [0.26;0.33] | |
| | 2014/15 | 4428 | 886 (20%) | 0.59 [0.54;0.66] | | 0.60 [0.54;0.67] | |
| | 2015/16 | 4780 | 829 (17%) | 0.50 [0.45;0.55] | | 0.50 [0.45;0.56] | |
| | 2016/17 | 4204 | 787 (19%) | 0.54 [0.49;0.60] | | 0.57 [0.51;0.63] | |
| | 2017/18 | 4158 | 855 (21%) | 0.61 [0.55;0.68] | | 0.64 [0.57;0.71] | |
| Sociodemographic characteristics | | | | | | | |
| Gender | Male | 9742 | 1623 (17%) | *Ref.* | $<10^{-4}$ | *Ref.* | $<10^{-4}$ |
| | Female | 14,911 | 3069 (21%) | 1.30 [1.20;1.40] | | 1.29 [1.20;1.40] | |
| Age | [45–65) | 9940 | 1959 (20%) | *Ref.* | $<10^{-4}$ | *Ref.* | $<10^{-4}$ |
| | [0–5) | 306 | 111 (36%) | 2.27 [1.78;2.89] | | 3.12 [2.05;4.68] | |
| | [5–15) | 1133 | 279 (25%) | 1.32 [1.12;1.55] | | 1.53 [1.17;2.00] | |
| | [15–45) | 5948 | 1309 (22%) | 1.15 [1.06;1.25] | | 1.13 [1.03;1.25] | |
| | [65–75) | 6003 | 885 (15%) | 0.71 [0.65;0.78] | | 0.88 [0.77;1.01] | |
| | ≥75 | 1300 | 146 (11%) | 0.54 [0.45;0.65] | | 0.70 [0.56;0.87] | |
| Household composition | Living alone | 3917 | 731 (19%) | Ref. | $<10^{-4}$ | NS | NS |
| | Living with ≥1 child | 7712 | 1714 (22%) | 1.25 [1.12;1.40] | | | |
| | Living with adults only | 12,953 | 2234 (17%) | 0.92 [0.83;1.02] | | | |
| Occupation | Working | 11,587 | 2432 (21%) | Ref. | $<10^{-4}$ | Ref. | $<10^{-2}$ |
| | Student | 2162 | 543 (25%) | 1.25 [1.11;1.41] | | 1.02 [0.83;1.26] | |
| | Unemployed | 575 | 138 (24%) | 1.18 [0.96;1.46] | | 1.18 [0.95;1.46] | |
| | Retired | 8932 | 1299 (15%) | 0.64 [0.59;0.70] | | 0.80 [0.70;0.91] | |
| | Stay at home/ Sick leave | 993 | 191 (19%) | 0.89 [0.74;1.07] | | 0.84 [0.70;1.02] | |
| Place of residency | Rural | 4840 | 906 (19%) | *Ref.* | 0.67 | – | – |
| | Urban | 19,813 | 3786 (19%) | 1.02 [0.93;1.12] | | | |



| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Lifestyle | | | | | | | |
| Use of public transport | No | 20,843 | 3855 (18%) | *Ref.* | $<10^{-4}$ | *Ref.* | $<10^{-2}$ |
| | Yes | 3810 | 837 (22%) | 1.22 [1.12;1.34] | | 1.17 [1.07;1.29] | |
| Pets at home | None | 13,492 | 2377 (17%) | *Ref.* | $<10^{-4}$ | *Ref.* | $<10^{-4}$ |
| | At least one | 11,118 | 2311 (20%) | 1.21 [1.13;1.30] | | 1.18 [1.09;1.27] | |
| Contacts with patients | No | 22,163 | 4154 (19%) | *Ref.* | 0.01 | NS | NS |
| | Yes | 2490 | 538 (22%) | 1.17 [1.04;1.30] | | | |
| Contacts with elderly | No | 22,153 | 4234 (19%) | *Ref.* | 0.41 | – | – |
| | Yes | 2496 | 458 (18%) | 0.96 [0.86;1.07] | | | |
| Contacts with a group of people | No | 16,795 | 3047 (18%) | *Ref.* | $<10^{-4}$ | NS | NS |
| | Yes | 7858 | 1645 (21%) | 1.17 [1.09;1.25] | | | |
| Contacts with children | No | 18,722 | 3342 (18%) | *Ref.* | $<10^{-4}$ | NS | NS |
| | Yes | 5931 | 1350 (23%) | 1.33 [1.23;1.43] | | | |
| Health characteristics | | | | | | | |
| Influenza vaccination (current season) | No | 16,083 | 3240 (20%) | *Ref.* | $<10^{-4}$ | *Ref.* | 0.04 |
| | Yes | 8554 | 1450 (17%) | 0.82 [0.76;0.88] | | 0.91 [0.84;0.99] | |
| Influenza vaccination (last season) | No | 15,190 | 3083 (20%) | *Ref.* | $<10^{-4}$ | NS | NS |
| | Yes | 9305 | 1575 (17%) | 0.81 [0.75;0.87] | | | |

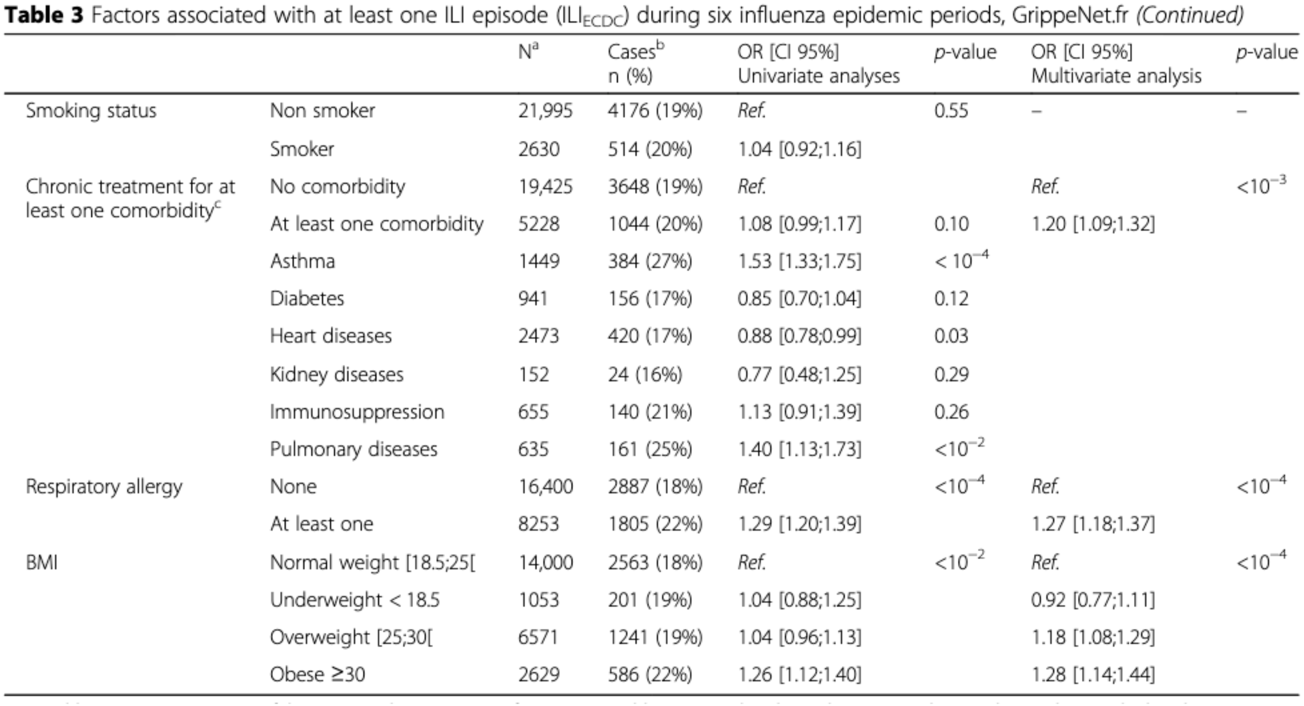
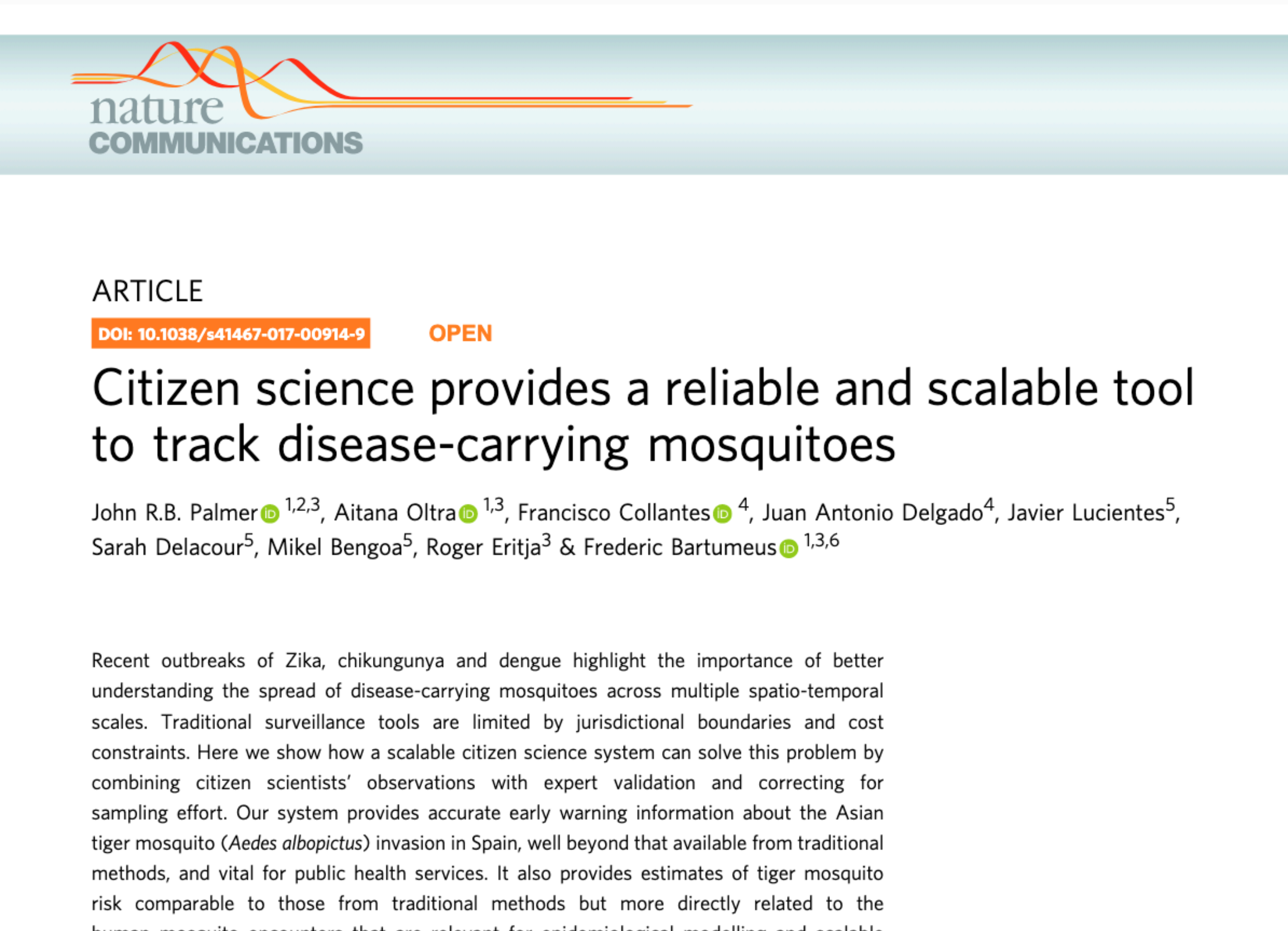


**Table 3** Factors associated with at least one ILI episode (ILI$_{ECDC}$) during six influenza epidemic periods, GrippeNet.fr *(Continued)*

| | | N[a] | Cases[b]<br>n (%) | OR [CI 95%]<br>Univariate analyses | *p*-value | OR [CI 95%]<br>Multivariate analysis | *p*-value |
|---|---|---|---|---|---|---|---|
| Smoking status | Non smoker | 21,995 | 4176 (19%) | *Ref.* | 0.55 | – | – |
| | Smoker | 2630 | 514 (20%) | 1.04 [0.92;1.16] | | | |
| Chronic treatment for at least one comorbidity[c] | No comorbidity | 19,425 | 3648 (19%) | *Ref.* | | *Ref.* | $<10^{-3}$ |
| | At least one comorbidity | 5228 | 1044 (20%) | 1.08 [0.99;1.17] | 0.10 | 1.20 [1.09;1.32] | |
| | Asthma | 1449 | 384 (27%) | 1.53 [1.33;1.75] | $< 10^{-4}$ | | |
| | Diabetes | 941 | 156 (17%) | 0.85 [0.70;1.04] | 0.12 | | |
| | Heart diseases | 2473 | 420 (17%) | 0.88 [0.78;0.99] | 0.03 | | |
| | Kidney diseases | 152 | 24 (16%) | 0.77 [0.48;1.25] | 0.29 | | |
| | Immunosuppression | 655 | 140 (21%) | 1.13 [0.91;1.39] | 0.26 | | |
| | Pulmonary diseases | 635 | 161 (25%) | 1.40 [1.13;1.73] | $<10^{-2}$ | | |
| Respiratory allergy | None | 16,400 | 2887 (18%) | *Ref.* | $<10^{-4}$ | *Ref.* | $<10^{-4}$ |
| | At least one | 8253 | 1805 (22%) | 1.29 [1.20;1.39] | | 1.27 [1.18;1.37] | |
| BMI | Normal weight [18.5;25[ | 14,000 | 2563 (18%) | *Ref.* | $<10^{-2}$ | *Ref.* | $<10^{-4}$ |
| | Underweight < 18.5 | 1053 | 201 (19%) | 1.04 [0.88;1.25] | | 0.92 [0.77;1.11] | |
| | Overweight [25;30[ | 6571 | 1241 (19%) | 1.04 [0.96;1.13] | | 1.18 [1.08;1.29] | |
| | Obese ≥30 | 2629 | 586 (22%) | 1.26 [1.12;1.40] | | 1.28 [1.14;1.44] | |



nature COMMUNICATIONS

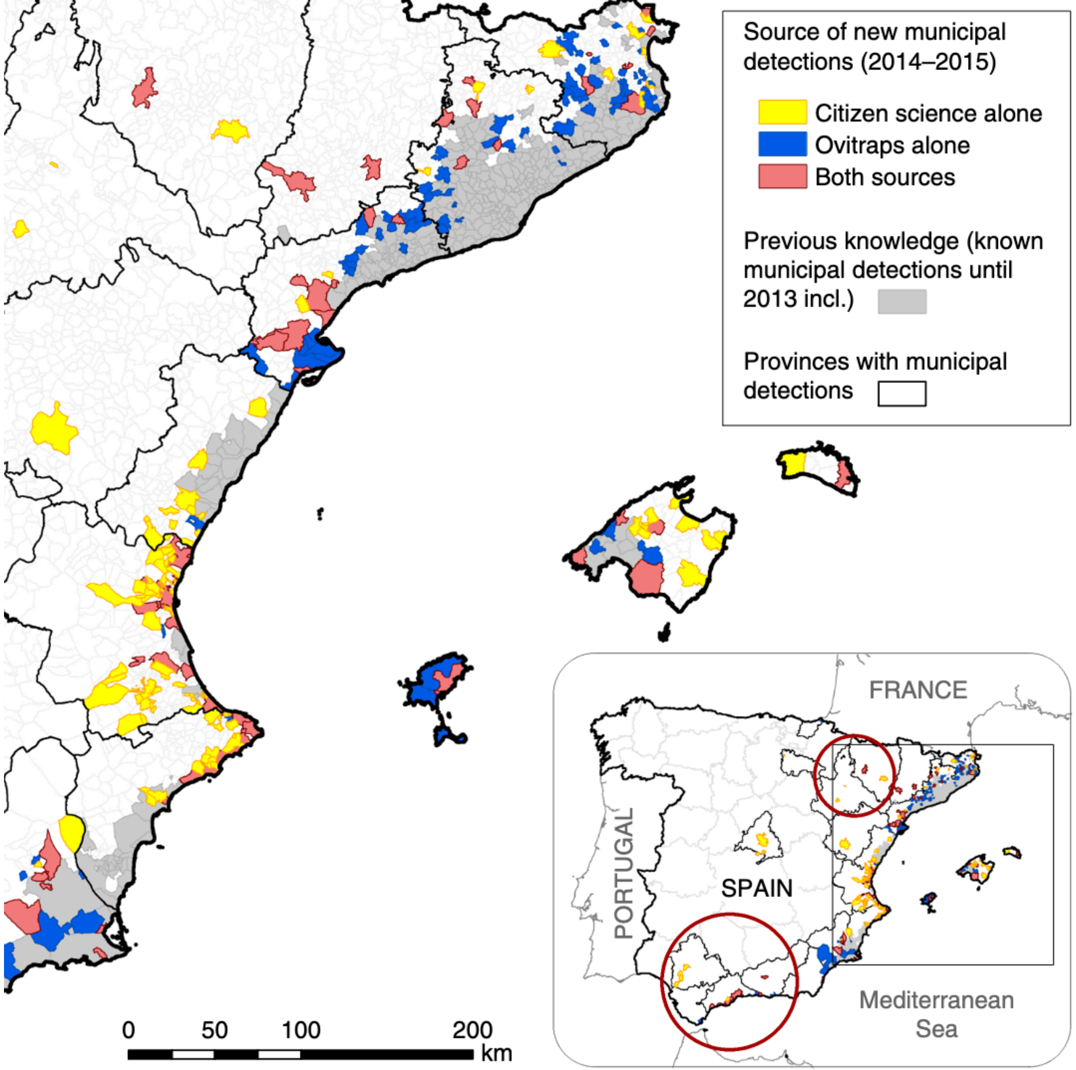
ARTICLE

DOI: 10.1038/s41467-017-00914-9 OPEN

# Citizen science provides a reliable and scalable tool to track disease-carrying mosquitoes

John R.B. Palmer[1,2,3], Aitana Oltra[1,3], Francisco Collantes[4], Juan Antonio Delgado[4], Javier Lucientes[5], Sarah Delacour[5], Mikel Bengoa[5], Roger Eritja[3] & Frederic Bartumeus[1,3,6]

Recent outbreaks of Zika, chikungunya and dengue highlight the importance of better understanding the spread of disease-carrying mosquitoes across multiple spatio-temporal scales. Traditional surveillance tools are limited by jurisdictional boundaries and cost constraints. Here we show how a scalable citizen science system can solve this problem by combining citizen scientists' observations with expert validation and correcting for sampling effort. Our system provides accurate early warning information about the Asian tiger mosquito (*Aedes albopictus*) invasion in Spain, well beyond that available from traditional methods, and vital for public health services. It also provides estimates of tiger mosquito risk comparable to those from traditional methods but more directly related to the

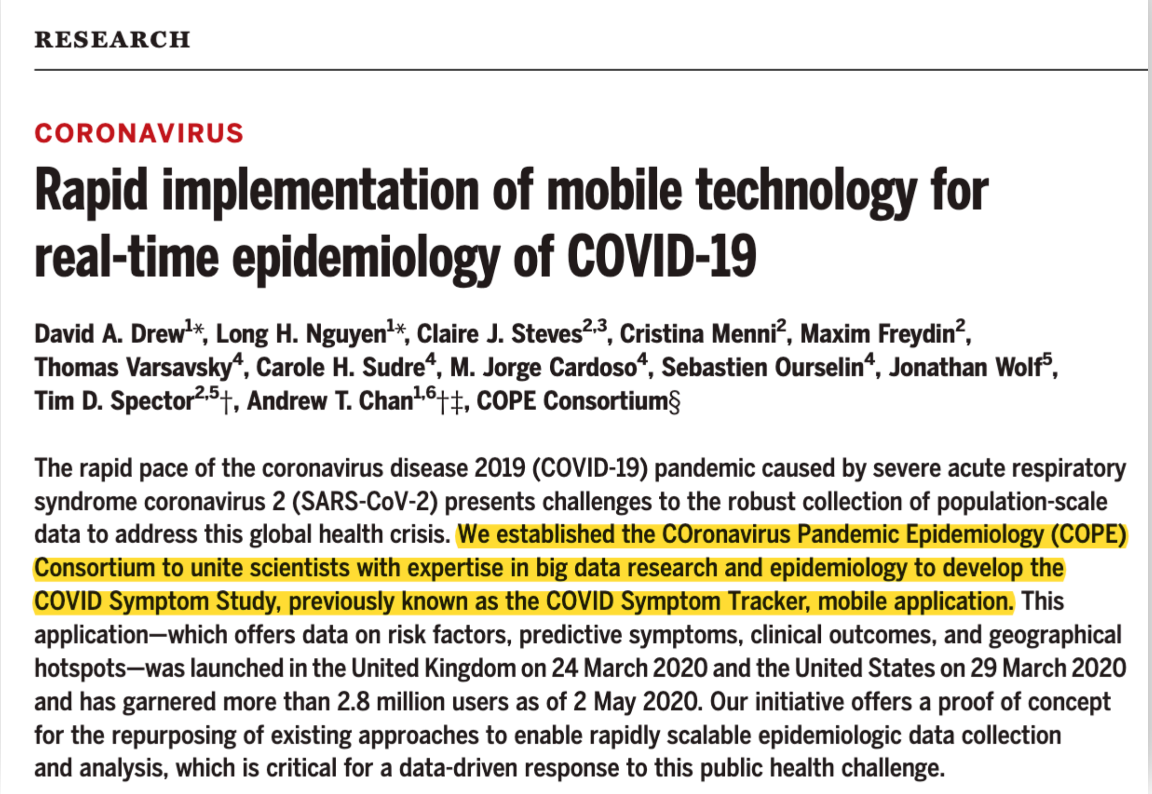

Source of new municipal detections (2014–2015)
Citizen science alone
Ovitraps alone
Both sources
Previous knowledge (known municipal detections until 2013 incl.)
Provinces with municipal detections
0
50
100
200
km
FRANCE
PORTUGAL
SPAIN
Mediterranean Sea

