



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Detecting COVID-19 infection hotspots in England using large-scale self-reported data from a mobile application: a prospective, observational study

Thomas Varsavsky*, Mark S Graham*, Liane S Canas, Sajaysurya Ganesh, Joan Capdevila Pujol, Carole H Sudre, Benjamin Murray, Marc Modat, M Jorge Cardoso, Christina M Astley, David A Drew, Long H Nguyen, Tove Fall, Maria F Gomez, Paul W Franks, Andrew T Chan, Richard Davies, Jonathan Wolf, Claire J Steves, Tim D Spector, Sebastien Ourselin



Summary

Background As many countries seek to slow the spread of COVID-19 without reimposing national restrictions, it has become important to track the disease at a local level to identify areas in need of targeted intervention.

Methods In this prospective, observational study, we did modelling using longitudinal, self-reported data from users of the COVID Symptom Study app in England between March 24, and Sept 29, 2020. Beginning on April 28, in England, the Department of Health and Social Care allocated RT-PCR tests for COVID-19 to app users who logged themselves as healthy at least once in 9 days and then reported any symptom. We calculated incidence of COVID-19 using the invited swab (RT-PCR) tests reported in the app, and we estimated prevalence using a symptom-based method (using logistic regression) and a method based on both symptoms and swab test results. We used incidence rates to estimate the effective reproduction number, $R(t)$, modelling the system as a Poisson process and using Markov Chain Monte-Carlo. We used three datasets to validate our models: the Office for National Statistics (ONS) Community Infection Survey, the Real-time Assessment of Community Transmission (REACT-1) study, and UK Government testing data. We used geographically granular estimates to highlight regions with rapidly increasing case numbers, or hotspots.

Findings From March 24 to Sept 29, 2020, a total of 2873726 users living in England signed up to use the app, of whom 2842732 (98.9%) provided valid age information and daily assessments. These users provided a total of 120192306 daily reports of their symptoms, and recorded the results of 169682 invited swab tests. On a national level, our estimates of incidence and prevalence showed a similar sensitivity to changes to those reported in the ONS and REACT-1 studies. On Sept 28, 2020, we estimated an incidence of 15841 (95% CI 14023–17885) daily cases, a prevalence of 0.53% (0.45–0.60), and $R(t)$ of 1.17 (1.15–1.19) in England. On a geographically granular level, on Sept 28, 2020, we detected 15 (75%) of the 20 regions with highest incidence according to government test data.

Interpretation Our method could help to detect rapid case increases in regions where government testing provision is lower. Self-reported data from mobile applications can provide an agile resource to inform policy makers during a quickly moving pandemic, serving as a complementary resource to more traditional instruments for disease surveillance.

Funding Zoe Global, UK Government Department of Health and Social Care, Wellcome Trust, UK Engineering and Physical Sciences Research Council, UK National Institute for Health Research, UK Medical Research Council and British Heart Foundation, Alzheimer's Society, Chronic Disease Research Foundation.

Copyright © 2020 The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

Introduction

The COVID-19 pandemic caused many countries to impose strict restrictions on their citizens' mobility and behaviour to curb the rapid spread of disease, often termed lockdowns. Since relaxing these restrictions, many countries sought to avoid their re-imposition through combinations of non-pharmaceutical interventions¹ and test-and-trace systems. Despite these efforts, many countries have had increases in infection since re-opening and have often re-imposed either regional² or national lockdowns. Regional lockdowns aim to contain the disease while minimising the severe economic effect of national lockdowns.

The effectiveness of regional interventions depends on the early detection of so-called infection hotspots.³ Large-scale, population-based testing can indicate regional hotspots, but at the cost of a delay between testing and actionable results. Moreover, accurately identifying changes in the infection rate requires sufficient testing coverage of a given population,⁴ which can be costly and requires substantial testing capacity. Regional variation in testing access can hamper the ability of public health organisations to detect rapid changes in infection rate. There is a high unmet need for tools and methods that can facilitate the timely and cost-effective identification

Lancet Public Health 2021;
6: e21–29

Published Online
December 3, 2020
[https://doi.org/10.1016/S2468-2667\(20\)30269-3](https://doi.org/10.1016/S2468-2667(20)30269-3)

*Contributed equally

School of Biomedical Engineering and Imaging Sciences (T Varsavsky MSc, M S Graham PhD, L S Canas PhD, C H Sudre PhD, B Murray MSc, M Modat PhD, M Jorge Cardoso PhD, Prof S Ourselin PhD) and Department of Twin Research and Genetic Epidemiology (C J Steves PhD, Prof T D Spector PhD), King's College London, London, UK; Zoe Global, London, UK (S Ganesh MSc, J Capdevila Pujol PhD, R Davies MA, J Wolf MA); Medical Research Council Unit for Lifelong Health and Ageing, Department of Population Science and Experimental Medicine, and Centre for Medical Image Computing, Department of Computer Science, University College London, London, UK (C H Sudre); Division of Endocrinology and Computational Epidemiology, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA (C M Astley MD); Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, USA (D A Drew PhD, L H Nguyen MD, Prof A T Chan MD); Department of Medical Sciences, Uppsala University, Uppsala, Sweden (Prof T Fall PhD); Department of Clinical Sciences (Prof M F Gomez PhD) and Genetic and Molecular Epidemiology Unit (Prof P W Franks PhD), Lund University Diabetes Centre, Malmö, Sweden

Correspondence to:
Dr Mark S Graham, School of
Biomedical Engineering and
Imaging Sciences, King's College
London, London SE1 7EH, UK
mark.graham@kcl.ac.uk

Research in context

Evidence before this study

To identify instances of the use of digital tools to perform COVID-19 surveillance, we searched PubMed for peer-reviewed articles published between Jan 1, and Oct 14, 2020, using the keywords "COVID-19" AND (["mobile application"] OR ["web tool"] OR ["digital survey"]). Of the 382 results, we found eight studies that utilised user-reported data to ascertain a user's COVID-19 status. None of these studies sought to provide disease surveillance on a national level or to compare these predictions with other tools to ascertain their accuracy. Furthermore, none of these papers sought to use their data to highlight geographical areas of concern.

