

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 Menni², Maxim Freydin², Thomas Varsavsky⁴, Carole H. Sudre⁴, M. Jorge Cardoso⁴, Sebastien Ourselin⁴, Jonathan Wolf⁵, Tim D. Spector^{2,5†}, Andrew T. Chan^{1,6†‡}, COPE Consortium§

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.

The exponentially increasing number of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infections has led to “an urgent need to expand public health activities in order to elucidate the epidemiology of the novel virus and characterize its potential impact” (1). Understanding risk factors for infection and predictors of subsequent outcomes is crucial to gain control of the coronavirus disease 2019 (COVID-19) pandemic (2). However, the speed at which the pandemic is unfolding poses an unprecedented challenge for the collection of exposure data to characterize the full breadth of disease severity, hampering efforts for timely dissemination of accurate information to affect public health planning and clinical management. Thus, there is an urgent need for an adaptable real-time data-capture platform to rapidly and prospectively collect actionable high-quality data that encompass the spectrum of subclinical and acute presentations and identify disparities in diagnosis, treatment, and clinical outcomes. Addressing this priority will allow for more accurate estimates of disease incidence, inform risk mitigation strat-

egies, facilitate allocation of scarce testing resources, and encourage appropriate quarantine and treatment of those afflicted.

An evolving body of literature suggests that COVID-19 incidence and outcomes vary according to age, sex, race, ethnicity, and underlying health status, with inconsistent evidence suggesting that commonly used medications—such as angiotensin-converting enzyme inhibitors, thiazolidinediones, and ibuprofen—may alter the natural disease course (3–9). Further, symptoms of COVID-19 vary widely, with fever and dry cough reportedly the most prevalent, though numerous investigations have demonstrated that asymptomatic carriage is a significant determinant of community spread (5–7, 10–13). In addition, the full spectrum of clinical presentation, which is still being characterized, may differ markedly across patient subgroups. Recent advisories from the American Gastroenterological Association, the American Academy of Otolaryngology–Head and Neck Surgery, and the British Geriatrics Society have emphasized the potential association between COVID-19 infection and previously underappreciated gastrointestinal symptoms (e.g., nausea, anorexia, and diarrhea), loss of taste and/or smell, and common geriatric syndromes (e.g., falls and delirium). The pandemic has considerably outpaced our collective efforts to fully characterize who is most at risk or may suffer the most serious sequelae of infection.

Mobile phone applications and web-based tools facilitate self-guided collection of population-level data at scale (14), the results of which can be rapidly redeployed to inform participants of urgent health information (14, 15). Both technologies are particularly advantageous when many individuals are advised to maintain physical distance from others (16). Such digital tools have already been applied

in more controlled research settings, and these studies benefit from greater lead time for field testing, question curation, and recruitment. Although many digital collection tools for COVID-19 are being developed and launched in the United States and abroad (see <http://mhealth-hub.org/mhealth-solutions-against-covid-19> for a continuously updated resource list from the European Union and the World Health Organization), including some in partnership with government health agencies such as the Centers for Disease Control and Prevention, most applications have largely been configured to offer a single assessment of symptoms to tailor semipersonalized recommendations for further evaluation. Infectious disease surveillance web-based tools (e.g., <http://flunearyou.org>) have been rapidly adapted for COVID-19–specific collection (e.g., <http://covidnearyou.org>). Alternatively, web portals have been developed for researchers to report patient-level information on behalf of participants already enrolled in clinical registries (e.g., ccc19.org). Integration with approaches that use remote data capture (e.g., wearable technology or symptom checkers such as real-time reporting thermometers) is also being considered. Although these approaches offer critical public health insights, they are often not tailored for the type of scalable longitudinal data capture that epidemiologists need to perform comprehensive, well-powered investigations.

To meet this challenge, we established a multinational collaboration, the COronavirus Pandemic Epidemiology (COPE) Consortium, composed of leading investigators from several large clinical and epidemiological cohort studies. COPE brings together a multidisciplinary team of scientists with expertise in big data research and translational epidemiology to investigate the COVID-19 pandemic in a large and diverse patient population. Several large cohorts have already agreed to join these efforts, including the Nurses' Health Study series, the Growing Up Today Study (GUTS), the Health Professionals Follow-Up Study (HPFS), TwinsUK, American Cancer Society Cancer Prevention Study 3 (CPS-3), the Multi-ethnic Cohort Study, the California Teachers Study (CTS), the Black Women's Health Study (BWHS), the Sister Study, Aspirin in Reducing Events in the Elderly (ASPREE), the Stanford Nutrition Studies, the Gulf Long-term Follow-up (GuLF) Study, the Agricultural Health Study, the National Institute of Environmental Health Sciences Environmental Polymorphisms Registry, the Predicting Progression of Developing Myeloma in a High-Risk Screened Population (PROMISE) study, and the Precursor Crowdsourcing (PCROWD) study. To aid in our data harmonization efforts in the United States, we developed the COVID Symptom Study (previously known as the COVID