# Digital Public Health Surveillance

## Participatory Surveillance

- For a long time, participatory surveillance systems were browser-based.
- Recent years have seen the emergence of mobile apps for participatory surveillance.
- During COVID, some (non-infectious disease) apps were repurposed for COVID. Example: ZOE

# Digital Public Health Surveillance
## Participatory Surveillance



RESEARCH

CORONAVIRUS

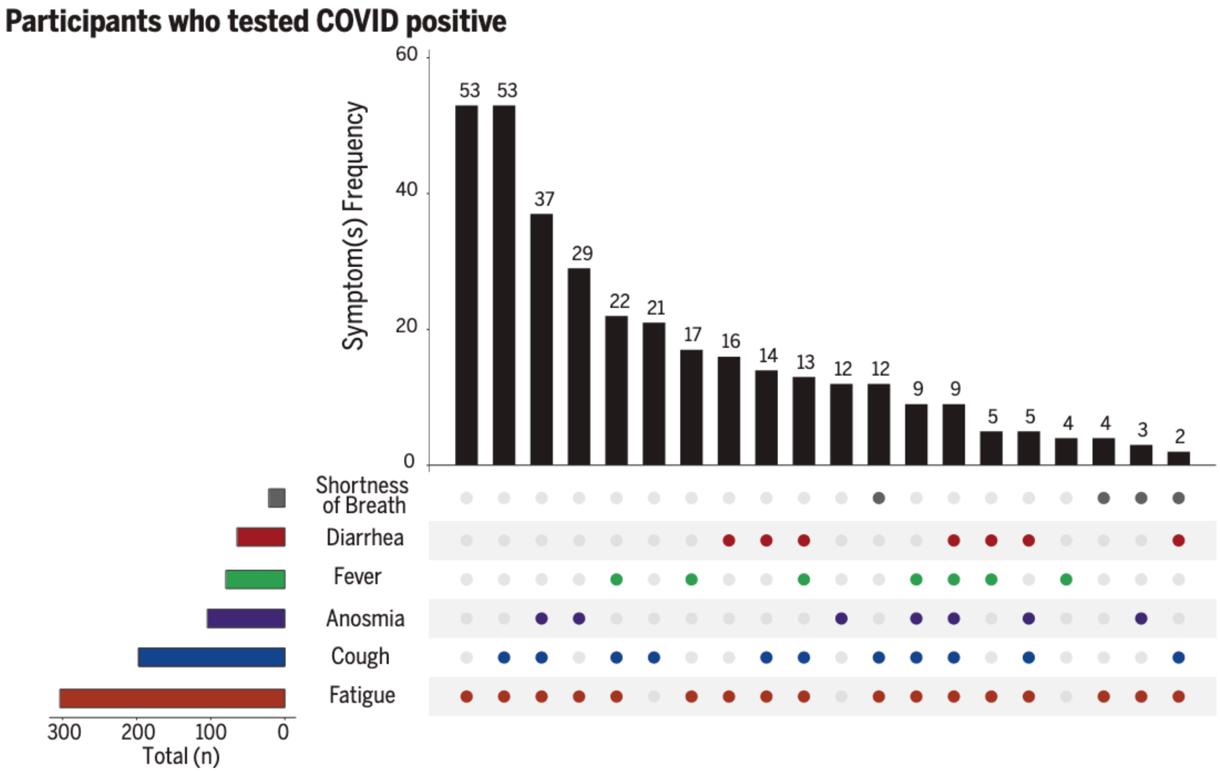
### Rapid implementation of mobile technology for real-time epidemiology of COVID-19

David A. Drew[1]*, Long H. Nguyen[1]*, Claire J. Steves[2,3], Cristina Menni[2], Maxim Freydin[2], Thomas Varsavsky[4], Carole H. Sudre[4], M. Jorge Cardoso[4], Sebastien Ourselin[4], Jonathan Wolf[5], Tim D. Spector[2,5]†, Andrew T. Chan[1,6]†‡, COPE Consortium§

The rapid pace of the coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) presents challenges to the robust collection of population-scale data to address this global health crisis. We established the COronavirus Pandemic Epidemiology (COPE) Consortium to unite scientists with expertise in big data research and epidemiology to develop the COVID Symptom Study, previously known as the COVID Symptom Tracker, mobile application. This application—which offers data on risk factors, predictive symptoms, clinical outcomes, and geographical hotspots—was launched in the United Kingdom on 24 March 2020 and the United States on 29 March 2020 and has garnered more than 2.8 million users as of 2 May 2020. Our initiative offers a proof of concept for the repurposing of existing approaches to enable rapidly scalable epidemiologic data collection and analysis, which is critical for a data-driven response to this public health challenge.

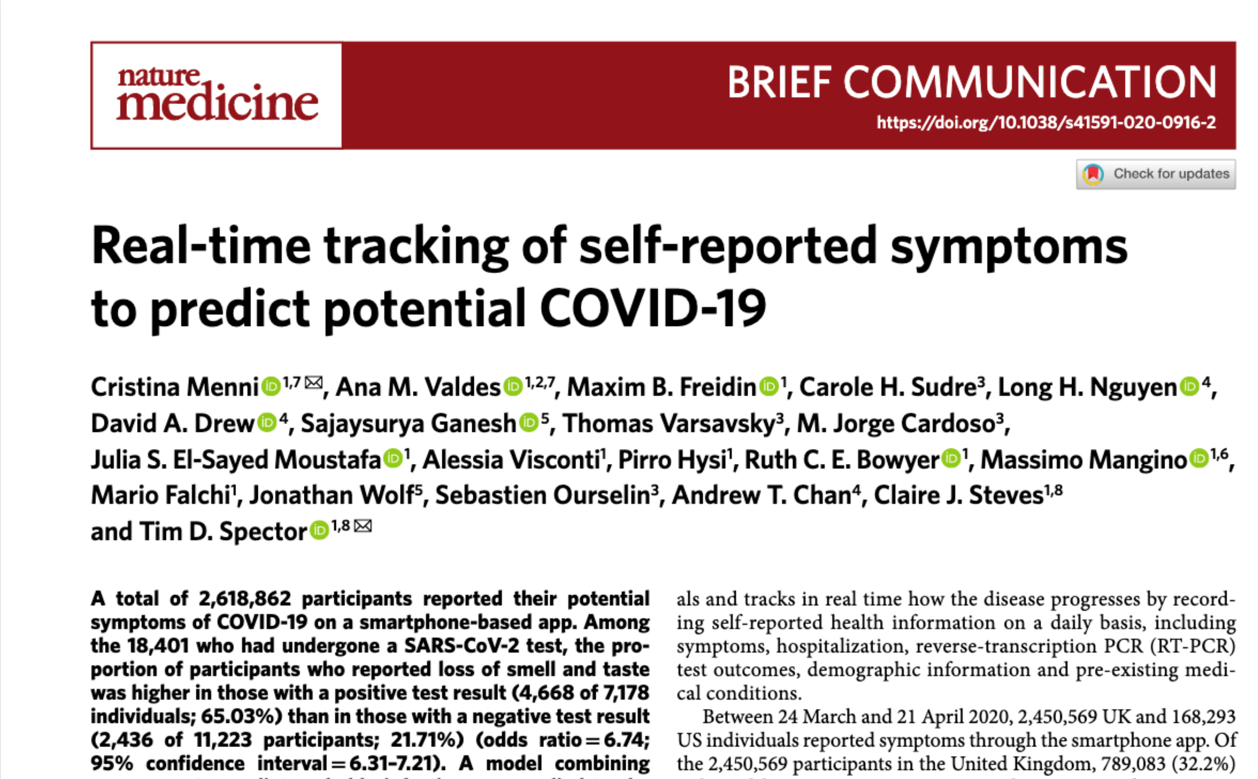
# Digital Public Health Surveillance

## Participatory Surveillance



# Digital Public Health Surveillance

## Participatory Surveillance



nature medicine

BRIEF COMMUNICATION

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

Check for updates

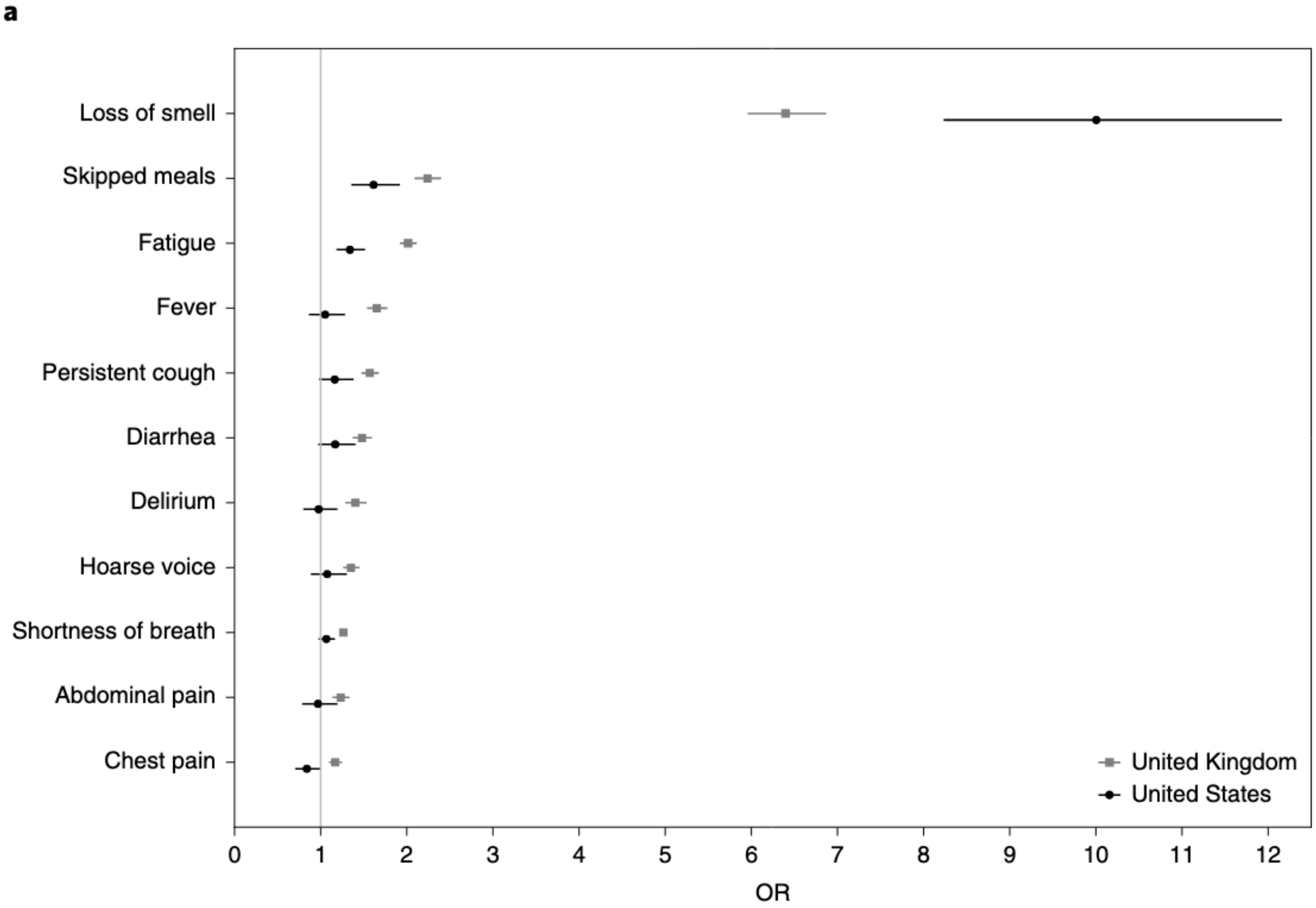
# Real-time tracking of self-reported symptoms to predict potential COVID-19

Cristina Menni[1,7] ✉, Ana M. Valdes[1,2,7], Maxim B. Freidin[1], Carole H. Sudre[3], Long H. Nguyen[4], David A. Drew[4], Sajaysurya Ganesh[5], Thomas Varsavsky[3], M. Jorge Cardoso[3], Julia S. El-Sayed Moustafa[1], Alessia Visconti[1], Pirro Hysi[1], Ruth C. E. Bowyer[1], Massimo Mangino[1,6], Mario Falchi[1], Jonathan Wolf[5], Sebastien Ourselin[3], Andrew T. Chan[4], Claire J. Steves[1,8] and Tim D. Spector[1,8] ✉

**A total of 2,618,862 participants reported their potential symptoms of COVID-19 on a smartphone-based app. Among the 18,401 who had undergone a SARS-CoV-2 test, the proportion of participants who reported loss of smell and taste was higher in those with a positive test result (4,668 of 7,178 individuals; 65.03%) than in those with a negative test result (2,436 of 11,223 participants; 21.71%) (odds ratio = 6.74; 95% confidence interval = 6.31–7.21). A model combining**

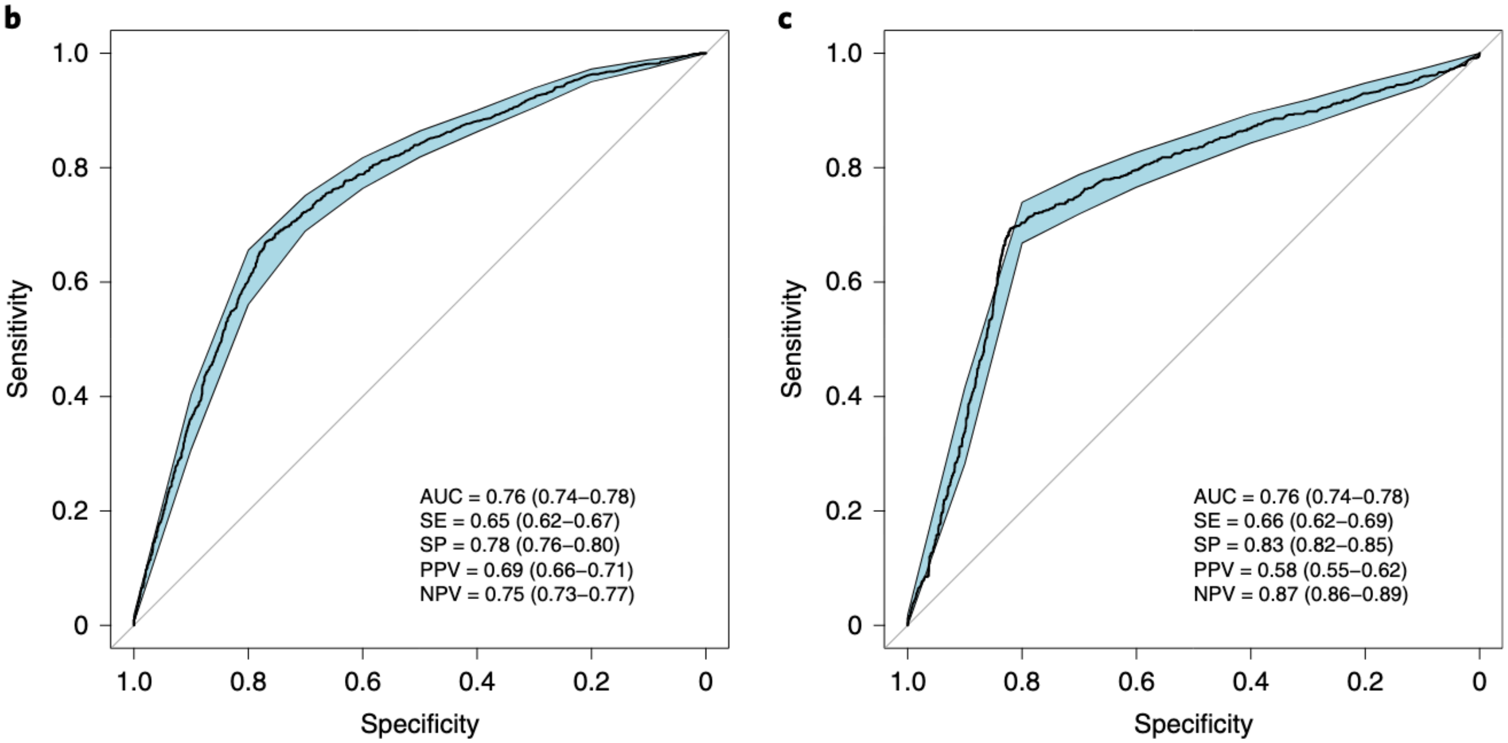
als and tracks in real time how the disease progresses by recording self-reported health information on a daily basis, including symptoms, hospitalization, reverse-transcription PCR (RT-PCR) test outcomes, demographic information and pre-existing medical conditions.

Between 24 March and 21 April 2020, 2,450,569 UK and 168,293 US individuals reported symptoms through the smartphone app. Of the 2,450,569 participants in the United Kingdom, 789,083 (32.2%)


a
Loss of smell
Skipped meals
Fatigue
Fever
Persistent cough
Diarrhea
Delirium
Hoarse voice
Shortness of breath
Abdominal pain
Chest pain
United Kingdom
United States
0
1
2
3
4
5
6
7
8
9
10
11
12
OR

