Detecting COVID-19 infection hotspots in England using large-scale self-reported data from a mobile application

Thomas Varsavsky[1*], Mark S. Graham[1*], Liane S. Canas[1], Sajaysurya Ganesh[2], Joan Capdevila Puyol[2], Carole H. Sudre[1], Benjamin Murray[1], Marc Modat[1], M. Jorge Cardoso[1], Christina M. Astley[4], David A Drew[4], Long H. Nguyen[4], Tove Fall[5], Maria F Gomez[6], Paul W. Franks[7], Andrew T. Chan[4], Richard Davies[2], Jonathan Wolf[2], Claire J. Steves[3], Tim D. Spector[3], Sebastien Ourselin[1]

1. School of Biomedical Engineering & Imaging Sciences, King's College London, London, UK
2. Zoe Global Limited, London, UK
3. Department of Twin Research and Genetic Epidemiology, King's College London, London, UK
4. Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, MA, USA
5. Department of Medical Sciences, Uppsala University, Uppsala, Sweden
6. Department of Clinical Sciences, Lund University Diabetes Centre, Sweden.
7. Genetic and Molecular Epidemiology Unit, Lund University Diabetes Centre, Sweden.

*Joint contribution

Corresponding Author:
Mark S. Graham, PhD
School of Biomedical Engineering and Imaging Sciences
King's College London,
Lambeth Palace Road,
SE1 7EH
mark.graham@kcl.ac.uk

Abstract

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

We performed modelling on longitudinal, self-reported data from users of the COVID Symptom Study app in England between 24 March and 29 September, 2020. Combining a symptom-based predictive model for COVID-19 positivity and RT-PCR tests provided by the Department of Health we were able to estimate disease incidence, prevalence and effective reproduction number. Geographically granular estimates were used to highlight regions with rapidly increasing case numbers, or hotspots.

Findings

More than 2.6 million app users in England provided 115 million daily reports of their symptoms, and recorded the results of 170,000 PCR tests. On a national level our estimates of incidence and prevalence showed similar sensitivity to changes as two national community surveys: the ONS and REACT studies. On a geographically granular level, our estimates were able to highlight regions before they were subject to local government lockdowns. Between 12 May and 29 September we were able to flag between 35-80% of regions appearing in the Government's hotspot list.

Interpretation

Self-reported data from mobile applications can provide a cost-effective and agile resource to inform a fast-moving pandemic, serving as an independent and complementary resource to more traditional instruments for disease surveillance.

Funding

Zoe Global Limited, Department of Health, Wellcome Trust, EPSRC, NIHR, MRC, Alzheimer's Society.

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 between 1 January and 14 October 2020, using the keywords COVID-19 AND ((mobile application) OR (web tool) OR (digital survey)). Of the 382 results, we found eight that utilised user-reported data to ascertain a user's COVID-19 status. Of these, none sought to provide disease surveillance on a national level, or to compare these predictions to 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 115 million reports from more than 2.6 million users across England, we estimate incidence, prevalence, and the effective reproduction number. We compare these estimates to those from national community surveys to understand the effectiveness of these digital tools. Furthermore, we demonstrate the large number of users can be used to provide disease surveillance with high geographical granularity, potentially providing a valuable source of information for policymakers 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 policymakers to make informed decisions in a fast-moving pandemic.

Introduction

The COVID-19 pandemic caused many countries to impose strict restrictions on their citizen's mobility and behaviour to curb the rapid spread of disease, often termed 'lockdowns'. Many countries have now relaxed these restrictions and hope to avoid the re-imposition of these measures. Governments are seeking to employ strategies to reduce transmission risk, such as the use of masks and social distancing, along with robust test-and-trace systems to detect new infections and reduce the risk of onwards transmission[1]. Despite these efforts, many countries have experienced localised increases in infections since reopening. Governments are increasingly relying on the enforcement of local lockdowns in these regions, with the aim of containing new outbreaks whilst minimising the severe economic impact of national lockdown[2].

The effectiveness of these drastic interventions depend on the early detection of so-called infection hotspots, regions with rapid increases in the number of COVID-19 cases[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 significant testing capacity. There is a high unmet need for tools and methods that can facilitate the timely and cost-effective identification of infection hotspots to enable policymakers to act with minimal delay[5].

In this work we use 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. Our predictions, published online on 23 July 2020, correctly identified four out of nine regions which were subject to a local lockdown by the UK government on 30 July. We believe that the wide adoption of an app-based symptom surveillance system can aid government and public health leaders identify areas of concern early and monitor the effects of targeted interventions.

Methods

Application data

Data were collected using the COVID Symptom Study app, developed by Zoe Global Ltd with input from King's College London, the Massachusetts General Hospital, and Lund and Uppsala Universities. The app guides participants through a set of enrolment questions, establishing baseline demographic information and comorbidities. Users are encouraged to record each day whether they feel physically normal, and if not, to log any symptoms experienced, along with details of any COVID-19 tests and their results. The study aims to enrol participants as broadly as possible in the countries where it is available. It was launched in the UK on 24 March, in the US on 29 March and in Sweden on 29 April 2020. Further details about the information collected through the app can be found in[6].

In England, the Department of Health and Social care allocated COVID-19 tests to users of the Study app, beginning on 28 April. Users who logged as healthy for an uninterrupted period of at least nine days and then reported any symptom (which we term 'newly sick') were sent invitations for testing, and asked to record the result of the test in the app. As some methods described in this paper make use of this testing capacity, we only include data from app users living in England. We include responses logged between 24 March and 29 September.