¹Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, 55 Fruit St., Boston, MA 02114, USA. ²Department of Twin Research and Genetic Epidemiology, King's College London, Westminster Bridge Road, London SE1 7EH, UK. ³Department of Ageing and Health, Guys and St. Thomas' NHS Foundation Trust, Lambeth Palace Road, London SE1 7EH, UK. ⁴School of Biomedical Engineering & Imaging Sciences, King's College London, 1 Lambeth Palace Road, London SE1 7EU, UK. ⁵Zoe Global Limited, 164 Westminster Bridge Road, London SE1 7RW, UK. ⁶Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, 665 Huntington Ave., Boston, MA 02114, USA.

*These authors contributed equally to this work.

†These authors contributed equally to this work.

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

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

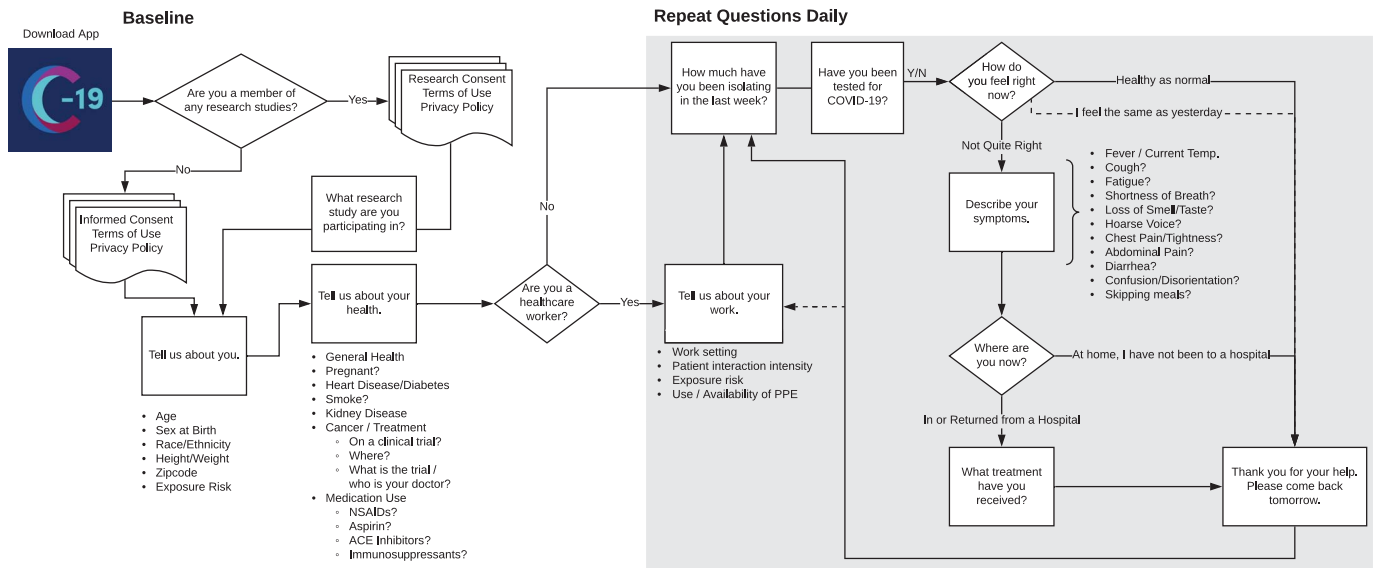


Fig. 1. Schematic of the participant workflow. After downloading the COVID Symptom Study app and providing consent, users are prompted to enter baseline demographic and clinical information and are serially queried about new or ongoing symptoms, testing results, and extent of isolation. Health care workers offer additional information about the intensity of their patient

interactions, potential exposure to infected patients, and use of PPE. With informed consent, users also participating in a variety of ongoing cohorts or clinical trials (Nurses' Health Study, TwinsUK, and others) have the option of linking COVID Symptom Study information to their extant research data.