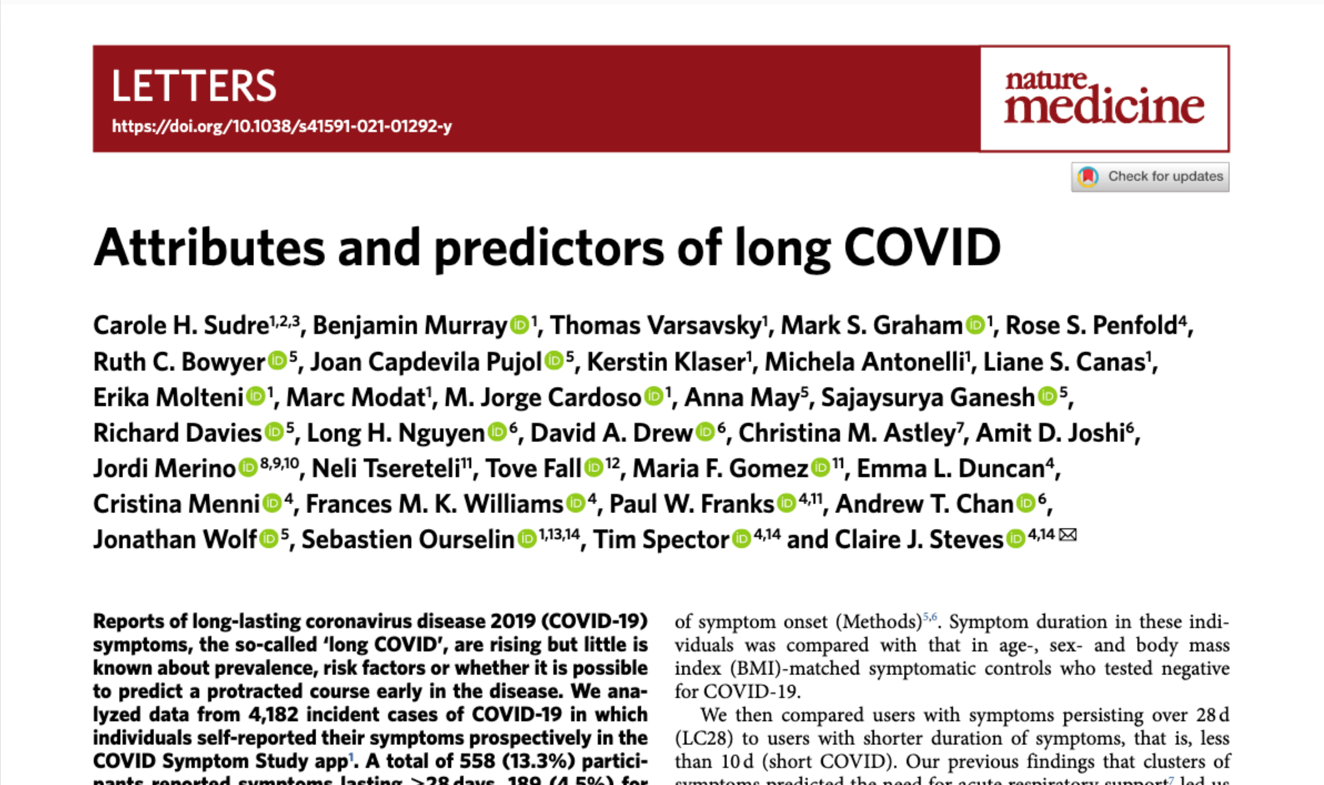
# Digital Public Health Surveillance

## Participatory Surveillance



# Digital Public Health Surveillance

## Participatory Surveillance



LETTERS
https://doi.org/10.1038/s41591-021-01292-y
nature medicine

Check for updates

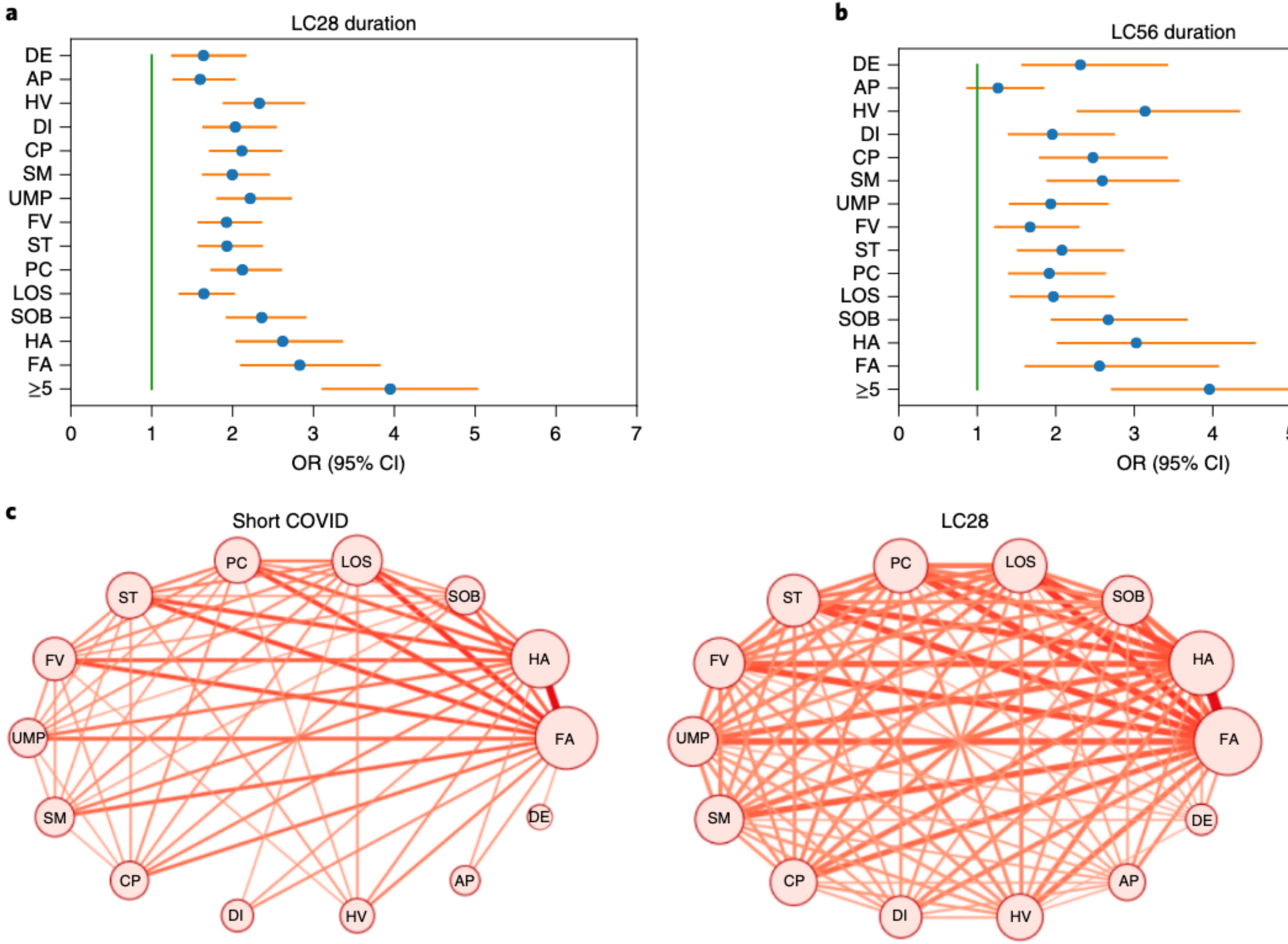
# Attributes and predictors of long COVID

Carole H. Sudre[1,2,3], Benjamin Murray[1], Thomas Varsavsky[1], Mark S. Graham[1], Rose S. Penfold[4], Ruth C. Bowyer[5], Joan Capdevila Pujol[5], Kerstin Klaser[1], Michela Antonelli[1], Liane S. Canas[1], Erika Molteni[1], Marc Modat[1], M. Jorge Cardoso[1], Anna May[5], Sajaysurya Ganesh[5], Richard Davies[5], Long H. Nguyen[6], David A. Drew[6], Christina M. Astley[7], Amit D. Joshi[6], Jordi Merino[8,9,10], Neli Tsereteli[11], Tove Fall[12], Maria F. Gomez[11], Emma L. Duncan[4], Cristina Menni[4], Frances M. K. Williams[4], Paul W. Franks[4,11], Andrew T. Chan[6], Jonathan Wolf[5], Sebastien Ourselin[1,13,14], Tim Spector[4,14] and Claire J. Steves[4,14] ✉

**Reports of long-lasting coronavirus disease 2019 (COVID-19) symptoms, the so-called 'long COVID', are rising but little is known about prevalence, risk factors or whether it is possible to predict a protracted course early in the disease. We analyzed data from 4,182 incident cases of COVID-19 in which individuals self-reported their symptoms prospectively in the COVID Symptom Study app[1]. A total of 558 (13.3%) participants reported symptoms lasting ≥28 days, 189 (4.5%) for**

of symptom onset (Methods)[5,6]. Symptom duration in these individuals was compared with that in age-, sex- and body mass index (BMI)-matched symptomatic controls who tested negative for COVID-19.

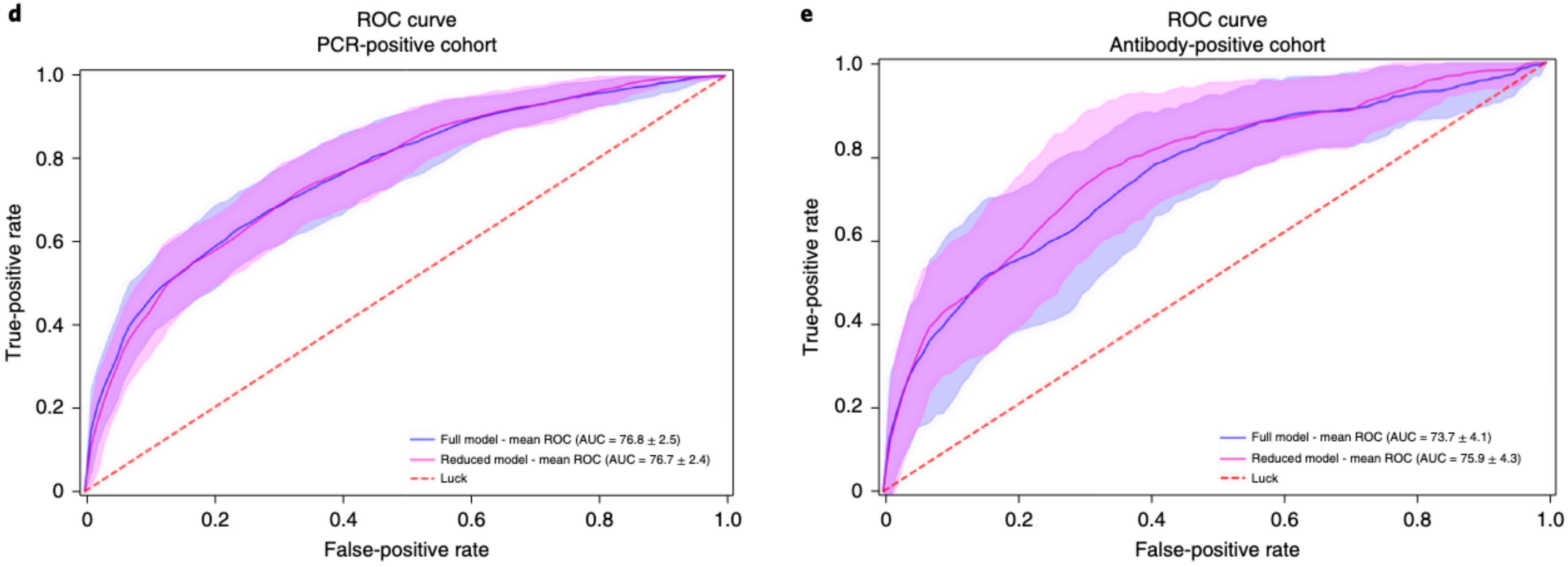
We then compared users with symptoms persisting over 28 d (LC28) to users with shorter duration of symptoms, that is, less than 10 d (short COVID). Our previous findings that clusters of symptoms predicted the need for acute respiratory support[7] led us



FA – fatigue
HA – headache
SOB – shortness of breath
LOS – loss of smell
PC – persistent cough
ST – sore throat
FV – fever
UMP – unusual muscle pain
SM – skipped meals
CP – chest pain
DI – diarrhea
HV – hoarse voice
AP – abdominal pain
DE – delirium

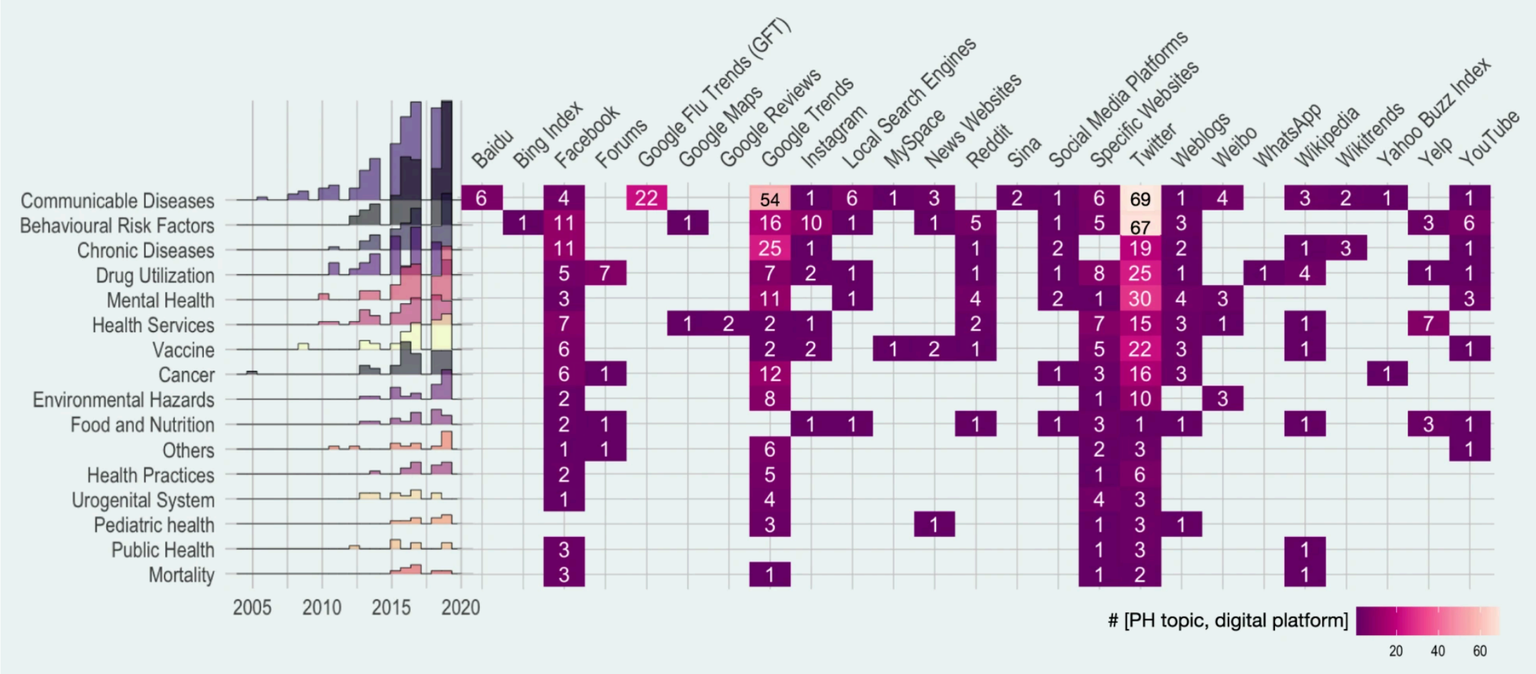
# Digital Public Health Surveillance

## Participatory Surveillance



# Digital Public Health Surveillance

## Social Media



| | Baidu | Bing Index | Facebook | Forums | Google Flu Trends (GFT) | Google Maps | Google Reviews | Google Trends | Instagram | Local Search Engines | MySpace | News Websites | Reddit | Sina | Social Media Platforms | Specific Websites | Twitter | Weblogs | Weibo | WhatsApp | Wikipedia | Wikitrends | Yahoo Buzz Index | Yelp | YouTube |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Communicable Diseases | 6 | | 4 | | 22 | | | 54 | 1 | 6 | 1 | 3 | | 2 | 1 | 6 | 69 | 1 | 4 | | 3 | 2 | 1 | | 1 |
| Behavioural Risk Factors | | 1 | 11 | | | 1 | | 16 | 10 | 1 | | 1 | 5 | | 1 | 5 | 67 | 3 | | | | | | 3 | 6 |
| Chronic Diseases | | | 11 | | | | | 25 | 1 | | | | 1 | | 2 | | 19 | 2 | | | 1 | 3 | | | 1 |
| Drug Utilization | | | 5 | 7 | | | | 7 | 2 | 1 | | | 1 | | 1 | 8 | 25 | 1 | | 1 | 4 | | | 1 | 1 |
| Mental Health | | | 3 | | | | | 11 | | 1 | | | 4 | | 2 | 1 | 30 | 4 | 3 | | | | | | 3 |
| Health Services | | | 7 | | | 1 | 2 | 2 | 1 | | | | 2 | | | 7 | 15 | 3 | 1 | | 1 | | | 7 | |
| Vaccine | | | 6 | | | | | 2 | 2 | | 1 | 2 | 1 | | | 5 | 22 | 3 | | | 1 | | | | 1 |
| Cancer | | | 6 | 1 | | | | 12 | | | | | | | 1 | 3 | 16 | 3 | | | | | 1 | | |
| Environmental Hazards | | | 2 | | | | | 8 | | | | | | | | 1 | 10 | | 3 | | | | | | |
| Food and Nutrition | | | 2 | 1 | | | | | 1 | 1 | | | 1 | | 1 | 3 | 1 | 1 | | | 1 | | | 3 | 1 |
| Others | | | 1 | 1 | | | | 6 | | | | | | | | 2 | 3 | | | | | | | | 1 |
| Health Practices | | | 2 | | | | | 5 | | | | | | | | 1 | 6 | | | | | | | | |
| Urogenital System | | | 1 | | | | | 4 | | | | | | | | 4 | 3 | | | | | | | | |
| Pediatric health | | | | | | | | 3 | | | | 1 | | | | 1 | 3 | 1 | | | | | | | |
| Public Health | | | 3 | | | | | | | | | | | | | 1 | 3 | | | | 1 | | | | |
| Mortality | | | 3 | | | | | 1 | | | | | | | | 1 | 2 | | | | 1 | | | | |

# Digital Public Health Surveillance
## Social Media

- Benefits of social media for DPHS:
  - content analysis
  - network
  - "ecological" assessment

- Twitter API freely accessible to researchers until early 2023.

# Digital Public Health Surveillance

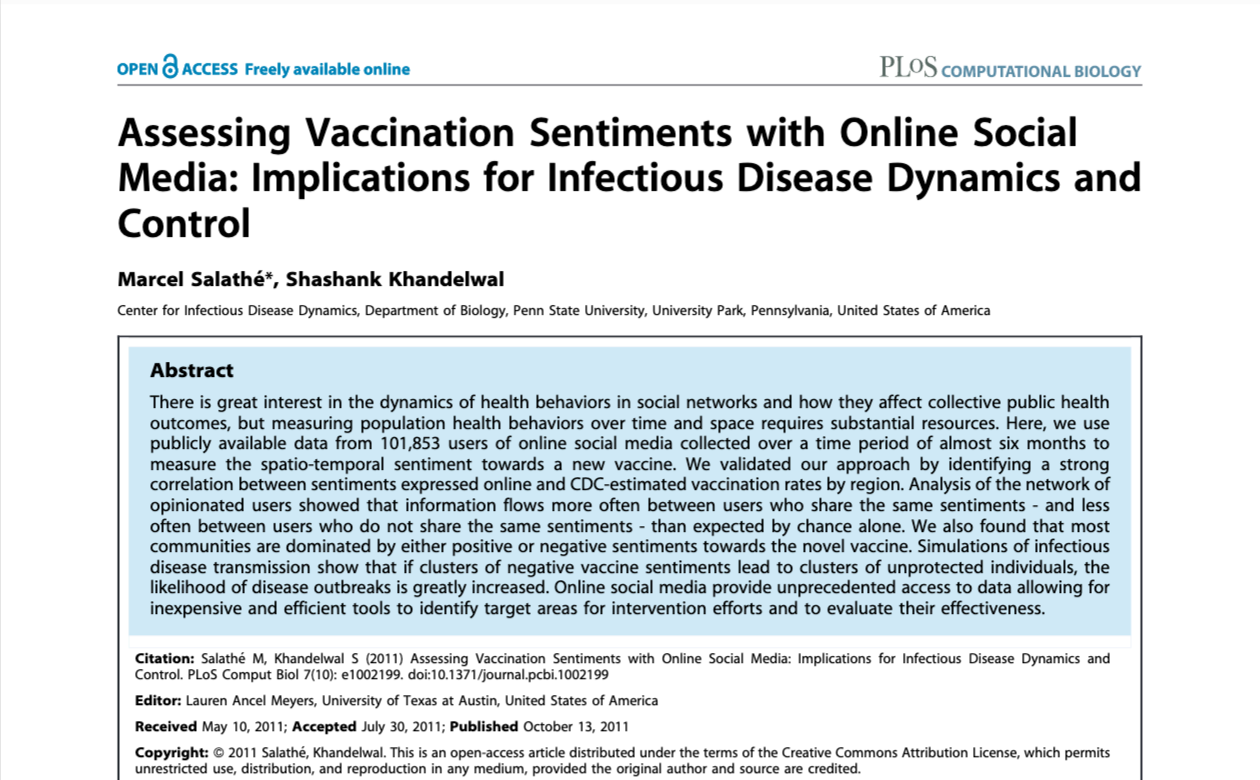
## Social Media

- Where will vaccine demand be lowest / highest
- 1. Collect Twitter data (define query)
  2. Analyze part of tweets and label them
  3. Train an ML model and apply to remaining data
  4. Analyze results

  (interesting if possible - connected follower relationships)

# Digital Public Health Surveillance

## Social Media



OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

### Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control

**Marcel Salathé*, Shashank Khandelwal**

Center for Infectious Disease Dynamics, Department of Biology, Penn State University, University Park, Pennsylvania, United States of America

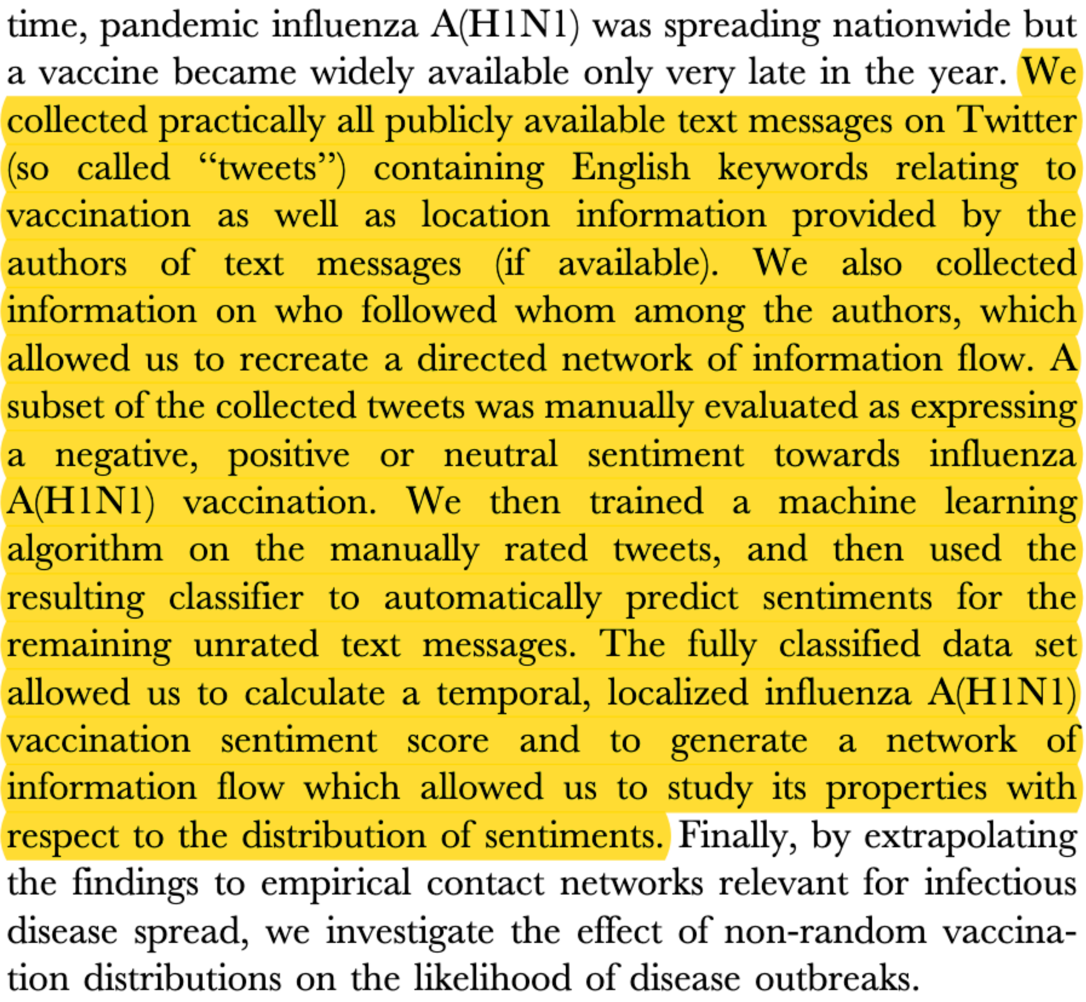
**Abstract**

There is great interest in the dynamics of health behaviors in social networks and how they affect collective public health outcomes, but measuring population health behaviors over time and space requires substantial resources. Here, we use publicly available data from 101,853 users of online social media collected over a time period of almost six months to measure the spatio-temporal sentiment towards a new vaccine. We validated our approach by identifying a strong correlation between sentiments expressed online and CDC-estimated vaccination rates by region. Analysis of the network of opinionated users showed that information flows more often between users who share the same sentiments - and less often between users who do not share the same sentiments - than expected by chance alone. We also found that most communities are dominated by either positive or negative sentiments towards the novel vaccine. Simulations of infectious disease transmission show that if clusters of negative vaccine sentiments lead to clusters of unprotected individuals, the likelihood of disease outbreaks is greatly increased. Online social media provide unprecedented access to data allowing for inexpensive and efficient tools to identify target areas for intervention efforts and to evaluate their effectiveness.

