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

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

David A. Drew^{1*}, Long H. Nguyen^{1*}, Claire J. Steves^{2,3}, Cristina Menni², Maxim Freydin², Thomas Varsavsky⁴, Carole H. Sudre⁴, M. Jorge Cardoso⁴, Sebastien Ourselin⁴, Jonathan Wolf⁴, Tim D. Spector^{2,4†}, Andrew T. Chan^{1,5,6†‡}, COPE Consortium§

¹Clinical & Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, 55 Fruit St., Boston, MA 02114, USA. ²Department of Twin Research and Genetic Epidemiology, King's College London, Westminster Bridge Road, London SE17EH, UK. ³Department of Ageing and Health, Guys and St. Thomas's NHS Foundation Trust, Lambeth Palace Road, London SE17EH, UK. ⁴School of Biomedical Engineering & Imaging Sciences, King's College London, 1 Lambeth Palace Road, London SE1 7EU, UK. ⁵Zoe Global Limited, 164 Westminster Bridge Road, London SE17RW, UK. ⁶Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, 665 Huntington Avenue, Boston, MA 02114, USA.

*These authors contributed equally to this work.

†These authors contributed equally to this work.

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

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

The rapid pace of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic (COVID-19) 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 bring together scientists with expertise in big data research and epidemiology to develop a COVID-19 Symptom Tracker mobile application that we launched in the UK on March 24, 2020 and the US on March 29, 2020 garnering more than 2.8 million users as of May 2, 2020. This mobile application offers data on risk factors, herald symptoms, clinical outcomes, and geographical hot spots. This initiative offers critical 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 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 critical 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 to collecting exposure data characterizing the full breadth of disease severity, hampering efforts to disseminate accurate information in a timely manner to impact 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 encompasses the spectrum of subclinical and acute presentations while identifying disparities in diagnosis, treatment, and clinical outcomes. Addressing this priority will allow for more accurate estimates of disease incidence, inform risk mitigation strategies, more effectively allocate still-scarce testing resources, and allow for appropriate quarantine and treatment of those afflicted.

An evolving body of literature suggests 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 (ACE) inhibitors, thiazolidinediones (TZD), 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 is still being characterized, which may significantly differ across patient subgroups, as evidenced by recent advisories by the American Gastroenterological Association (AGA) and the American Academy of Otolaryngology - Head and Neck Surgery (AAO-HNS), and British Geriatric Society (BGS) on the potential importance of previously underappreciated gastrointestinal symptoms (e.g., nausea, anorexia, and diarrhea) or loss of taste and/or smell associated with COVID-19 infection, as well as common geriatric syndromes (e.g., falls and delirium). The pandemic has dramatically outpaced our collective efforts to fully characterize who is most at-risk or may suffer the most serious sequelae of infection.

Mobile phone applications or web-based tools facilitate self-guided collection of population-level data at scale (14), the results of which can then be rapidly redeployed to inform participants of urgent health information (14, 15). Both are particularly advantageous when many Americans are advised to physically distance (16). Such digital tools have already been applied in more controlled research settings which benefit from greater lead time for field testing, question curation, and recruitment. Although an increasing number of digital collection tools for COVID-19 are being developed and launched in the U.S. and abroad (see <http://mhealth-hub.org/mhealth-solutions-against-covid-19> for a continuously updated resource list from the European Union and WHO), including some in partnership with government health agencies such as the Centers for Disease Control and Prevention (CDC), most applications have largely been configured to offer a single assessment of symptoms to tailor semi-personalized 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, others have developed web portals for researchers to report patient-level information on behalf of participants already enrolled in clinical registries (e.g., ccc19.org). Integration with approaches that utilize remote data capture (e.g., wearables or symptom checkers such as real-time reporting thermometers) are also being considered. Although each of 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, comprised 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 interrogate the COVID-19 pandemic in the largest and most diverse patient population assembled to-date. Several large cohorts have already agreed to join these efforts, including the Nurses' Health Study (NHS), NHSII, NHS3, the Growing Up Today Study (GUTS), the Health Professionals Follow-Up Study (HPFS), TwinsUK, American Cancer Society Cancer Prevention Study 3 (CPS-3), the Multiethnic 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 NIEHS Environmental Polymorphisms Registry, and the Predicting Progression of Developing Myeloma in a High-Risk Screened Population (PROMISE) and Precursor Crowdsourcing (PCROWD)

Studies. To aid in our data harmonization efforts in the US, we co-developed the COVID Symptom Tracker mobile app in collaboration with in-kind contributions from Zoe Global Ltd, a digital healthcare 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 Tracker was launched in the UK on March 24, 2020, and became available in the U.S. on March 29, 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 a part of the Coronavirus Census Collective (17).