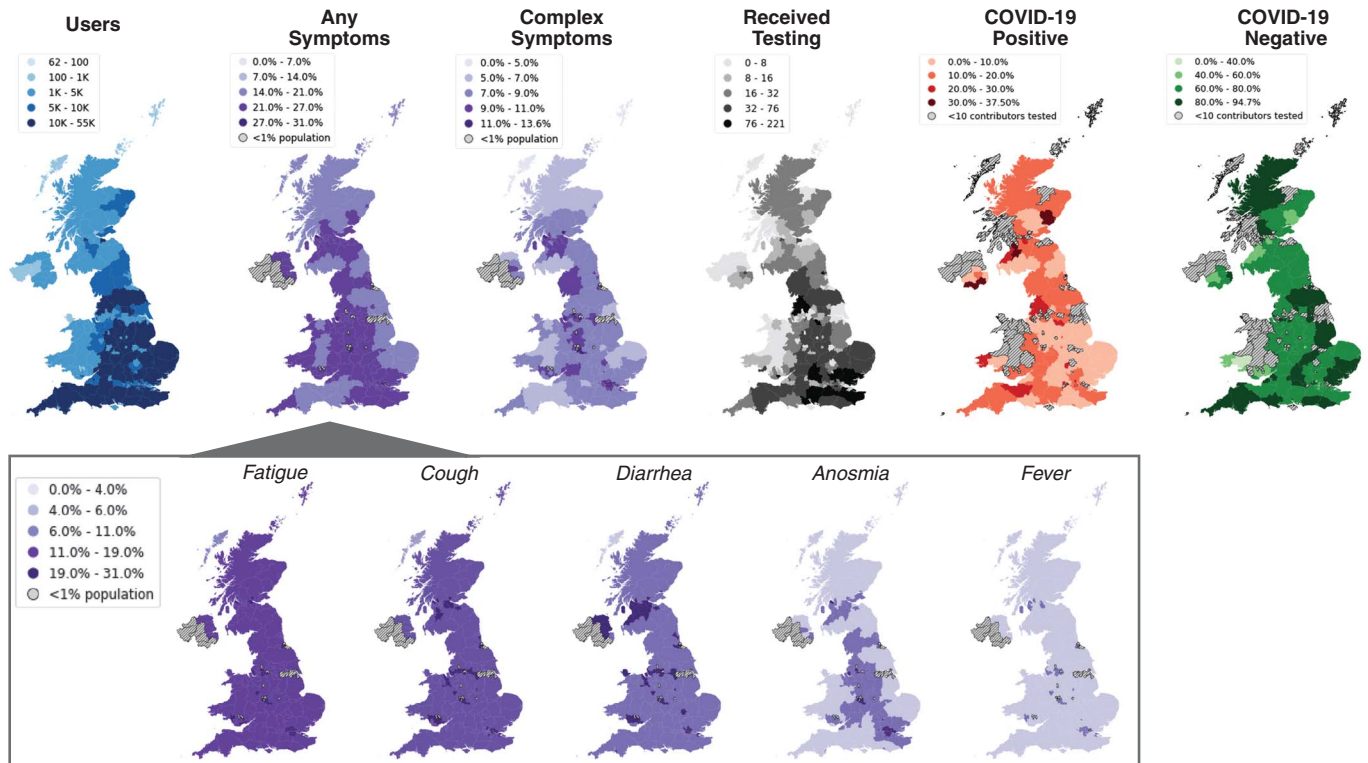
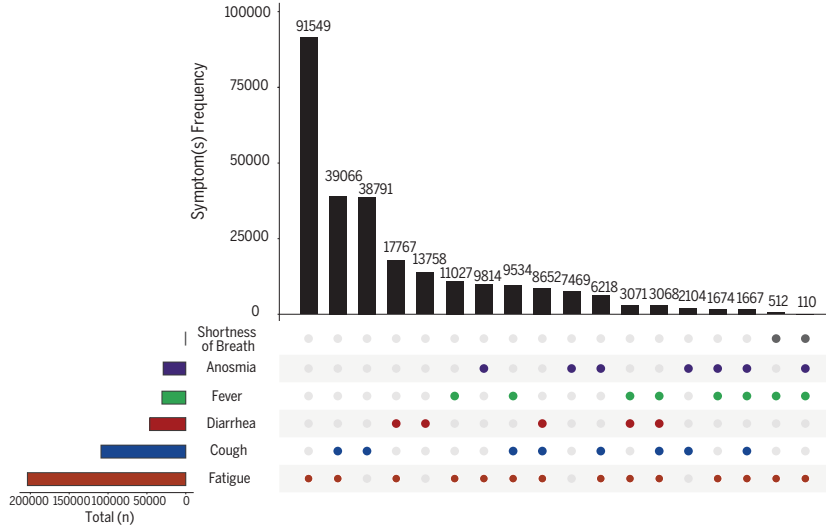


