

# The epidemiological impact of the NHS COVID-19 App

Chris Wymant<sup>\*1</sup>, Luca Ferretti<sup>\*1</sup>, Daphne Tsallis<sup>2</sup>, Marcos Charalambides<sup>3</sup>, Lucie Abeler-Dörner<sup>1</sup>, David Bonsall<sup>1</sup>, Robert Hinch<sup>1</sup>, Michelle Kendall<sup>1,4</sup>, Luke Milsom<sup>5</sup>, Matthew Ayres<sup>3</sup>, Chris Holmes<sup>1,3,6</sup>, Mark Briers<sup>3</sup>, Christophe Fraser<sup>†1</sup>

\*equal contribution

†correspondence to [christophe.fraser@bdi.ox.ac.uk](mailto:christophe.fraser@bdi.ox.ac.uk)

<sup>1</sup>Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department for Medicine, University of Oxford, Oxford, UK

<sup>2</sup>Zühlke Engineering Ltd., London, UK

<sup>3</sup>The Alan Turing Institute, London, UK

<sup>4</sup>Department of Statistics, University of Warwick, Coventry, UK

<sup>5</sup>Department of Economics, University of Oxford, Oxford, UK

<sup>6</sup>Department of Statistics, University of Oxford, Oxford, UK

Last updated 9 Feb 2021 - For public release

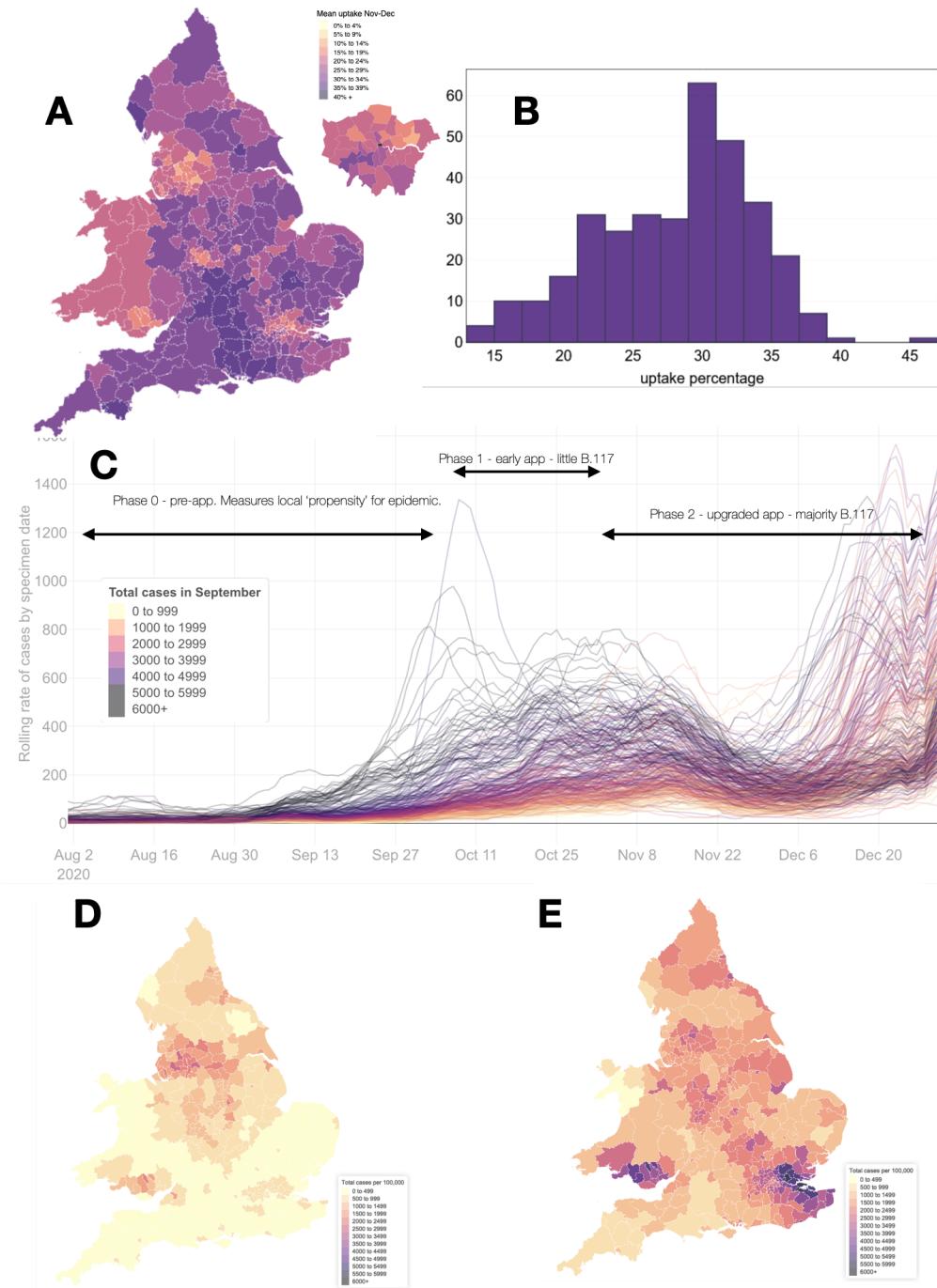
## Abstract

Since its launch on 24 September 2020, the NHS COVID-19 app has been downloaded to over 21 million phones, and used regularly by approximately 16.5 million users in England and Wales, which is 49% of the eligible population with compatible phones, and 28% of the total population. The main epidemiological impact of the app to date has been from the exposure notification function, which automates contact tracing from confirmed cases based on digital measurements of proximity events between phones. The app sent 1.7 million exposure notifications: 3.2 per index case, or 4.4 per index case who consented to be contact traced. We estimated that 6.1% of app-notified individuals subsequently tested positive (the secondary attack rate, SAR), comparable to the SAR for manual contact tracing (7.3% for close contacts and 13.5% for direct contacts). We estimated infected cases averted by the app in October-December 2020 using two conceptually complementary approaches. Modelling based on the observed notifications and SAR yielded 284,000 (224,000-344,000) averted cases, whilst statistical comparison of matched neighbouring local authorities yielded 594,000 (317,000-914,000) averted cases, i.e. about one case averted for each case consenting to notification of their contacts through the app. Improvements in the app notification system gave increased epidemiological effectiveness of the app, further supporting evidence for an effect of the app. We estimated that for every 1% increase in app users, the number of infections can be reduced by 0.8% (from modelling) or 2.3% (from statistical analysis).

The United Kingdom has been heavily hit by the COVID-19 pandemic, recording one of the highest confirmed death rates in the world in 2020 and experiencing two waves in spring and autumn, and a third in winter due to the B.1.1.7 SARS-CoV-2 variant. In order to help control the spread of the disease, the NHS COVID-19 app was launched in England and Wales on 24 September 2020, early in the second wave. Out of 33.7 million eligible people with compatible

smartphones, the app has been downloaded on 21 million unique devices, and is regularly used by at least 16.5 million users. The app provides a number of functions that were expected to reduce the spread of COVID-19. The main function is ‘digital contact tracing’ (Ferretti, Wymant, et al. 2020; Kretzschmar et al. 2020; Lunz, Batt, and Ruess 2020; Abueg et al. 2020; Lambert 2020; Cencetti et al. 2020) using the Google Apple Exposure Notification (GAEN) system. GAEN is embedded in the Android and iOS operating systems and allows privacy-preserving contact tracing (Google/Apple 2020; Troncoso et al. 2020). App users can be notified if they had a contact with another user later confirmed to have COVID-19. If the exposure has characteristics that exceeded a risk threshold, the user will receive an exposure notification and an instruction to quarantine. App-based contact tracing is a novel public health measure, with limited empirical validation or estimates of epidemiological impact. Users also check into venues using a custom QR code scanner, and may receive notifications that they have visited risky venues. This feature enables a degree of “backwards” contact tracing, where a common source for clusters of cases can be found (Endo et al. 2020). In addition, the app includes a symptom checker linked to the booking of PCR tests, automated reporting of a majority of test results in the app, and locally appropriate information on COVID-19 prevention.

When installing the app, users enter their postcode district (the first half of the postcode), allowing analysis of geographic variation in app use. For consistency between data sources, we aggregated data at the level of lower tier local authorities (LTLAs), of which there are 338 in England and Wales. We defined app uptake as the mean number of active users over November and December, divided by the population size. App uptake was variable between LTLAs (Figures 1A and 1B), with an interquartile range of 24.8 - 33.2%. We defined three phases for the analysis, shown in Figure 1C: phase 0 before app launch, phase 1 from 1 October to early November 2020 (first version of app) and phase 2 from early November to 31 December 2020 (improved version of app). These are described in more detail in Table 1. Phases in the app precede phases in the resulting cases: there is a lag between changes in transmission rates and changes in confirmed cases, we assumed by 8 days. Other factors beside the app changed during these phases. Phase 1 saw increases and local targeted and tiered control measures, phase 2 saw first a reduction due to the November lockdown and then a rise in December, mostly driven by the new variant B.1.1.7 (Figure 1C-E).

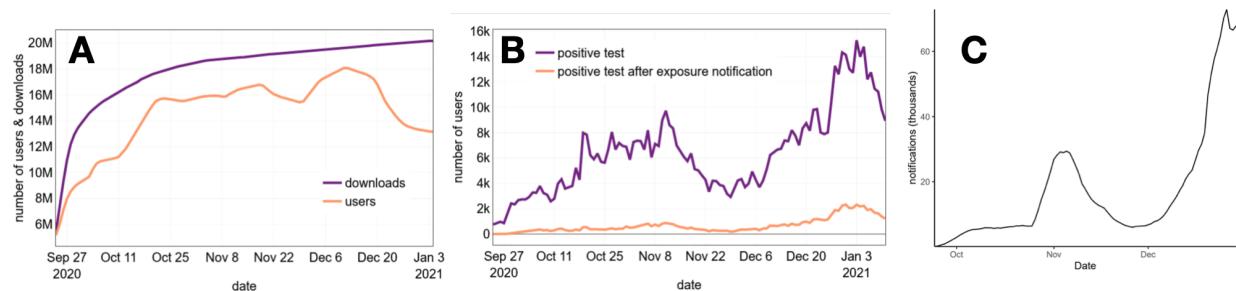


**Figure 1: geographical variability, at the Lower Tier Local Authority (LTLA) level, of both app uptake and number of COVID cases.** Panel A: mean app uptake by LTLA in phase 2/3, as a map. Panel B: mean app uptake by LTLA in phase 2/3, as a histogram. Panel C: the seven-day rolling mean of the number of daily confirmed COVID-19 cases per 100,000 population, by date of specimen, coloured by the cumulative per-capita number of cases in September. Each LTLA is shown as one line. Bold black horizontal arrows indicate the different phases of the app intervention, and the appearance of the new B.117 viral variant of concern. Panel D: cumulative number of confirmed cases (by specimen date) in phase 1, by LTLA. Panel

E: as Panel D but in Phase 2. Panels C-E are based on public data on the total number of people tested positive for COVID, not app users specifically.