The COVID Symptom Tracker enables self-report 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. In 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. Importantly, participants enrolled in ongoing epidemiologic studies, clinical cohorts, or clinical trials can provide informed consent to link survey data collected through the app in a Health Insurance Portability and Accountability Act (HIPAA)- and General Data Protection Regulation (GDPR)-compliant manner to their pre-existing study cohort data and any relevant biospecimens. A specific module is also provided for participants who identify as healthcare 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 critical 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 that may emerge as hot spots for outbreaks. An early snapshot of the first 1.6 million users in the UK over the first five 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 were a mean age of 41 with a range from 18 to 90 years, with 75% female. Graphic visualization of initial results (Fig. 3) demonstrates that among those reporting symptoms by March 27, 2020 (n = 265,851) the most common symptoms were fatigue and cough, followed by diarrhea, fever, and

anosmia. Shortness of breath was relatively rarely reported. Only 0.4% ($n = 1,176$) of individuals reporting possible COVID-19 symptoms reported receiving a qPCR test for COVID-19.

Comparing users with symptoms who reported testing within the initial launch period generated several hypotheses for future study using the growing dataset. The frequency of cough and fatigue alone or in combination appeared to commonly lead to testing, but did not appear to be particularly sensitive for a positive test. Similarly, no individuals reporting diarrhea in the absence of other symptoms tested positive. Interestingly, 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 compared to negative results. In particular, anosmia may be a more sensitive symptom as it was more common than fever in individuals 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). In contrast, fever alone was not particularly discriminatory; however, in combination with lesser appreciated symptoms, a greater frequency of positive tests was observed. These findings suggest that individuals with complex or multiple (3 or more) symptomatic presentation perhaps 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 received testing, representing a substantial population who appear to be at greater risk for the disease. Additional work is warranted to confirm if complex or multiple (3 or more) symptomatic cases may accurately predict COVID incidence.

Based on these initial findings, our team subsequently developed a weighted prediction model based on these symptoms trained on more than 2 million individuals using the app (18). Using this prediction model, we demonstrate the potential utility of the COVID Symptom Tracker to collect data not only for long-term studies, but also for immediate public health planning. In Southern Wales in the United Kingdom, users reported symptoms that predicted, five to seven days in advance, two spikes in the number of individuals reported by public health authorities to be confirmed with COVID (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 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, they 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 novel patterns that emerge 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 uniquely positioned to investigate long-term outcomes 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 healthcare worker surveillance programs. Healthcare workers are a particularly vulnerable population 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 healthcare workforce and assess their efficacy.

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 which relies on voluntary participation. However, our approach has the benefit of allowing rapid deployment across a large cross-section of the population during an unprecedented health crisis. With time and continued use, the large number of participants will include a sufficient number of users within key subgroups that will allow for adjustment for potential sources of confounding. By engaging cohorts with underrepresented populations, such as the BWHS in the U.S., we also hope to leverage existing investigator-participant relationships to encourage enrollment of individuals who are traditionally more challenging to recruit. Moreover, by encouraging longitudinal, prospective data collection, we can capture associations based on within-person variation over time, a significant advantage over repeated cross-sectional surveys that introduce significant 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 for clinical researchers to facilitate local Institutional Review Board (IRB) and regulatory approval to facilitate deployment within research studies (www.monganinstitute.org/cope-consortium). This toolkit includes full detail of the questions, consent documents, privacy policies, and terms of use for the mobile app. With broader implementation, data generated from the COVID Symptom Tracker app have become increasingly linked to the public health response within the National Health Service in the UK. The app is endorsed by the Welsh Government, NHS Wales, the Scottish Government and NHS Scotland. Our scientific team update the UK Chief Scientific Officer daily. We are working to develop a similar approach in the US. However, the lack of a national healthcare 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 state-wide surveillance to support public health decision making, especially as their state government begins softening mitigation strategies.

In summary, our novel approach demonstrates critical proof-of-concept for rapid repurposing of existing data collection approaches to implement scalable real-time collection of population-level data during a fast-moving global health crisis and National Emergency. We call upon our colleagues to work with us so that we may deploy all the tools at our disposal to address this unprecedented public health challenge.