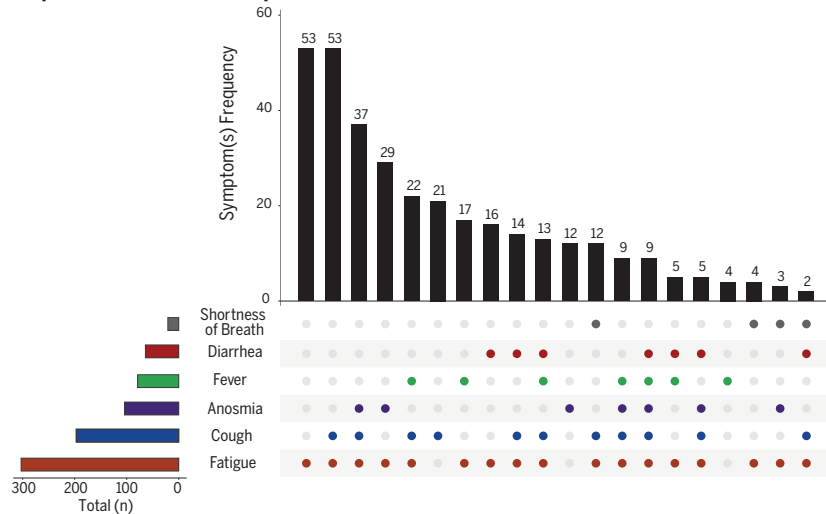
Fig. 2. COVID Symptom Study use, reported symptoms, and testing results according to geographic location in the United Kingdom. Between 24 and 29 March 2020, more than 1.6 million unique individuals downloaded the application and shared clinical and demographic information, as well as daily symptoms and high-intensity occupational exposures (blue map). Population density of those presenting with any symptoms (left purple map) varied according to region, with widespread reports of fatigue, cough, and diarrhea, followed by anosmia and, relatively infrequently, fever (inset). Examination of individuals who reported complex symptoms (right purple map), defined as having cough or fever and at least

one other symptom (diarrhea, anosmia, or fever), reveals areas that potentially need more testing. For the subset of the population that received a COVID-19 test (black map), areas with larger proportions of positive tests (orange map) appear to coincide with areas in which high proportions of the population reported complex symptoms. By contrast, some areas with low prevalence of complex symptoms have received higher rates of testing and, consequently, more negative tests (green map). This example of real-time visualization of data captured by the COVID Symptom Study may help public health and government officials reallocate resources, identify areas with unmet testing needs, and detect emerging hotspots.

All participants who have reported any COVID-related symptoms



Participants who tested COVID positive



Participants who tested COVID negative

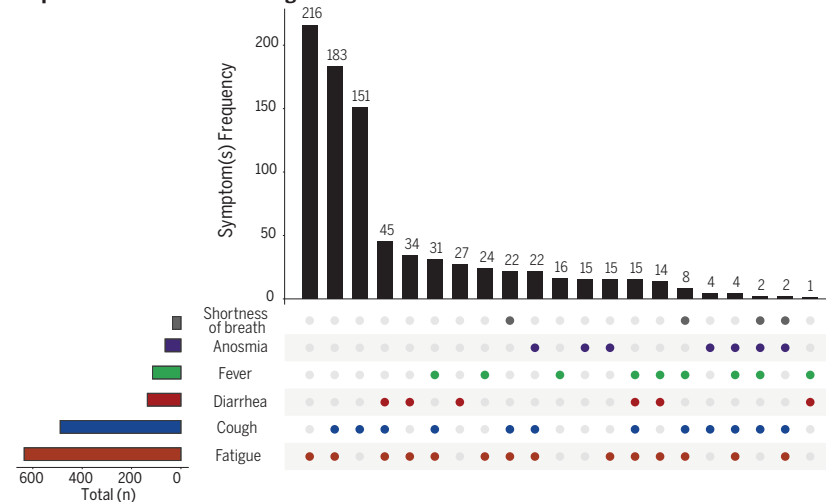


Fig. 3. Symptoms reported through the COVID Symptom Study app. By 27 March 2020, 265,851 individuals in the United Kingdom reported any symptom potentially associated with COVID-19 (top). Participants provided data on whether they were tested for COVID-19, as well as the test result. 1176 individuals reported having received a COVID-19 test (0.4% of those with symptoms). Symptom frequencies among those who tested positive (middle; $n = 340$) versus negative (bottom; $n = 836$) are shown.

Symptom Tracker) mobile app with in-kind contributions from Zoe Global Ltd., a digital health care company, and academic scientists from Massachusetts General Hospital and King’s College London. By leveraging the established digital backbone of an application used for personal nutrition studies, the COVID Symptom Study app was launched in the United Kingdom on 24 March 2020 and became available in the United States on 29 March 2020 (<https://covid.joinzoe.com/us>). The COPE Consortium is committed to the shared international pursuit of combating COVID-19 and has worked with scientific collaborators and thought leaders in real-time epidemiology to prioritize data harmonization and sharing as part of the Coronavirus Census Collective (17).