**Citation:** Salathé M, Khandelwal S (2011) Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. PLoS Comput Biol 7(10): e1002199. doi:10.1371/journal.pcbi.1002199

**Editor:** Lauren Ancel Meyers, University of Texas at Austin, United States of America

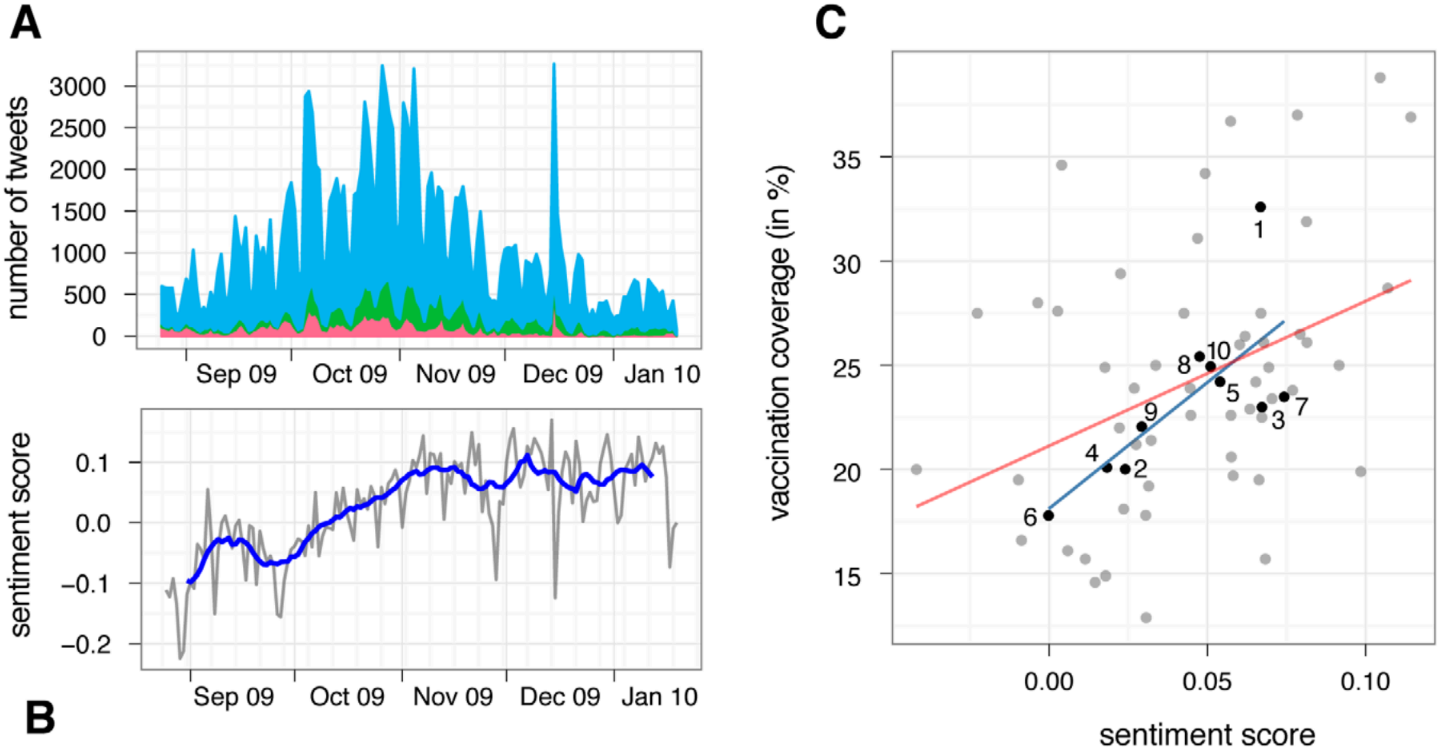
**Received** May 10, 2011; **Accepted** July 30, 2011; **Published** October 13, 2011



time, pandemic influenza A(H1N1) was spreading nationwide but a vaccine became widely available only very late in the year. We collected practically all publicly available text messages on Twitter (so called "tweets") containing English keywords relating to vaccination as well as location information provided by the authors of text messages (if available). We also collected information on who followed whom among the authors, which allowed us to recreate a directed network of information flow. A subset of the collected tweets was manually evaluated as expressing a negative, positive or neutral sentiment towards influenza A(H1N1) vaccination. We then trained a machine learning algorithm on the manually rated tweets, and then used the resulting classifier to automatically predict sentiments for the remaining unrated text messages. The fully classified data set allowed us to calculate a temporal, localized influenza A(H1N1) vaccination sentiment score and to generate a network of information flow which allowed us to study its properties with respect to the distribution of sentiments. Finally, by extrapolating the findings to empirical contact networks relevant for infectious disease spread, we investigate the effect of non-random vaccination distributions on the likelihood of disease outbreaks.

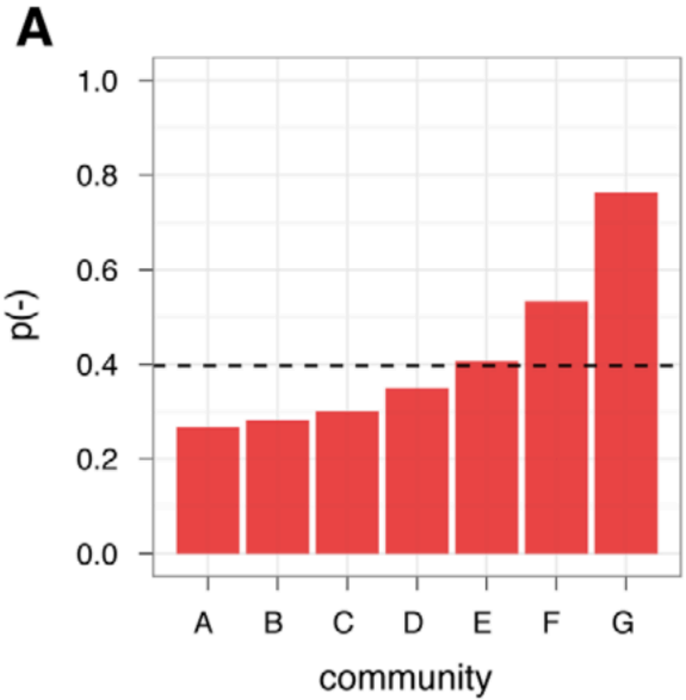
# Digital Public Health Surveillance

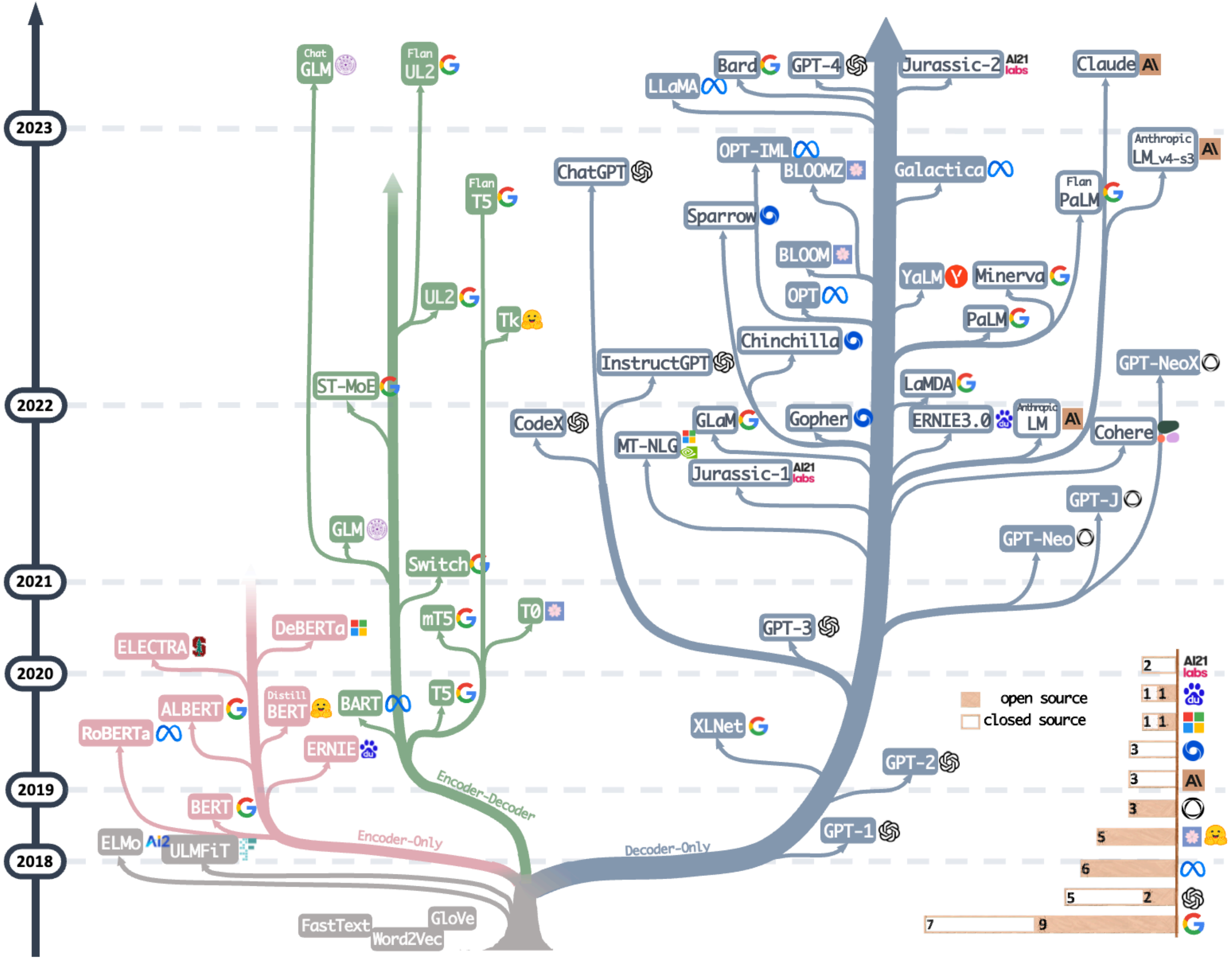
## Social Media



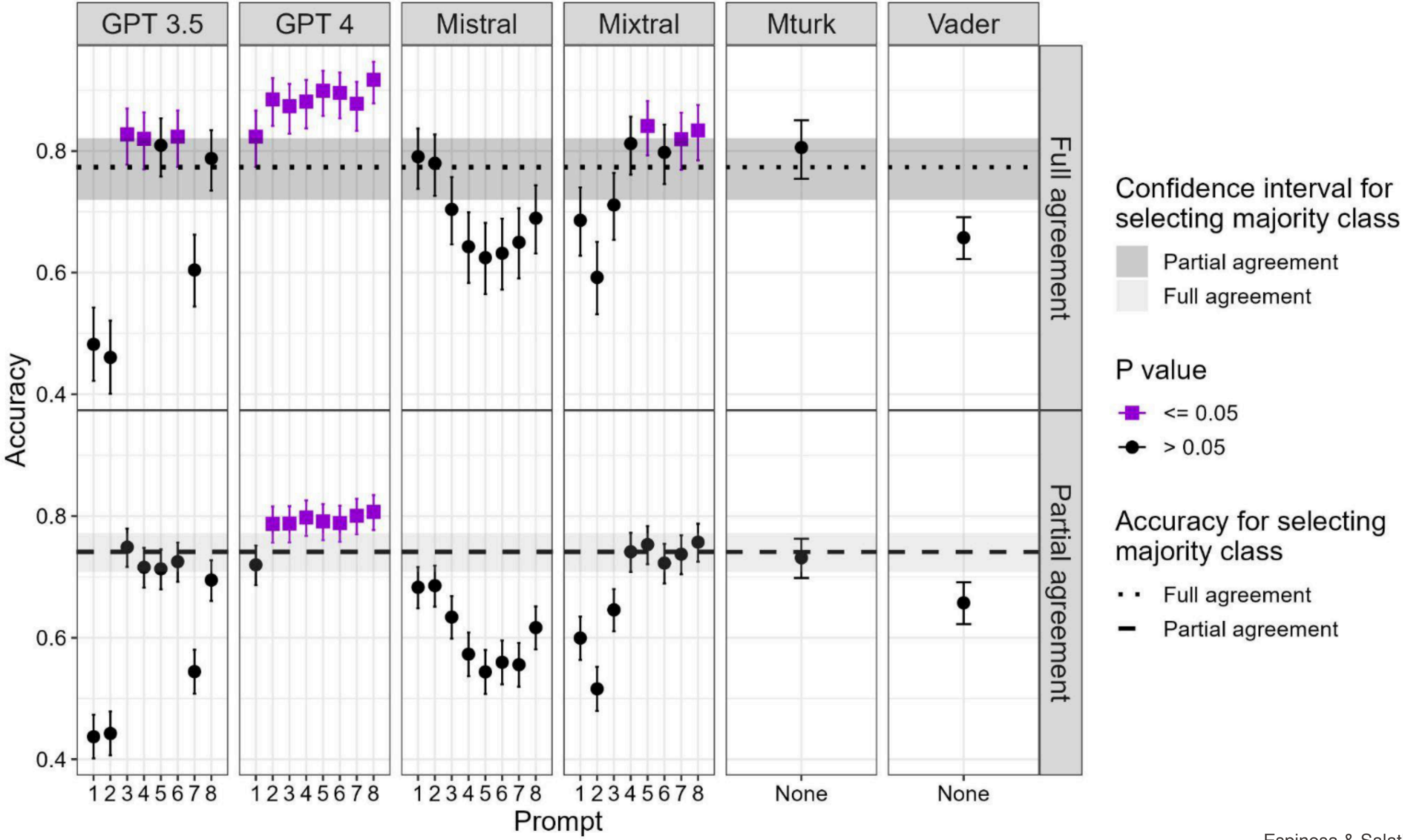
# Digital Public Health Surveillance

## Social Media

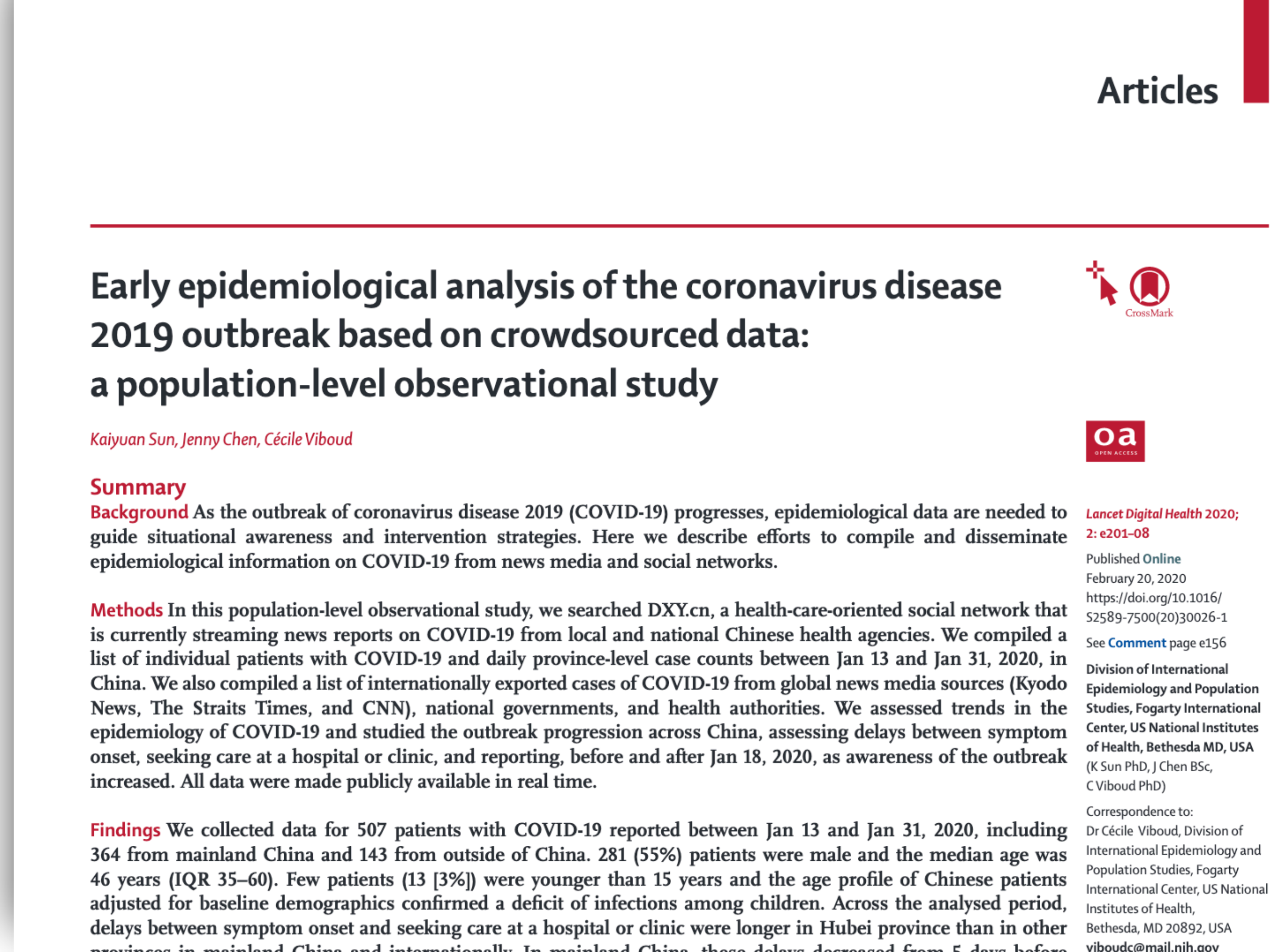




Source: https://github.com/Mooler0410/LLMsPracticalGuide



Espinosa & Salathé, 2024



# Early epidemiological analysis of the coronavirus disease 2019 outbreak based on crowdsourced data: a population-level observational study

*Kaiyuan Sun, Jenny Chen, Cécile Viboud*

## Summary

**Background** As the outbreak of coronavirus disease 2019 (COVID-19) progresses, epidemiological data are needed to guide situational awareness and intervention strategies. Here we describe efforts to compile and disseminate epidemiological information on COVID-19 from news media and social networks.

**Methods** In this population-level observational study, we searched DXY.cn, a health-care-oriented social network that is currently streaming news reports on COVID-19 from local and national Chinese health agencies. We compiled a list of individual patients with COVID-19 and daily province-level case counts between Jan 13 and Jan 31, 2020, in China. We also compiled a list of internationally exported cases of COVID-19 from global news media sources (Kyodo News, The Straits Times, and CNN), national governments, and health authorities. We assessed trends in the epidemiology of COVID-19 and studied the outbreak progression across China, assessing delays between symptom onset, seeking care at a hospital or clinic, and reporting, before and after Jan 18, 2020, as awareness of the outbreak increased. All data were made publicly available in real time.