Phase	Date range for cases*	Rationale	Additional context
0	1 Aug 2020 - 30 Sept 2020	Two months of baseline data before the introduction of the app. Used for adjustment and stratification of analyses. We ignore the effects of pilot studies in Newham and the Isle of Wight.	Infection rates grew approximately exponentially.
Deployment	1 Oct 2020 - 7 Oct 2020	The number of app users increased rapidly in the first week after launch (24-30 Sep). This phase is excluded from our statistical analysis (which assumes a steady-state number of users) but not from our modelling approach.	
1	8 Oct 2020 - 6 Nov 2020	The phase begins with stabilisation of number of app users, and ends with release of version 3.9 of the app (29 Oct). During this phase the app used a simple implementation of GAEN with a conservative (high) threshold for triggering exposure notifications.	Local alert levels (3-tier system) in England from 14 Oct. Firebreak lockdown in Wales 23 Oct - 9 Nov.
2	7 Nov 2020 - 31 Dec 2020	The phase begins with release of version 3.9 of the app, when a custom risk scoring system for exposure notification was introduced in England and Wales and the number of notifications per index case increased immediately by a factor of 2. The phase ends when the threshold risk score for exposure notification was reduced, leading to a further immediate increase in the number of notifications.	England lockdown until 2 Dec, 3-tier system afterwards. Emergence of the new B.1.1.7 viral variant. Tier 4 introduced on 20 Dec.
3	1 Jan 2021 onwards	We exclude this phase from our analysis, because the app was sufficiently different from phase 2 to be separate, but there were not enough data for reliable analysis.	
App	8 Oct 2020 - 11 Jan 2021	Combined period for the overall effect of the app, at the present time of analysis	

**Table 1.** Phases of the analysis, and their justification in terms of changes to the NHS COVID-19 App. \*Changes in case numbers resulting from changes to the app will lag behind in time; we assumed by 8 days, hence for example the change in the app occurring on 23 December is not expected to be reflected in case numbers until 31 December.



**Figure 2: characterising the behaviour of the app.** A: the total number of app downloads and active users over time. Fluctuations in app users are artifactual, driven by reporting issues on Android handsets, such that the estimate of ‘active users’ is a lower bound estimate. B: the daily number of app users receiving a confirmed positive test, and the daily number of those who had

already been advised to quarantine due to receiving an exposure notification. C. The total number of notifications triggered each day.

Roughly 1.75 million exposure notifications were sent as a result of 560,000 app users testing positive over the whole time period: a mean number of notifications per index case of 3.2. Since 72% of app-using index cases consented to app-based contact tracing after receiving a positive test, the number of notifications per contact tracing event was 4.4. For comparison, the mean number of contacts traced per index case by the central manual tracing system in the same period is 1.8 in phases 1 and 2 combined.

(<https://www.gov.uk/government/publications/nhs-test-and-trace-england-statistics-24-december-to-30-december>). Unlike for manual contact tracing, we do not know any details of the nature of the contacts that resulted in a contact tracing notification. We can however indirectly infer that they are likely in large part non-overlapping with those found by manual contact tracing. The majority of contacts found by manual contact tracing live in the same household as the index case (68% according to (Lee et al. 2021)). In England and Wales, each adult lives with an average of 0.9 other adults, and therefore, given that users of the app must be aged over 16, it seems highly likely that most of the 4.4 adults contact traced by the app do not live in the same household as the index case. The total number of notifications is shown in Figure 2.

An important question regards the probability that an app user who receives an exposure notification is actually infected. We partially address this by reporting the secondary attack rate (SAR) in app users, which is the probability that someone who receives a notification later tests positive within two weeks of being notified. We estimated this to be 6.1%.

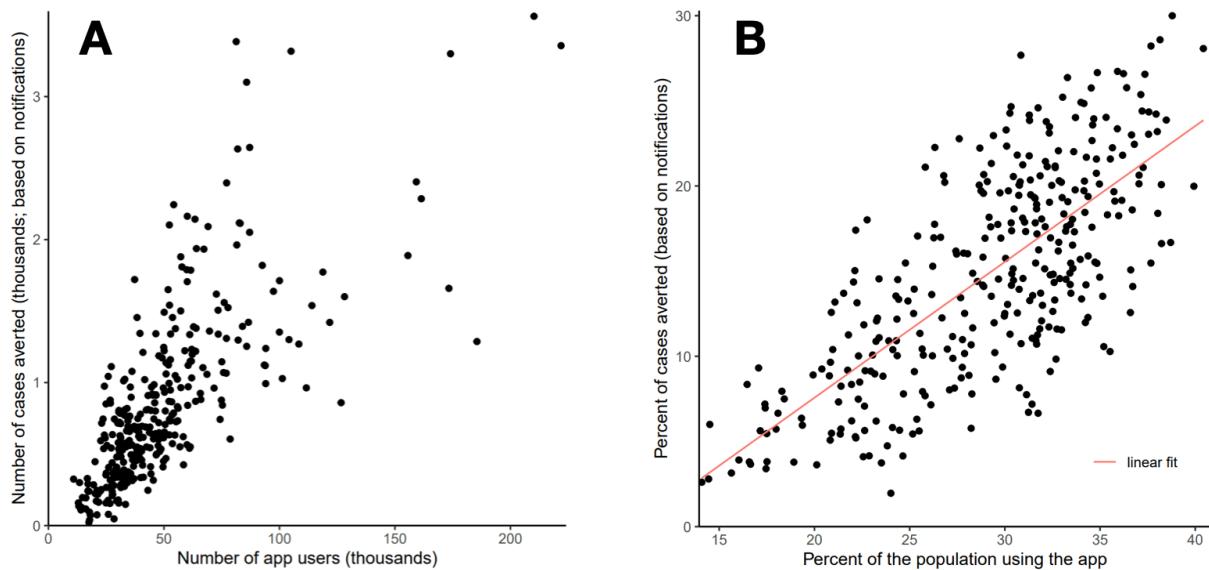
This estimate does not count individuals who were infected, but who did not test, or who did test but did not enter the result in the app; only individuals who were notified and received a positive test contribute to this, whilst all notified individuals are counted in the denominator. Data to estimate this was available only for phase 3, in January 2021, after a change in the threshold for risk calculation which meant that more notifications were being sent than during phases 1 and 2. Our estimate is commensurate with the SAR observed in encounters with relatives and social contacts, at the workplace and in school (Koh et al. 2020; Public Health England 2020b; Lee et al. 2021). The SAR for close contacts identified by manual contact tracing during December 2020 and January 2021 was 7.2% for close contacts and 13.0% for direct contacts (Public Health England 2020b). An in depth investigation of the SAR for manual contact tracing in an earlier time period (Lee et al. 2021), perhaps better corresponding to our period of analysis, suggested an overall SAR of 6.4%, and a SAR of 4.6% for contacts not with other members of the household. In summary, contact tracing with the app identifies contacts who are approximately equally likely to be infected than with manual contact tracing.

These results indicate that the app is functioning at a technical level: sending an appropriate number of exposure notifications, and to the right people. Technical functioning of the Swiss and Spanish COVID-19 tracing app have recently been demonstrated (Salathé et al. 2020; Rodríguez et al. 2021). The epidemiological functioning of the app -- the number of infections

averted -- depends on the degree of complementarity of the app with other interventions, the timeliness of exposure notifications (see Supplementary figure 3), and adherence to quarantine.

To evaluate this epidemiological effect we used two complementary approaches. First, we modelled the number of cases averted using the number of observed notifications and the SAR in app users. Second, we performed statistical comparisons of cumulative cases in neighbouring LTLAs with a similar baseline infection rate.

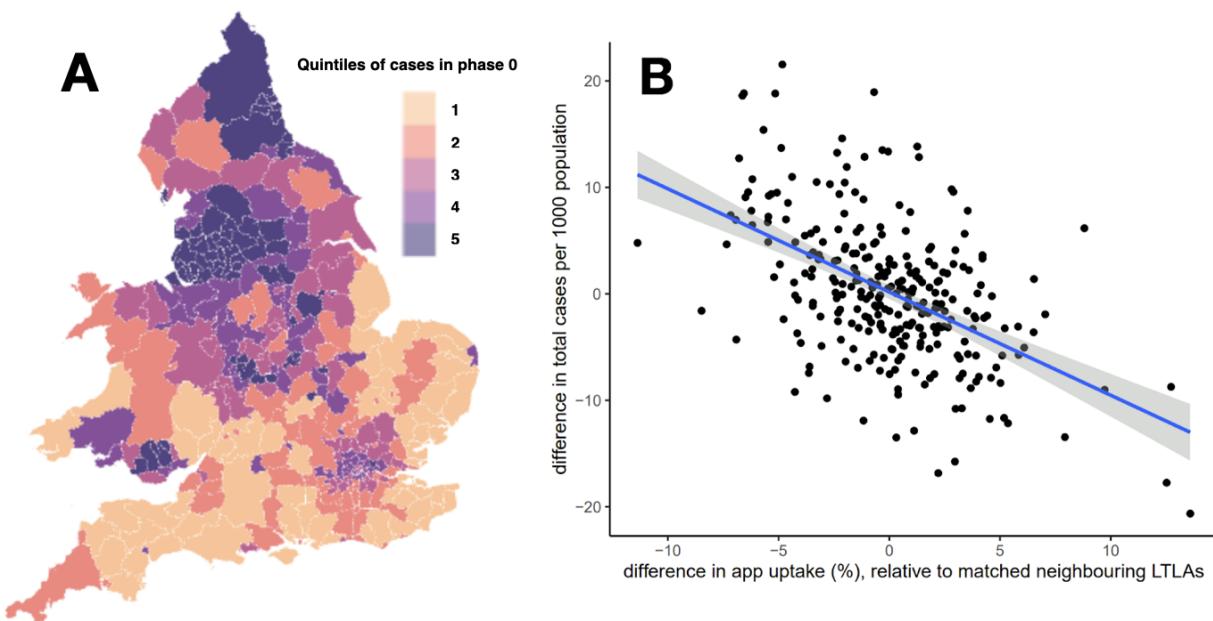
To model the impact of the app, we started from the number of notifications in each local area (Figure 2C and 2D) and modelled the likely impact in averting cases downstream. For a notification to prevent infections the notified person must be infected, and so calculating the number of cases averted requires multiplying the number of notifications by the SAR. The number of transmissions averted depends also on the timeliness of the notification relative to onwards transmission (Ferretti, Wymant, et al. 2020), which can be quantified in terms of the delay between exposure and notification. Adherence to instructions to quarantine is critical but hard to assess reliably. UK surveys found only 11% of individuals in quarantine declared proper adherence to quarantine rules, but 65% of individuals intended to adhere to quarantine (Smith et al. 2020), albeit imperfectly. In this more pessimistic scenario, the total number of predicted averted cases during the period 24 Sep - 29 Dec would be 176,000, assuming 50% adherence to imperfect quarantine (95% credibility interval: 56,000-295,000). Recent surveys found a high adherence to quarantine (greater than 80%) (Fancourt et al. 2020), and this behaviour may be more representative of app users. In this more optimistic scenario, the number of averted cases is approximately 382,000 (95% credibility interval: 364,000-399,000). We considered an intermediate scenario of 61% overall effectiveness of quarantine in preventing transmissions as our central estimate, corresponding to 279,000 cases averted (95% credibility interval: 210,000-347,000). The predicted number of cases averted was higher in areas of high app uptake (Figure 3). The slope of the regression in Figure 3B indicates that the fraction of cases averted increases by 0.8% for every 1% increase in app uptake (Table 2).