External datasets

Three datasets were used for validation of our models: the Office for National Statistics (ONS) survey, the REACT study, and Government testing data. The ONS Infection survey[7] is a parallel effort to track the spread of COVID-19 in the UK over time. The survey sends self-swab PCR testing kits to individuals selected to be a representative sample of private households. The results give estimates of prevalence and incidence over time, with the first estimates being on 30 April. The sample size has increased over the course of the study: the first report from 12 May tested 10,705 people in 5,276 households in England, and by September they had recorded 151,000 participants who had provided at least one swab result. The REACT study is another community survey, relying on self-administered PCR tests from a representative sample of the population in England[8,9]. The sample size ranges between 120,000 and 160,000 people in each round of data collection. The Government swab-test data is made up of two 'Pillars' of testing: Pillar 1 covers those with clinical need and health and care workers, and Pillar 2 testing covers the wider population who meet Government guidelines for testing.

Incidence

We define incidence as the number of new COVID-19 cases per day. It is calculated using the swab tests reported in the app. We took 14-day averages starting on 12 May to calculate the percentage of positive cases per NHS region in England. We exploited the following Bayesian relationship to estimate the number of symptomatic COVID-19 cases in the general population,

$$P(C) = P(C|S)\, P(S) \,/\, P(S|C)$$

P(S) is the probability of being newly symptomatic according to the data entered in the app, defined as somebody who remains healthy for at least nine days before reporting any symptoms. These newly sick users are invited to take a swab test. P(C|S) is the probability of a user being COVID-19 positive given they are newly symptomatic on the app. This is estimated as the percentage of positive cases in the newly symptomatic users that accept the test invite, with the assumption that the number of positive cases in this population is representative of the full population of newly symptomatic users. P(S|C) is the probability of developing symptoms given that one has COVID-19. We set P(S|C) = 1 as we focus on the prediction of symptomatic cases. We extrapolated to the populations of each of these regions based on these ratios to produce our swab-based incidence estimate, I_S.

Prevalence

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

Symptom-based

The symptom-based approach uses a previously validated logistic regression model[10] to predict whether a user is COVID-19 positive or not, based on their reported symptoms. For a given day, each user's most recent symptom report from the previous seven days is used for prediction. The proportion of positive users are used to estimate prevalence. A user that is predicted COVID-19 positive for more than 30 days is considered long-term sick and no longer infectious, and removed from the calculation. In order to extrapolate these prevalence estimates to the general population, the app population should be representative of the general population. As noted in previous work[11] there is a disparity in COVID-19 prevalence between regions of higher index of multiple deprivation (IMD) and those of lower IMD. In the UK we stratify and predict prevalence by Upper Tier Local Authority (UTLA), Index of Multiple Deprivation (IMD) tertile, and age bands (in decades), then combine our predicted percentage of COVID 19 positive cases per strata with census data to produce our population prevalence estimate. We term this prevalence P_A.

Symptom and swab based

To estimate prevalence using the swab data, we make use of our incidence estimates and the relationship

$$P_{t+1} = P_t + I_t - M_t$$

where M_t represents the number of patients that recover at time t and I_t the number of new COVID-19 cases at time t. We estimate M_t from our data. We looked at users who reported a positive RT-PCR test and were also predicted positive from our symptom-based model. We defined onset as the first appearance of any symptom that occurred less than seven days before a positive test or being predicted positive. We defined recovery as either seven days of uninterrupted healthy reporting or the date of a negative RT-PCR test, selecting the smallest value if both occurred. For our prevalence estimate, we truncate the model of recovery at 30 days, meaning we consider anyone who has suffered from COVID-19 for greater than 30 days as a long-term COVID-19 patient who is no longer infectious. This agrees with[12] where the authors suggest that although patients can test positive on RT-PCR beyond 28 days, they are unlikely to be infectious. This gives a probability distribution for the number of days it takes to recover from COVID-19.

Using the incidence estimates per NHS region, I_S, we can produce prevalence estimates per NHS region using the following dot-product:

$$P = I_{t:t+30} \bullet (1 - CDF(M))$$

Where $I_{t:t+30}$ is a vector of incidence estimates in the 30 days up to time t and CDF(M) is the cumulative distribution function of the model of recovery, M.

These prevalence estimates make use of the swab 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 per-NHS region prevalence we estimate. We term this hybrid method, making use of 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.

Calculating R(t)

It is possible to retrieve R(t) from incidence rates by combining them with known values of the serial interval[13]. Briefly, we use the relationship

$$I_{t+1} = I_t \exp(\mu (R(t) -1))$$

Where $1/\mu$ is the serial interval. We model the system as a Poisson process and use Markov Chain Monte-Carlo (MCMC) to estimate R(t). In our probabilistic modelling we assumed that the serial interval was drawn from a Gamma distribution with α=6.0 and β=1.5 as in[14]. By sampling successive chains from the system we obtain a distribution over R values at each timepoint t, which allows us to report a median and 95% confidence intervals. These estimates of uncertainty do not take into account the uncertainty in the estimate of incidence. PyMC3 was used for the probabilistic modelling.

Hotspot detection

A hotspot is defined as a sudden increase in the number of cases in a specific geographic region. To be useful for policymakers, these regions should be as circumscribed as possible to allow for highly localised interventions that minimise the total disruption to the country. We produce two rankings of UTLAs in England. The first ranks each by their estimated prevalence $P_{H.}$ This has the advantage of being pre-registered; a list of the top ten UTLAs according to P_H has been published online since 23 July. However, this does not allow the direct identification of areas of concern, i.e. areas with a large number of new cases, and so we 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, containing a mean of 370,000 people. We used the Government data to produce daily reference rankings of each UTLA., based on seven-day moving averages of daily cases per UTLA. We included all tests performed 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 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 historic reference using two metrics. The first, recall at 20, is the number of UTLAs in our top twenty that appear in the reference top twenty. The second, the normalised mean reciprocal rank at 20, measures the agreement between ranks of our top twenty list.