**Findings** We collected data for 507 patients with COVID-19 reported between Jan 13 and Jan 31, 2020, including 364 from mainland China and 143 from outside of China. 281 (55%) patients were male and the median age was 46 years (IQR 35–60). Few patients (13 [3%]) were younger than 15 years and the age profile of Chinese patients adjusted for baseline demographics confirmed a deficit of infections among children. Across the analysed period, delays between symptom onset and seeking care at a hospital or clinic were longer in Hubei province than in other

*Lancet Digital Health* 2020; 2: e201–08

https://doi.org/10.1016/S2589-7500(20)30026-1

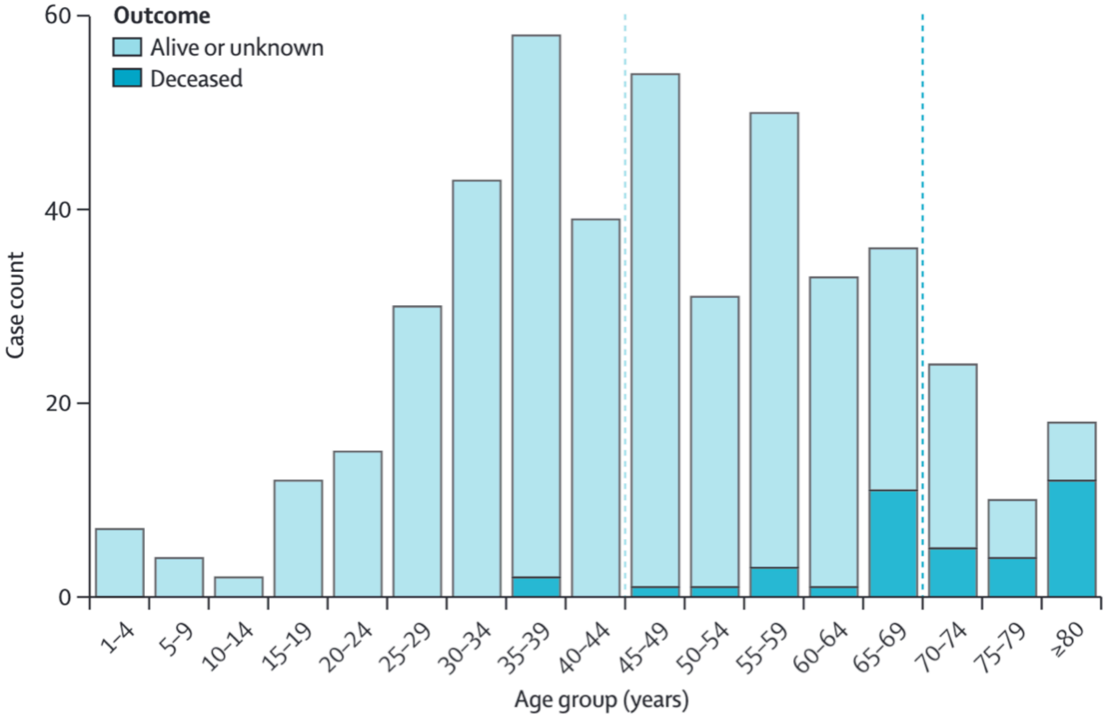
See **Comment** page e156

**Division of International Epidemiology and Population Studies, Fogarty International Center, US National Institutes of Health, Bethesda MD, USA** (K Sun PhD, J Chen BSc, C Viboud PhD)

Correspondence to:
Dr Cécile Viboud, Division of International Epidemiology and Population Studies, Fogarty International Center, US National Institutes of Health, Bethesda, MD 20892, USA
viboudc@mail.nih.gov

# Digital Public Health Surveillance

## Social Media



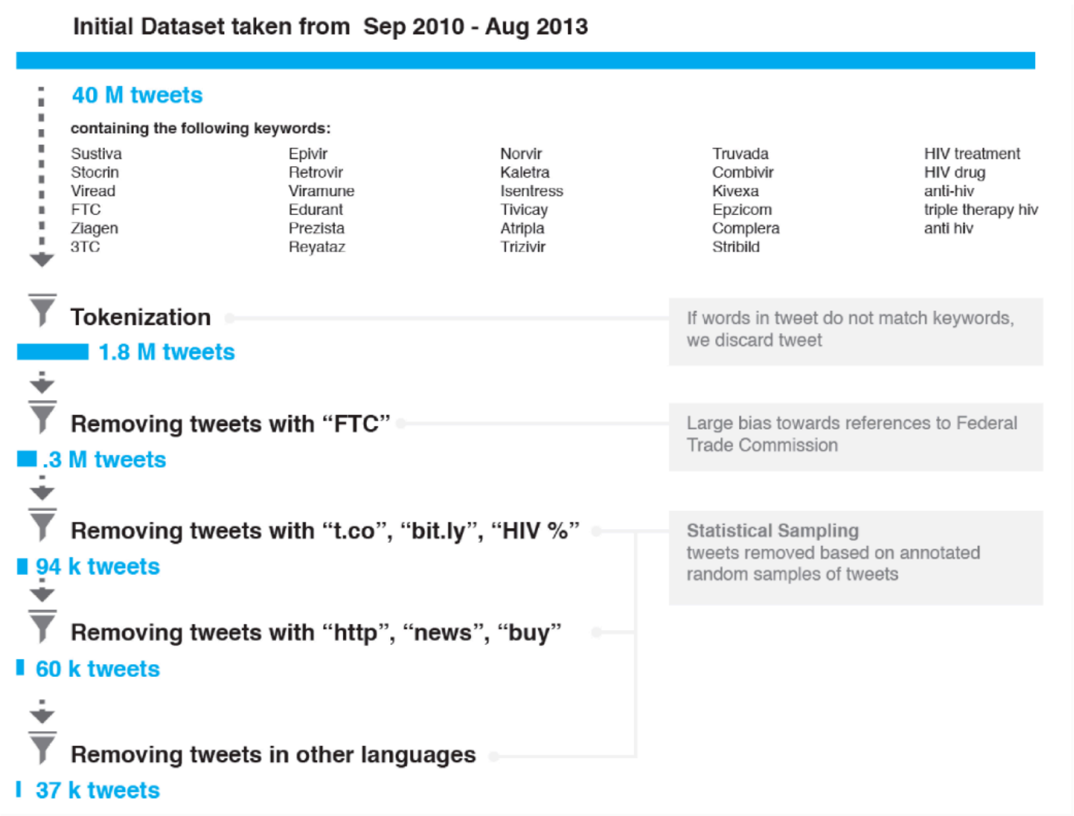
# Digital Public Health Surveillance

## Social Media

- Content analysis allows us to go beyond pure disease surveillance
- Example: Pharmacovigilance

# Digital Public Health Surveillance

## Social Media



**Initial Dataset taken from Sep 2010 - Aug 2013**

**40 M tweets**

**containing the following keywords:**

Sustiva
Stocrin
Viread
FTC
Ziagen
3TC

Epivir
Retrovir
Viramune
Edurant
Prezista
Reyataz

Norvir
Kaletra
Isentress
Tivicay
Atripla
Trizivir

Truvada
Combivir
Kivexa
Epzicom
Complera
Stribild

HIV treatment
HIV drug
anti-hiv
triple therapy hiv
anti hiv

**Tokenization**

If words in tweet do not match keywords, we discard tweet

**1.8 M tweets**

**Removing tweets with "FTC"**

Large bias towards references to Federal Trade Commission

**.3 M tweets**

**Removing tweets with "t.co", "bit.ly", "HIV %"**

**Statistical Sampling**
tweets removed based on annotated random samples of tweets

**94 k tweets**

**Removing tweets with "http", "news", "buy"**

**60 k tweets**

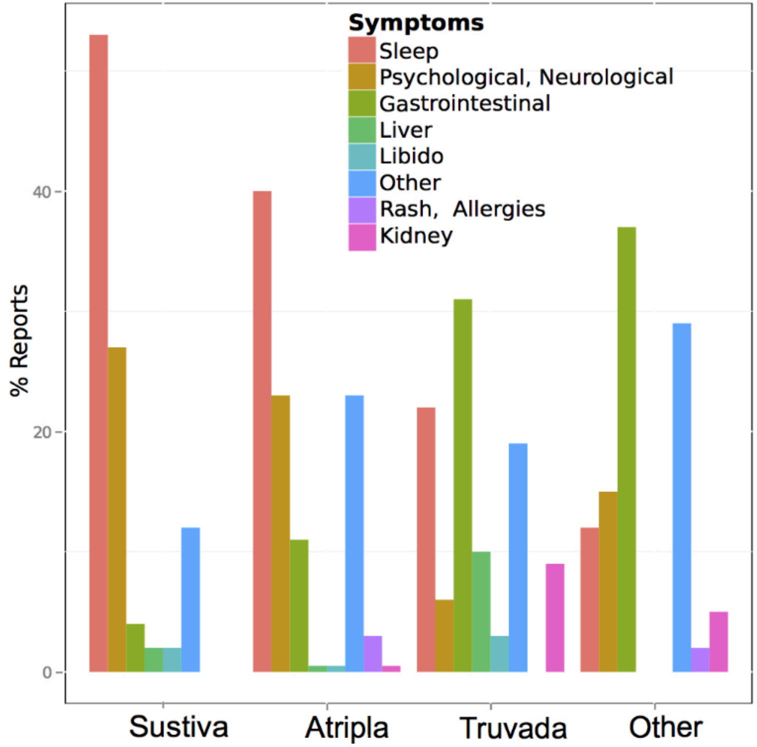
**Removing tweets in other languages**

**37 k tweets**

Source: Adrover et al. 2015

# Digital Public Health Surveillance

## Social Media



Source: Adrover et al. 2015

# Digital Public Health Surveillance

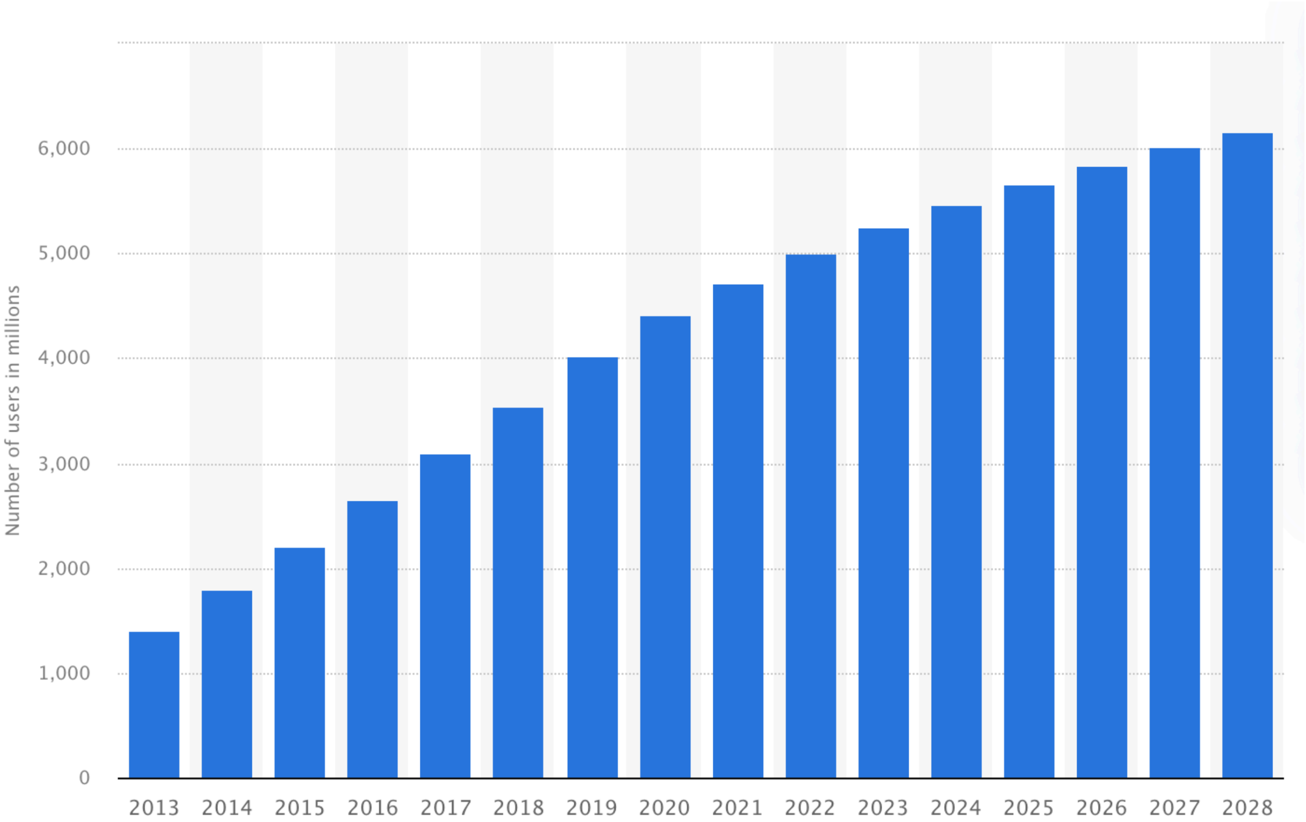
## Mobile phones

- In this section, we will focus exclusively on location data

- Location data is important to infer movement, which is critical for infectious disease dynamics.

# Digital Public Health Surveillance

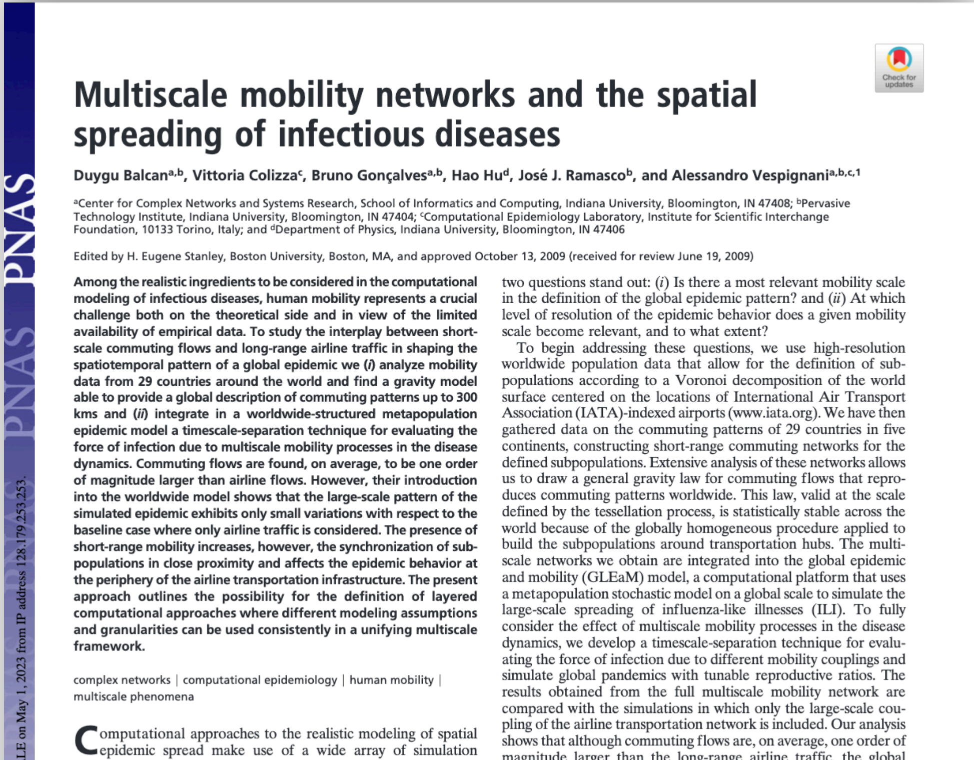
## Mobile Phones



Source: Statista

# Digital Public Health Surveillance

## Mobile Phones



PNAS

### Multiscale mobility networks and the spatial spreading of infectious diseases

Duygu Balcan[a,b], Vittoria Colizza[c], Bruno Gonçalves[a,b], Hao Hu[d], José J. Ramasco[b], and Alessandro Vespignani[a,b,c,1]

[a]Center for Complex Networks and Systems Research, School of Informatics and Computing, Indiana University, Bloomington, IN 47408; [b]Pervasive Technology Institute, Indiana University, Bloomington, IN 47404; [c]Computational Epidemiology Laboratory, Institute for Scientific Interchange Foundation, 10133 Torino, Italy; and [d]Department of Physics, Indiana University, Bloomington, IN 47406

Edited by H. Eugene Stanley, Boston University, Boston, MA, and approved October 13, 2009 (received for review June 19, 2009)

**Among the realistic ingredients to be considered in the computational modeling of infectious diseases, human mobility represents a crucial challenge both on the theoretical side and in view of the limited availability of empirical data. To study the interplay between short-scale commuting flows and long-range airline traffic in shaping the spatiotemporal pattern of a global epidemic we (*i*) analyze mobility data from 29 countries around the world and find a gravity model able to provide a global description of commuting patterns up to 300 kms and (*ii*) integrate in a worldwide-structured metapopulation epidemic model a timescale-separation technique for evaluating the force of infection due to multiscale mobility processes in the disease dynamics. Commuting flows are found, on average, to be one order of magnitude larger than airline flows. However, their introduction into the worldwide model shows that the large-scale pattern of the simulated epidemic exhibits only small variations with respect to the baseline case where only airline traffic is considered. The presence of short-range mobility increases, however, the synchronization of subpopulations in close proximity and affects the epidemic behavior at the periphery of the airline transportation infrastructure. The present approach outlines the possibility for the definition of layered computational approaches where different modeling assumptions and granularities can be used consistently in a unifying multiscale framework.**

complex networks | computational epidemiology | human mobility | multiscale phenomena

Computational approaches to the realistic modeling of spatial epidemic spread make use of a wide array of simulation

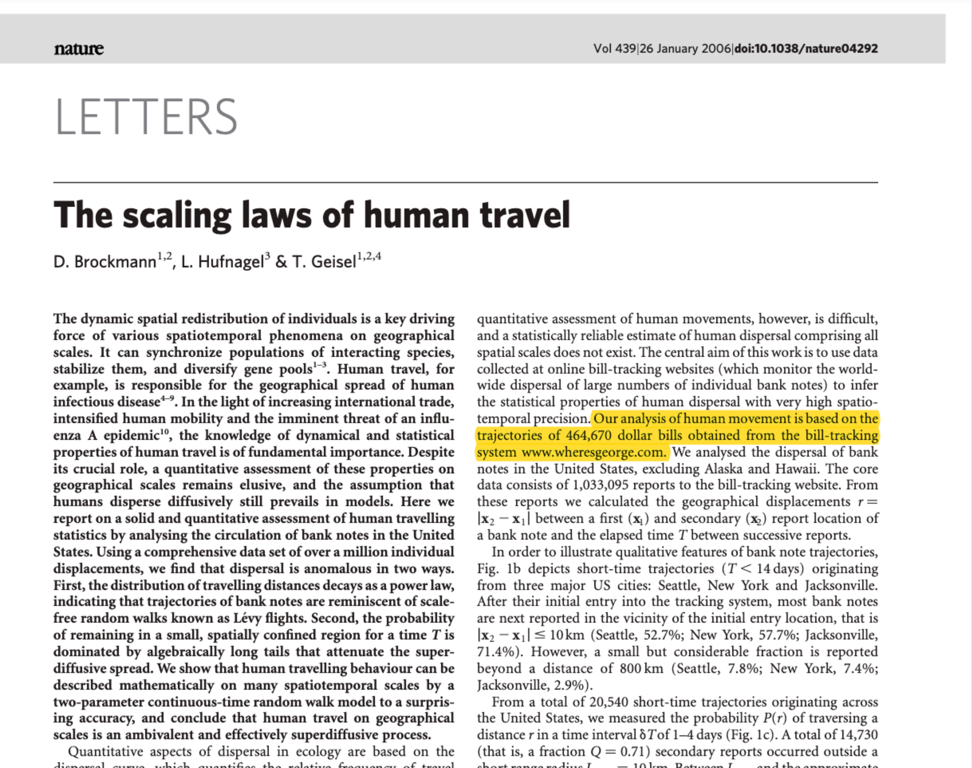
two questions stand out: (*i*) Is there a most relevant mobility scale in the definition of the global epidemic pattern? and (*ii*) At which level of resolution of the epidemic behavior does a given mobility scale become relevant, and to what extent?