REFERENCES AND NOTES

1. M. Lipsitch, D. L. Sverdlow, L. Finelli, Defining the Epidemiology of Covid-19 - Studies Needed. *N. Engl. J. Med.* **382**, 1194–1196 (2020). [doi:10.1056/NEJMp2002125](https://doi.org/10.1056/NEJMp2002125) Medline
2. G. A. FitzGerald, Misguided drug advice for COVID-19. *Science* **367**, 1434 (2020). [doi:10.1126/science.abb8034](https://doi.org/10.1126/science.abb8034) Medline
3. M. Day, Covid-19: Ibuprofen should not be used for managing symptoms, say doctors and scientists. *BMJ* **368**, m1086 (2020). [doi:10.1136/bmj.m1086](https://doi.org/10.1136/bmj.m1086) Medline
4. L. Fang, G. Karakiulakis, M. Roth, Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir. Med.* **8**, e21 (2020). [doi:10.1016/S2213-2600\(20\)30116-8](https://doi.org/10.1016/S2213-2600(20)30116-8) Medline
5. W. J. Guan, Z. Y. Ni, Y. Hu, W. H. Liang, C. Q. Ou, J. X. He, L. Liu, H. Shan, C. L. Lei, D. S. C. Hui, B. Du, L. J. Li, G. Zeng, K.-Y. Yuen, R. C. Chen, C. L. Tang, T. Wang, P. Y. Chen, J. Xiang, S. Y. Li, J. L. Wang, Z. J. Liang, Y. X. Peng, L. Wei, Y. Liu, Y. H. Hu, P. Peng, J. M. Wang, J. Y. Liu, Z. Chen, G. Li, Z. J. Zheng, S. Q. Qiu, J. Luo, C. J. Ye, S. Y. Zhu, N. S. Zhong; China Medical Treatment Expert Group for Covid-19, Clinical Characteristics of Coronavirus Disease 2019 in China. *N. Engl. J. Med.* **382**, 1708–1720 (2020). [doi:10.1056/NEJMoa2002032](https://doi.org/10.1056/NEJMoa2002032) Medline
6. D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, Z. Peng, Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* **323**, 1061 (2020). [doi:10.1001/jama.2020.1585](https://doi.org/10.1001/jama.2020.1585) Medline
7. C. Wu, X. Chen, Y. Cai, J. Xia, X. Zhou, S. Xu, H. Huang, L. Zhang, X. Zhou, C. Du, Y. Zhang, J. Song, S. Wang, Y. Chao, Z. Yang, J. Xu, X. Zhou, D. Chen, W. Xiong, L. Xu, F. Zhou, J. Jiang, C. Bai, J. Zheng, Y. Song, Risk Factors Associated With Acute Respiratory Distress Syndrome and Death in Patients With Coronavirus Disease 2019 Pneumonia in Wuhan, China. *JAMA Intern. Med.* **10.1001/jamainternmed.2020.0994** (2020). [doi:10.1001/jamainternmed.2020.0994](https://doi.org/10.1001/jamainternmed.2020.0994) Medline
8. X. Yang, Y. Yu, J. Xu, H. Shu, J. Xia, H. Liu, Y. Wu, L. Zhang, Z. Yu, M. Fang, T. Yu, Y. Wang, S. Pan, X. Zou, S. Yuan, Y. Shang, Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: A single-centered, retrospective, observational study. *Lancet Respir. Med.* **10.1016/S2213-2600(20)30079-5** (2020). [doi:10.1016/S2213-2600\(20\)30079-5](https://doi.org/10.1016/S2213-2600(20)30079-5) Medline
9. F. Zhou, T. Yu, R. Du, G. Fan, Y. Liu, Z. Liu, J. Xiang, Y. Wang, B. Song, X. Gu, L. Guan, Y. Wei, H. Li, X. Wu, J. Xu, S. Tu, Y. Zhang, H. Chen, B. Cao, Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *Lancet* **395**, 1054–1062 (2020). [doi:10.1016/S0140-6736\(20\)30566-3](https://doi.org/10.1016/S0140-6736(20)30566-3) Medline
10. N. Chen, M. Zhou, X. Dong, J. Qu, F. Gong, Y. Han, Y. Qiu, J. Wang, Y. Liu, Y. Wei, J. Xia, T. Yu, X. Zhang, L. Zhang, Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *Lancet* **395**, 507–513 (2020). [doi:10.1016/S0140-6736\(20\)30211-7](https://doi.org/10.1016/S0140-6736(20)30211-7) Medline