Added value of this study

To our knowledge, we provide the first demonstration of mobile technology to provide national-level disease

surveillance. Using over 120 million reports from more than 2.8 million users across England, we estimated incidence, prevalence, and the effective reproduction number. We compared these estimates with those from national community surveys to understand the effectiveness of these digital tools. Furthermore, we showed that the large number of users can be used to provide disease surveillance with high geographical granularity, potentially providing a valuable source of information for policy makers who are seeking to understand the spread of the disease.

Implications of all the available evidence

Our findings suggest that mobile technology can be used to provide real-time data on the national and local state of the pandemic, enabling policy makers to make informed decisions in a quickly moving pandemic.

of infection hotspots to enable policy makers to act with minimal delay.⁵

In this study, we used self-reported, population-wide data, obtained from a mobile application (the COVID Symptom Study app), combined with targeted PCR testing to provide geographical estimates of disease prevalence and incidence. We further show how these estimates can be used to provide timely identification of infection hotspots.

Methods

Study design and participants

For this prospective observational study, data were collected using the COVID Symptom Study app, developed by Zoe Global with input from King's College London (London, UK), the Massachusetts General Hospital (Boston, MA, USA), and Lund and Uppsala Universities (Sweden). The app guides users through a set of enrolment questions, establishing baseline demographic information. Users are asked to record each day whether they feel physically normal, and if not, to log any symptoms and keep a record of any COVID-19 tests and their results (appendix pp 3–7). More details about the app can be found in a study by Drew and colleagues,⁶ which also contains a preliminary demonstration of how symptom data can be used to estimate prevalence.

For the current study, eligible participants were required to have valid age information and to have logged at least one daily assessment. This study included only app users living in England, because some of the methods make use of England-specific testing capacity. In England, the Department of Health and Social Care (DHSC) allocated RT-PCR (swab) tests for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to users of the COVID Symptom Study app, beginning on April 28, 2020. Users who logged themselves as healthy at least once in a 9-day period and then reported any symptom (which we term newly sick) were sent invitations to book a test through

the DHSC national testing programme, and they were then asked to record the result of the test in the app (note that individuals could be invited multiple times; appendix p 11). We included responses from all users in England (whether invited for a test or not) logged between March 24 and Sept 29, 2020. Our study was approved by King's College London ethics committee (REMAS ID 18210; LRS-19/20-18210), and all participants provided consent through the app.

Procedures

We used three datasets to validate our models: the Office for National Statistics (ONS) Community Infection Survey, the Real-time Assessment of Community Transmission (REACT-1) study, and UK Government testing data. The ONS survey⁷ is a longitudinal survey of individuals selected to be a representative sample of private households (excluding eg, care homes and student accommodation), which began on April 26, 2020. Individuals in the ONS survey are supervised while they self-administer nose and throat swabs. The results give estimates of prevalence and incidence over time. Data are released weekly, with each release covering 7–14 days before the release date, with the first release on May 10, 2020. The ONS survey swab 150 000 participants per fortnight. The REACT-1 study began on May 1, 2020, and is a cross-sectional community survey, relying on self-administered swab tests from a sample of the population in England.^{8,9} The sample is randomly selected in each round of data collection. Data releases are intermittent and cover periods of several weeks. The UK Government swab test data are made up of two so-called pillars of testing: pillar 1 covers those with clinical need and health-care workers, and pillar 2 testing covers the wider population who meet government guidelines for testing.¹⁰ We used the ONS and REACT-1 surveys to compare our national estimates of incidence and prevalence, and we used the UK Government testing data to validate our geographically granular list of hotspots.

For the COVID Symptom Study
app see <https://covid.joinzoe.com/>

See Online for appendix

Participants (n=2 842 732)	
Age, years	43.3 (17.3)
≤18	196 143 (6.9%)
19–64	2 296 614 (80.8%)
≥65	349 975 (12.3%)
Sex	..
Female	1 739 929 (61.2%)
Male	1 101 145 (38.7%)
Intersex	150 (<0.1%)
Prefer not to say	1 508 (0.1%)
Preconditions*	..
Kidney disease	18 674 (0.7%); n=2 842 028
Lung disease	302 657 (12.1%); n=2 506 549
Heart disease	66 445 (2.3%); n=2 842 028
Diabetes (type 1 and type 2)	79 143 (2.8%); n=2 842 028
Cancer†	19 543 (1.3%); n=1 508 505
Index of multiple deprivation	..
Bottom tertile (most deprived)	530 216 (18.7%)
Middle tertile	990 122 (34.8%)
Upper tertile (least deprived)	1 322 394 (46.5%)
Smokers	276 044 (9.7%)
Race or ethnicity‡	
Asian or Asian British (Indian, Pakistani, Bangladeshi, or other)	49 237 (0.6%); n=2 144 799
Black or Black British (Caribbean, African, or other)	13 556 (0.6%); n=2 144 799
Mixed race (White and Black or Black British)	13 556 (0.6%); n=2 144 799
Mixed race (other)	31 627 (1.5%); n=2 144 799
White (British, Irish, or other)	1 001 553 (93.3%); n=2 144 799
Chinese or Chinese British	7 823 (0.4%); n=2 144 799
Middle Eastern or Middle Eastern British (Arab, Turkish, or other)	8 880 (0.4%); n=2 144 799
Other ethnic group	10 464 (0.5%); n=2 144 799
Prefer not to say	8 103 (0.4%); n=2 144 799

Data are mean (SD) or n (%). *Not all participants answered questions on preconditions. The numbers reported are the absolute number of users that reported having the precondition, and percentages are calculated among all participants who answered each precondition question. †Question asked from March 29, 2020. ‡Question asked from April 18, 2020.

Table: Characteristics of all app users in England who signed up between March 24, and Sept 29, 2020

We calculated incidence using the invited swab tests reported in the app. We took 14-day averages, starting on May 12, 2020, to calculate the percentage of positive tests among newly sick users according to National Health Service (NHS) regions in England. We then combined this percentage with the proportion of users who report as newly sick in that region to produce the probability that a randomly selected person in that region is infected with COVID-19 on a given day (appendix p 2). We multiplied this probability by the population of each region to produce our swab-based incidence estimates, which we term I_s . Incidence values are released daily, with a 4-day reporting lag.