Results

Data

Table 1 reports the characteristics of the subjects in the population. From 24 March to 29 September, 2,666,419 participants signed up to use the app in England, completing a total of 115,268,569 daily assessments. Between 28 April and 29 September, 851,250 invitations for swab tests were sent out. Of these 169,682 people reported a swab test on the app, of which 1,912 tested positive.

		N	Percentage
Users		2,666,419	---
Daily reports		115,268,569	
Age in years		45.39 (15.56)	---
Female		1,649,959	61.88%
Pre-conditions*	Kidney disease	30,910	0.71%
	Lung disease	459,783	11.92%
	Heart disease	116,641	2.68%
	Diabetes	129,773	2.99%
	Cancer**	35,673	1.33%
Smokers		399,469	8.60 %

Table 1. Characteristics of all app users in England that signed up between 24 March and 29 September, 2020.
*Not all participants answered questions on preconditions. Numbers reported are absolute number of users that reported having the precondition, percentages are calculated amongst all participants that answered each precondition question.
**Question asked from 29 March

Incidence

Figure 1A compares England-wide incidence estimates I_S to Government testing data, and the ONS survey[7]. We include two estimates from the ONS: the official reports, released every week, and the results from time-series modelling. The reports represent the ONS best estimate at the time of release, whilst the times-series model can evolve and lead to revision of previous estimates in response to new data (see Figure S1 for an example of this). The Government figures are consistently lower than other estimates because there is no attempt to produce 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 7 July and 5 August and who did not get tested; we found the number to be 59%. We used this 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, with a trend in agreement with the Government and ONS. All three estimates show an increase in the number of daily cases from mid-August. Estimates of incidence per NHS region are shown in Figure S2.

Prevalence

Figure 1B compares our England-wide prevalence estimates P_A and P_H to prevalence reported by the ONS (both modelled and reported) and REACT studies. The app-based assessments P_A indicate a continuous drop in the number of cases from 1 April, following the lockdown measures instigated on 23 March, plateauing in mid-June and beginning to rise again sharply from early September. The trends observed for P_A agree with data from the ONS survey, which begins on 26 April, and the REACT study, which begins on 1 May. There is some divergence in late September, when both P_A and REACT show a sharper rise in prevalence than the ONS study. The absolute values are similar for the three studies, but it is important to note that P_A only captures symptomatic cases, whilst the ONS and REACT also capture asymptomatic cases which are thought to account for 40-45% of the total cases[15] - taking this into account, P_A is slightly higher than ONS and REACT.

The combined app and swab prevalence estimate, $P_{H,}$ can only be estimated from mid-June due to its reliance on invited testing and the model of recovery. The model agrees with the trends in the other estimates, and predicts a rise in cases in late September at a similar rate to P_A and REACT. The absolute values are consistently lower than other estimates. The recovery model used to calculate P_H is shown in Figure 2. Whilst most users recover in 7-10 days, the curve shows there is a significant minority who take longer than three weeks to recover from COVID-19.

The size of our dataset allows us to estimate prevalence for more granular geographic regions than the ONS (see Supplementary Figure S3). We considered that our estimates of prevalence might be biased by a user's health-seeking behaviour; that is, a user may be more likely to begin using the app if they suspect they have COVID-19. We sought to assess the influence of this factor by removing from the analysis all users who reported being sick upon sign-up. Results are shown in Figure S4.