The COVID Symptom Study enables self-reporting of data related to COVID-19 exposure and infections (Fig. 1). On first use, the app queries location, age, and core health risk factors. Daily prompts query for updates on interim symptoms, health care visits, and COVID-19 testing results. For those self-quarantining or seeking health care, the level of intervention and related outcomes are collected. Individuals without obvious symptoms are also encouraged to use the app. Through pushed software updates, we can add or modify questions in real time to test emerging hypotheses about COVID-19 symptoms and treatments. Notably, participants enrolled in ongoing epidemiologic studies, clinical cohorts, or clinical trials can provide informed consent to link survey data collected through the app to preexisting study cohort data and any relevant biospecimens in a Health Insurance Portability and Accountability Act (HIPAA)-compliant and General Data Protection Regulation (GDPR)-compliant manner. A specific module is also provided for health care workers to determine the intensity and type of their direct patient care experiences, the availability and use of personal protective equipment (PPE), and work-related stress and anxiety.

Through rapid deployment of this tool, we can gain key insights into population dynamics of the disease (Fig. 2). By collecting participant-reported geospatial data, highlighted as a critical need for pandemic epidemiologic research (15), we can rapidly identify populations with highly prevalent symptoms in regions that may emerge as outbreak hotspots. An early snapshot of the first 1.6 million users in the United Kingdom over the first 5 days of use confirms the variability in symptoms reported across suspected COVID-19 cases and is useful for generating and testing broader hypotheses. At the time, users had a mean age of 41, ranged from 18 to 90 years old, and were 75% female. Graphic visualization of our initial results (Fig. 3) demonstrates that among those reporting symptoms by 27 March 2020 ($n =$