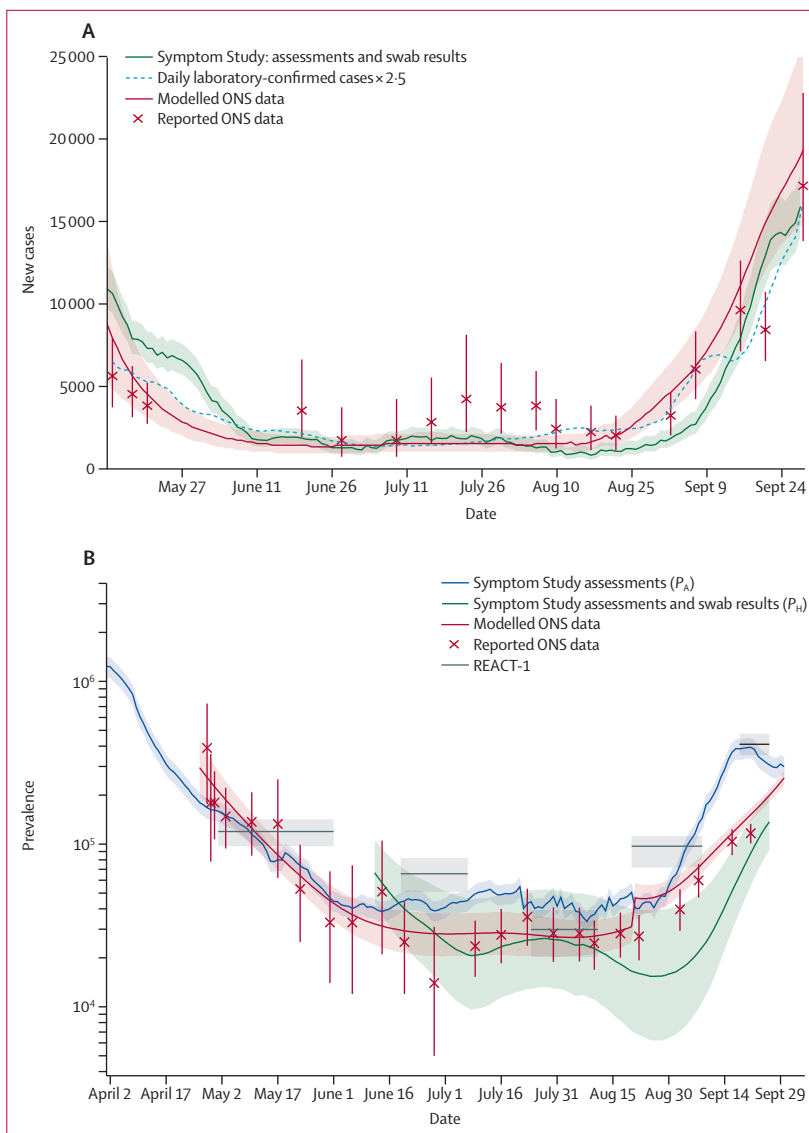


Figure 1: Daily incidence in the UK since May 12, 2020, compared with daily laboratory-confirmed cases and the ONS study (A) and daily prevalence in the UK compared with the ONS and REACT-1 studies (B)
 ONS data are taken from the report released on Oct 9, 2020, and ONS report dates are taken as the midpoint for the date range covered by the estimate. For the Symptom Study and REACT-1, the shaded areas represent 95% CIs. For the ONS, the shaded areas and error bars represent 95% credible intervals. ONS=Office for National Statistics. REACT-1=Real-time Assessment of Community Transmission.

We describe two methods for estimating prevalence. The first is symptom-based, primarily making use of self-reported symptoms and a predictive, symptom-based model for COVID-19. The second is both symptom-based and swab-based, and seeks to further integrate the information from swab test results collected in the app.

The symptom-based approach uses a previously validated logistic regression model¹¹ to predict whether a user is SARS-CoV-2 positive on the basis of their reported symptoms (appendix p 2). For a given day, each user's most recent symptom report from the previous 7 days is used for prediction. If a user reports a positive COVID-19