**Figure 3: the predicted cases averted in each LTLA, in Phases 1 and 2 combined, as a function of local app uptake.** A: predicted number of cases averted. B: predicted percent of cases averted (among all cases observed or averted).

Our second approach to evaluate the impact of the app was to link the variation in app uptake between LTLAs with the variation in cumulative cases. We took care to account for strong confounding factors. We therefore used a stratified approach, only comparing LTLAs with similar socio-economic and geographic properties. We used several different ways of grouping LTLAs into comparable units, and found similar results; one is described below, with the others in Supplementary Information.

People who use the app live in LTLAs which, compared to the national average, have lower levels of poverty, are more rural, and have higher local GDP (see Supplementary Information section on confounding). We included these variables in the regression to measure an effect of the app adjusting for these known confounders. We expected that app uptake would also be associated with other factors that affect epidemic dynamics but which were not directly measured, such as adherence to social distancing measures, use of face masks, etc. Since these factors affected case numbers before app release, app uptake should have some correlation with case numbers even before app release. To test this, we regressed the number of cases in Phase 0 (before release of the app) on a number of covariates including the later uptake of the app. App uptake was indeed a major predictor of Phase 0 cases (see Supplementary Information). To account for this confounding effect, we stratified LTLAs into quintiles based on their Phase 0 cases, and only compared within these strata. With this stratification, the correlation between app uptake and pre-app cases disappeared, i.e. this adjusted for unmeasured confounders (see Supplementary Information, “Placebo regression”). The number of confirmed cases in an LTLA is also confounded by those in neighbouring LTLAs, therefore we only compared neighbouring LTLAs (those sharing a border). We found that the difference in number of cases per capita between neighbouring LTLAs, matched by Phase 0 case number quintile, was strongly and robustly associated with differences in app use, whether or not we adjusted for other demographic confounders (Figure 4, Table 2, and Supplementary Information for Robustness analyses).



**Figure 4.** A: a map showing the quintiles of cases per capita LTLAs during Phase 0. We only compared each LTLA to other neighbouring LTLAs within the same quintile. B: the unadjusted relationship between difference in app uptake and difference in number of cases per capita in Phases 1-3 combined.

Disaggregating the effect by phase, we found that it was larger during the second phase (Table 2). This is consistent with the sudden increase in the number of notifications sent out by the app during phase 2 (Figure 2C). Table 2 shows the estimated effect size replicated in a number of different statistical analyses (see Supplementary Information for details).

	Percent reduction in cases for every percent increase in app use		
Analysis	Phase 1	Phase 2	Overall
<b>Modelling</b>	0.32 (0.30 - 0.35)	0.92 (0.84 - 1.0)	0.79 (0.71 - 0.87)
<b>Matched neighbours regression</b>	1.09 (0.04 - 2.14) (bootstrap: 0.15 - 2.16)	2.66 (1.75 - 3.56) (bootstrap: 0.80 - 4.71)	2.26 (1.50 - 3.00) (bootstrap: 1.60 - 3.19)
Secondary analyses	Phase 1	Phase 2	Overall
Stratified linear regression in clusters	-1.05 (-2.08 - -0.04)	3.34 (2.53 - 4.14)	2.76 (2.16 - 3.35)
Matched pairs regression*	5.08 (1.77 - 8.40)	3.89 (1.05 - 6.74)	4.39 (1.70 - 7.08)
Matched pairs regression adjusted for local efficiency of manual contact tracing*	4.49 (0.21 - 8.77)	3.11 (-0.14 - 6.35)	3.67 (0.31 - 7.02)

**Table 2.** The estimated effect of the app, measured as the percent reduction in cases for every percent increase in app use, based on different analytical approaches. The two primary approaches -- modelling and matched neighbours -- are described in the main text. The three robustness checks for the statistical analysis are described in the Supplementary Information. Ranges shown are 95% confidence intervals (CIs) for regressions, and a sensitivity analysis exploring 2.5%-97.5% effectiveness of imperfect quarantine for the modelling. \*Is restricted to England only, and uses data on the national NHS Test & Trace contact tracing program aggregated by Upper Tier Local Authority. Measure of manual contact tracing quality is the proportion of contacts reached per case.

We translated both the modelling and statistical findings into estimates of cases averted in Table 3. These estimates are comparable to the number of app users who tested positive and consented for notifications to be sent: roughly 400,000. This finding suggests that on average, each confirmed case who consented to notification of their contacts through the app prevented one person from becoming infected.

Analysis	Cases averted in phases 1 and 2 combined	Deaths averted in phases 1 and 2 combined
Modelling	279,000 (210,000 - 347,000)	4,100 (3,100 - 5,100)
Matched neighbours	594,000 (317,000 - 914,000)	8,700 (4,700 - 13,500)

**Table 3.** Cases averted, as predicted by the different approaches. The total number of cases and deaths that actually occurred in phases 1 and 2 combined was 1,892,000 and 32,500, respectively. Ranges shown are 95% CIs for regressions, and a sensitivity analysis exploring 2.5-97.5% effectiveness of imperfect quarantine for the modelling. Deaths averted were estimated using the empirical case fatality rate.

Finally, we extrapolated the findings to explore different ways in which the app could be improved, by re-running scenarios with different parameters (Table 4). Even though these are retrospective projections, the expected reductions in number of cases are relevant when considering forward projections.

	Percent reduction in total case burden in phase 2 (in addition to reductions observed for the current implementation of the app)	
Analysis	Modelling	Statistical extrapolation
Increase uptake to 35.9% - current 90th percentile - for all LTLAs (Improve equity)	11% (8-13%)	21.0% (14.5-26.8%)
Increase uptake across the board by 20 percentage points (Mass improvement)	23% (18-29%)	41.5% (29.5-51.5%)
Switch to opt-out notification (5% drop-off)*	6.5% (4.9-8.1%)	NA
Improve adherence to quarantine by 20 percentage points	6.8% (5.1-8.5%)	NA
Reduce time to test result by one day**	3.6% (2.7-4.5%)	NA

**Table 4.** Scenarios for improvements, in terms of the percent reduction in total case burden that would have occurred during phase 2. This is the further reduction relative to the cases that actually occurred, not relative to cases inferred in the absence of the app. Ranges shown are 95% CIs for regressions, 2.5-97.5% effectiveness of imperfect quarantine for the modelling.

\*Currently, the app requires consent after the receipt of a positive test for contact tracing to be initiated, which is provided by 72% of users. We assume that changing to opt-out consent, e.g. by consent at registration, would increase this to 95%. \*\*Reducing test turnaround time has many benefits not modelled here, here the effect is only that mediated through the app.

## Discussion

Our analysis suggests a relatively large number of COVID-19 cases were averted by contact tracing via the NHS app, ranging from approximately 200,000 to 900,000 depending on the details of the method, compared to the 1.9 million cases that actually arose. Averted cases were concentrated in Phase 2, covering November and December 2020, after a major upgrade to the app's risk scoring function (Lovett et al. 2020). These findings can be compared to prior results from modelling, such as our report to NHSx (Hinch et al. 2020). Applying our individual-based simulation of a contact tracing app (Hinch R and Probert WJM et al. 2020), with a similar mechanism calibrated to a different setting, Abueg et al estimated that a 30% uptake of the app would result in approximately 1 infection averted for every 4 infections arising for 4.5 months of action (Abueg et al. 2020). Our estimates are in line with these expectations, and also with an approximately exponential dependence of the number of cases averted on uptake over the observed range of uptakes. Problems with contact tracing can cause large increases in infections: one estimate attributed 127,000 additional cases to the failure to manually trace contacts of 16,000 index cases in a timely manner (Fetzer and Graeber 2020). Equivalently, improvements to contact tracing -- such as adding app-based proximity detection and notification to the traditional manual system -- can cause large reductions in infections.

Though it is informative to estimate effects on the time-varying reproduction number  $R$ , as done for example in (Flaxman et al. 2020), we did not pursue this analysis here. The epidemic dynamics of individual LTLAs are difficult to interpret: the time period of analysis coincided with staggered introductions of locally targeted restrictions, a short national lockdown, the Christmas holiday season, and the emergence of the B.117 variant genotype, which is more infectious and has rapidly become the majority variant spreading nationally across the country (Volz et al. 2021; Davies et al. 2020; Vöhringer et al. 2020; Public Health England 2020a). Future work could perhaps model all of these effects in a single hierarchical model, permitting joint estimation of the app effect over LTLAs with linked drivers and dynamics. Careful construction of counterfactual time series could also be considered (Brodersen et al. 2015). Our simpler approaches have the benefit of transparency, and we hypothesise that under negative-feedback dynamics, appropriately constructed comparisons of total case counts over a sufficiently long time period may reveal underlying propensity for disease spread.

The main limitation of our analysis is that it is an observational study: no randomized or systematic experiment resulted in different app uptake in different places. Interpreting observational analyses requires particular care due to the risk of confounding. To mitigate inherent limitations, we combined a modelling approach based on the number of exposure notifications sent and the observed secondary attack rate, with a statistical approach that was stratified to focus on differences between directly comparable areas only. Our modelling approach was based on the mechanistic function of the app. Our statistical approach was based on emulating how a cluster randomised trial would have been conducted (Hernán and Robins 2016). To test for the tendency of the statistical method to be biased by confounding factors, we conducted a placebo analysis (see Supplementary Information). The method was found, reassuringly, to give a null association for the placebo analysis. Nonetheless, it remains possible that changes in app use over time and across geographies reflect changes in other interventions, and that our analysis incorrectly attributes the effect to the app. Our modelling approach is less prone to this bias than our statistical approach. A second limitation of our analysis is that we did not have data on the combined effect of manual and digital contact tracing. Based on the analysis of delays (Supplementary Information), and on the rate of notifications per index case, we may reasonably infer that the app had a separate and additional effect to manual contact tracing. We confirmed this with an analysis which included adjustment of a metric of quality of manual contact tracing, which did not affect our conclusions. Our statistical approach is less prone to this second source of bias than the modelling approach.