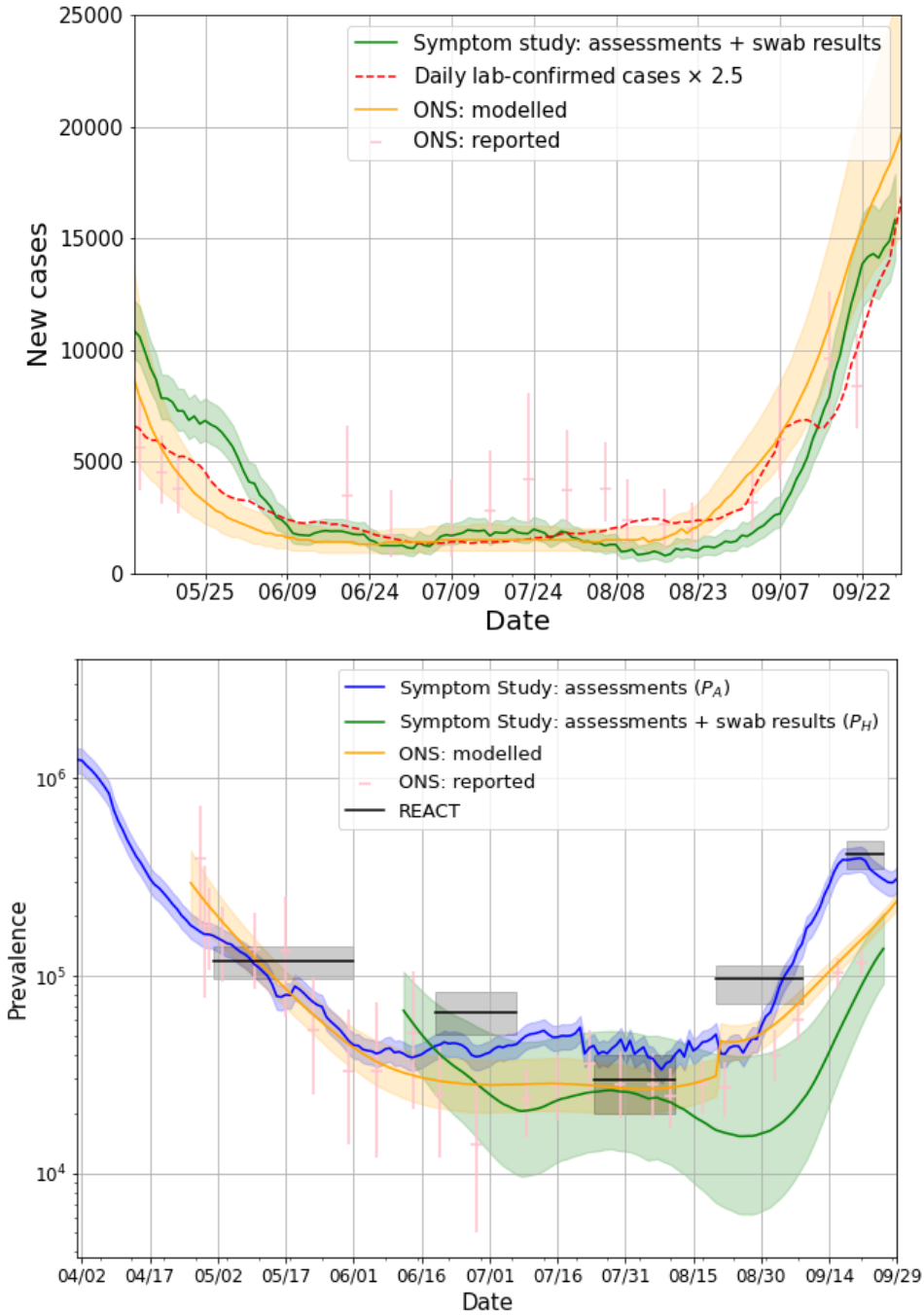

Figure 1. A) Daily incidence since 12 May in the UK compared against daily lab-confirmed cases and the ONS study **B)** Daily prevalence in the UK, compared with the ONS and REACT studies. ONS data is taken from the report released on 9 October 2020. ONS report dates are taken as the midpoint for the date range covered by the estimate.

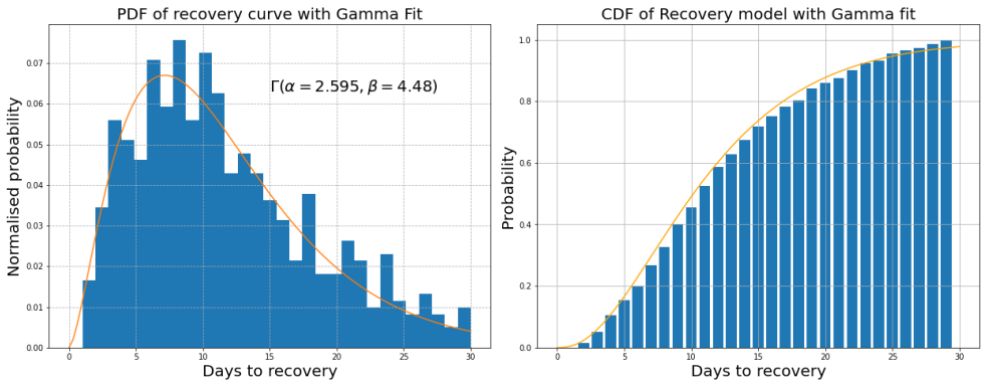

Figure 2. Left: Empirical PDF of days to recovery along with a Gamma fit. Right: Empirical CDF of days to recovery along with the same Gamma fit.

Estimating R(t)

Figure 3 shows estimates of R(t) for each of the NHS regions in England between 24 June and 28 September, compared with the consensus estimates provided by the UK Government's Science Pandemic Influenza Modelling group (SPI-M)[16]. The estimates both agree that R(t) has been above 1 from early September. The Government estimates are much smoother, likely because they are derived from a consensus from the R(t) estimates of many groups.