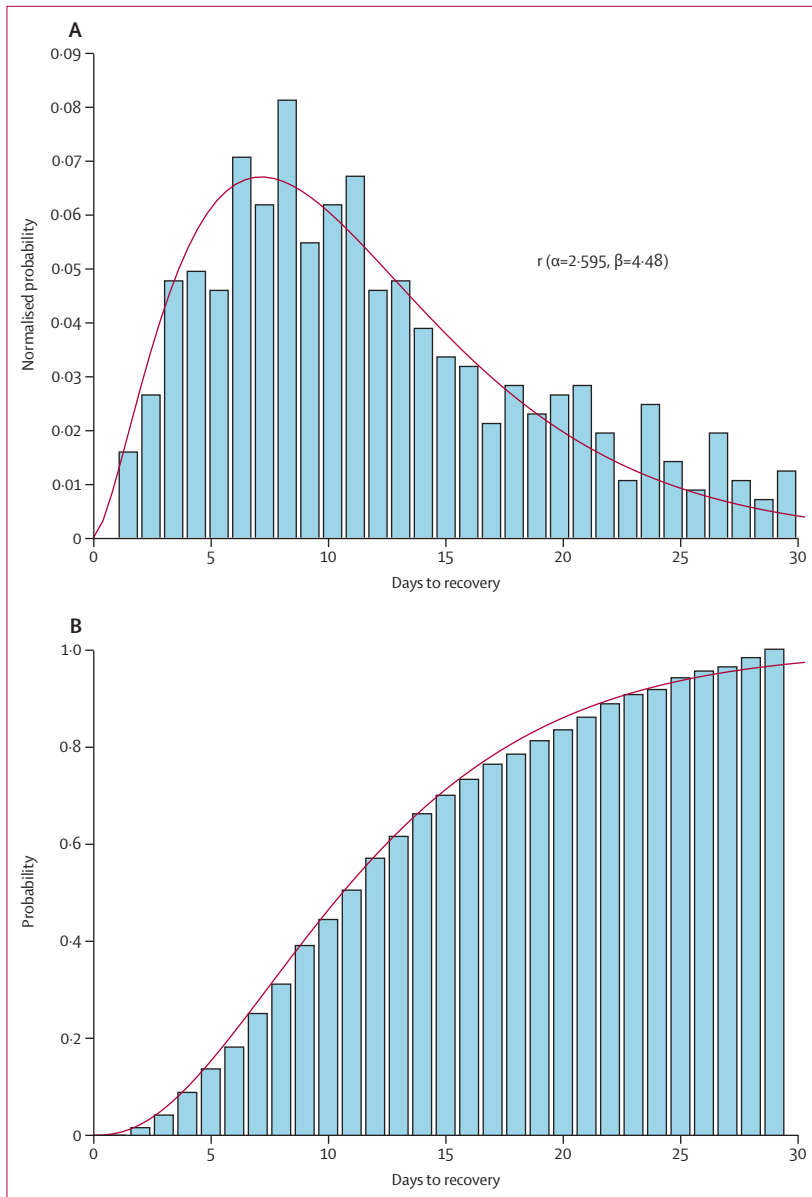


Figure 2: Empirical probability density function of days to recovery with a gamma fit (A) and empirical cumulative density function of days to recovery with the same gamma fit (B). The bars represent empirical findings and the red line is the gamma fit.

test in that 7-day period, the test result is used to override the user’s symptom-based estimate. The proportion of positive users is used to estimate prevalence. A user who is predicted to be COVID-19-positive for more than 30 days is considered long-term sick and no longer infectious, and they are then removed from the calculation. We sought to extrapolate these prevalence estimates to the general population. As noted in a previous study,¹² there is a disparity in COVID-19 prevalence between regions of higher index of multiple deprivation (IMD), a measure of the relative deprivation of geographical regions,¹³ and those of lower IMD. We

stratified users by Upper Tier Local Authority (UTLA), IMD tertile, and age bands (in decades), and we predicted the percentage prevalence per stratum. We then multiplied our predicted percentage of positive cases per stratum with that stratum’s population size, according to census data, to estimate cases per stratum. These estimates were then summed to produce our population prevalence estimate, which we term P_A . We examined the sensitivity of P_A to health-seeking bias by removing all users reporting sick at sign-up from the analysis.

To estimate prevalence using the swab test data, we used our incidence estimates and the relationship $P_{t+1}=P_t+I_t-M_t$ where P represents prevalence, M_t represents the number of patients that recover at time t , and I_t is the number of new COVID-19 cases at time t . We estimated M_t from our data (appendix p 3).

These prevalence estimates make use of the swab test results but lack geographical granularity, being per NHS region. We can increase the granularity by taking the symptom-based estimates, which are calculated per UTLA, and rescaling all the estimates that make up an NHS region such that the total prevalence across those UTLAs matches the estimated per-NHS region prevalence. We term this hybrid method, which uses both symptom reports and swab tests, as P_H . It is possible to produce granular incidence estimates by applying the model of recovery to these granular prevalence estimates; we term these estimates I_H .

It is possible to retrieve the effective reproduction number, known as $R(t)$, from incidence rates by combining them with known values of the serial interval.¹⁴ Briefly, we used the relationship $I_{t+1}=I_t \exp(\mu(R(t)-1))$, where $1/\mu$ is the serial interval. We modelled the system as a Poisson process and used Markov Chain Monte-Carlo to estimate $R(t)$. In our probabilistic modelling, we assumed that the serial interval was drawn from a gamma distribution with $\alpha=6.0$ and $\beta=1.5$ as in the study by Nishiura and colleagues.¹⁵ By sampling successive chains from the system, we obtained a distribution over $R(t)$, which allowed us to report a median and 95% credible intervals. These estimates of uncertainty do not account for the uncertainty in the estimate of incidence, which we found to be mostly systematic and smaller than the other forms of uncertainty modelled.

A hotspot is defined as a sudden increase in the number of cases in a specific geographical region. We produced two rankings of UTLAs in England. The first ranks each by their estimated prevalence, P_H . This ranking has the advantage of being preregistered; a list of the top ten UTLAs according to P_H has been published online since July 23, 2020. However, it does not allow the direct identification of areas of concern; ie, areas with a large number of new cases, and we therefore also report a second ranking using I_H .

We compared our rankings to those obtained by ranking according to government testing data. England contains 149 UTLAs, each containing a mean of

370 000 people. We used the government data to produce daily reference rankings of each UTLA, based on 7-day moving averages of daily cases per UTLA. We included all tests done on a given day to produce the ranking for that day, even if that test took several days to have its result returned, to produce the most accurate gold standard ranking that we could. We used 7-day moving averages of P_H and I_H to produce our predicted rankings of each UTLA. We then evaluated these predictions against the historical reference using two metrics. The first, recall at 20, is the number of UTLAs in our top 20 that appear in the reference top 20. The second, the normalised mean reciprocal rank at 20, measures the agreement between ranks of our top 20 list. We estimated uncertainty by drawing 100 samples of P_H , I_H , and the government testing data for each UTLA and day, making use of errors calculated using the Wilson interval approximation for the binomial distribution. These samples were ranked and metrics re-computed to produce 95% CIs for each metric.

Statistical analysis

Power analysis found that 320 000 weekly active users are needed to detect an increase from 5000 to 7500 daily cases with a significance of 0.05. All analysis was done with Python 3.7. We used ExeTera 0.2.7 for dataset processing and analysis.¹⁶ The COVID Symptom Study app is registered with ClinicalTrials.gov, NCT04331509.

Role of the funding source

Zoe Global developed the app for data collection. The funders had no role in the study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to all the data and the corresponding author had final responsibility for the decision to submit for publication.