The app is best understood as part of a system of non-pharmaceutical interventions, not in isolation. The specific role of app-based contact tracing is to speed up contact tracing, and to reach more people per index case. Contact tracing requires case finding, and so is a follow up to effective, widespread and rapid testing. Contact tracing is not a substitute for social distancing or face covering; control of the epidemic requires all available interventions to work together. Isolation and quarantine can only be effective when financially supported; replacing quarantine by daily testing could improve both specificity and adherence. Outbreak response and backwards contact tracing are ways of increasing the effectiveness of the intervention. Better use of location-specific QR code scanning could help: notifications for 226 risky venue events

have been issued as of 20 Jan 2021. App-based contact tracing may free up resources for specialist contact tracers to focus on investigations of complex local epidemics and outbreaks.

The app is working as a public health tool. Improvements were implemented to the risk scoring model that resulted in increased accuracy and efficacy (Lovett et al. 2020). These improvements are currently specific to the NHS app, but are open source and could be incorporated into the Google Apple Exposure Notification system to improve digital contact tracing around the world.

Users should be informed that if they receive an exposure notification and are asked to quarantine, the risk that they are infected is similar to that when they were identified by manual contact tracing as a ‘close contact’, i.e. 7.3%. The secondary attack rate of 6.1% is the proportion of people notified by the app who go on to themselves test positive during the period when they were asked to isolate, or in the 14 days after. Since only a fraction of infected people ever test to become confirmed cases, it is not a measure of the proportion of these notified people who are infected. The proportion of contacts that are infected may be twice as high as the secondary attack rate, both for manual and app based contact tracing.

The surest way to increase the effectiveness of the programme is to increase uptake, and to provide material support to individuals undergoing isolation and quarantine. Special efforts may be needed to reach underserved communities. Faster testing, perhaps using point-of-care antigen tests, and integration of self-testing with the app could result in better disease control. Widespread vaccination will eventually reduce the need for non-pharmaceutical interventions, but vaccination will unlikely have global reach within the coming months, during which time improved non-pharmaceutical interventions could still prevent many infections (Galanti et al. 2020; Moore et al. 2021). Smartphone use is already global, and thus privacy-preserving contact tracing apps should be further integrated into the public health toolkit.

## Methods

**Estimating app uptake.** To monitor the safe function of the app and enable its evaluation, a limited amount of data are shared with a secure NHS server. Each active app sends a single data packet daily. The fields in these packets contain no sensitive or identifying information, and are approved and publicly listed by the Information Commissioner

[<https://www.gov.uk/government/publications/nhs-covid-19-app-privacy-information>]. For the reported numbers of downloads, repeat downloads to the same phone are counted only once. The number of *active* users each day is defined as the number of data packets received by the NHS server; for a single representative value of this quantity, we took the mean over all days in November and December (earlier data was deemed less reliable). We note that there continue to be unexplained fluctuations in reported user numbers on Android phones. To estimate uptake within an LTLA, each postcode district was mapped to the LTLA in which the majority of its population reside, and we took the ratio (number of active users in postcode districts mapped to this LTLA) / (total population in postcode districts mapped to this LTLA). The population of England and Wales is 59.91 million, of whom 47.48 are over 16 and thus eligible to use the app

[ONS]. Assuming that England and Wales is representative of the UK, we estimate that 82% of people aged 16+ have smartphones [OFCOM], and that of smartphones in circulation, 87% support the Google Apple Exposure Notification system [Department of Health and Social Care, personal communication]. The denominators for measuring uptake at the national level are therefore 59.91 million (total population) and 33.87 million (eligible population with compatible phones).

**Defining numbers of cases.** The COVID-19 case numbers per day we used here are those reported at <https://coronavirus.data.gov.uk/>, by specimen date and LTLA. We obtained per-capita case numbers at the LTLA level by dividing by LTLA populations reported by ONS. Testing has been available through the NHS Test and Trace system in all areas throughout the period, with a median delay of less than 2 days from booking a test to receiving the result. Testing capacity has mostly exceeded demand, except for two weeks in early September. We assumed that case ascertainment has been relatively constant over the period of analysis, an assumption qualitatively supported by the unbiased ONS and REACT studies (Steel and Donnarumma 2020; Riley et al. 2020).

**Estimation of the secondary attack rate amongst exposure notified people.** We focussed on a period in January when these data were linked, rather than earlier periods when isolations and test positivity were available at the population level. The SAR at time  $t_1$  was computed by comparing the number of cases in exposed users (i.e. those that become positive after being asked to isolate due to risky contact)  $I_N$  to the number of exposure notifications  $N$  over the same

period:  $SAR(t) = \sum_y p_{NP}(t|y)I_N(y)/N(t)$ , where  $p_{NP}(t_N|t_P)$  is the probability of a user being notified at time  $t_N$  conditional on getting a positive test at a later time  $t_P$ .

**Modelling cases averted based on exposure notifications and secondary attack rate.** The effect of notifications received at time  $t$  on cases averted can be modelled as the product of (i) the number of notifications, (ii) the secondary attack rate, i.e. the probability that notified individuals are actually infected, (iii) the expected fraction of transmissions preventable by strict quarantine of an infectious individual after a notification, (iv) the actual adherence to quarantine, and (v) the expected size of the full transmission chain that would be originated by the contact if not notified. Before each notification, the contact's app sends a request for permission to the central NHS server. We estimated the total number of notifications per day on each Operating System (OS: Android or iOS) from these requests. We estimated the number of notifications per LTLA from the number of partial days of quarantine (typically corresponding to the first day of quarantine, i.e. the day of notification) per day, OS and LTLA, rescaling it by a time- and OS-dependent factor to match the number of notifications per day and OS. The delay between last exposure and notification is assumed to follow a normal distribution, with time-dependent parameters estimated via Least Squares from the daily number of notifications and individuals in quarantine. The fraction of preventable transmissions is estimated from the delay distribution using the generation time distribution in (Ferretti, Ledda, et al. 2020) with mean 5.5 days. We assume 97.5% effectiveness of quarantine in preventing transmission with complete adherence,

and 50% as central value for quarantine with imperfect compliance. Finally, the size of the epidemic chain triggered by a single case is computed assuming that local epidemics do not mix and that the extra cases do not affect the epidemic dynamic.

**Statistical analysis.** Each statistic of interest, app uptake, log cases per capita for phase X, % population living in poverty, rural/urban score and local GDP, is collated by LTLA. Let  $y$  label a local authority, and let  $M(y)$  label the subset set of neighbouring (adjacent) LTLAs that are in the same quintile of cases for phase 0. Each statistic  $f_y$  measured in  $y$  is compared to the population-weighted mean in matched neighbours  $f_M(y)$ , such that  $D_f_y$  is the difference in  $f$  between  $x$  and its matched neighbours. Effects are estimated by regressing  $D_{logcases\_y}$  for each phase with predictor variables  $D_{uptake\_y}$ ,  $D_{poverty\_y}$ ,  $D_{rural\_y}$  and  $D_{GDP\_y}$ . Full results of regression are shown in Supplementary Materials. Predictions for cases averted are found using the regression coefficient  $\beta_{uptake}$  for  $D_{uptake\_y}$ , and predicting the effect of uptake being reduced to 15% (or kept as is if less than 15%) by linear extrapolation. We assume that there is little benefit of app uptake below 15%. A quadratic regression with intercept at 0 provides very similar findings (not shown). Cases averted are multiplied by the crude case fatality rate to predict deaths averted.

## References

- Abueg, Matthew, Robert Hinch, Neo Wu, Luyang Liu, William J. M. Probert, Austin Wu, Paul Eastham, et al. 2020. “Modeling the Combined Effect of Digital Exposure Notification and Non-Pharmaceutical Interventions on the COVID-19 Epidemic in Washington State.” *bioRxiv*. medRxiv. <https://doi.org/10.1101/2020.08.29.20184135>.
- Brodersen, Kay H., Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. 2015. “Inferring Causal Impact Using Bayesian Structural Time-Series Models.” *The Annals of Applied Statistics* 9 (1): 247–74.
- Cencetti, G., G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, S. Lehmann, M. Salathé, and B. Lepri. 2020. “Digital Proximity Tracing in the COVID-19 Pandemic on Empirical Contact Networks.” *bioRxiv*. medRxiv. <https://doi.org/10.1101/2020.05.29.20115915>.
- Davies, Nicholas G., Rosanna C. Barnard, Christopher I. Jarvis, Adam J. Kucharski, James Munday, Carl A. B. Pearson, Timothy W. Russell, et al. 2020. “Estimated Transmissibility and Severity of Novel SARS-CoV-2 Variant of Concern 202012/01 in England.” *bioRxiv*. medRxiv. <https://doi.org/10.1101/2020.12.24.20248822>.
- Endo, Akira, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Quentin J. Leclerc, Gwenan M. Knight, Graham F. Medley, Katherine E. Atkins, Sebastian Funk, and Adam J. Kucharski. 2020. “Implication of Backward Contact Tracing in the Presence of Overdispersed Transmission in COVID-19 Outbreaks.” *Wellcome Open Research* 5 (October): 239.
- Fancourt, D., F. Bu, H. Wan Mak, and A. Steptoe. 2020. “Covid-19 Social Study.” Results Release 28.
- Ferretti, Luca, A. Ledda, C. Wymant, L. Zhao, V. Ledda, L. Abeler-Dorner, and Others. 2020. “The Timing of COVID-19 Transmission.” *MedRxiv*. Publisher: Cold Spring Harbor Laboratory Press. <https://www.medrxiv.org/content/10.1101/2020.09.04.20188516v2>.
- Ferretti, Luca, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, and Christophe Fraser. 2020. “Quantifying SARS-CoV-2 Transmission Suggests Epidemic Control with Digital Contact Tracing.” *Science* 368 (6491).