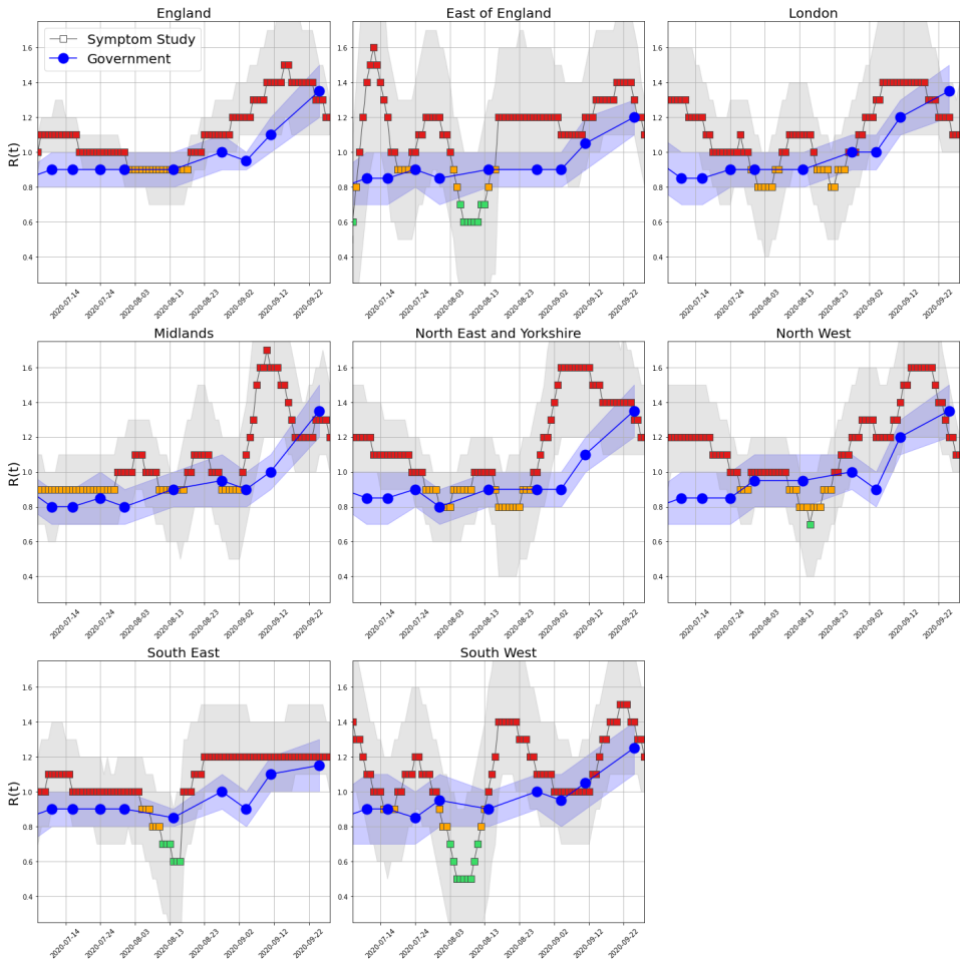

Figure 3. Estimated R(t), for NHS regions in England between 24 June and 28 September. The R values translating a high rate of infections are highlighted by the red colour, whilst dark yellow and green represent medium to low rates of infections, respectively. The blue line represents the UK government estimates published every 7-12 days from 12 June.

Hotspot detection

Figure 4 shows the results for hotspot detection in England. Ranking based on incidence consistently outperforms the prevalence-based ranking. The incidence-based ranking produces recall scores between 0.35-0.80, indicating we can predict between 7-16 of the regions in the Government's top twenty. Figure 5 compares agreement of weekly cases in each UTLA between Symptom Study and Government numbers against the number of Government Pillar 2 tests carried out. The correlation indicates that the two estimates agree best when the Government carries out more tests. This means that disagreements between the two rankings may be partially explained by insufficient testing, indicating that our results could provide valuable forecasting in regions with poor testing provision. Predictions from our hotspot predictor have shown good agreement with Government policy. On 24 June, our hotspot detection method suggested Leicester as an area of concern. On 29 June, Leicester became the first English region to be placed under local lockdown. Among the ten highest-ranking UTLAs on 23 July, five were subject to local lockdown by the UK government on 30 July (out of nine).

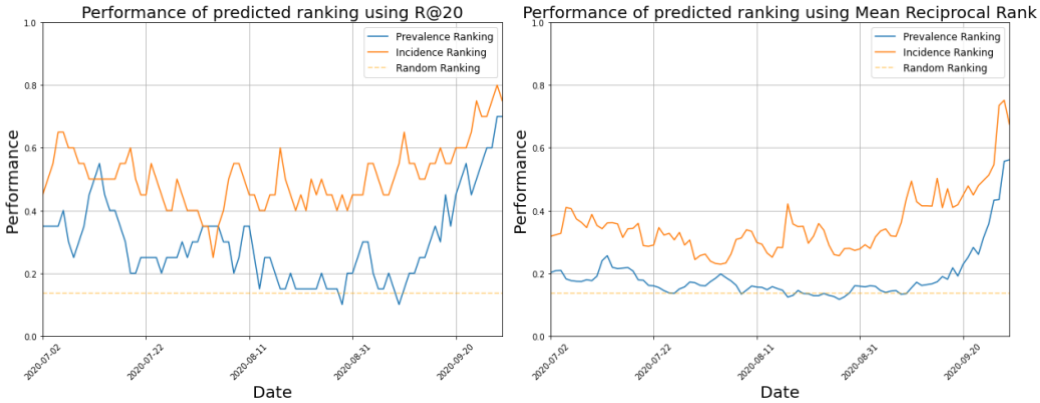

Figure 4. Performance of our two ranking methods: ranking by prevalence and incidence ranking on two metrics, Recall @ 20 and the Normalised Mean Reciprocal Rank.

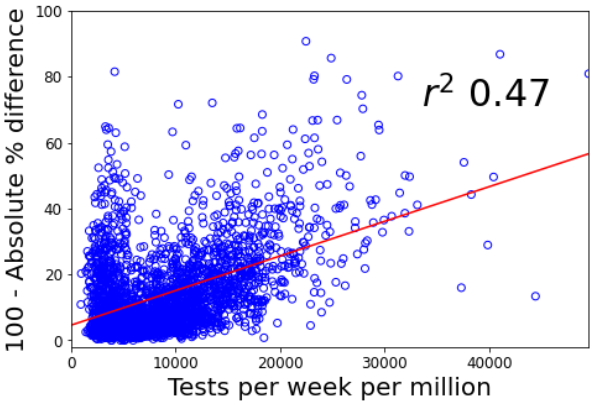

Figure 5. Agreement between Symptom Study and Government case numbers per week and UTLA, against the number of Government Pillar 2 tests carried out. These results show there is more agreement when the Government carries out more tests.