11. C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, B. Cao, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **395**, 497–506 (2020). [doi:10.1016/S0140-6736\(20\)30183-5](https://doi.org/10.1016/S0140-6736(20)30183-5) Medline
12. L. Pan, M. Mu, P. Yang, Y. Sun, R. Wang, J. Yan, P. Li, B. Hu, J. Wang, C. Hu, Y. Jin, X. Niu, R. Ping, Y. Du, T. Li, G. Xu, Q. Hu, L. Tu, Clinical characteristics of COVID-19 patients with digestive symptoms in Hubei, China: A descriptive, cross-sectional, multicenter study. *Am. J. Gastroenterol.* **115**, 766–773 (2020). [doi:10.14309/ajg.0000000000000620](https://doi.org/10.14309/ajg.0000000000000620) Medline
13. X. W. Xu, X.-X. Wu, X.-G. Jiang, K.-J. Xu, L.-J. Ying, C.-L. Ma, S.-B. Li, H.-Y. Wang, S. Zhang, H.-N. Gao, J.-F. Sheng, H.-L. Cai, Y.-Q. Qiu, L.-J. Li, Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: Retrospective case series. *BMJ* **368**, m606 (2020). [doi:10.1136/bmj.m606](https://doi.org/10.1136/bmj.m606) Medline
14. J. S. Brownstein, C. C. Freifeld, L. C. Madoff, Digital disease detection—Harnessing the Web for public health surveillance. *N. Engl. J. Med.* **360**, 2153–2157 (2009). [doi:10.1056/NEJMp0900702](https://doi.org/10.1056/NEJMp0900702) Medline
15. B. Xu, M. U. G. Kraemer; Open COVID-19 Data Curation Group, Open access epidemiological data from the COVID-19 outbreak. *Lancet Infect. Dis.* **20**, 534 (2020). [doi:10.1016/S1473-3099\(20\)30119-5](https://doi.org/10.1016/S1473-3099(20)30119-5) Medline
16. P. J. Lyons, in *The New York Times* (New York, New York, 2020).
17. E. Segal, F. Zhang, X. Lin, G. King, O. Shalem, S. Shilo, W. E. Allen, Y. H. Grad, C. S. Greene, F. Alquaddoomi, S. Anders, R. Balicer, T. Bauman, X. Bonilla, G. Booman, A. T. Chan, O. Cohen, S. Coletti, N. Davidson, Y. Dor, D. A. Drew, O. Elemento, G. Evans, P. Ewels, J. Gale, A. Gavrieli, B. Geiger, I. Hajirasouliha, R. Jerala, A. Kahles, O. Kallioniemi, A. Keshet, G. Landua, T. Meir, A. Muller, L. H. Nguyen, M. Oresic, S. Ovchinnikova, H. Peterson, J. Rajagopal, G. Ratsch, H. Rossman, J. Rung, A. Sboner, A. Sigaras, T. Spector, R. Steinherz, I. Stevens, J. Vilo, P. Wilmes; CCC (Coronavirus Census Collective, Building an International Consortium for Tracking Coronavirus Health Status. medRxiv [Preprint]. 6 April 2020. <https://doi.org/10.1101/2020.04.02.20051284>.
18. C. Menni, A. M. Valdes, M. B. Freidin, S. Ganesh, J. S. El-Sayed Moustafa, A. Visconti, P. Hysi, R. C. E. Bowyer, M. Mangino, M. Falchi, J. Wolf, C. J. Steves, T. D. Spector, Loss of smell and taste in combination with other symptoms is a strong predictor of COVID-19 infection. medRxiv 7 April 2020. <https://doi.org/10.1101/2020.04.05.20048421>.
19. D. A. Drew, L. H. Nguyen, C. J. Steves, J. Wolf, T. D. Spector, A. T. Chan, COPE Consortium, Rapid implementation of mobile technology for real-time epidemiology of COVID-19. 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