- [https://doi.org/10.1126/science.abb6936.](https://doi.org/10.1126/science.abb6936)
- Fetzer, Thiemo, and Thomas Graeber. 2020. "Does Contact Tracing Work? Quasi-Experimental Evidence from an Excel Error in England." *bioRxiv*. medRxiv.  
<https://doi.org/10.1101/2020.12.10.20247080>.
- Flaxman, Seth, Swapnil Mishra, Axel Gandy, H. Juliette T. Unwin, Thomas A. Mellan, Helen Coupland, Charles Whittaker, et al. 2020. "Estimating the Effects of Non-Pharmaceutical Interventions on COVID-19 in Europe." *Nature* 584 (7820): 257–61.
- Galanti, Marta, Sen Pei, Teresa K. Yamana, Frederick J. Angulo, Apostolos Charos, David L. Swerdlow, and Jeffrey Shaman. 2020. "The Importance of Continued Non-Pharmaceutical Interventions during the Upcoming SARS-CoV-2 Vaccination Campaign." *bioRxiv*. medRxiv.  
<https://doi.org/10.1101/2020.12.23.20248784>.
- Google/Apple. 2020. "Exposure Notifications: Helping Fight COVID-19 - Google."  
[https://www.google.com/intl/en\\_us/covid19/exposurenotifications/](https://www.google.com/intl/en_us/covid19/exposurenotifications/).
- Hernán, Miguel A., and James M. Robins. 2016. "Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available." *American Journal of Epidemiology* 183 (8): 758–64.
- Hinch R and Probert WJM, Anel Nurtay, Michelle Kendall, Chris Wymant, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, et al. 2020. "OpenABM-Covid19 - an Agent-Based Model for Non-Pharmaceutical Interventions against COVID-19 Including Contact Tracing." medRxiv.  
<https://doi.org/10.1101/2020.09.16.20195925>.
- Hinch, Robert, W. Probert, A. Nurtay, M. Kendall, C. Wymant, M. Hall, K. Lythgoe, et al. 2020. "Effective Configurations of a Digital Contact Tracing App: A Report to NHSX."  
[https://github.com/BDI-pathogens/covid-19\\_instant\\_tracing/blob/master/Report%20-%20Effective%20Configurations%20of%20a%20Digital%20Contact%20Tracing%20App.pdf](https://github.com/BDI-pathogens/covid-19_instant_tracing/blob/master/Report%20-%20Effective%20Configurations%20of%20a%20Digital%20Contact%20Tracing%20App.pdf).
- Koh, Wee Chian, Lin Naing, Liling Chaw, Muhammad Ali Rosledzana, Mohammad Fathi Alikhan, Sirajul Adli Jamaludin, Faezah Amin, et al. 2020. "What Do We Know about SARS-CoV-2 Transmission? A Systematic Review and Meta-Analysis of the Secondary Attack Rate and Associated Risk Factors." *PloS One* 15 (10): e0240205.
- Kretzschmar, Mirjam E., Ganna Rozhnova, Martin C. J. Bootsma, Michiel van Boven, Janneke H. H. M. van de Wijgert, and Marc J. M. Bonten. 2020. "Impact of Delays on Effectiveness of Contact Tracing Strategies for COVID-19: A Modelling Study." *The Lancet. Public Health* 5 (8): e452–59.
- Lambert, Amaury. 2020. "A Mathematically Rigorous Assessment of the Efficiency of Quarantining and Contact Tracing in Curbing the COVID-19 Epidemic." *bioRxiv*. medRxiv.  
<https://doi.org/10.1101/2020.05.04.20091009>.
- Lee, Lennard Y. W., Stefan Rozmanowski, Matthew Pang, Andre Charlett, Charlotte Anderson, Gareth J. Hughes, Matthew Barnard, et al. 2021. "An Observational Study of SARS-CoV-2 Infectivity by Viral Load and Demographic Factors and the Utility Lateral Flow Devices to Prevent Transmission." *NSMS Oxford*.  
[http://modmedmicro.nsms.ox.ac.uk/wp-content/uploads/2021/01/infectivity\\_manuscript\\_20210119\\_merged.pdf](http://modmedmicro.nsms.ox.ac.uk/wp-content/uploads/2021/01/infectivity_manuscript_20210119_merged.pdf).
- Lovett, Tom, Mark Briers, Marcos Charalambides, Radka Jersakova, James Lomax, and Chris Holmes. 2020. "Inferring Proximity from Bluetooth Low Energy RSSI with Unscented Kalman Smoothers." *arXiv [eess.SP]*. arXiv. <http://arxiv.org/abs/2007.05057>.
- Lunz, Davin, Gregory Batt, and Jakob Ruess. 2020. "To Quarantine, or Not to Quarantine: A Theoretical Framework for Disease Control via Contact Tracing." *Epidemics* 34 (December): 100428.
- Moore, Sam, Edward M. Hill, Mike J. Tildesley, Louise Dyson, and Matt J. Keeling. 2021. "Vaccination and Non-Pharmaceutical Interventions: When Can the UK Relax about COVID-19?" *bioRxiv*. medRxiv. <https://doi.org/10.1101/2020.12.27.20248896>.
- Pellis, Lorenzo, Neil M. Ferguson, and Christophe Fraser. 2008. "The Relationship between

- Real-Time and Discrete-Generation Models of Epidemic Spread." *Mathematical Biosciences* 216 (1): 63–70.
- Public Health England. 2020a. "Investigation of Novel SARS-CoV-2 Variant - Variant of Concern 202012/01 - TC2." Technical briefing 2.  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/949639/Technical\\_Briefing\\_VOC202012-2\\_Briefing\\_2\\_FINAL.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/949639/Technical_Briefing_VOC202012-2_Briefing_2_FINAL.pdf).
- \_\_\_\_\_. 2020b. "Investigation of Novel SARS-CoV-2 Variant - Variant of Concern 202012/01 - TC3." Technical briefing 3.  
[https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/950823/Variant\\_of\\_Concern\\_VOC\\_202012\\_01\\_Technical\\_Briefing\\_3\\_-\\_England.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950823/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3_-_England.pdf)
- Riley, Steven, Caroline E. Walters, Huawei Wang, Oliver Eales, Kylie E. C. Ainslie, Christina Atchinson, Claudio Fronterre, et al. 2020. "REACT-1 Round 7 Updated Report: Regional Heterogeneity in Changes in Prevalence of SARS-CoV-2 Infection during the Second National COVID-19 Lockdown in England." *bioRxiv*. medRxiv.  
<https://doi.org/10.1101/2020.12.15.20248244>.
- Rodríguez, Pablo, Santiago Graña, Eva Elisa Alvarez-León, Manuela Battaglini, Francisco Javier Darias, Miguel A. Hernán, Raquel López, et al. 2021. "A Population-Based Controlled Experiment Assessing the Epidemiological Impact of Digital Contact Tracing." *Nature Communications* 12 (1): 587.
- Salathé, Marcel, Christian Althaus, Nanina Anderegg, Daniele Antonioli, Tala Ballouz, Edouard Bugnon, Srdjan Ćapkun, et al. 2020. "Early Evidence of Effectiveness of Digital Contact Tracing for SARS-CoV-2 in Switzerland." *Swiss Medical Weekly* 150 (December): w20457.
- Smith, Louise E., Henry W. W. Potts, Richard Amlot, Nicola T. Fear, Susan Michie, and James Rubin. 2020. "Adherence to the Test, Trace and Isolate System: Results from a Time Series of 21 Nationally Representative Surveys in the UK (the COVID-19 Rapid Survey of Adherence to Interventions and Responses [CORSAIR] Study)." medRxiv.  
<https://doi.org/10.1101/2020.09.15.20191957>.
- Steel, Kara, and Hannah Donnarumma. 2020. "Coronavirus (COVID-19) Infection Survey, UK: 8 January 2021." ONS.  
<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionssurveypilot/8january2021>.
- Troncoso, Carmela, Mathias Payer, Jean-Pierre Hubaux, Marcel Salathé, James Larus, Edouard Bugnion, Wouter Lueks, et al. 2020. "Decentralized Privacy-Preserving Proximity Tracing." *arXiv [cs.CR]*. arXiv. <http://arxiv.org/abs/2005.12273>.
- Vöhringer, Harald, Matthew Sinnott, Roberto Amato, Inigo Martincorena, Dominic Kwiatkowski, Jeffrey C. Barrett, and Moritz Gerstung. 2020. "Lineage-Specific Growth of SARS-CoV-2 B.1.1.7 during the English National Lockdown." <https://virological.org/t/lineage-specific-growth-of-sars-cov-2-b-1-1-7-during-the-english-national-lockdown/575>.
- Volz, Erik, Swapnil Mishra, Meera Chand, Jeffrey C. Barrett, Robert Johnson, Lily Geidelberg, Wes R. Hinsley, et al. 2021. "Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from Linking Epidemiological and Genetic Data." *bioRxiv*. medRxiv.  
<https://doi.org/10.1101/2020.12.30.20249034>.

## Acknowledgements

We thank Richard Hayes, Moritz Kraemer, Samir Bhatt, Swapnil Mishra, Jamie Scott, Seth Flaxman, Axel Gandy, Chris Jewell and Will Probert for helpful comments.

**Funding.** This work was funded by a Li Ka Shing Foundation award to CF, and by research grant funding from the UK Department of Health and Social Care (DHSC) to MB, CH and CF.

### **Role of the funders.**

The funders played no role in the design and conduct of the analysis. DHSC runs the app, and manages the secure data environment where the analyses conducted here were performed. DHSC led dissemination of the findings in the UK.

## **Supplementary Information**

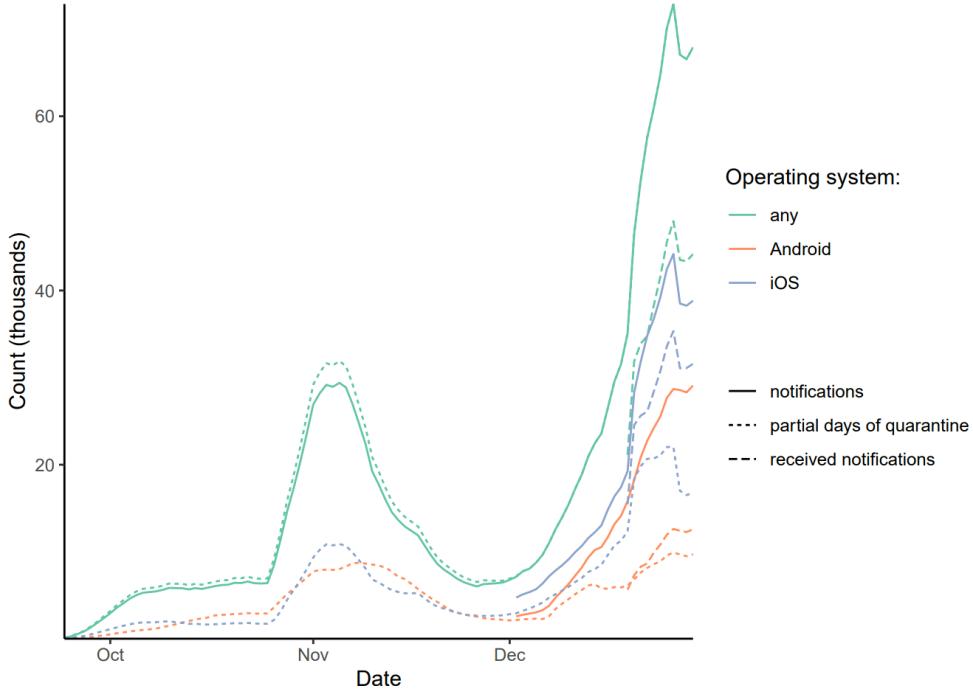
### **Estimate of cases averted from modelling**

#### **Estimate of the number of notifications**

There are three sources for the actual number of notifications by date ( $t$ ), Operating system (OS) and LTLA ( $x$ ):

- The daily number of notifications requested:  $N(t)$  (available by OS, i.e.  $N_{iOS}(t)$  and  $N_{Android}(t)$ , for December only). This is an upper bound on the number of notifications. It could slightly overestimate the number of notifications if app users are exposed to multiple infected app users in a short time, because requests can be submitted when the user is already isolating, if the app records further exposures to other infected individuals.
- The daily number of notifications received, by LTLA  $x$  and OS:  $R_{iOS}(x, t)$  and  $R_{Android}(x, t)$  (available from December 17th only). This is a lower bound on the number of notifications, because not all phones send packets daily. The number of notifications received is less than the number requested, because requests to notify someone who is currently in quarantine due to a previous recent notification are not registered as received.
- The daily number of users isolating for a partial day in LTLA  $x$ :  $P_{iOS}(x, t)$  and  $P_{Android}(x, t)$ . A partial day of isolation usually corresponds to the first day of isolation (i.e. the day of notification). This number is also affected by the missing daily packets.