Discussion

In this work, we have demonstrated the use of population-wide data reported through the COVID Symptom Study app, a mobile application to predict the prevalence, incidence, R(t), and regional outbreaks of COVID-19. Using over 115 million daily reports from more than 2.6 million users, we were able to apply a symptom-based model to estimate the disease prevalence time-course in England. By further combining our symptom reports with swab-test results reported through the app, we were able to estimate the daily incidence and R(t). Other digital surveys have been used to provide valuable real time information about the pandemic. These include How We Feel[17], Corona Israel[5], the Facebook Survey, and CovidNearYou. However, we believe we are the first to provide national-level disease surveillance, and find good agreement with traditional, representative community surveys.

We further demonstrate how our data could be used to produce geographically granular estimates and produce a list of potential hotspots. The list consistently flags a number of regions highlighted by the Government's testing data. Whilst we compare to data from Government testing in our results, it must be noted these cannot be considered ground truth estimates of COVID-19 cases. The Government data is an incomplete sampling of new COVID-19 cases. Results from our app indicate that only 40% of those who report classic COVID-19 symptoms (a persistent cough, fever, or loss of taste or smell) go on to receive a test, and this is likely an overestimate given that the app directly contacts users who report symptoms and invites them to take a test. Furthermore, testing capacity is not uniform across all UTLAs. Our results indicate that our case estimates agree best with Government estimates in areas where their levels of testing per capita are high, suggesting that our estimates could prove a valuable resource in regions with limited testing. Furthermore, our approach requires modest numbers of PCR tests: we recorded 170,000 tests across the whole study period, or 1,100 tests per day whilst the Government performed 21 million tests over the same period, or an average of 137,000 per day[18]. This suggests our approach could prove valuable in countries where testing infrastructure is more limited.

Other efforts to track the national progression of COVID-19 rely on repeated self-swabbing from a cross-sectional community cohort. Two such efforts exist in England: the ONS study[7] and the REACT study[8,9]. These studies directly contact participants and invite them to participate. 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 currently report between 120,000-160,000 participants in England, whilst the app reports over 2,600,000 users enrolled in England. The ability to use self-reported symptom data from this large cohort enables us to make predictions of more geographically granular regions than either the ONS or REACT studies, enabling us to predict COVID-19 hotspots at the UTLA level. Our estimates should thus be viewed as independent and complementary to those provided by the ONS and REACT 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. For instance, there is a clear shift in age and gender compared to the general population, with the app users tending to be younger and more female, and our users tend to live in less deprived area[11]. We account for these differences when producing prevalence estimates using specific census adjusted population strata. However, some will not have been addressed. For instance, differences in reported symptoms across age groups[19] would likely lead to different prediction models of COVID-19 positivity. Furthermore, the app population is less racially and ethnically diverse than the general population, and people from ethnic minorities are at elevated risk of COVID-19 infection[20]. Other sources of error include collider bias[21] 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 how our estimates may be affected by app usage being affected by worries about being sick, but acknowledge there are many other biases that may affect our results and that our results must be interpreted with this in mind.

Conclusion

We have presented a means of combining app-based symptom reports and targeted testing from over 2.6 million users to estimate incidence, prevalence and R(t) in England. Using this data, we were able to compile a list of hotspots that can highlight regions which may have concerning increases in COVID-19 cases. We suggest this approach could be an effective way for Governments to monitor the spread of COVID-19 and identify a potential area of concern that may require intervention.

Ethics

Ethics has been approved by KCL Ethics Committee REMAS ID 18210, review reference LRS-19/20-18210 and all participants provided consent.

Data sharing

Data collected in the COVID-19 Symptom Study smartphone application are being shared with other health researchers through the UK National Health Service-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 (https://web.www.healthdatagateway.org/dataset/fddcb382-3051-4394-8436-b92295f14259). US investigators are encouraged to coordinate data requests through the Coronavirus Pandemic Epidemiology Consortium (https://www.monganinstitute.org/cope-consortium).

Author Contributions

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

Declaration of interests

JW, RD, JCP, SG are employees of Zoe Global Ltd. TDS is a consultant to Zoe Global Ltd.

Acknowledgements

Zoe provided in kind support for all aspects of building, running and supporting the app and service to all users worldwide. Support for this study was provided by the NIHR-funded Biomedical Research Centre based at GSTT NHS Foundation Trust. Investigators also received support from the Wellcome Trust, the MRC/BHF, Alzheimer's Society, EU, NIHR, CDRF, and the NIHR-funded BioResource, Clinical Research Facility and BRC based at GSTT NHS Foundation Trust in partnership with KCL, the UK Research and Innovation London Medical Imaging & Artificial Intelligence Centre for Value Based Healthcare, and the Wellcome Flagship Programme (WT213038/Z/18/Z). LHN is supported by the American Gastroenterological Association Research Scholars Award. DAD is supported by the National Institute of Diabetes and Digestive and Kidney Diseases K01DK120742. ATC was supported in this work through a Stuart and Suzanne Steele MGH Research Scholar Award.The Massachusetts Consortium on Pathogen Readiness (MassCPR) and Mark and Lisa Schwartz supported MGH investigators (LHN, DAD, ADJ, CGG, WM, CHL, ATC). Special thank you to Catherine Burrows.

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? 2020. https://www.bmj.com/content/370/bmj.m2679.abstract.

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**: 425–9.

4 Peto J, Alwan NA, Godfrey KM, *et al.* Universal weekly testing as the UK COVID-19 lockdown exit strategy. *Lancet* 2020; **395**: 1420–1.

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**: 634–8.

6 Drew DA, Nguyen LH, Steves CJ, *et al.* Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science* 2020; published online May 5. DOI:10.1126/science.abc0473.

7 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. https://www.medrxiv.org/content/10.1101/2020.07.06.20147348v1.abstract.