Results

From March 24 to Sept 29, 2020, a total of 4644227 participants signed up to use the app, of which 2873726 (61.9%) reported living in England. After excluding a further 4508 (0.1%) participants with invalid age information and 26486 (0.6%) without any daily assessments logged, we report 2842732 users that participated in this study (table). In total, these participants completed 120192306 daily assessments, with a median of 16 logs (IQR 3–68) per user, corresponding to logged data for 22.4% of the total possible person-logs across the study period. Compared with the population of England, we had fewer participants aged 18 or younger (196143 [6.9%] vs 22.5%) and fewer aged 65 or older (349975 [12.3%] vs 18.4%), more female than male participants (1739929 [61.2%] vs 50.6%), and fewer participants living in deprived regions (1322394 [46.5%] in the least deprived tercile vs 33.3%; table). Between April 28 and Sept 29, a total of 851250 invitations for swab tests were sent out. 169682 (19.9%) of these invitations were responded to by

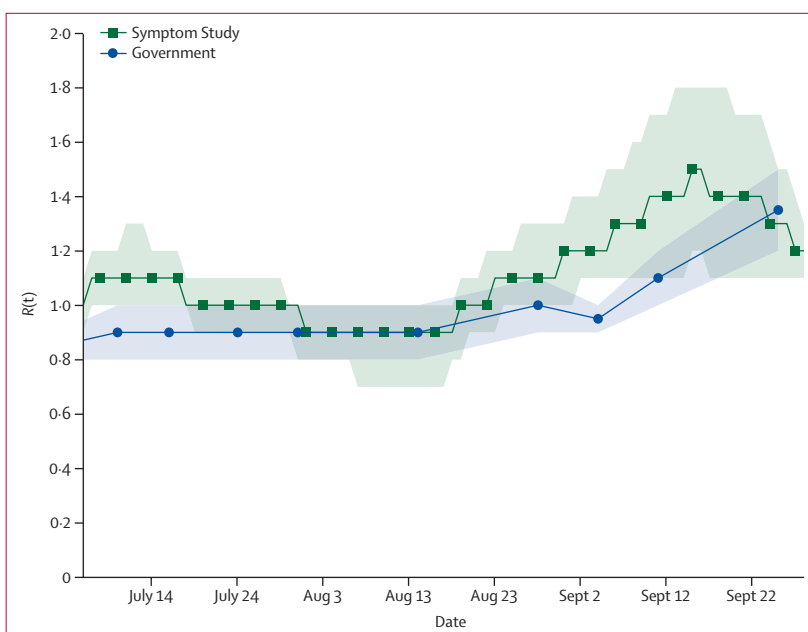


Figure 3: Estimated $R(t)$ for England between June 24 and Sept 28, 2020

The shaded area for Symptom Study data represents 95% credible intervals and for government data represents 95% CIs. UK Government estimates published every 7–12 days from June 12, 2020. $R(t)$ =effective reproduction number.

a person reporting a swab test on the app, of which 1912 (1.1%) tested positive. In addition, users reported the results of 689426 non-invited tests in the app, of which 25663 (3.7%) were positive. 621031 (0.5%) of 120192306 symptom reports were classified as COVID-19 positive by the symptom model.

In our comparison of England-wide incidence estimates (I_s) with government testing data and the ONS survey¹⁷ (figure 1A), we included two estimates from the ONS: the official reports, released every week, and the results from time-series modelling. The reports represent the best estimate of the ONS at the time of release, whereas the time-series model can evolve, leading to revision of previous estimates in response to new data (appendix p 7). The government values are consistently lower than other estimates because they are not a representative figure for the population. To account for this, we looked at the number of people who reported classic symptoms (fever, loss of smell, and persistent cough) for the first time between July 7 and Aug 5, 2020, and who did not get tested; we found the number to be 50499 (62%). We used this percentage to scale the government data by a factor of 2.5, our best estimate of the systematic undercounting of new cases.

Our results predict a steep decline in incidence until the middle of July, a trend in agreement with the government data and ONS estimates. All three estimates show an increase in the number of daily cases from mid-August throughout September; on Sept 28, 2020, we estimated that there were 15841 (95% CI 14023–17885) daily cases. Estimates of incidence per NHS region are in