The relationship in time between the above statistics is shown in Supplementary Figure 1, and the geographical relationship in Supplementary Figure 2.



**Supplementary Figure 1:** time dependence of statistics related to notifications (rolling 7-day average).

To estimate the number of notifications by date and LTLA, we assumed that the biases on  $P_{iOS/Android}(x, t)$  do not depend on LTLA, and its counts are binomial extractions of the actual numbers of notifications  $P \sim Bin(N, c)$  with a time- and OS-dependent (but not LTLA-dependent) correction factor  $c_{iOS/Android}(t)$ . To compute this factor, we took its Maximum Likelihood estimate

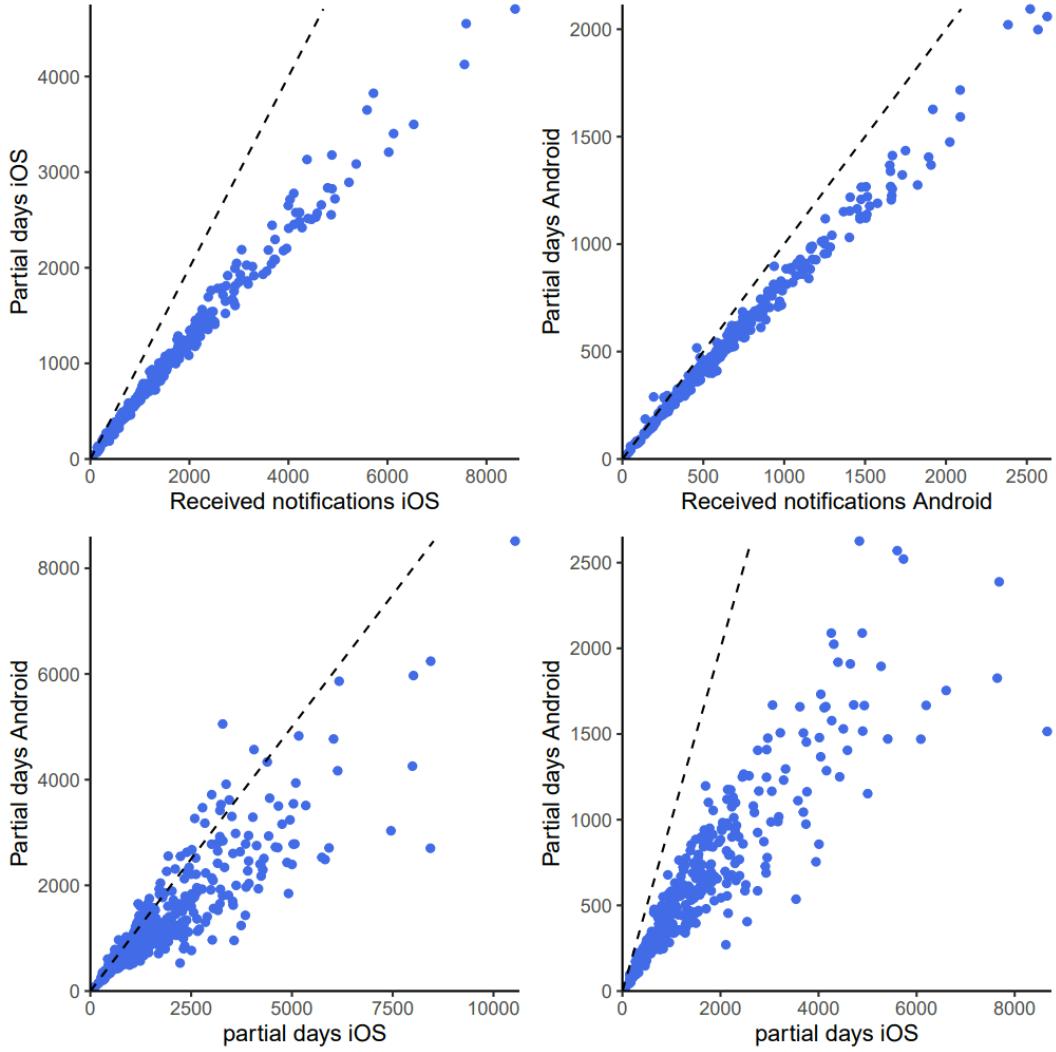
$$c_{iOS}(t) = \frac{\sum_x P_{iOS}(x, t)}{N_{iOS}(t)}, \quad c_{Android}(t) = \frac{\sum_x P_{Android}(x, t)}{N_{Android}(t)}$$

For the dates when  $N_{iOS/Android}(t)$  are not defined, we defined them as

$$N_{iOS}(t) = N(t) \frac{\bar{N}_{iOS}}{\bar{N}_{iOS} + \bar{N}_{Android}}, \quad N_{Android}(t) = N(t) \frac{\bar{N}_{Android}}{\bar{N}_{iOS} + \bar{N}_{Android}}$$

Our final estimate of the number of notifications received on date  $t$  in LTLA  $x$  is therefore

$$\hat{N}(x, t) = \frac{P_{iOS}(x, t)}{c_{iOS}(t)} + \frac{P_{Android}(x, t)}{c_{Android}(t)}$$



**Supplementary Figure 2:** relationship between cumulative values by LTLA of statistics related to notifications. Dashed lines correspond to a 1:1 relation.

### Estimate of Secondary Attack Rate

The SAR was computed by comparing the number of cases in exposed users (i.e. those that become positive after being asked to isolate due to risky contact)  $I_N(t)$  with the number of exposure notifications  $N(t)$ . We used only data from iPhones, excluding Android phones, for more stable daily numbers of analysis packets.

Let the probability that an individual would test positive  $t$  days after being notified, conditional on a notification and positive test occurring, assuming a constant rate of notification in the population, be  $f_{NP}(t)$ . We restrict it to tests received after notifications, i.e.  $f_{NP}(t)=0$  if  $t<0$ . Weighting by a

time-varying rate of notification in the population,  $N(t)$ , the probability of being notified at time  $t_1$  conditional on getting a positive test at time  $t_2$  is

$$p_{NP}(t_1|t_2) = \frac{f_{NP}(t_2-t_1)N(t_1)}{\sum_x f_{NP}(t_2-x)N(x)}$$

The expected number of cases notified at time  $t_1$  and getting a positive test at time  $t_2$  is therefore  $p_{NP}(t_1|t_2)I_N(t_2)$  and the secondary attack rate at time  $t_1$  is

$$SAR(t_1) = \frac{\sum_y p_{NP}(t_1|y)I_N(y)}{N(t_1)} = \sum_y \frac{f_{NP}(y-t_1)I_N(y)}{\sum_x f_{NP}(y-x)N(x)}$$

To estimate the distribution  $f_{NP}(t_P-t_N)$ , we consider the events from the perspective of a user who gets exposed (E), then notified (N) and becomes symptomatic (S), then receives a positive test result after symptoms (P). Given the relation  $t_P-t_N=(t_P-t_S)+(t_S-t_E)-(t_N-t_E)=t_{SP}+t_{ES}-t_{EN}$  between the waiting times, the distribution  $f_{NP}$  is then the convolution

$$f_{NP}(t_{NP}) = \int_0^\infty dt_{EN} f_{EN}(t_{EN}) \int_0^\infty dt_{ES} f_{ES}(t_{ES}) f_{SP}(t_{NP} - t_{ES} + t_{EN})$$

of  $f_{EN}$  (distribution of time from exposure to notification; we assume a gamma distribution with the median values of mean and standard deviation for December, derived below),  $f_{ES}$  (incubation period distribution, given by lognormal with mean 5.42 days and s.d. 2.7 days, McAlloon et al *BMJ Open* 2020) and  $f_{SP}$  (distribution of time from symptoms to positive test result). The latter can be modelled from testing statistics: the delay from booking a test to receiving results in January 2021 had mean=1.5 days with sd=0.5 days

(<https://www.gov.uk/government/collections/nhs-test-and-trace-statistics-england-weekly-reports>), while the delay between symptoms and booking a test is assumed to have mean=1.5 days and sd=1.5 days (<https://www.gov.uk/guidance/coronavirus-covid-19-getting-tested>), so we assume  $f_{SP}$  is gamma-distributed with mean=3 days and variance=2.5 days<sup>2</sup>.

### Estimate of delay from exposure to exposure notification

For the period considered in this paper, the advised duration of quarantine in England and Wales was for  $q(t)=14$  full days after the day of last exposure to a case until December 14th, and  $q(t)=10$  full days after. The number of full days of quarantine advised by the app (i.e. after the partial day on which the notification is received) is then  $d=q-\delta$  where  $\delta$  is the number of days' delay from exposure to exposure notification.

Therefore, under these assumptions, the delay distribution can be obtained directly from the distribution of duration of app-based quarantine. We assumed that the latter distribution is Gaussian

with mean  $\mu(t)$  and standard deviation  $\sigma(t)$ . We modelled the time-dependence of these quantities as natural splines interpolating the values  $\mu_i, \sigma_i$  at 10 equally spaced time points  $t_i$  along phases 1 and 2, plus a term that accounts for the change in quarantine length on December 14th. We also imposed a minimum standard deviation  $\sigma(t) > 1$  to avoid inferring unrealistically low standard deviations.