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. https://www.medrxiv.org/content/10.1101/2020.09.30.20204727v1.abstract.

10 Menni C, Valdes AM, Freidin MB, *et al.* Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat Med* 2020; published online May 11. DOI:10.1038/s41591-020-0916-2.

11 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. Epidemiology. 2020; published online April 27. DOI:10.1101/2020.04.23.20076521.

12 Wajnberg A, Mansour M, Leven E, Bouvier NM, Patel G. Humoral immune response and prolonged PCR positivity in a cohort of 1343 SARS-CoV 2 patients in the New York City region. *medRxiv* 2020. https://www.medrxiv.org/content/10.1101/2020.04.30.20085613v1.abstract.

13 Bettencourt LMA, Ribeiro RM. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS One* 2008; **3**: e2185.

14 Nishiura H, Linton NM, Akhmetzhanov AR. Serial interval of novel coronavirus (COVID-19) infections. *Int J Infect Dis* 2020; **93**: 284–6.

15 Oran DP, Topol EJ. Prevalence of Asymptomatic SARS-CoV-2 Infection : A Narrative Review. *Ann Intern Med* 2020; **173**: 362–7.

16 UK Government Published R Estimates. https://www.gov.uk/guidance/the-r-number-in-the-uk. (accessed Sept 30, 2020).

17 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.

18 Government testing dashboard. https://coronavirus.data.gov.uk/testing (accessed Oct 20, 2020).

19 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 and Ageing. 2020. DOI:10.1093/ageing/afaa223.

20 Lo C-H, Nguyen LH, Drew DA, *et al.* Racial and ethnic determinants of Covid-19 risk. *medRxiv* 2020. https://www.medrxiv.org/content/10.1101/2020.06.18.20134742v1?rss=1%22.

21 Griffith G, Morris TT, Tudball M, *et al.* Collider bias undermines our understanding of COVID-19 disease risk and severity. *medRxiv* 2020. https://www.medrxiv.org/content/10.1101/2020.05.04.20090506v3.abstract.

Supplementary material

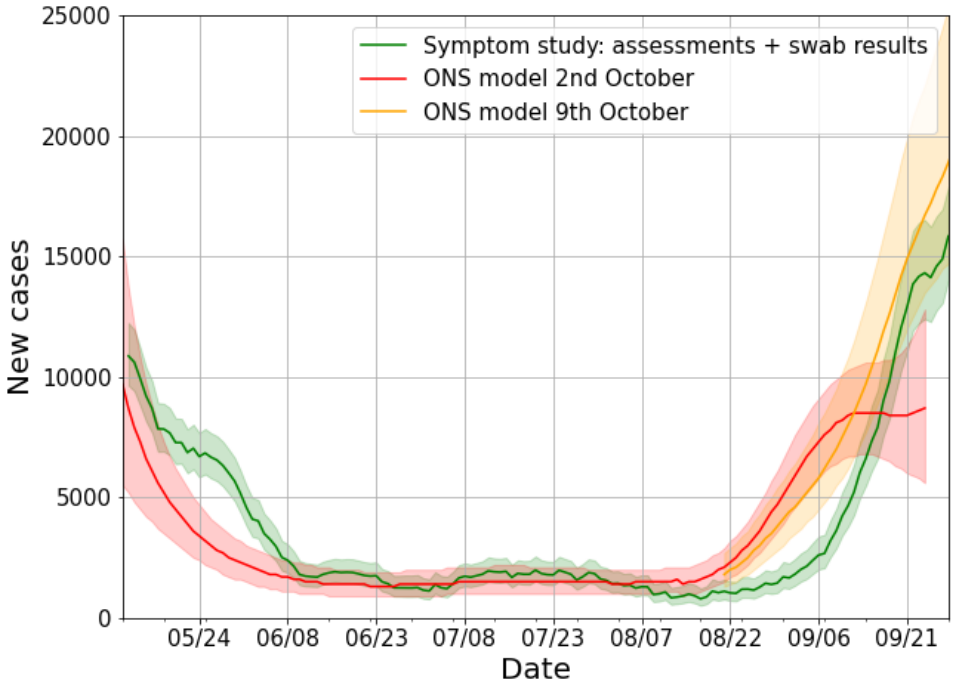

Figure S1: Comparison of our incidence data with two ONS models, released on 2 October and 9 October. The model released on 2 October showed incidence levelling off in

September, in disagreement with our released incidence estimates. By contrast the 9 October model showed a rapid increase in daily cases throughout September.