To begin addressing these questions, we use high-resolution worldwide population data that allow for the definition of subpopulations according to a Voronoi decomposition of the world surface centered on the locations of International Air Transport Association (IATA)-indexed airports (www.iata.org). We have then gathered data on the commuting patterns of 29 countries in five continents, constructing short-range commuting networks for the defined subpopulations. Extensive analysis of these networks allows us to draw a general gravity law for commuting flows that reproduces commuting patterns worldwide. This law, valid at the scale defined by the tessellation process, is statistically stable across the world because of the globally homogeneous procedure applied to build the subpopulations around transportation hubs. The multiscale networks we obtain are integrated into the global epidemic and mobility (GLEaM) model, a computational platform that uses a metapopulation stochastic model on a global scale to simulate the large-scale spreading of influenza-like illnesses (ILI). To fully consider the effect of multiscale mobility processes in the disease dynamics, we develop a timescale-separation technique for evaluating the force of infection due to different mobility couplings and simulate global pandemics with tunable reproductive ratios. The results obtained from the full multiscale mobility network are compared with the simulations in which only the large-scale coupling of the airline transportation network is included. Our analysis shows that although commuting flows are, on average, one order of magnitude larger than the long-range airline traffic, the global

- uses mobility data provided by countries and airlines

# Digital Public Health Surveillance

## Mobile Phones



nature Vol 439|26 January 2006|doi:10.1038/nature04292

LETTERS

### The scaling laws of human travel

D. Brockmann[1,2], L. Hufnagel[3] & T. Geisel[1,2,4]

**The dynamic spatial redistribution of individuals is a key driving force of various spatiotemporal phenomena on geographical scales. It can synchronize populations of interacting species, stabilize them, and diversify gene pools[1–3]. Human travel, for example, is responsible for the geographical spread of human infectious disease[4–9]. In the light of increasing international trade, intensified human mobility and the imminent threat of an influenza A epidemic[10], the knowledge of dynamical and statistical properties of human travel is of fundamental importance. Despite its crucial role, a quantitative assessment of these properties on geographical scales remains elusive, and the assumption that humans disperse diffusively still prevails in models. Here we report on a solid and quantitative assessment of human travelling statistics by analysing the circulation of bank notes in the United States. Using a comprehensive data set of over a million individual displacements, we find that dispersal is anomalous in two ways. First, the distribution of travelling distances decays as a power law, indicating that trajectories of bank notes are reminiscent of scale-free random walks known as Lévy flights. Second, the probability of remaining in a small, spatially confined region for a time *T* is dominated by algebraically long tails that attenuate the superdiffusive spread. We show that human travelling behaviour can be described mathematically on many spatiotemporal scales by a two-parameter continuous-time random walk model to a surprising accuracy, and conclude that human travel on geographical scales is an ambivalent and effectively superdiffusive process.**

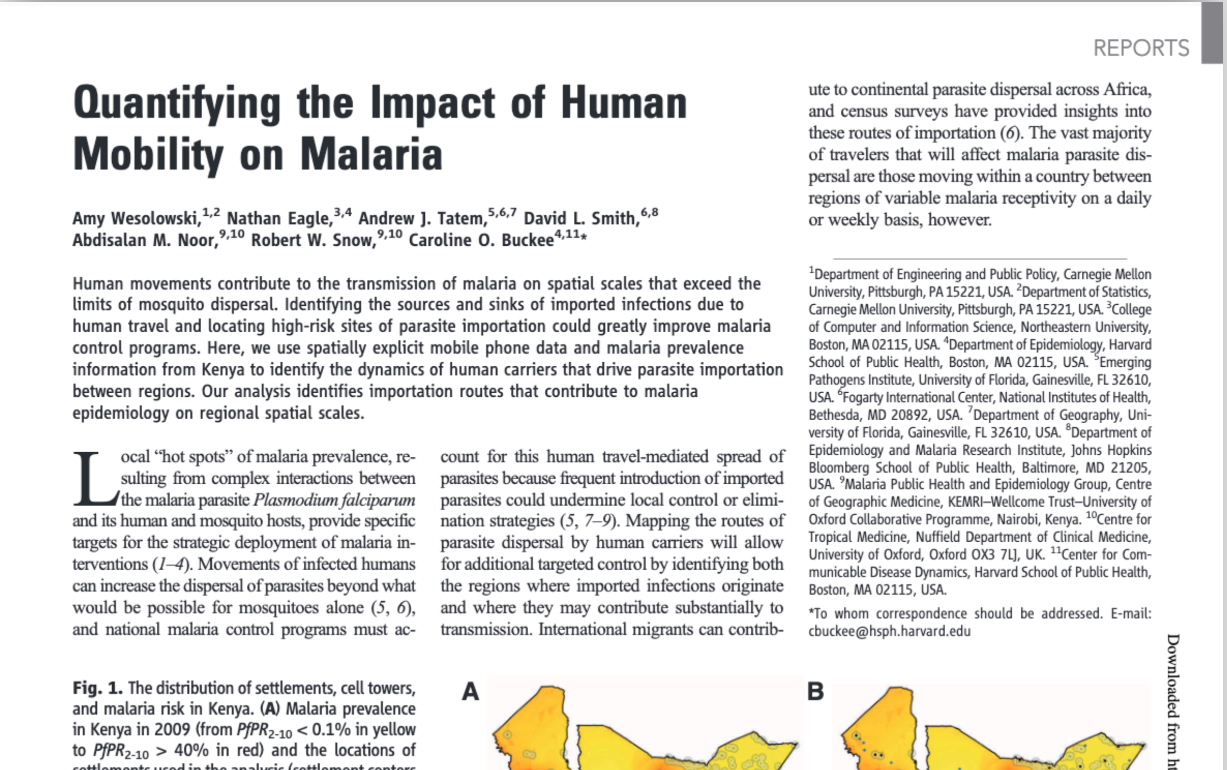
Quantitative aspects of dispersal in ecology are based on the quantitative assessment of human movements, however, is difficult, and a statistically reliable estimate of human dispersal comprising all spatial scales does not exist. The central aim of this work is to use data collected at online bill-tracking websites (which monitor the worldwide dispersal of large numbers of individual bank notes) to infer the statistical properties of human dispersal with very high spatiotemporal precision. Our analysis of human movement is based on the trajectories of 464,670 dollar bills obtained from the bill-tracking system www.wheresgeorge.com. We analysed the dispersal of bank notes in the United States, excluding Alaska and Hawaii. The core data consists of 1,033,095 reports to the bill-tracking website. From these reports we calculated the geographical displacements $r = |\mathbf{x}_2 - \mathbf{x}_1|$ between a first ($\mathbf{x}_1$) and secondary ($\mathbf{x}_2$) report location of a bank note and the elapsed time $T$ between successive reports.

In order to illustrate qualitative features of bank note trajectories, Fig. 1b depicts short-time trajectories ($T < 14$ days) originating from three major US cities: Seattle, New York and Jacksonville. After their initial entry into the tracking system, most bank notes are next reported in the vicinity of the initial entry location, that is $|\mathbf{x}_2 - \mathbf{x}_1| \leq 10$ km (Seattle, 52.7%; New York, 57.7%; Jacksonville, 71.4%). However, a small but considerable fraction is reported beyond a distance of 800 km (Seattle, 7.8%; New York, 7.4%; Jacksonville, 2.9%).

From a total of 20,540 short-time trajectories originating across the United States, we measured the probability $P(r)$ of traversing a distance $r$ in a time interval $\delta T$ of 1–4 days (Fig. 1c). A total of 14,730 (that is, a fraction $Q = 0.71$) secondary reports occurred outside a

# Digital Public Health Surveillance

## Mobile Phones



### Quantifying the Impact of Human Mobility on Malaria

Amy Wesolowski,[1,2] Nathan Eagle,[3,4] Andrew J. Tatem,[5,6,7] David L. Smith,[6,8]
Abdisalan M. Noor,[9,10] Robert W. Snow,[9,10] Caroline O. Buckee[4,11]*

**Human movements contribute to the transmission of malaria on spatial scales that exceed the limits of mosquito dispersal. Identifying the sources and sinks of imported infections due to human travel and locating high-risk sites of parasite importation could greatly improve malaria control programs. Here, we use spatially explicit mobile phone data and malaria prevalence information from Kenya to identify the dynamics of human carriers that drive parasite importation between regions. Our analysis identifies importation routes that contribute to malaria epidemiology on regional spatial scales.**

Local "hot spots" of malaria prevalence, resulting from complex interactions between the malaria parasite *Plasmodium falciparum* and its human and mosquito hosts, provide specific targets for the strategic deployment of malaria interventions (*1*–*4*). Movements of infected humans can increase the dispersal of parasites beyond what would be possible for mosquitoes alone (*5*, *6*), and national malaria control programs must account for this human travel-mediated spread of parasites because frequent introduction of imported parasites could undermine local control or elimination strategies (*5*, *7*–*9*). Mapping the routes of parasite dispersal by human carriers will allow for additional targeted control by identifying both the regions where imported infections originate and where they may contribute substantially to transmission. International migrants can contribute to continental parasite dispersal across Africa, and census surveys have provided insights into these routes of importation (*6*). The vast majority of travelers that will affect malaria parasite dispersal are those moving within a country between regions of variable malaria receptivity on a daily or weekly basis, however.

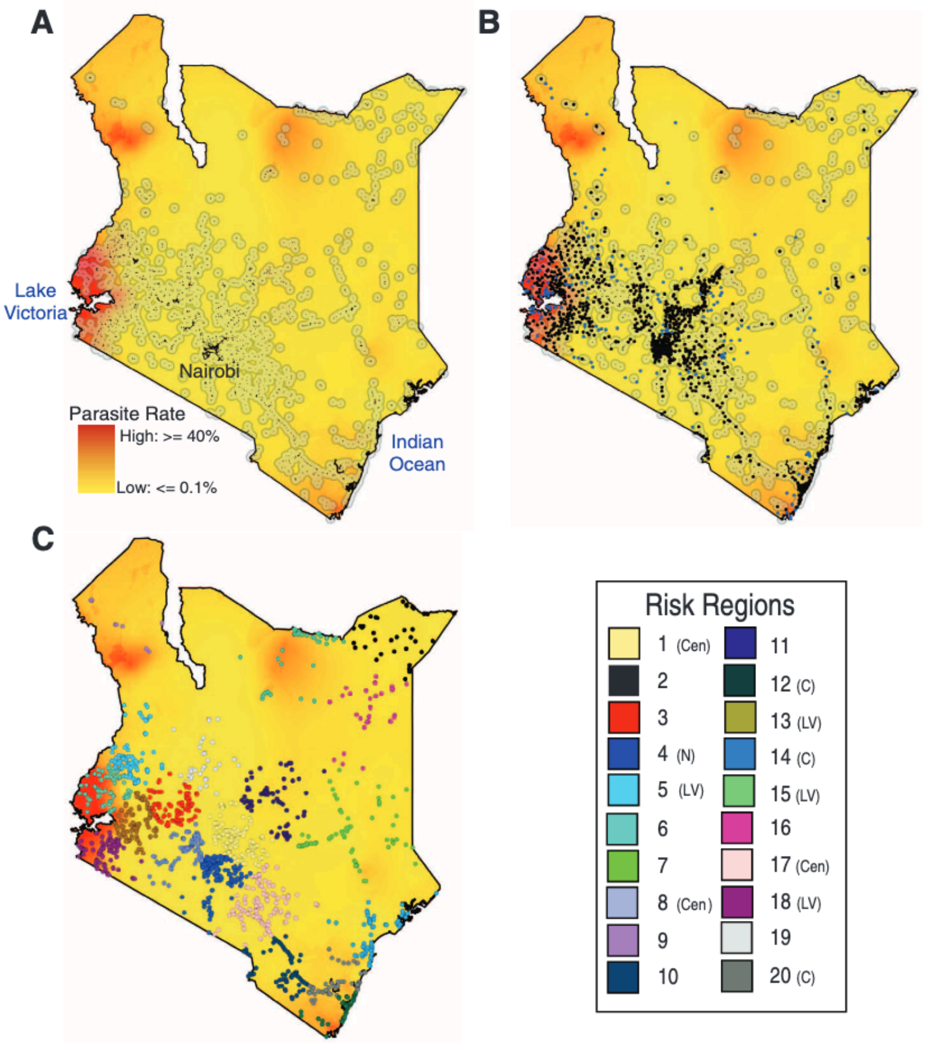
**Fig. 1.** The distribution of settlements, cell towers, and malaria risk in Kenya. (**A**) Malaria prevalence in Kenya in 2009 (from $PfPR_{2\text{-}10} < 0.1\%$ in yellow to $PfPR_{2\text{-}10} > 40\%$ in red) and the locations of settlements used in the analysis (settlement centers

---

[1] Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15221, USA. [2] Department of Statistics, Carnegie Mellon University, Pittsburgh, PA 15221, USA. [3] College of Computer and Information Science, Northeastern University, Boston, MA 02115, USA. [4] Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA. [5] Emerging Pathogens Institute, University of Florida, Gainesville, FL 32610, USA. [6] Fogarty International Center, National Institutes of Health, Bethesda, MD 20892, USA. [7] Department of Geography, University of Florida, Gainesville, FL 32610, USA. [8] Department of Epidemiology and Malaria Research Institute, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA. [9] Malaria Public Health and Epidemiology Group, Centre of Geographic Medicine, KEMRI–Wellcome Trust–University of Oxford Collaborative Programme, Nairobi, Kenya. [10] Centre for Tropical Medicine, Nuffield Department of Clinical Medicine, University of Oxford, Oxford OX3 7LJ, UK. [11] Center for Communicable Disease Dynamics, Harvard School of Public Health, Boston, MA 02115, USA.

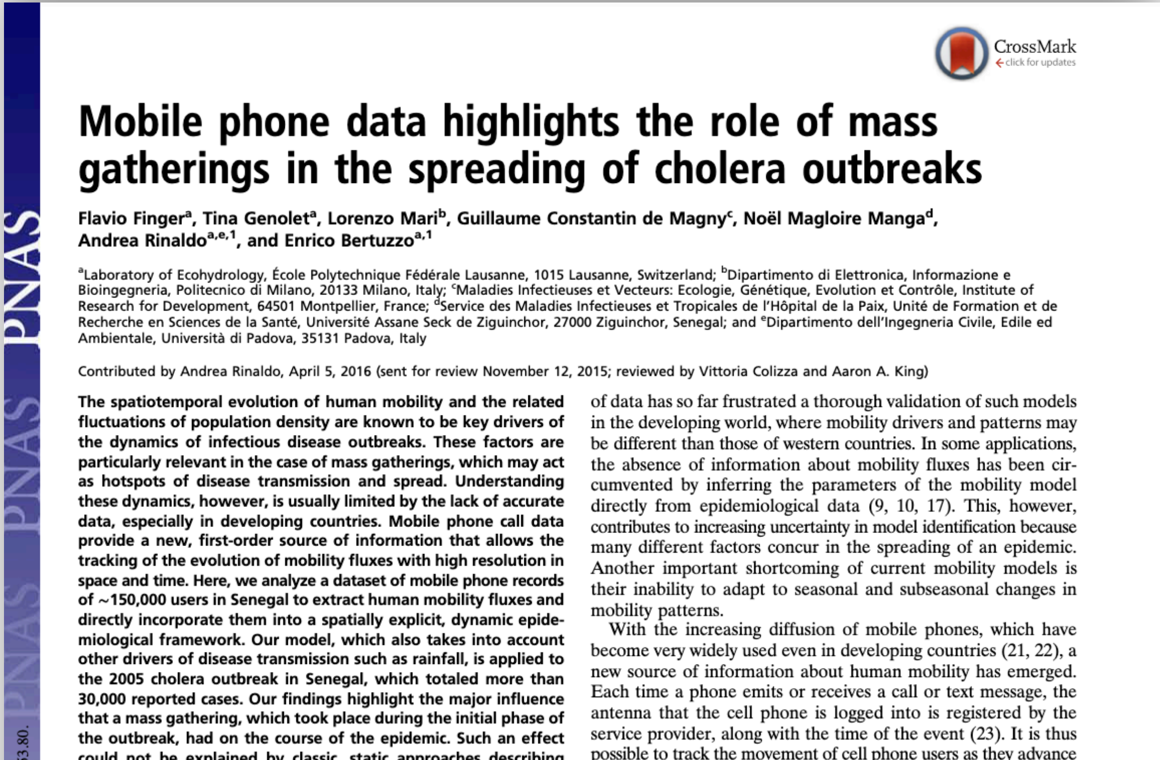
*To whom correspondence should be addressed. E-mail: cbuckee@hsph.harvard.edu

**Fig. 1.** The distribution of settlements, cell towers, and malaria risk in Kenya. **(A)** Malaria prevalence in Kenya in 2009 (from $PfPR_{2\text{-}10} < 0.1\%$ in yellow to $PfPR_{2\text{-}10} > 40\%$ in red) and the locations of settlements used in the analysis (settlement centers are shown in black, and mapped with a 10-km extent around the perimeter of the settlement in gray). **(B)** The location of mobile phone towers (black or blue dots) and the extended settlement boundaries. Towers that fall within a settlement are shown in black, and those excluded from the analysis are shown in blue. **(C)** Regions used for visual mapping of transmission routes. Each settlement was allocated to 1 of 20 regions by a clustering algorithm (*14*) on the basis of homogenous malaria risk and geography, as shown. Regions near Lake Victoria (LV), in Nairobi (Nairobi), the central areas (Cen), and along the coast (C) are labeled accordingly.



# Digital Public Health Surveillance

## Mobile Phones



CrossMark
click for updates

# Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks

Flavio Finger[a], Tina Genolet[a], Lorenzo Mari[b], Guillaume Constantin de Magny[c], Noël Magloire Manga[d], Andrea Rinaldo[a,e,1], and Enrico Bertuzzo[a,1]

[a]Laboratory of Ecohydrology, École Polytechnique Fédérale Lausanne, 1015 Lausanne, Switzerland; [b]Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milano, Italy; [c]Maladies Infectieuses et Vecteurs: Ecologie, Génétique, Evolution et Contrôle, Institute of Research for Development, 64501 Montpellier, France; [d]Service des Maladies Infectieuses et Tropicales de l'Hôpital de la Paix, Unité de Formation et de Recherche en Sciences de la Santé, Université Assane Seck de Ziguinchor, 27000 Ziguinchor, Senegal; and [e]Dipartimento dell'Ingegneria Civile, Edile ed Ambientale, Università di Padova, 35131 Padova, Italy

Contributed by Andrea Rinaldo, April 5, 2016 (sent for review November 12, 2015; reviewed by Vittoria Colizza and Aaron A. King)

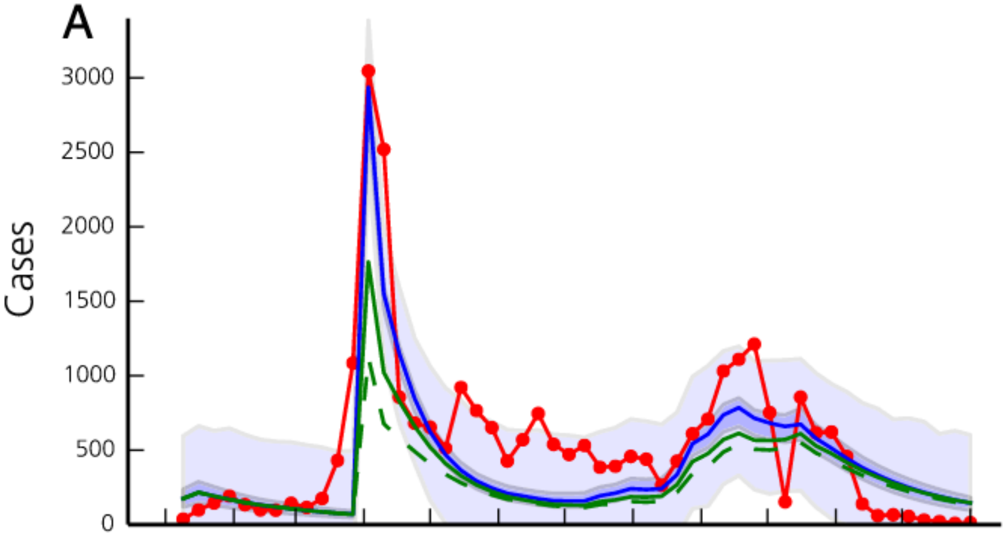
**The spatiotemporal evolution of human mobility and the related fluctuations of population density are known to be key drivers of the dynamics of infectious disease outbreaks. These factors are particularly relevant in the case of mass gatherings, which may act as hotspots of disease transmission and spread. Understanding these dynamics, however, is usually limited by the lack of accurate data, especially in developing countries. Mobile phone call data provide a new, first-order source of information that allows the tracking of the evolution of mobility fluxes with high resolution in space and time. Here, we analyze a dataset of mobile phone records of ~150,000 users in Senegal to extract human mobility fluxes and directly incorporate them into a spatially explicit, dynamic epidemiological framework. Our model, which also takes into account other drivers of disease transmission such as rainfall, is applied to the 2005 cholera outbreak in Senegal, which totaled more than 30,000 reported cases. Our findings highlight the major influence that a mass gathering, which took place during the initial phase of the outbreak, had on the course of the epidemic. Such an effect could not be explained by classic, static approaches describing**

of data has so far frustrated a thorough validation of such models in the developing world, where mobility drivers and patterns may be different than those of western countries. In some applications, the absence of information about mobility fluxes has been circumvented by inferring the parameters of the mobility model directly from epidemiological data (9, 10, 17). This, however, contributes to increasing uncertainty in model identification because many different factors concur in the spreading of an epidemic. Another important shortcoming of current mobility models is their inability to adapt to seasonal and subseasonal changes in mobility patterns.