To infer the parameters  $\mu_i, \sigma_i$  from the data, we made use of the fact that the app packets report if users are isolating due to exposure notification that day, and if it is a partial or full day of isolation. We rescaled both partial and full days by the correction factors presented above, i.e.

$$\hat{P}(x, t) = \frac{P_{iOS}(x, t)}{c_{iOS}(t)} + \frac{P_{Android}(x, t)}{c_{Android}(t)}, \quad \hat{F}(x, t) = \frac{F_{iOS}(x, t)}{c_{iOS}(t)} + \frac{F_{Android}(x, t)}{c_{Android}(t)}$$

In the ideal scenario, these quantities are expected to satisfy the equation

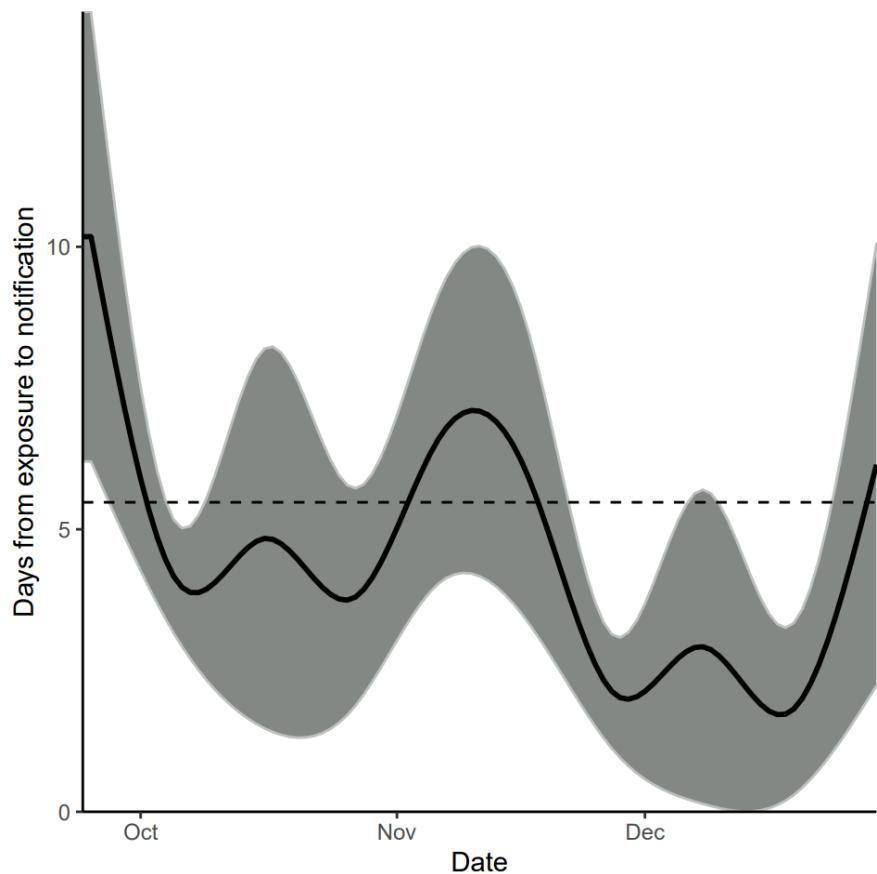
$$F(x, t) = \sum_{\tau=1}^{q_{max}} P(x, t - \tau) \sum_{l=\tau}^{q(t-\tau)} p(l|t - \tau)$$

where  $p(l|t)$  is the distribution of quarantine lengths  $l$  for individuals who get notified on day  $t$ , which is modelled as a discretisation of the Gaussian distribution discussed above. In words, this equation states that the number of individuals in full quarantine today equals the number who started quarantine  $\tau$  days ago, multiplied by the probability they are still in quarantine now (i.e. that their assigned number of full quarantine days was at least  $\tau$ ), summed over  $\tau$  (to a point far enough in the past that no individuals before that would still be in quarantine now, i.e. to a time  $t - q_{max}$ , where  $q_{max}$  is the longest relevant  $q$  which is here 14 days).

In practice, given the noise in  $\hat{P}(x, t), \hat{F}(x, t)$  we solved for this distribution by Least Squares, assuming the delays to be independent of LTAs. This corresponds to the minimisation of

$$\sum_t \left[ \sum_x \left( \hat{F}(x, t) - \sum_{\tau=1}^{14} \hat{P}(x, t - \tau) \sum_{l=\tau}^{q(t-\tau)} p(l|t - \tau) \right) \right]^2 \text{ with respect to the parameters } \mu_i, \sigma_i.$$

The resulting mean and sd for delays from exposure to notification are shown in Supplementary Figure 3. The time from exposure to notification seems to decrease during the second half of November, before confirmed cases began increasing rapidly at the end of November. This is compatible with a shortening of certain delay distributions (those from an infection to a subsequent event, such as notification, symptoms, hospitalisation etc.) during a growing epidemic, when an increased number of observations are of people infected more recently.



**Supplementary Figure 3:** inferred mean and standard deviation of the delay from exposure to notification. The dashed line illustrates the average generation time, i.e. the average time from exposure to transmission.

### Modelling the effectiveness of quarantine

To understand the effect of quarantine, we rely on two surveys for the UK. The first one (Smith et al. 2020) found that 11% of individuals in quarantine actually adhered properly to quarantine rules, but another 54% of individuals intended to adhere to quarantine. We assume 100% effectiveness of quarantine for individuals who quarantined properly, and a partial effectiveness of quarantine  $Q$  for the ones who declared they intended to adhere. The second more recent and optimistic survey (Fancourt et al. 2020) found a high adherence to quarantine, with 80% of individuals declaring that they were adhering to quarantine for the full duration advised by the public health authorities, and another 8% adhering only for a fraction of that duration. We assume 100% effectiveness for individuals who comply for the whole period, and  $Q$  for those who quarantine partially. This scenario might be more representative of app users, given their initial compliance to public health advice demonstrated by installing and using the app.

As our central scenario, we considered an intermediate scenario corresponding to the average adherence of the two studies above, with 45.5% of individuals fully adhering to quarantine, and 31% adhering with effectiveness Q. We assumed a central value of Q=50% adherence to imperfect quarantine, running sensitivity analyses with Q=2.5%, 25%, 75%, 97.5%.

### **Modelling the size of the transmission chain from a single case**

Having estimated the expected number of cases *directly* averted by a notification on day  $t$  in LTLA  $x$ , we inferred the *total* number of cases averted by taking the weekly moving average of cases  $C(x, t)$  in the LTLA  $x$  and assuming that the impact of an additional case in  $x$  on the size of the local epidemic at some later day  $T$ ,  $C(x, T)$ , would be the same as the impact of one of the cases on day  $t$ , i.e.  $C(x, T)/C(x, t)t_g$  cases where  $t_g$  is the generation time. This is equivalent to the assumptions that (a) the same NPIs would have been in place at the same times even without the app; (b) the additional number of infections would have been small enough that the saturation of the epidemic would not have changed; (c) transmission of the virus between LTLAs can be neglected.

Assumption (c) implies that we can describe the epidemic growth in terms of the local effective reproduction number  $R_t(x)$  (or equivalently, the growth rate  $r_t(x)$ ), while (a) and (b) ensure that the underlying  $r_t(x)$  would not change with the additional cases. The size of the whole chain caused by

a single transmission at time  $t$  could therefore be quantified by  $\frac{1}{t_g} \exp\left[\int_t^T dt' r_{t'}(x)\right]$ , and since

$C(x, T) = C(x, t)\exp\left[\int_t^T dt' r_{t'}(x)\right]$ , we obtain a number of cases at time  $T$  equal to a factor

$C(x, T)/C(x, t)t_g$  for each transmission at time  $t$ . This approach can also be interpreted in line with Ludwig's argument (Pellis, Ferguson, and Fraser 2008).

### **Final estimate for the total number of cases averted**

Given the assumptions in the last section, the expected decrease in the number of cases at any time  $T$  due to the effect of notifications received at time  $t < T$  can be estimated as the number of notifications  $\hat{N}(x, t)$ , multiplied by the number of cases averted  $C(x, T)/C(x, t)t_g$  as a consequence of a single case averted at time  $t$ , multiplied by the expected reduction in transmissions due to a single notification.

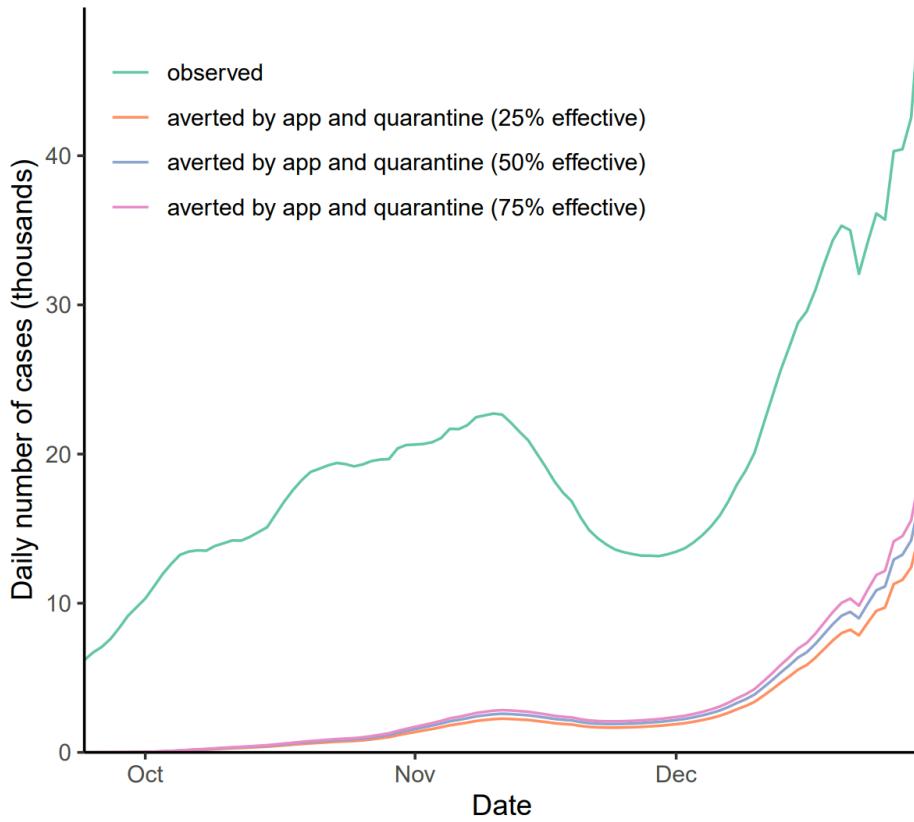
Since the virus can be transmitted only by infected individuals, the latter is the product of the secondary attack rate among contacts - or, more precisely, the probability that a notified contact is infected - and the relative reduction in transmissions from an infected individual being notified by the

app. In turn, the fraction of transmissions averted from an infected individual is the product of the fraction of transmissions potentially occurring after a notification, and the effectiveness of quarantine itself in preventing transmission.

The fraction of transmissions potentially occurring after notification can be obtained combining the cumulative generation time distribution  $W(\tau)$ , which correspond to the (cumulative) distribution of the timing of transmission  $\tau$  with respect to the time of exposure, and the distribution  $p(\delta|t)$  of the delay  $\delta$  from exposure to notification. The expected fraction of transmissions after notification is given by the delay-weighted average of the complementary cumulative generation time distribution,

i.e. 
$$\sum_{\tau=0}^{14} (1 - W(\tau))p(\tau|t)$$
. Final results for cases averted in time are shown in Supplementary Figure 4.

4.