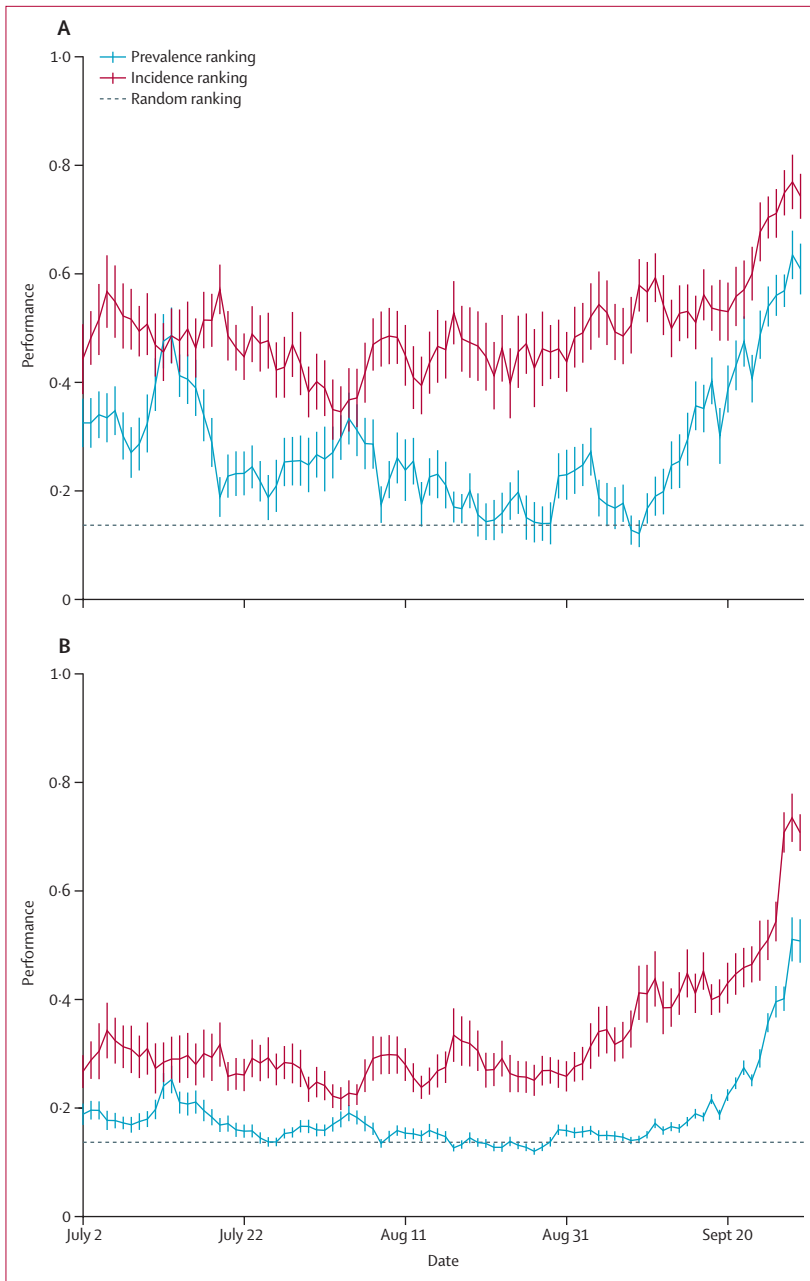


Figure 4: Performance of our two ranking methods: ranking by prevalence and incidence on two metrics, recall at 20 (A) and the normalised mean reciprocal rank at 20 (B). Vertical lines represent 95% CIs.

the appendix (p 8), as are maps of our most granular incidence, I_H (appendix p 8).

We also compared our England-wide prevalence estimates, P_A and P_H , with prevalence reported by the ONS and REACT-1 studies (figure 1B). The symptom-based assessments, P_A , indicate a continuous drop in the number of cases from April 1, following the lockdown measures instigated on March 23, which then plateaued in mid-June and began to rise again sharply from early

September. The trends observed for P_A agree with data from the ONS survey and the REACT-1 study, although there is some divergence in September, when both P_A and REACT-1 data show a sharper rise in prevalence than the ONS data (figure 1B). It should be noted that P_A captures only symptomatic cases, whereas ONS and REACT-1 data also capture asymptomatic cases, which are thought to account for 40–45% of the total cases.¹⁸ Taking this factor into account, P_A is slightly higher than ONS and REACT-1 data. On Sept 28, 2020, we estimated a prevalence of 0.53% (95% CI 0.45–0.60).

P_H agrees with the trends in the other estimates, and predicts a rise in cases in September at a similar rate to P_A and REACT-1 data (figure 1B). However, the absolute values are consistently lower than other estimates. The recovery model used to calculate P_H is shown in figure 2. Although most users recover in 7–10 days, the curve shows there is a substantial minority who take longer than 3 weeks to recover from COVID-19.

The size of our dataset allows us to estimate prevalence for more granular geographical regions than the ONS (appendix p 9). We considered that our estimates of prevalence might be biased by a user's health-seeking behaviour, and we sought to assess the influence of this factor by removing from the analysis all users who reported being sick upon sign-up (appendix p 10).

We compared our estimates of $R(t)$ for England between June 24 and Sept 28, 2020, with the consensus estimates provided by the UK Government Science Pandemic Influenza Modelling group (figure 3).¹⁹ Estimates for each of the NHS regions are in the appendix (p 12). The estimates agree that $R(t)$ has been above 1 from early-mid-September, and we estimate that $R(t)$ in England was 1.17 (95% CI 1.15–1.19) on Sept 28, 2020. The government estimates are much smoother than our estimates, which is probably because they are derived from a consensus of the $R(t)$ estimates from the models produced by many groups.

Concerning hotspot detection in England, ranking based on incidence consistently outperformed the prevalence-based ranking (figure 4). Performance varied over time; the best incidence-based ranking produced recall scores of up to 0.75, with a minimum of 0.35, indicating that we can predict between 7 and 15 of the regions in the UK Government's top 20. On a geographically granular level, on Sept 28, 2020, we detected 15 (75%) of the 20 regions with highest incidence according to government test data. The ranking performed best in late September when cases began to rise, due to a greater difference between regional case numbers. We also compared the agreement of weekly cases in each UTLA between our data and government data against the number of government pillar 2 tests carried out (figure 5). The correlation indicates that the two estimates agree best when the government carries out more tests. This finding means that disagreements between the two rankings could be partially explained by

poorer ranking in regions with limited government testing.

Discussion

In this study, we have demonstrated the use of population-wide data reported through the COVID Symptom Study app, drawing on more than 120 million daily reports and about 170 000 invited swab tests from more than 2·8 million users to estimate prevalence, incidence, and $R(t)$. Other digital surveys have been used to provide valuable real-time information about the pandemic, including How We Feel,²⁰ Corona Israel,⁵ the Facebook Survey, and CovidNearYou. However, to our knowledge we are the first to provide national-level disease surveillance, and we have found good agreement with traditional, representative community surveys.

Furthermore, we have used our data to produce geographically granular estimates and a list of potential hotspots. The list consistently flags a relatively high number of regions highlighted by the UK Government testing data. Although we compared our data with that from government testing in our results, it must be noted these cannot be considered ground truth estimates of COVID-19 cases. The government data are an incomplete sampling of new COVID-19 cases; results from our app indicate that only about 40% of those who report classic COVID-19 symptoms go on to receive a test. Furthermore, testing capacity is not uniform across all UTAs.²¹ Our results indicate that our case estimates agree best with government estimates in areas with high levels of testing per capita, suggesting that our estimates could prove a valuable resource for forecasting in regions with poor testing provision. There are other reasons why our list might differ from the Government list. First, the two methods might have different uptake in some higher-risk groups, such as students in provided accommodation, thus showing different sensitivity to hotspots based on the demographic make-up of a region. The two methods might therefore be complementary, and we suggest that our hotspot detection could be most beneficial as an additional indication of regions where increased testing might be best focused. The modest reliance on PCR tests suggests our approach could prove valuable in countries where testing infrastructure is less developed, although further work is required to assess our approach in other locations.