With the increasing diffusion of mobile phones, which have become very widely used even in developing countries (21, 22), a new source of information about human mobility has emerged. Each time a phone emits or receives a call or text message, the antenna that the cell phone is logged into is registered by the service provider, along with the time of the event (23). It is thus possible to track the movement of cell phone users as they advance

# Digital Public Health Surveillance

## Mobile Phones

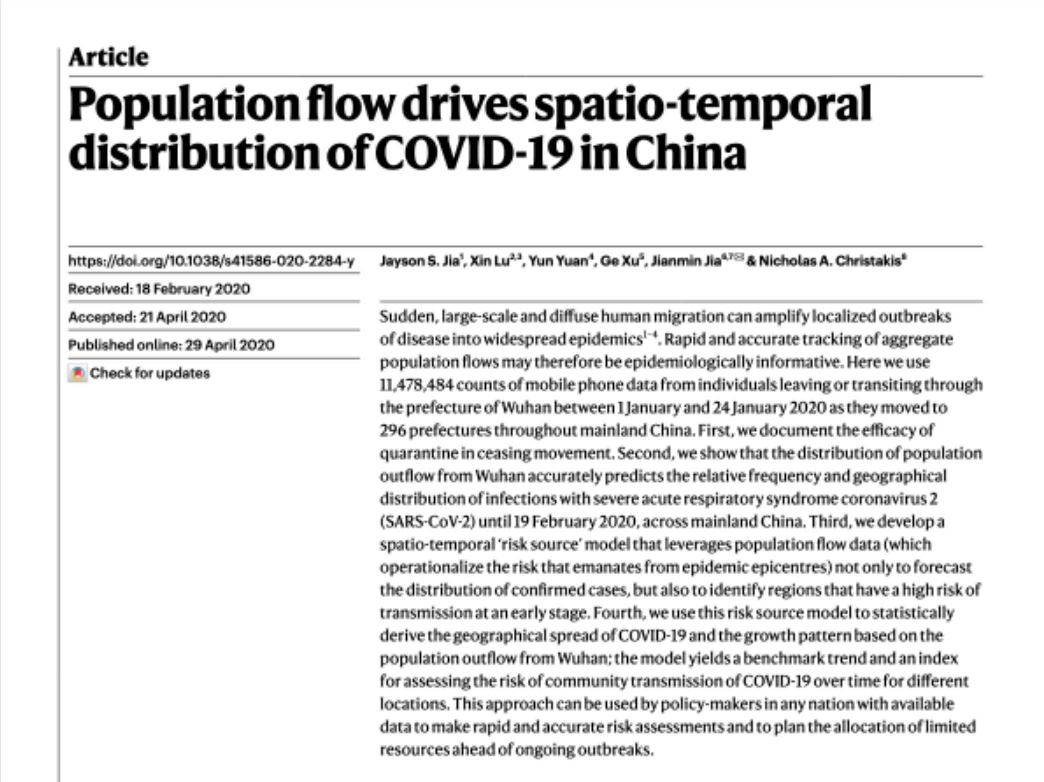


Red: reported

Blue: modeled

# Digital Public Health Surveillance

## Mobile Phones



**Article**

# Population flow drives spatio-temporal distribution of COVID-19 in China

https://doi.org/10.1038/s41586-020-2284-y

Received: 18 February 2020

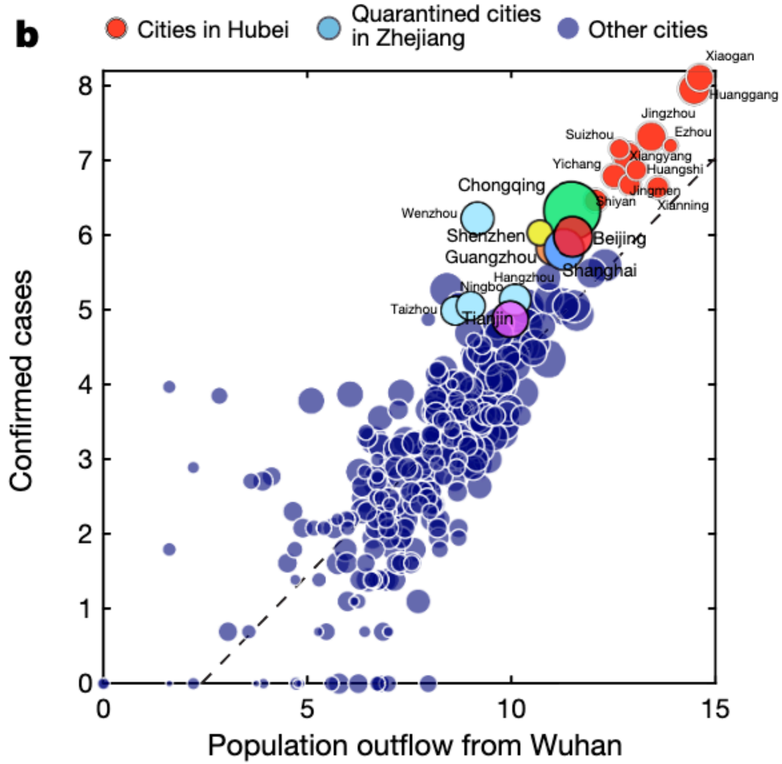
Accepted: 21 April 2020

Jayson S. Jia[1], Xin Lu[2,3], Yun Yuan[4], Ge Xu[5], Jianmin Jia[6,7]✉ & Nicholas A. Christakis[8]

Sudden, large-scale and diffuse human migration can amplify localized outbreaks of disease into widespread epidemics[1–4]. Rapid and accurate tracking of aggregate population flows may therefore be epidemiologically informative. Here we use 11,478,484 counts of mobile phone data from individuals leaving or transiting through the prefecture of Wuhan between 1 January and 24 January 2020 as they moved to 296 prefectures throughout mainland China. First, we document the efficacy of quarantine in ceasing movement. Second, we show that the distribution of population outflow from Wuhan accurately predicts the relative frequency and geographical distribution of infections with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) until 19 February 2020, across mainland China. Third, we develop a spatio-temporal 'risk source' model that leverages population flow data (which operationalize the risk that emanates from epidemic epicentres) not only to forecast the distribution of confirmed cases, but also to identify regions that have a high risk of transmission at an early stage. Fourth, we use this risk source model to statistically derive the geographical spread of COVID-19 and the growth pattern based on the population outflow from Wuhan; the model yields a benchmark trend and an index for assessing the risk of community transmission of COVID-19 over time for different locations. This approach can be used by policy-makers in any nation with available data to make rapid and accurate risk assessments and to plan the allocation of limited resources ahead of ongoing outbreaks.

# Digital Public Health Surveillance

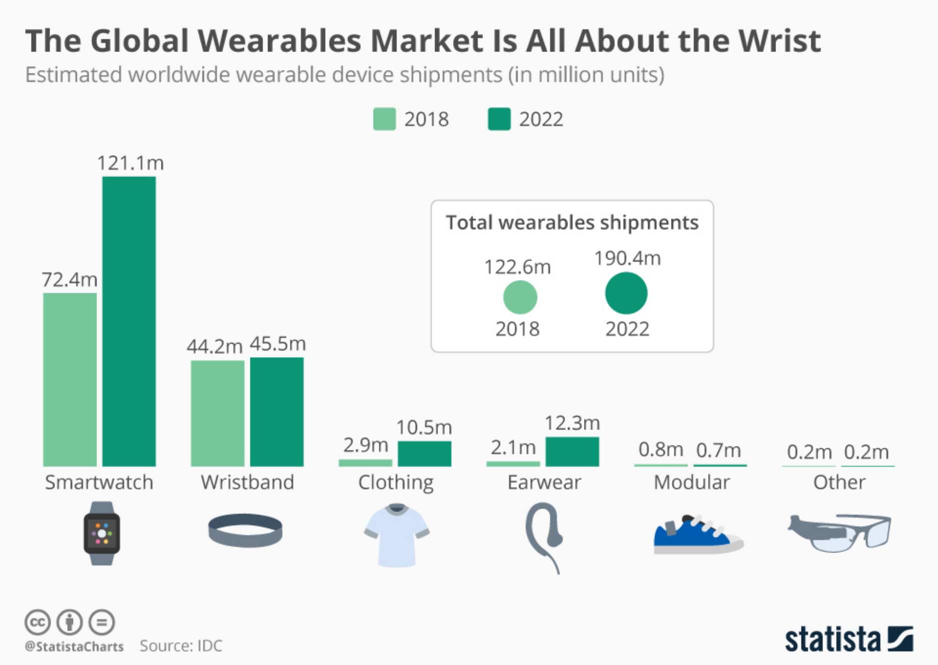
## Mobile Phones



# Digital Public Health Surveillance

## Wearables

- Rapidly growing segment



The Global Wearables Market Is All About the Wrist

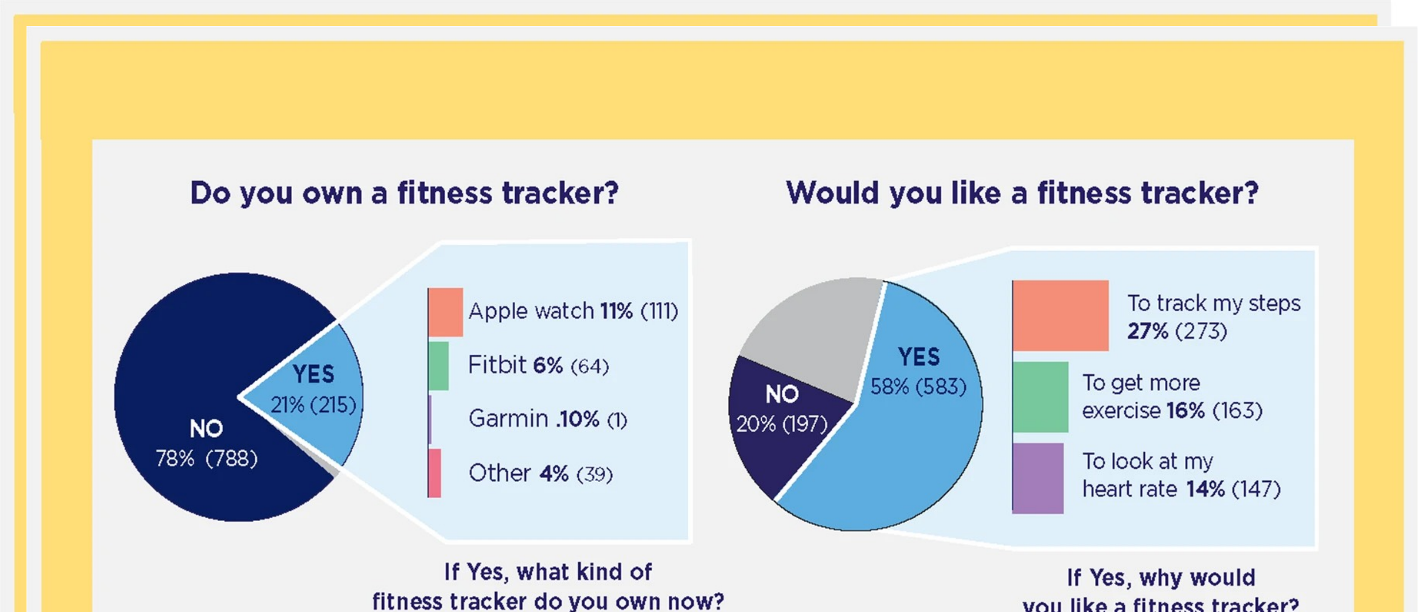
Estimated worldwide wearable device shipments (in million units)

Source: IDC

# Digital Public Health Surveillance

## Wearables

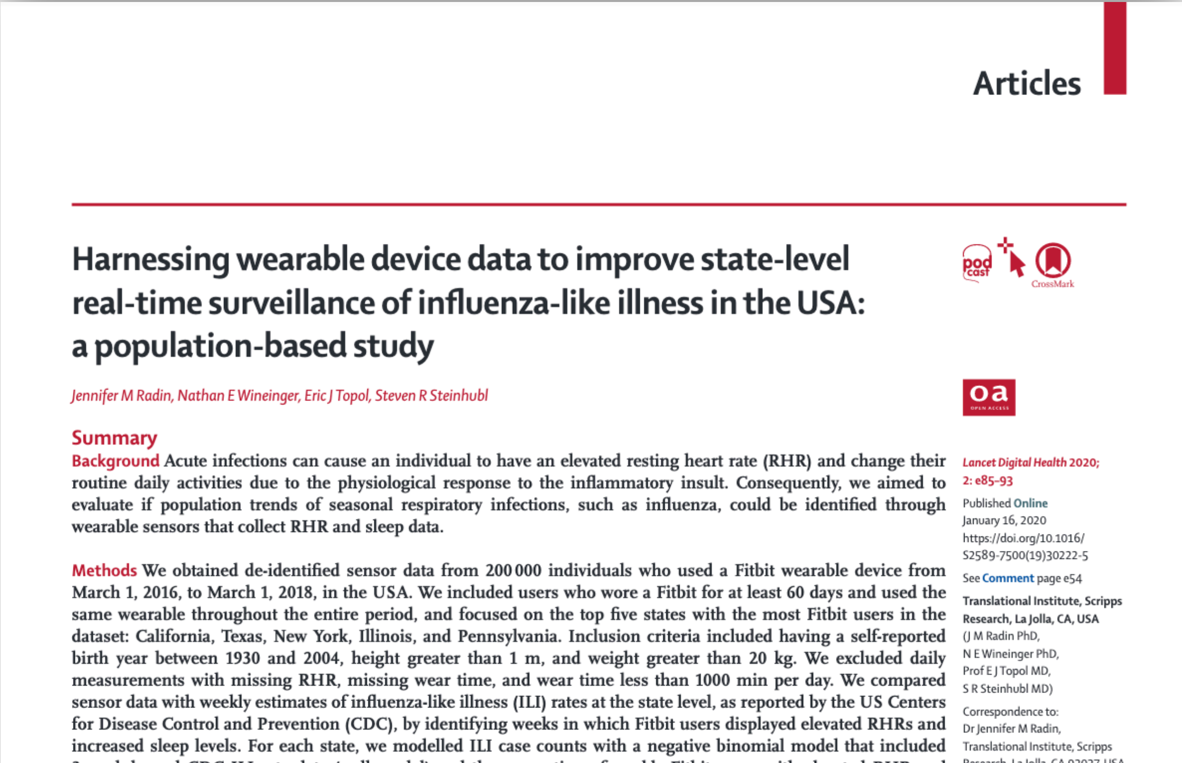
- Rapidly growing segment



ce: Holko et al, 2022

# Digital Public Health Surveillance

## Wearables



Articles

### Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: a population-based study

*Jennifer M Radin, Nathan E Wineinger, Eric J Topol, Steven R Steinhubl*

**Summary**

**Background** Acute infections can cause an individual to have an elevated resting heart rate (RHR) and change their routine daily activities due to the physiological response to the inflammatory insult. Consequently, we aimed to evaluate if population trends of seasonal respiratory infections, such as influenza, could be identified through wearable sensors that collect RHR and sleep data.

**Methods** We obtained de-identified sensor data from 200 000 individuals who used a Fitbit wearable device from March 1, 2016, to March 1, 2018, in the USA. We included users who wore a Fitbit for at least 60 days and used the same wearable throughout the entire period, and focused on the top five states with the most Fitbit users in the dataset: California, Texas, New York, Illinois, and Pennsylvania. Inclusion criteria included having a self-reported birth year between 1930 and 2004, height greater than 1 m, and weight greater than 20 kg. We excluded daily measurements with missing RHR, missing wear time, and wear time less than 1000 min per day. We compared sensor data with weekly estimates of influenza-like illness (ILI) rates at the state level, as reported by the US Centers for Disease Control and Prevention (CDC), by identifying weeks in which Fitbit users displayed elevated RHRs and increased sleep levels. For each state, we modelled ILI case counts with a negative binomial model that included

*Lancet Digital Health* 2020; 2: e85–93

https://doi.org/10.1016/S2589-7500(19)30222-5

See Comment page e54

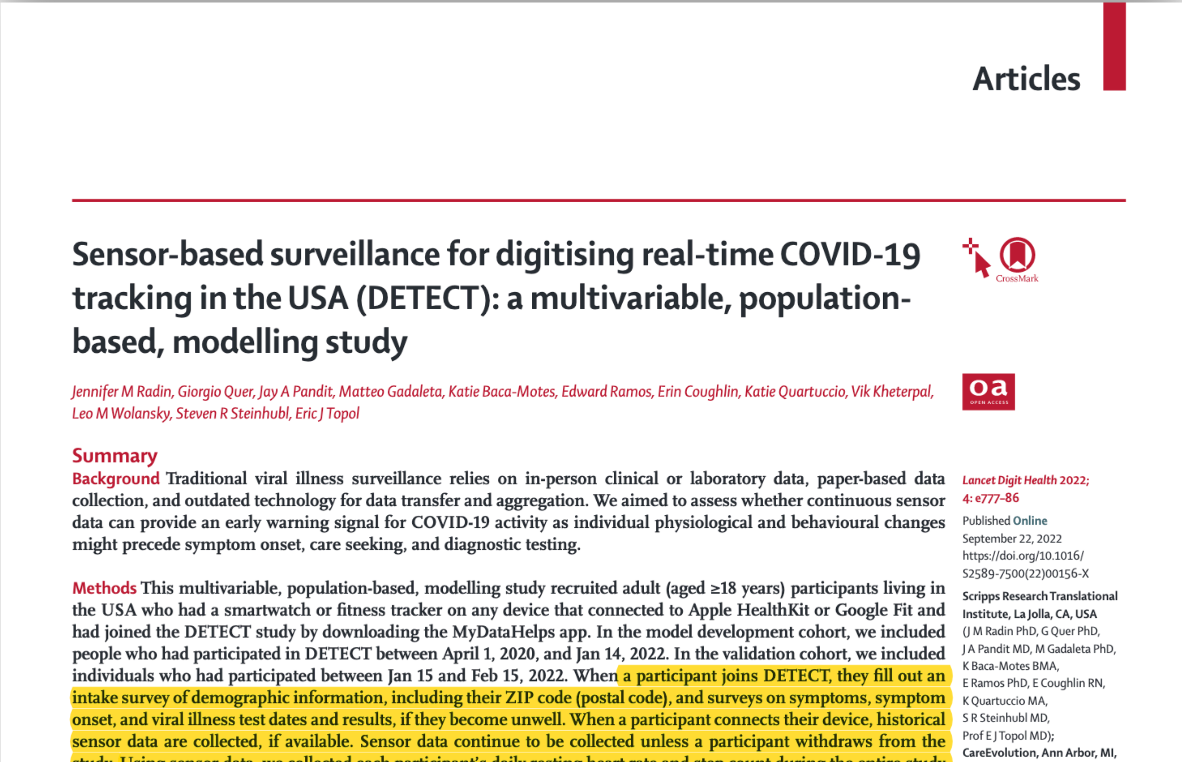
**Translational Institute, Scripps Research, La Jolla, CA, USA**
(J M Radin PhD,
N E Wineinger PhD,
Prof E J Topol MD,
S R Steinhubl MD)

Correspondence to:
Dr Jennifer M Radin,
Translational Institute, Scripps

- 200,000 ppl

# Digital Public Health Surveillance

## Wearables



### Sensor-based surveillance for digitising real-time COVID-19 tracking in the USA (DETECT): a multivariable, population-based, modelling study

*Jennifer M Radin, Giorgio Quer, Jay A Pandit, Matteo Gadaleta, Katie Baca-Motes, Edward Ramos, Erin Coughlin, Katie Quartuccio, Vik Kheterpal, Leo M Wolansky, Steven R Steinhubl, Eric J Topol*

#### Summary

**Background** Traditional viral illness surveillance relies on in-person clinical or laboratory data, paper-based data collection, and outdated technology for data transfer and aggregation. We aimed to assess whether continuous sensor data can provide an early warning signal for COVID-19 activity as individual physiological and behavioural changes might precede symptom onset, care seeking, and diagnostic testing.