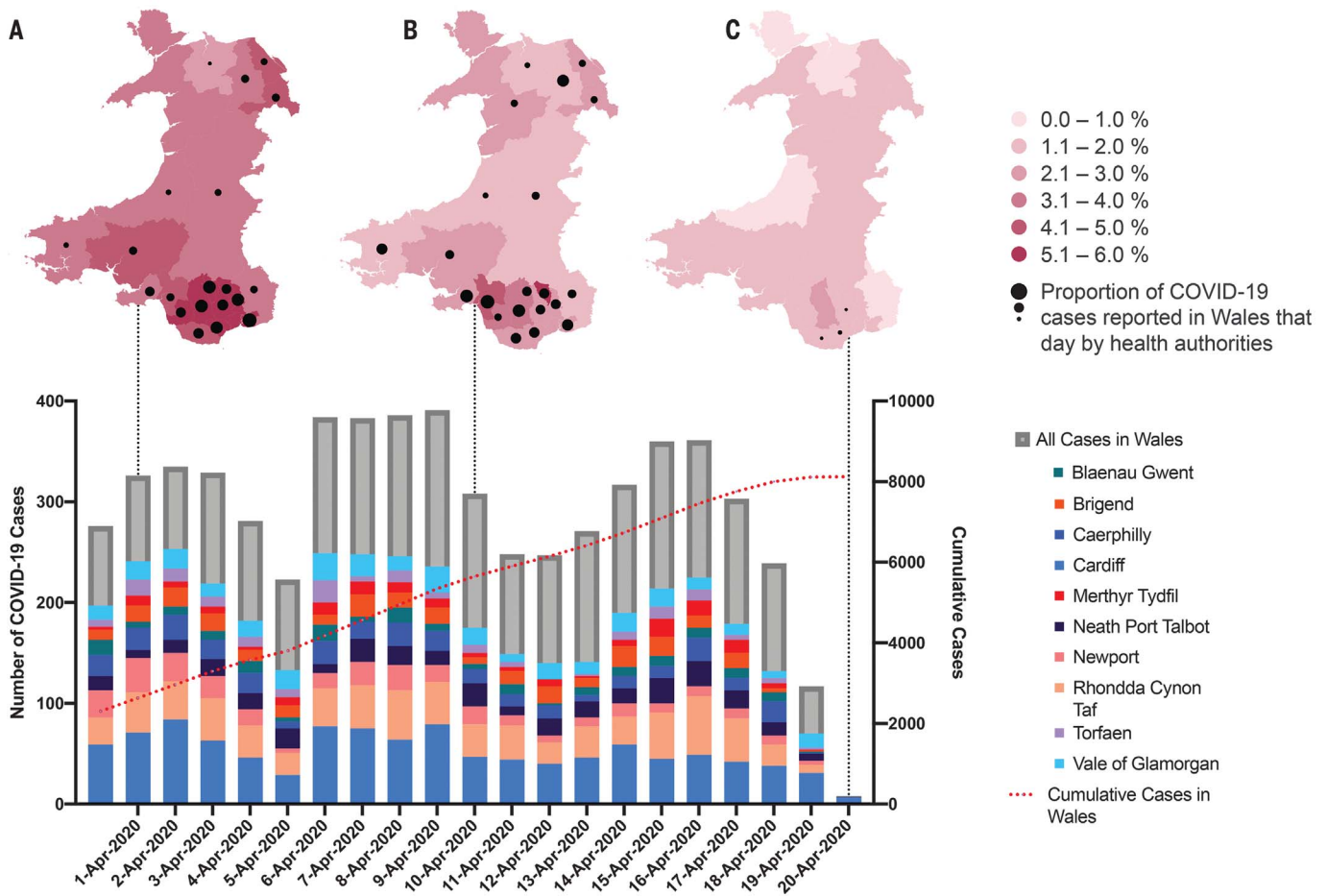


Fig. 4. Predicting COVID-19 cases on the basis of real-time symptom reporting in Wales, United Kingdom. This time series (bar graph) displays the number of new confirmed cases (gray bars) reported by the Public Health Wales NHS Trust between 31 March 2020 and 20 April 2020. After 2 April, case numbers appear to have declined through 5 April. However, our symptom-based prediction model (18), developed from symptom reports from untested users of the COVID Symptom Study app, showed a high proportion of predicted COVID-19 cases in southern Wales on 1 April [dark red areas in (A)]. Six days later, Welsh health authorities reported a subsequent peak in cases over a 4-day period (6 to 9 April), driven primarily by these southern regions (colored bars). By 10 April, new confirmed cases across Wales declined. However, on the basis of reported symptoms (B), regions in South Wales still had high predicted levels of COVID-19, which became apparent as a

second spike in confirmed cases between 15 and 16 April. As of 20 April (C), predicted COVID-19 prevalence across Wales according to symptom reporting appears to be low, which corresponds to a flattening of the cumulative incidence curve. However, several regions in southern Wales still have relatively high reports of symptoms and appear at risk for subsequent cases of COVID-19. Black dots on the maps represent the relative proportion of positive tests reported by health authorities across Wales that day by region. The prediction mapping included data from 1,339,670 users of the COVID Symptom Study on 1 April; 998,244 users on 10 April; and 1,234,918 users on 20 April. Public Health Wales NHS Trust data were current as of 21 April 2020 at 13:00 local time, taken from the “Rapid COVID-19 Virology - Public” dashboard (accessed via <https://phw.nhs.wales/>), and downloaded on 22 April 2020 at 12:30 p.m. Eastern Standard Time.

265,851 individuals), the most common symptoms were fatigue and cough, followed by diarrhea, fever, and anosmia. Shortness of breath was reported relatively rarely. Only 0.4% ($n = 1176$) of individuals reporting possible COVID-19 symptoms reported receiving a quantitative polymerase chain reaction test for COVID-19.

A comparison of symptomatic users who reported receiving a test within the initial launch period generated several hypotheses for future study with the growing dataset. The frequency of cough or fatigue alone or in com-

bination commonly led to testing but was not a particularly accurate predictor of a positive test. Similarly, no individuals who reported diarrhea in the absence of other symptoms tested positive. Notably, more complex presentations with cough and/or fatigue and at least one additional symptom, including less commonly appreciated complaints such as diarrhea and anosmia, appeared to be enriched among those with positive test results relative to those with negative results. In particular, anosmia may be a more predictive symptom, as it was more common than fever in indi-

viduals who tested positive. Indeed, in subsequent analyses with a larger sample set, we have shown that anosmia appears to be a strong predictor for COVID-19 (18). By contrast, fever alone was not particularly discriminatory. However, when fever was present in combination with less appreciated symptoms, a greater frequency of positive tests was observed. These findings suggest that perhaps individuals with complex or multiple-symptom (three or more) presentations should be prioritized for testing. Concerningly, 20% of individuals reported complex symptoms

(cough and/or fatigue plus at least one of anosmia, diarrhea, or fever) but had not yet been tested, representing a substantial population that appears to be at elevated risk for the disease. Additional work is warranted to confirm whether complex or multiple-symptom cases can accurately predict COVID-19 incidence.