Other efforts to track the national progression of COVID-19 rely on self-swabbing from community cohorts. Two such efforts exist in England: the ONS study¹⁷ and the REACT-1 study.^{8,9} These studies have the advantage of being more representative of the population, and their design enables the detection of asymptomatic cases. However, they are smaller than the COVID Symptom Study; the ONS and REACT-1 studies currently report 120 000–175 000 participants in England, whereas the app reports over 2 800 000 users in England. The ability to use self-reported symptom data from this

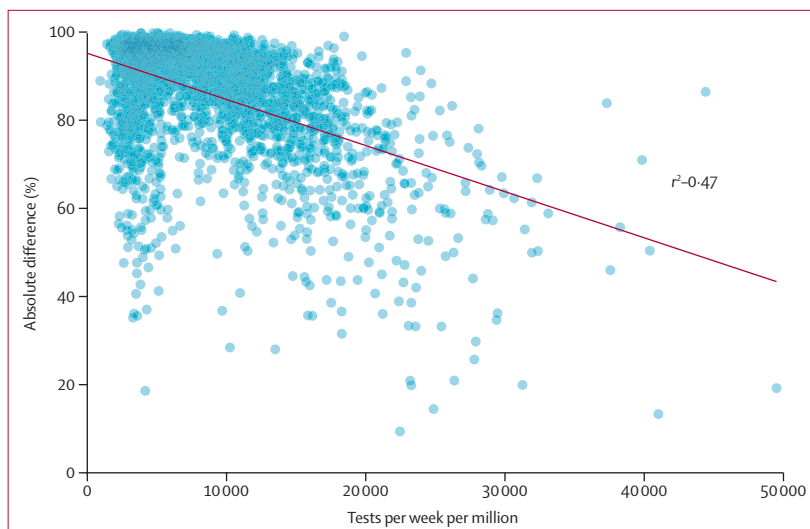


Figure 5: Agreement between COVID Symptom Study and UK Government case numbers per week and UTLA, against the number of government pillar 2 tests carried out
UTLA=Upper Tier Local Authority.

large cohort enables us to make predictions of more geographically granular regions than do either the ONS or REACT-1 studies, allowing us to predict COVID-19 hotspots at the UTLA level. Our estimates should therefore be viewed as independent and complementary to those provided by the ONS and REACT-1 studies.

Several limitations to our work must be acknowledged. The app users are not a representative sample of the wider population for which we aim to make an inference. There is a clear shift in age and gender compared with the general population, our users tend to live in less deprived areas,¹² and we have few users reporting from key sites such as care homes and hospitals. We accounted for some population differences when producing prevalence estimates using specific census adjusted population strata, but the number of invited tests does not allow us to do this when calculating incidence. Differences in reported symptoms across age groups²² would probably lead to different prediction models of COVID-19 positivity, and the performance of the model will vary with the prevalence of other infections with symptoms that overlap with COVID-19, such as influenza. Furthermore, the app population is less ethnically diverse than the general population.²³ Reliance on user self-reporting can also introduce bias into our results; for instance, users who are very sick might be less likely to report than those with mild symptoms. Other sources of error include collider bias²⁴ arising from a user's probability of using the app being dependent on their likelihood of having COVID-19, potentially biasing our estimates of incidence and prevalence. We showed a sensitivity analysis that attempts to understand the effect of health-seeking behaviour, but we acknowledge that there are many other biases that might affect our results—for example, our users might be more risk-averse than the general

For the Facebook Survey see <https://cmu-delphi.github.io/delphi-epidata/symptom-survey/>

For CovidNearYou see <https://covidnearyou.org/>

For the UK government testing dashboard see <https://coronavirus.data.gov.uk/testing>

population—and that our results must be interpreted with this in mind.

We have presented a means of combining app-based symptom reports and targeted testing from over 2.8 million users to estimate incidence, prevalence, and $R(t)$ in England. By integrating symptom reports with PCR test results, we were able to highlight regions which might have concerning increases in COVID-19 cases. This approach could be an effective, complementary way for governments to monitor the spread of COVID-19 and identify potential areas of concern.

Contributors

TV, MSG, TF, MFG, PWF, JW, CJS, TDS, and SO contributed to the study concept and design. SG, JCP, CHS, DAD, LHN, ATC, RD, JW, CJS, TDS, and SO contributed to the acquisition of data. TV, MSG, LSC, SG, JCP, CHS, and BM contributed to data analysis and have verified the underlying data. TV, MSG, and LSC contributed to the initial drafting of the manuscript. ATC, CJS, TDS, and SO contributed to study supervision. All authors contributed to the interpretation of data and critical revision of the manuscript.

Declaration of interests

SG, JCP, RD, and JW are employees of Zoe Global. DAD and ATC previously served as investigators on a clinical trial of diet and lifestyle using a separate smartphone application that was supported by Zoe Global. TF reports grants from the European Research Council, Swedish Research Council, Swedish FORTE Research Council, and the Swedish Heart–Lung Foundation, outside the submitted work. MFG reports financial and in-kind support within the Innovative Medicines Initiative project BEAR-DKD from Bayer, Novo Nordisk, Astellas, Sanofi-Aventis, AbbVie, Eli Lilly, JDRF International, and Boehringer Ingelheim; personal consultancy fees from Lilly; financial and in-kind support within a project funded by the Swedish Foundation for Strategic Research on precision medicine in diabetes from Novo-Nordisk, Pfizer, Follicum, Abcentra; in-kind support on that project from Probi and Johnson and Johnson; and a grant from the EU, outside the submitted work. ATC reports grants from Massachusetts Consortium on Pathogen Readiness, during the conduct of the study, and personal fees from Pfizer and Boehringer Ingelheim and grants and personal fees from Bayer, outside the submitted work. TDS is a consultant to Zoe Global. All other authors declare no competing interests.

Data sharing

Data collected in the COVID Symptom Study smartphone application are being shared with other health researchers through the UK NHS-funded Health Data Research UK (HDRUK) and Secure Anonymised Information Linkage consortium, housed in the UK Secure Research Platform (Swansea, UK). Anonymised data are available to be shared with HDRUK researchers according to their protocols in the public interest. US investigators are encouraged to coordinate data requests through the Coronavirus Pandemic Epidemiology Consortium.