**Methods** This multivariable, population-based, modelling study recruited adult (aged ≥18 years) participants living in the USA who had a smartwatch or fitness tracker on any device that connected to Apple HealthKit or Google Fit and had joined the DETECT study by downloading the MyDataHelps app. In the model development cohort, we included people who had participated in DETECT between April 1, 2020, and Jan 14, 2022. In the validation cohort, we included individuals who had participated between Jan 15 and Feb 15, 2022. When a participant joins DETECT, they fill out an intake survey of demographic information, including their ZIP code (postal code), and surveys on symptoms, symptom onset, and viral illness test dates and results, if they become unwell. When a participant connects their device, historical sensor data are collected, if available. Sensor data continue to be collected unless a participant withdraws from the

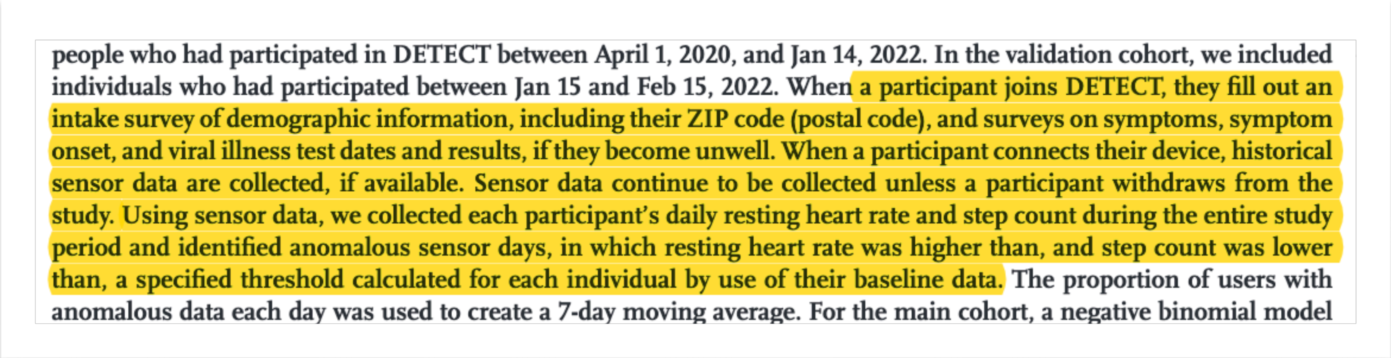
*Lancet Digit Health* 2022;
4: e777–86

https://doi.org/10.1016/
S2589-7500(22)00156-X

**Scripps Research Translational Institute, La Jolla, CA, USA**
(J M Radin PhD, G Quer PhD,
J A Pandit MD, M Gadaleta PhD,
K Baca-Motes BMA,
E Ramos PhD, E Coughlin RN,
K Quartuccio MA,
S R Steinhubl MD,
Prof E J Topol MD);
**CareEvolution, Ann Arbor, MI,**

# Digital Public Health Surveillance

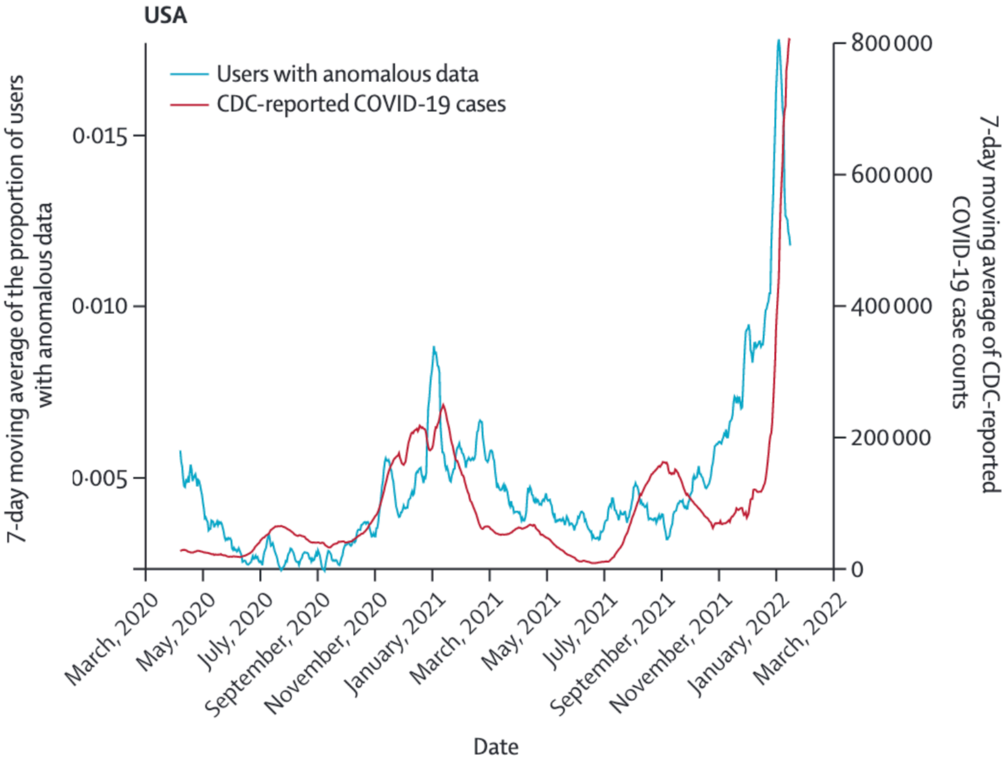
## Wearables



people who had participated in DETECT between April 1, 2020, and Jan 14, 2022. In the validation cohort, we included individuals who had participated between Jan 15 and Feb 15, 2022. When a participant joins DETECT, they fill out an intake survey of demographic information, including their ZIP code (postal code), and surveys on symptoms, symptom onset, and viral illness test dates and results, if they become unwell. When a participant connects their device, historical sensor data are collected, if available. Sensor data continue to be collected unless a participant withdraws from the study. Using sensor data, we collected each participant's daily resting heart rate and step count during the entire study period and identified anomalous sensor days, in which resting heart rate was higher than, and step count was lower than, a specified threshold calculated for each individual by use of their baseline data. The proportion of users with anomalous data each day was used to create a 7-day moving average. For the main cohort, a negative binomial model

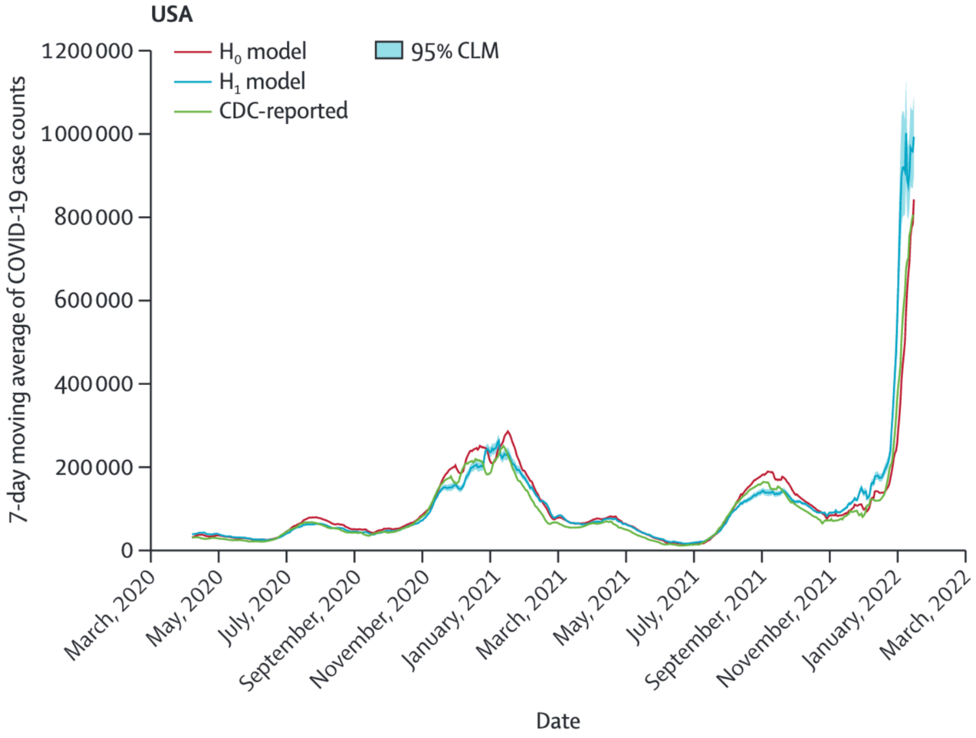
# Digital Public Health Surveillance

## Wearables



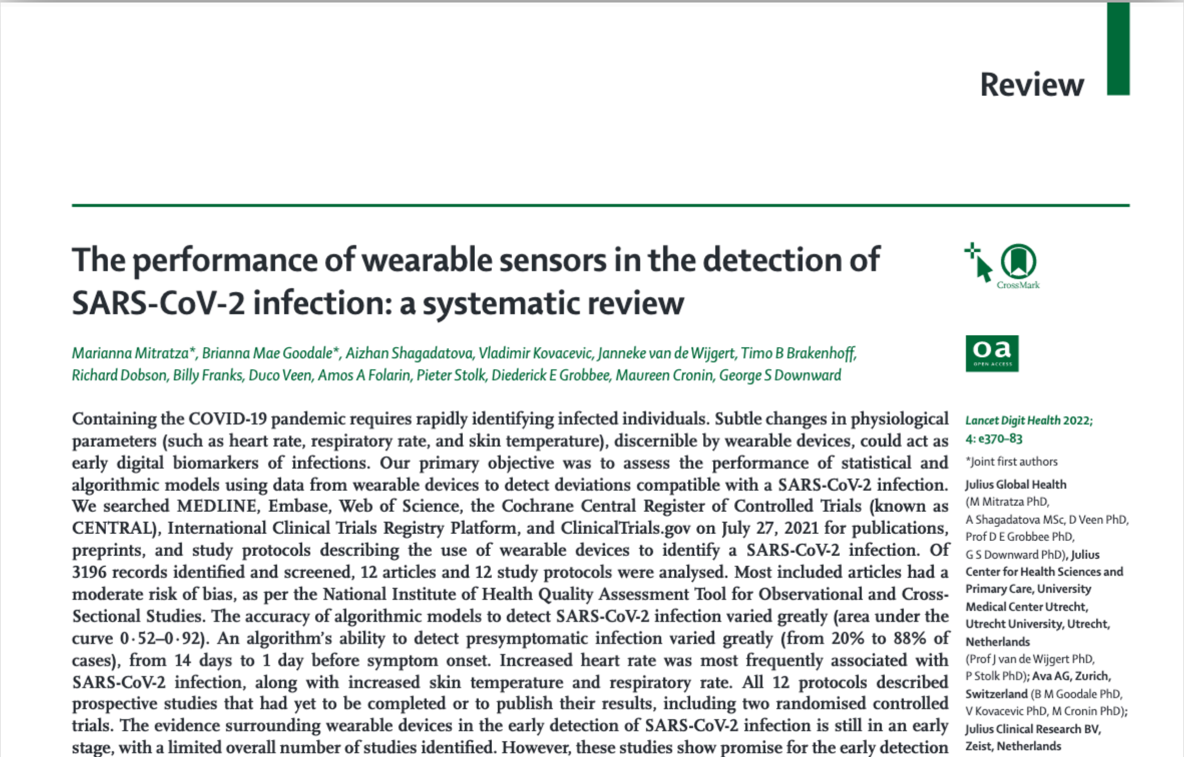
# Digital Public Health Surveillance

## Wearables



# Digital Public Health Surveillance

## Wearables



Review

### The performance of wearable sensors in the detection of SARS-CoV-2 infection: a systematic review

Marianna Mitratza*, Brianna Mae Goodale*, Aizhan Shagadatova, Vladimir Kovacevic, Janneke van de Wijgert, Timo B Brakenhoff, Richard Dobson, Billy Franks, Duco Veen, Amos A Folarin, Pieter Stolk, Diederick E Grobbee, Maureen Cronin, George S Downward

Containing the COVID-19 pandemic requires rapidly identifying infected individuals. Subtle changes in physiological parameters (such as heart rate, respiratory rate, and skin temperature), discernible by wearable devices, could act as early digital biomarkers of infections. Our primary objective was to assess the performance of statistical and algorithmic models using data from wearable devices to detect deviations compatible with a SARS-CoV-2 infection. We searched MEDLINE, Embase, Web of Science, the Cochrane Central Register of Controlled Trials (known as CENTRAL), International Clinical Trials Registry Platform, and ClinicalTrials.gov on July 27, 2021 for publications, preprints, and study protocols describing the use of wearable devices to identify a SARS-CoV-2 infection. Of 3196 records identified and screened, 12 articles and 12 study protocols were analysed. Most included articles had a moderate risk of bias, as per the National Institute of Health Quality Assessment Tool for Observational and Cross-Sectional Studies. The accuracy of algorithmic models to detect SARS-CoV-2 infection varied greatly (area under the curve 0·52–0·92). An algorithm's ability to detect presymptomatic infection varied greatly (from 20% to 88% of cases), from 14 days to 1 day before symptom onset. Increased heart rate was most frequently associated with SARS-CoV-2 infection, along with increased skin temperature and respiratory rate. All 12 protocols described prospective studies that had yet to be completed or to publish their results, including two randomised controlled trials. The evidence surrounding wearable devices in the early detection of SARS-CoV-2 infection is still in an early stage, with a limited overall number of studies identified. However, these studies show promise for the early detection

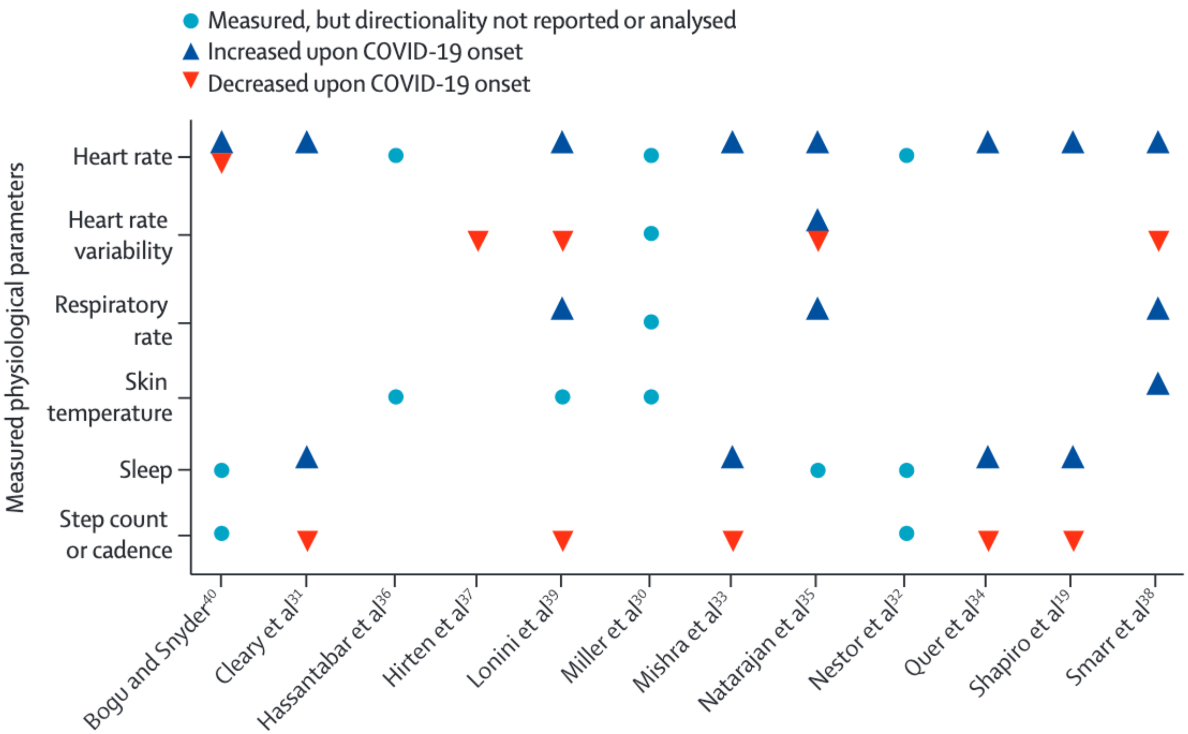
*Lancet Digit Health* 2022; 4: e370–83

*Joint first authors

**Julius Global Health** (M Mitratza PhD, A Shagadatova MSc, D Veen PhD, Prof D E Grobbee PhD, G S Downward PhD), **Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht University, Utrecht, Netherlands** (Prof J van de Wijgert PhD, P Stolk PhD); **Ava AG, Zurich, Switzerland** (B M Goodale PhD, V Kovacevic PhD, M Cronin PhD); **Julius Clinical Research BV, Zeist, Netherlands**

# Digital Public Health Surveillance

## Wearables



# Digital Public Health Surveillance

## Challenges

- We have seen many successful examples of digital public health surveillance. Many are academic demonstrations, and the integration into public health has been slow. Why?

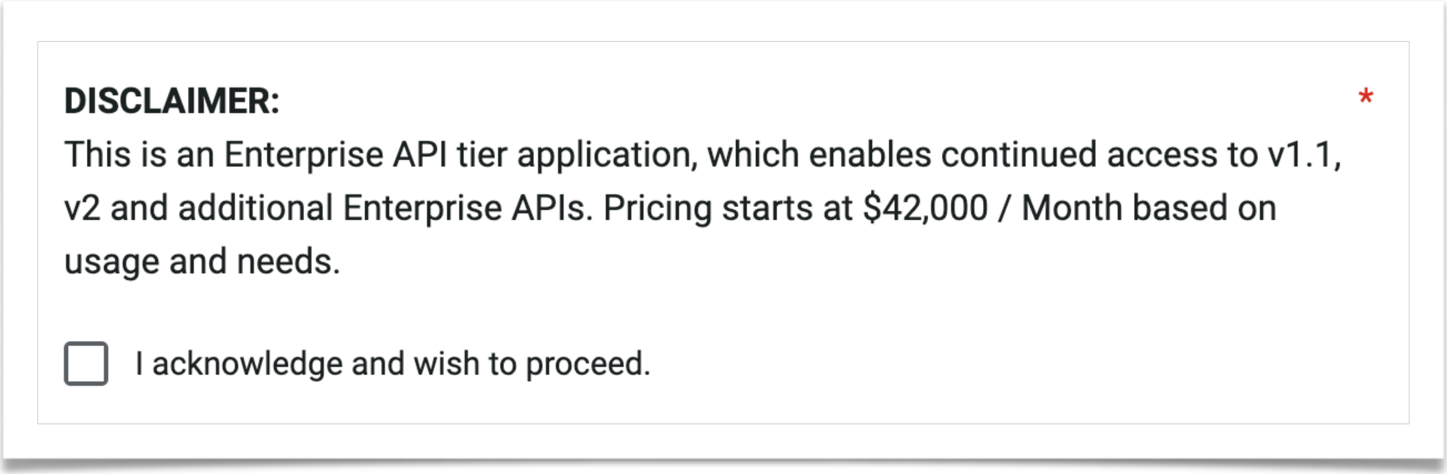
# Digital Public Health Surveillance

## Bringing PH & IC together

- The multidisciplinary challenge: different languages; different view on problems (e.g. noise in data).
- The stability challenge: public health wants to see demonstrated, continuous usability.
- The data challenge: public health does not like to depend on data sources that could be gone tomorrow.

# Digital Public Health Surveillance
## Data Availability

- Data availability: What if a company board decided tomorrow that access costs 500’000 per year?



**DISCLAIMER:** *

This is an Enterprise API tier application, which enables continued access to v1.1, v2 and additional Enterprise APIs. Pricing starts at $42,000 / Month based on usage and needs.

☐ I acknowledge and wish to proceed.

Source: Twitter

# Digital Public Health Surveillance

## Data Availability

- Data availability: Even if a company would be willing to share all data with public health, *should* they?

# Digital Public Health Surveillance

## Data & Algorithm Transparency

- Data & algorithm transparency: would you trust a system for your public health surveillance if the provider didn’t make the underlying algorithm available?

# Digital Public Health Surveillance

## Bias

- Bias: “a systematic distortion in the sampled data that compromises its representativeness”

- Selection bias
- Apps can be algorithmically confounded
- Our online self does not necessarily reflect our true self

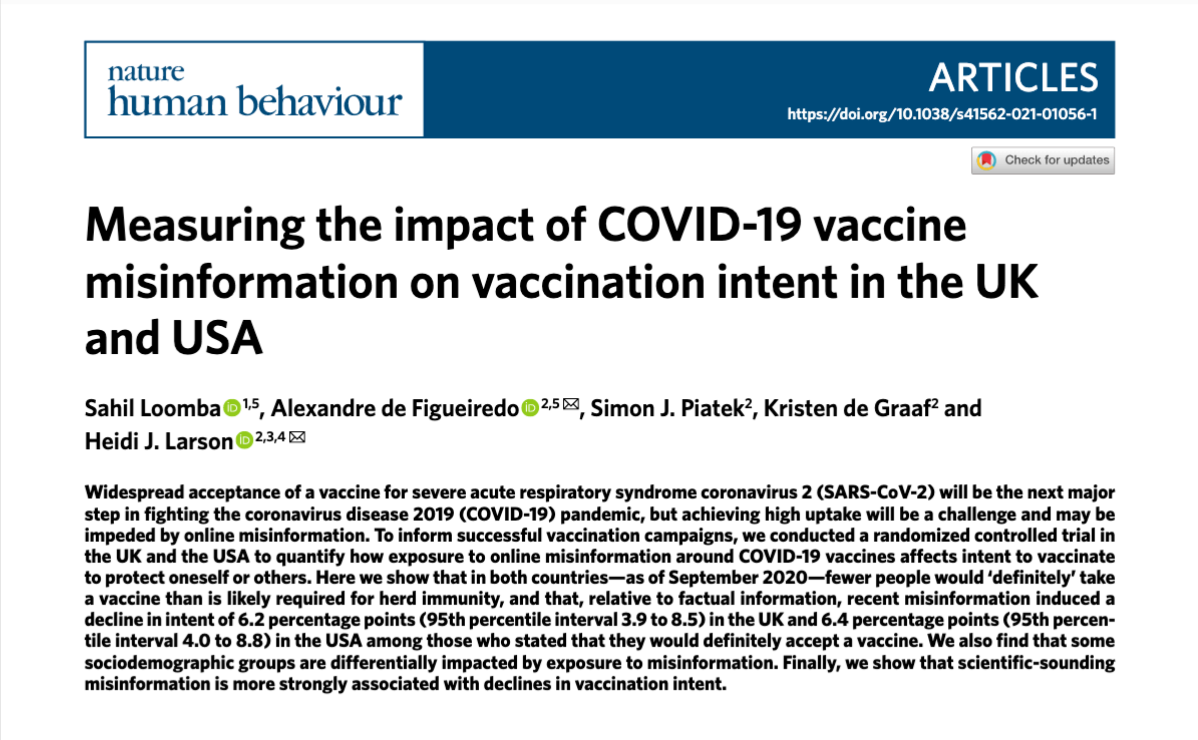
# Digital Public Health Surveillance
## Disinformation

- more than 1/3 (pre ChatGPT!) internet traffic is from bots

- “trolls”: account to sow discord

# Digital Public Health Surveillance

## Disinformation



nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-021-01056-1

Check for updates

### Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA

Sahil Loomba[1,5], Alexandre de Figueiredo[2,5] ✉, Simon J. Piatek[2], Kristen de Graaf[2] and Heidi J. Larson[2,3,4] ✉

**Widespread acceptance of a vaccine for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) will be the next major step in fighting the coronavirus disease 2019 (COVID-19) pandemic, but achieving high uptake will be a challenge and may be impeded by online misinformation. To inform successful vaccination campaigns, we conducted a randomized controlled trial in the UK and the USA to quantify how exposure to online misinformation around COVID-19 vaccines affects intent to vaccinate to protect oneself or others. Here we show that in both countries—as of September 2020—fewer people would 'definitely' take a vaccine than is likely required for herd immunity, and that, relative to factual information, recent misinformation induced a decline in intent of 6.2 percentage points (95th percentile interval 3.9 to 8.5) in the UK and 6.4 percentage points (95th percentile interval 4.0 to 8.8) in the USA among those who stated that they would definitely accept a vaccine. We also find that some sociodemographic groups are differentially impacted by exposure to misinformation. Finally, we show that scientific-sounding misinformation is more strongly associated with declines in vaccination intent.**

# Digital Public Health Surveillance

## The future

- AI everywhere
- Miniaturization
- Stronger connectivity