Building on these initial findings, our team subsequently developed a weighted prediction model based on the symptoms of more than 2 million individual app users (18). By using this prediction model, we demonstrate the potential utility of the COVID Symptom Study app to collect data for long-term studies as well as for immediate public health planning. In southern Wales in the United Kingdom, users reported symptoms that predicted, 5 to 7 days in advance, two spikes in the number of confirmed positive COVID-19 cases reported by public health authorities (Fig. 4). Conversely, a decline in reports of symptoms preceded a drop in confirmed cases by several days. These results demonstrate that this app prospectively captures the dynamics of COVID-19 incidence days in advance of traditional measures, such as positive tests, hospitalizations, or mortality. We are currently planning additional studies using a broadly representative sample of individuals who will undergo uniform COVID-19 testing to further validate our approach to symptom-based modeling of incidence. These data demonstrate compelling evidence for the potential predictive power of our approach, which will improve as more data are collected to inform the model. Further, our data highlight the potential utility of real-time symptom tracking to help guide allocation of resources for testing and treatment as well as recommendations for lockdown or easement in specific areas.

With additional data collection, we will also apply big data approaches (e.g., machine learning) to identify emerging patterns in dynamic settings of exposure, onset of symptoms, disease trajectory, and clinical outcomes. Our launch of the app within several large epidemiology cohorts that have previously gathered longitudinal data on lifestyle, diet and health factors, and genetic information will allow investigation of a much broader range of putative risk factors for COVID-19 outcomes. With additional follow-up, we will also be positioned to investigate long-term effects of COVID-19, including mental health, disability, mortality, and financial outcomes. Mobile technology can also supplement recently launched clinical trials or biobanking protocols already embedded within clinical settings. In collaboration with the Stand Up to Cancer foundation, we have also developed a strategy to track information among individuals living with cancer, including those

enrolled in clinical trials. At the Massachusetts General Hospital and Brigham and Women's Hospital, we are deploying the tool within several clinical studies, centralized biobanking efforts, and health care worker surveillance programs. Health care workers are particularly vulnerable to COVID-19's effects beyond infection, including work hazards from PPE shortages, emotional stress, and absenteeism. Real-time data generation focused within these populations will be critical to optimally allocate resources to protect our health care workforce and assess its efficacy.

Even so, our approach has limitations. We recognize that a smartphone application does not represent a random sampling of the population. However, this is an inherent limitation of any epidemiologic study that relies on voluntary participation. Our approach has the benefit of allowing rapid deployment across a large cross section of the population during a major public health crisis. With time and continued use, the large number of participants will include a sufficient quantity of users within key subgroups such that we can adjust our methodology for potential sources of confounding. By engaging cohorts with underrepresented populations, such as the BWHS in the United States, we also hope to leverage existing investigator-participant relationships to encourage enrollment of individuals from populations that have traditionally been challenging to recruit. Moreover, by encouraging longitudinal, prospective data collection, we can capture associations based on within-person variation over time, a notable advantage over repeated cross-sectional surveys that introduce considerable between-person variation. In the near future, we hope to release our app as fair-use open source software to facilitate translation and development in other regions. We have begun working with colleagues in Canada, Australia, and Sweden to implement this tool within their countries. We have also developed a practical toolkit to assist clinical researchers with local institutional review board and regulatory approval to facilitate deployment within research studies (www.monganinstitute.org/cope-consortium). This toolkit includes full details of the mobile app's questions, consent documents, privacy policies, and terms of use. With broader implementation, data generated from the COVID Symptom Study app are increasingly being linked to the public health response within the National Health Service (NHS) in the United Kingdom. The app is endorsed by the Welsh government, NHS Wales, the Scottish government, and NHS Scotland, and our scientific team updates the U.K. chief scientific officer daily. We are working to develop a similar approach in the United States. However, the lack of a national health care system has required a strategy focused

on engaging local public health leaders. For example, we have partnered with the University of Texas School of Public Health to conduct statewide surveillance to support public health decision-making, especially as the Texas state government begins softening mitigation strategies.

Our approach demonstrates a proof of concept for rapid repurposing of existing data collection methods to implement scalable real-time collection of population-level data during a fast-moving global health crisis. We call on our colleagues to work with us so that we may deploy all of the tools at our disposal to address this unprecedented public health challenge.