Acknowledgments

Zoe Global provided in-kind support for all aspects of building, running, and supporting the app, and provided service to all users worldwide. Support for this study was provided by the National Institute for Health Research (NIHR)-funded Biomedical Research Centre (BRC) based at Guys' and St Thomas' NHS Foundation Trust. Investigators also received support from the Wellcome Trust, the UK Medical Research Council and British Heart Foundation, Alzheimer's Society, the EU, the UK NIHR, Chronic Disease Research Foundation, and the NIHR-funded BioResource, Clinical Research Facility and BRC based at Guys' and St Thomas' NHS Foundation Trust, in partnership with King's College London, the UK Research and Innovation London Medical Imaging and Artificial Intelligence Centre for Value Based Healthcare, the Wellcome Flagship Programme (WT213038/Z/18/Z), the Chronic Disease Research Foundation, and the DHSC. CMA is supported by the National Institute of Diabetes and Digestive and Kidney Diseases (K23 DK120899), as is DAD (K01DK120742). DAD and LHN are supported by the American Gastroenterological Association–Takeda COVID-19 Rapid Response

Research Award (AGA2021-5102). The Massachusetts Consortium on Pathogen Readiness and Mark and Lisa Schwartz supported DAD, LHN, and ATC. LHN is supported by the American Gastroenterological Association Research Scholars Award and the National Institute of Diabetes and Digestive and Kidney Diseases (K23 DK125838). TF holds a European Research Council Starting Grant. ATC was supported in this work through a Stuart and Suzanne Steele MGH Research Scholar Award. Investigators from the COVID Symptom Study Sweden were funded in part by grants from the Swedish Research Council, Swedish Heart–Lung Foundation, and the Swedish Foundation for Strategic Research (LUDC-IRC 15-0067). We thank Catherine Burrows (Bulb, UK) for assistance with database querying.

References

- 1 Flaxman S, Mishra S, Gandy A, et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* 2020; **584**: 257–61.
- 2 Mahase E. COVID-19: how does local lockdown work, and is it effective? *BMJ* 2020; **370**: m2679.
- 3 Lavezzo E, Franchin E, Ciavarella C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'. *Nature* 2020; **584**: 425–29.
- 4 Peto J, Alwan NA, Godfrey KM, et al. Universal weekly testing as the UK COVID-19 lockdown exit strategy. *Lancet* 2020; **395**: 1420–21.
- 5 Rossman H, Keshet A, Shilo S, et al. A framework for identifying regional outbreak and spread of COVID-19 from one-minute population-wide surveys. *Nat Med* 2020; **26**: 634–38.
- 6 Drew DA, Nguyen LH, Steves CJ, et al. Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020; **368**: 1362–67.
- 7 Pouwels KB, House T, Pritchard E, et al. Community prevalence of SARS-CoV-2 in England during April to September 2020: results from the ONS Coronavirus Infection Survey. *Lancet* 2020 (in press).
- 8 Riley S, Atchison C, Ashby D, et al. REal-time Assessment of Community Transmission (REACT) of SARS-CoV-2 virus: study protocol. *Wellcome Open Research* 2020; **5**: 200.
- 9 Riley S, Ainslie KEC, Eales O, Walters CE, Wang H. High prevalence of SARS-CoV-2 swab positivity in England during September 2020: interim report of round 5 of REACT-1 study. *medRxiv* 2020; published online Oct 2. <https://www.medrxiv.org/content/10.1101/2020.09.30.20204727v1.abstract> (preprint).
- 10 UK Department of Health and Social Care. Guidance: COVID-19 testing data: methodology note. Aug 21, 2020. <https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note> (accessed Sept 11, 2020).
- 11 Menni C, Valdes AM, Freidin MB, et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020; **26**: 1037–40.
- 12 Bowyer R, Varsavsky T, Sudre CH, et al. Geo-social gradients in predicted COVID-19 prevalence and severity in Great Britain: results from 2 266 235 users of the COVID-19 Symptoms Tracker app. *medRxiv* 2020; published online April 27. <https://www.medrxiv.org/content/10.1101/2020.04.23.20076521v1> (preprint).
- 13 McLennan D, Noble S, Noble M, Plunkett E, Wright G, Gutacker N. The English Indices of Deprivation 2019: technical report. September, 2019. https://dera.ioe.ac.uk/34259/1/loD2019_Technical_Report.pdf (accessed Nov 16, 2020).
- 14 Bettencourt LMA, Ribeiro RM. Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One* 2008; **3**: e2185.
- 15 Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis* 2020; **93**: 284–86.
- 16 Murray B, Kerfoot E, Graham MS, et al. Accessible data curation and analytics for international-scale citizen science datasets. *ArXiv*. 2020; published online Nov 2. <http://arxiv.org/abs/2011.00867> (preprint).
- 17 Pouwels KB, House T, Robotham JV, Birrell P. Community prevalence of SARS-CoV-2 in England: results from the ONS coronavirus infection survey pilot. *medRxiv* 2020; published online July 7. <https://www.medrxiv.org/content/10.1101/2020.07.06.20147348v1.abstract> (preprint).

For data sharing see <https://web.www.healthdatagateway.org/dataset/fddcb382-3051-4394-8436-b92295f14259>

For the Coronavirus Pandemic Epidemiology Consortium see <https://www.monganinstitute.org/cope-consortium>

- 18 Oran DP, Topol EJ. Prevalence of asymptomatic SARS-CoV-2 infection: a narrative review. *Ann Intern Med* 2020; 173: 362–67.
- 19 UK Government. The R number and growth rate in the UK. May 15, 2020. <https://www.gov.uk/guidance/the-r-number-in-the-uk>. (accessed Sept 30, 2020).
- 20 Allen WE, Altae-Tran H, Briggs J, et al. Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nat Hum Behav* 2020; 4: 972–82.
- 21 The Lancet Respiratory Medicine. COVID-19 testing in the UK. *Lancet Respir Med* 2020; 8: 1061.
- 22 Zazzara MB, Penfold RS, Roberts AL, et al. Probable delirium is a presenting symptom of COVID-19 in frail, older adults: a cohort study of 322 hospitalised and 535 community-based older adults. *Age Ageing* 2020; published online Sept 28. <https://doi.org/10.1093/ageing/afaa223>.
- 23 Lo C-H, Nguyen LH, Drew DA, et al. Racial and ethnic determinants of COVID-19 risk. *MedRxiv* 2020; published online June 20. <https://www.medrxiv.org/content/10.1101/2020.06.18.20134742v1?rss=1%22> (preprint).
- 24 Griffith G, Morris TT, Tudball M, et al. Collider bias undermines our understanding of COVID-19 disease risk and severity. *Nature Commun* 2020; 11: 5749.