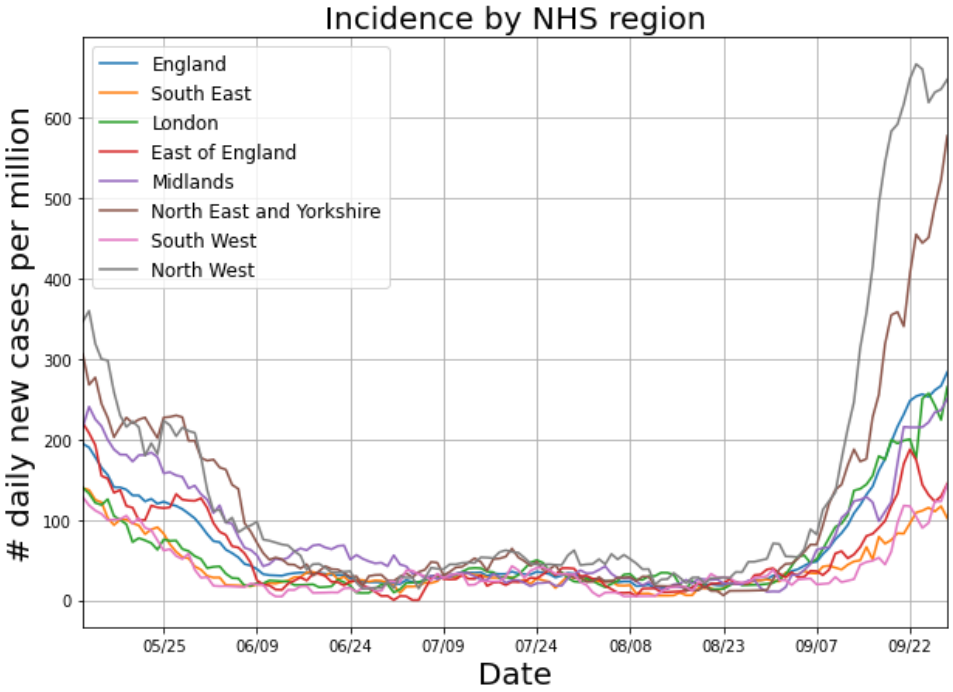

Figure S2: Estimated incidence for each NHS region in England.

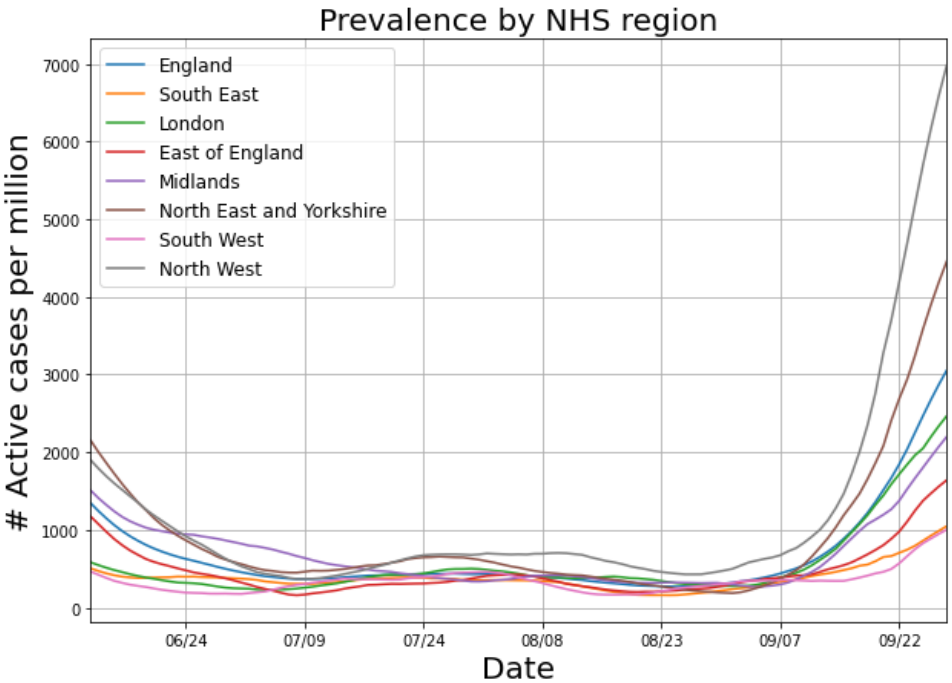

Figure S3: Estimated prevalence for each NHS region in England.

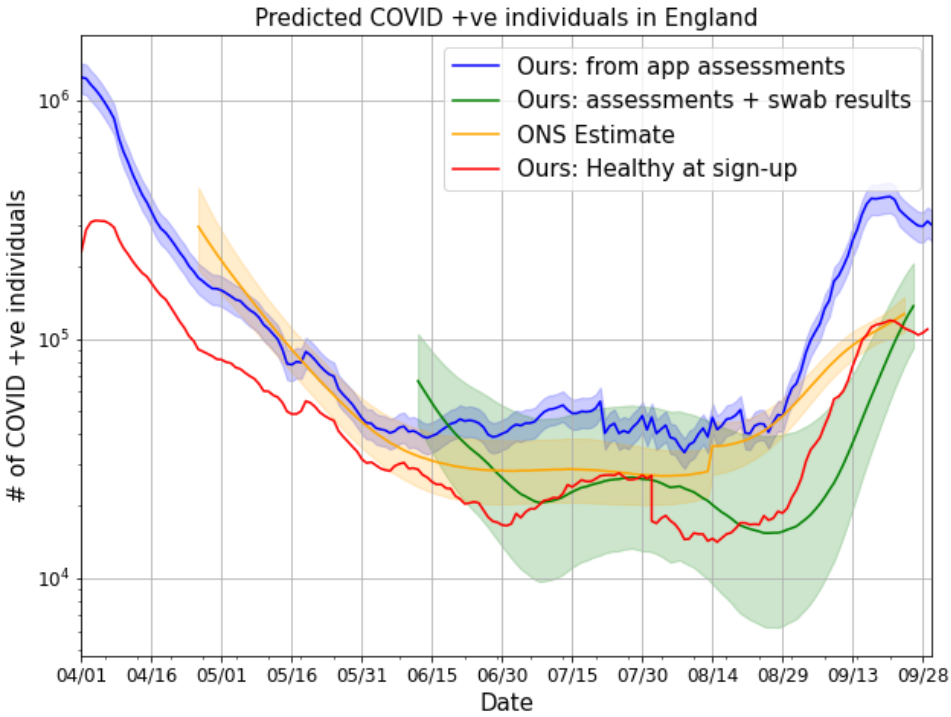

Figure S4: Estimates of prevalence, including an estimate obtained when all users who are sick upon sign-up are dropped from the analysis.