This manuscript was previously made available as a pre-print on medRxiv.org under the DOI: 10.1101/2020.04.02.20051334 (19). 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. Code used to generate maps is published and available in Github (20). We also specifically acknowledge Joan Capdevila Pujol of Zoe Global Ltd. for his contributions to this work. **Funding:** Zoe provided in kind support for all aspects of building, running and supporting the tracking app and service to all users worldwide. CJS, CM, MF, TV, CHS, MJC, SO, TDS were supported by the Wellcome Trust and EPSRC (WT212904/Z/18/Z, WT203148/Z/16/Z, 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). ATC is the Stuart and Suzanne Steele MGH Research Scholar and Stand Up to Cancer scientist. LHN, DAD, and ATC were supported by the Massachusetts Consortium on Pathogen Readiness (MassCPR) and Mark and Lisa Schwartz. **Author contributions:** DAD and LHN drafted the manuscript. All authors contributed to the conceptualization, methodology, formal analysis, investigation, resources, data curation, and review and editing of the manuscript. ATC and TDS supervised the study and acquired funding.

Competing interests: TDS is a consultant to Zoe Global Ltd. JW is an employee of Zoe Global Ltd. DAD and ATC previously served as investigators on a clinical trial of diet and lifestyle using a distinct mobile application that was supported by Zoe Global Ltd. Other authors have no conflict of interest to declare. **Data and materials availability:** Data collected in the app is being shared with other health researchers through the NHS funded Health Data Research UK (HDRUK)/SAIL consortium, housed in the UK Secure Research Platform (UKSeRP) in Swansea.

Anonymized data are available to be shared with bona fide researchers HDRUK according to their protocols in the public interest. See <https://healthdatagateway.org/detail/9b604483-9cdc-41b2-b82c-14ee3dd705f6>. US 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 timestamped datasets by email 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.

according to their protocols in the public interest. See <https://healthdatagateway.org/detail/9b604483-9cdc-41b2-b82c-14ee3dd705f6>. US 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 timestamped datasets by email 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/cgi/content/full/science.abc0473/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

Fig. 1. A schematic of participant workflow. After downloading the COVID Symptom Tracker and providing consent, users are prompted to provide baseline demographic and clinical information and are serially queried about new or ongoing symptoms, testing results, and how much they have been isolating. Health care workers offer additional information about the intensity of their patient interactions, potential exposure to infected patients, and usage of personal protective equipment. With informed consent, users already engaging in active research as participants in a variety of ongoing cohorts or clinical trials (i.e., Nurses' Health Study, TwinsUK, etc.) have the option of linking COVID Symptom Tracker information to their extant research data.

Fig. 2. COVID Symptom Tracker use, reported symptoms, and testing results according to geographic location in the United Kingdom. Between March 24 and March 29, 2020, more than 1.6 million unique individuals downloaded the application and shared clinical and demographic information, as well daily symptoms and high-intensity occupational exposures across the United Kingdom (UK) (*blue map*). Population density of those presenting with any symptoms varied according to region with widespread reports of fatigue, cough, and diarrhea, followed by anosmia, and relatively, infrequently, fever (*inlay*). Examining those individuals who reported complex symptoms, defined as having cough or fever and at least one other of diarrhea, anosmia, and fever, reveals areas of the UK in potential need for more testing. Among the subset of the population that reported receiving a COVID-19 test (*black map*), areas with larger proportions of positive tests (*orange map*) appear to coincide with areas with high proportions of their population reporting complex symptoms, whereas some areas with low-complex symptom prevalence 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 Symptoms Tracker may assist public health and government officials in reallocating resources, identifying areas with unmet testing needs, and detect emerging hot spots earlier.

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 Tracker app. By March 27, 2020, 265,851 individuals reported any change in recent symptoms in the United Kingdom (left). Participants provided data on whether they were tested for COVID-19 and the test result. 1,176 individuals reported having received a COVID-19 test (0.4% of those with symptoms). Symptom frequencies among those who tested COVID-19 positive (upper right, n = 340) vs. negative (lower right, n = 836).

Fig. 4. Predicting COVID-19 cases based on real-time symptom reporting in Wales, United Kingdom. This time series (bar graph) displays the number of new confirmed cases (gray bar) reported by Public Health Wales NHS Trust between 31 March 2020 and 20 April 2020. After April 2, case numbers appeared to be declining through April 5. However, our symptom-based prediction model (see Menni *et al.*) based on symptoms reported from untested users in the COVID Symptom Tracker app on April 1 showed a high proportion of predicted COVID cases in the southern regions of Wales (dark red in Panel A). Six days later, Welsh health authorities reported a subsequent peak in cases over a three day-period (April 6 to 9) driven primarily by these southern regions (colored bars). By April 10, new confirmed cases across Wales declined. However, based on reported symptoms (Panel B), regions in South Wales still had high predicted COVID cases, which became apparent as a second spike in confirmed COVID cases between April 15 to 16. As of April 20 (Panel C), predicted COVID 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 the south of Wales still have relatively high reports of symptoms and appear at risk for subsequent cases of COVID. 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 COVID Symptom Tracker on April 1; 998,244 on April 10; and 1,234,918 on April 20. Public Health Wales NHS Trust data current as of Apr 21 2020 13:00 local time; “Rapid COVID-19 Virology - Public” accessed via publichealthwales.org and downloaded on Apr 22 2020 12:30 PM EST.