REFERENCES AND NOTES

1. M. Lipsitch, D. L. Swerdlow, L. Finelli, *N. Engl. J. Med.* **382**, 1194–1196 (2020).
2. G. A. FitzGerald, *Science* **367**, 1434 (2020).
3. M. Day, *BMJ* **368**, m1086 (2020).
4. L. Fang, G. Karakiulakis, M. Roth, *Lancet Respir. Med.* **8**, e21 (2020).
5. W. J. Guan et al., *N. Engl. J. Med.* **382**, 1708–1720 (2020).
6. D. Wang et al., *JAMA* **323**, 1061 (2020).
7. C. Wu et al., *JAMA Intern. Med.* 10.1001/jamainternmed.2020.0994 (2020).
8. X. Yang et al., *Lancet Respir. Med.* 10.1016/S2213-2600(20)30079-5 (2020).
9. F. Zhou et al., *Lancet* **395**, 1054–1062 (2020).
10. N. Chen et al., *Lancet* **395**, 507–513 (2020).
11. C. Huang et al., *Lancet* **395**, 497–506 (2020).
12. L. Pan et al., *Am. J. Gastroenterol.* **115**, 766–773 (2020).
13. X. W. Xu et al., *BMJ* **368**, m606 (2020).
14. J. S. Brownstein, C. C. Freifeld, L. C. Madoff, *N. Engl. J. Med.* **360**, 2153–2157 (2009).
15. B. Xu, M. U. G. Kraemer; Open COVID-19 Data Curation Group, *Lancet Infect. Dis.* **20**, 534 (2020).
16. P. J. Lyons, "Coronavirus Briefing: What Happened Today," *The New York Times* (30 March 2020).
17. E. Segal et al., *Nat. Med.* 10.1038/s41591-020-0929-x (2020).
18. C. Menni et al., *Nat. Med.* 10.1038/s41591-020-0916-2 (2020).
19. D. A. Drew et al., medRxiv [Preprint], 6 April 2020; <https://doi.org/10.1101/2020.04.02.20051334>.
20. MGHcteu, MGHcteu/ScienceCOPEMethodsCode: Science v1.0.1, Version 1.0.1, Zenodo (2020); <https://doi.org/10.5281/zenodo.3765955>.

ACKNOWLEDGMENTS

Use of the COVID-19 app in cohorts and informed consent as described was approved by the Partners Human Research Committee (Protocol 2020P000909) and is registered on ClinicalTrials.gov as NCT04331509. We acknowledge J. Capdevila Pujol of Zoe Global Ltd. for his contributions to this work. **Funding:** Zoe Global Ltd. provided in kind support for all aspects of building, running, and supporting the tracking app and service to all users worldwide. C.J.S., C.M., M.F., T.V., C.H.S., M.J.C., S.O., and T.D.S. were supported by the Wellcome Trust and EPSRC (WT212904/Z/18/Z, WT203148/Z/16/Z, and WT213038/Z/18/Z), the NIHR GSTT/KCL Biomedical Research Centre, MRC/BHF (MR/M016560/1), the NIHR, and the Alzheimer's Society (AS-JF-17-011). A.T.C. is the Stuart and Suzanne Steele MGH Research Scholar and Stand Up to Cancer scientist. L.H.N., D.A.D., and A.T.C. were supported by the Massachusetts Consortium on Pathogen Readiness (MassCPR), M. Schwartz, and L. Schwartz. **Author contributions:** D.A.D. and L.H.N. drafted the manuscript. All authors contributed to the conceptualization, methodology, formal analysis, investigation, resources, data curation, and review and editing of the manuscript. A.T.C. and T.D.S. supervised the study and acquired funding. **Competing interests:** T.D.S. is a consultant for Zoe Global Ltd. J.W. is an employee of Zoe Global Ltd. D.A.D. and A.T.C. previously served as investigators in a clinical trial of diet and lifestyle using a

distinct mobile application that was supported by Zoe Global Ltd. The other authors have no conflicts of interest to declare. **Data and materials availability:** This manuscript was previously made available as a preprint on medRxiv.org (19). Code used to generate maps is published and available in Zenodo (20). Data collected in the app are being shared with other health researchers through the NHS-funded Health Data Research UK (HDRUK)/SAIL Consortium, housed in the UK Secure e-Research Platform (UKSeRP) in Swansea, Wales. Anonymized data are available to be shared with bona fide researchers through HDRUK according to their protocols in the public interest. See <https://healthdatagateway.org/detail/9b604483-9cdc-41b2-b82c-14ee3dd705f6>. U.S. investigators are encouraged to coordinate data

requests through the COPE Consortium (www.monganinstitute.org/cope-consortium). Data updates can be found on <https://covid.joinzoe.com>. For aggregate deidentified "snapshot" datasets used for analyses provided here, please request time-stamped datasets by email to predict@mgh.harvard.edu for data files, "COVIDSymptomTrackerData_03292020" (Figs. 2 and 3), or "COVIDSymptomTrackerData_04/21/20" (Fig. 4). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>. This license does not apply to figures/photos/artwork or other content included in the

article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6497/1362/suppl/DC1
Figs. S1 to S9
COPE Consortium Members
MDAR Reproducibility Checklist
2 April 2020; accepted 30 April 2020
Published online 5 May 2020
10.1126/science.abc0473