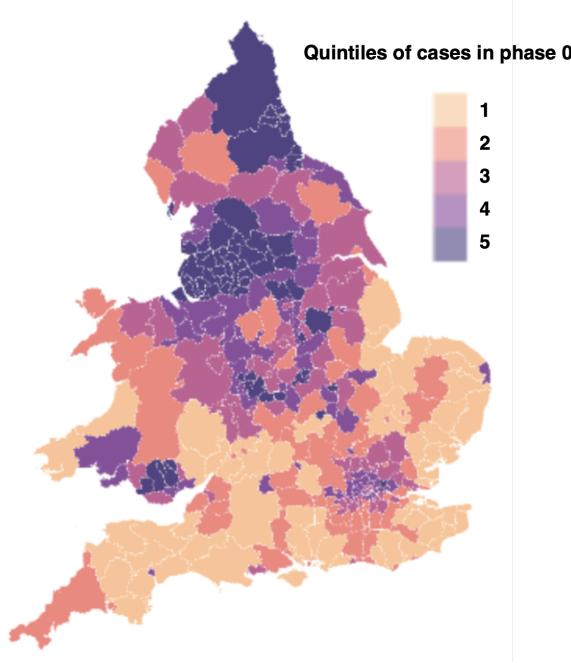


**Supplementary Figure 4:** rolling 7-day average of the number of cases observed and predicted number of cases averted thanks to the notifications sent by the app, for different values of adherence to imperfect quarantine (25%, 50%, 75%) for the 31% of notified individuals who we assumed adhere imperfectly, in addition to 45.5% of notified individuals assumed to adhere perfectly.

## Full results of the statistical analysis

### Main regression results within matched neighbours

The main statistical analysis compares each LTLA statistics to those of its neighbours that are in the same quintile of the number of cases in phase 0. Stratification into quintiles (as opposed to deciles etc.) was chosen to balance power and sufficient adjustment. (No other possibility was tried, to guard against investigator bias.)

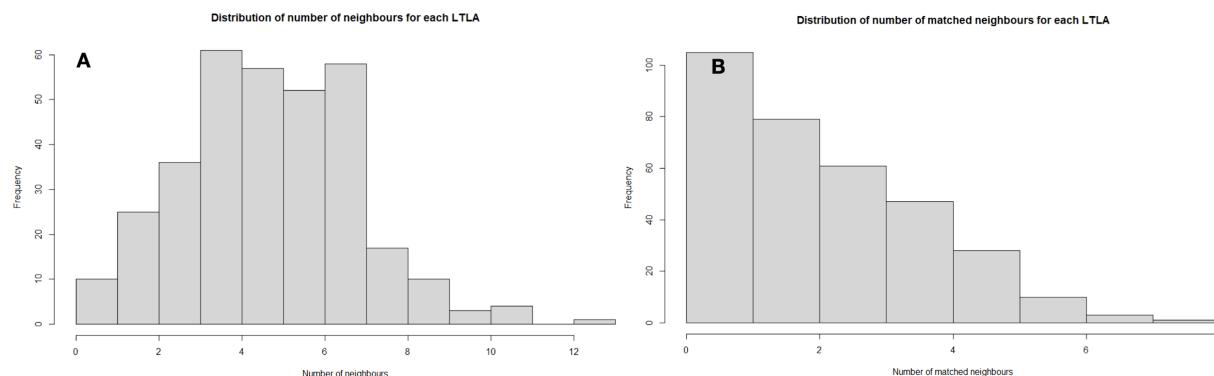


**Supplementary Figure 6.** Map showing local authorities coloured according to the number of cases during phase 0. The statistical analysis focuses on differences between neighbouring local areas that are in the same quintile.

For the statistical analysis, each area is compared to its matching neighbours. So consider an area labelled  $x$  with matching neighbours  $y_1, y_2, y_3 \dots$ . Then each statistic  $f$  is in turn averaged for the neighbours, weighting by population size, to obtain a value  $f_y(x)$ , the mean value in the

neighbours of  $x$ . This is compared to the statistic for  $x$ , denoted  $f_x$ . Linear regression is carried out on differences  $f_x - f_y(x)$ . Statistics included are number of confirmed cases in each phase, the number of active app users, a measure of rural/urban mix on a scale of 1 to 5, a measure of local GDP per capita adjusted for rural/urban score, and a measure of % of population living in poverty (before housing costs). In a few cases, this approach produces redundancy, when  $y$  and  $x$  are unique neighbours of each other, or when three local authorities are neighbours of each other. To guard against this, we repeated the analysis via bootstrapping, including only non-redundant pairs and assessing confidence intervals from 10000 bootstraps.

The number of neighbours and matched neighbours is shown in Supplementary Figure 2.



**Supplementary Figure 7.** The number of neighbours (A) and matched neighbours (B) for each local authority. Over 100 local authorities were not included in the analysis as they had no matched neighbours.

The main regression is

$\log(\text{difference in cases in phase } X) =$

$$\begin{aligned} & \text{beta\_rural\_urban} * (\text{difference in local rural/urban score}) + \\ & \text{beta\_gdp\_band} * (\text{difference in local GDP band}) + \\ & \text{beta\_poverty} * (\text{difference in percent of population living in poverty}) + \\ & \text{beta\_users} * (\text{difference in percent of population using the app}) + \\ & \text{epsilon\_residual} \end{aligned}$$

Results shown for each phase.



### Phase 1 and 2 combined

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.070	0.052 - 0.088	1e-12
beta_gdp_band	-0.00096	-0.0078 - 0.0059	0.79
beta_poverty	0.002	-0.0076 - 0.012	0.68
beta_users*	-0.023	-0.030 - -0.015	1e-8

\*when negative, beta\_users is the decrease in log(cases) per 1% increase in app users.

### Phase 1

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.078	0.052 - 0.10	6e-9
beta_gdp_band	0.0060	-0.0035 - 0.016	0.21
beta_poverty	0.021	0.0076 - 0.034	0.002
beta_users*	-0.011	-0.022 - -0.003	0.04

\*when negative, beta\_users is the decrease in log(cases) per 1% increase in app users.

### Phase 2

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.063	0.041 - 0.085	5e-8
beta_gdp_band	-0.0002	-0.0083 - 0.0078	0.95
beta_poverty	-0.0098	-0.021 - 0.0018	0.09
beta_users*	-0.027	-0.036 - -0.018	4e-8

\*when negative, beta\_users is the decrease in log(cases) per 1% increase in app users.

## **Naive regression**

We do not interpret a naive linear regression as indicative of app effectiveness, due to confounding (see next section), but it does provide some indication of the signal present in the data. The slope, beta\_users, of the naive regression log(cases per capita in phases 1 and 2 combined) = beta\_users \* (fraction of population using the app) is -0.042, with 95% confidence interval (-0.049 - -0.036), p-value < 2e-16, with adjusted R2 of 32%. For phase 1 it is -0.049, with 95% confidence interval (-0.061 - -0.038), p-value = 3e-16, with adjusted R2 of 18%, and for phase 2 it is -0.036, with 95% confidence interval (-0.044 - -0.029), p-value < 2e-16, with adjusted R2 of 20%.

## **Predictors of app uptake & confounding**

The fraction of the population using the app, app uptake, is correlated with socio-demographic factors that are structural drivers of the epidemic. This causes confounding, which we sought to correct for in our analysis. Rural/urban score, local GDP, and proportion of the population in poverty were structural drivers of the epidemic, and also predict the app uptake. We also hypothesised that additional factors not captured by these indicators would be reflected in the number of COVID-19 cases per capita in phase 0, the time period before the introduction of the app, and that this last factor could serve as an additional measure of the confounding propensity of each area to experience a larger number of cases, and to have lower uptake of the app. We tested this with a linear model that predicts the local uptake of the app.

Fraction of population using the app =

```
gamma_rural_urban * (local rural/urban score) +  
gamma_gdp_band * (local GDP band) +  
gamma_poverty * (% of population living in poverty) +  
gamma_phase0_cases * log(total cases per capita during phase 0) +  
epsilon_residual
```

### Predictor of local app uptake

Coefficient	Estimate	95% confidence interval	P value
gamma_rural_urban	-0.013	-0.016 - -0.010	4e-15
gamma_gdp_band	0.0020	0.0005 - 0.0036	0.010
gamma_poverty	-0.0039	-0.005 - -0.0024	2e-6
gamma_phase0_cases	-0.023	-0.029 - -0.016	5e-10

That predictors of low app uptake are also predictors of increased cases, confounding, is reflected in the correlation between our measure of local app uptake, (fraction of population using the app), which is measured as a mean value November through December, and the number of cases per capita in phase 0. The regression slope of this naive regression is -0.083, with 95% confidence interval (-0.095 - -0.071), p-value < 2e-16, with adjusted R2 of 36%.

### Placebo regression

To address this confounding, we restrict our analysis to neighbouring LTLAs, and match on the quintile of number of cases per capita during phase 0, thus resulting in a stratified analysis of local differences. A measure of how successful this approach is the so-called placebo analysis, reproducing our main analysis, but predicting phase 0 cases. The placebo regression is  
 $\log(\text{difference in cases in phase 0}) =$

$$\begin{aligned} & \text{beta\_rural\_urban} * (\text{difference in local rural/urban score}) + \\ & \text{beta\_gdp\_band} * (\text{difference in local GDP band}) + \\ & \text{beta\_poverty} * (\text{difference in percent of population living in poverty}) + \\ & \text{beta\_users} * (\text{difference in percent of population using the app}) + \\ & \text{epsilon\_residual} \end{aligned}$$

## Phase 0 placebo regression

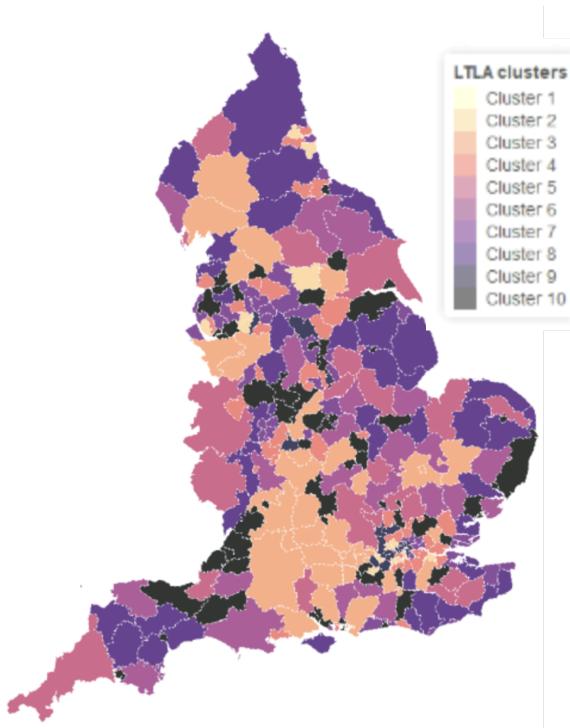
Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.054	0.030 - 0.079	2e-5
beta_gdp_band	0.0097	0.0005 - 0.019	0.038
beta_poverty	0.039	0.026 - 0.052	1e-8
beta_users*	-0.0062	-0.016 - 0.0041	0.24

\*when negative, beta\_users is the decrease in log(cases) per 1% increase in app users.

We find that this approach removes the strong correlation observed in the naive placebo regression, changing the unadjusted regression coefficient from -0.083 ( $p < 2e-16$ ) to -0.0062 ( $p = 0.24$ ).

## Stratified linear regression in clusters

As a robustness check, we took an alternative approach, grouping local authorities into ten clusters. We used a k-means clustering with a range of indicators: proportion aged under 30, proportion aged over 70, ethnicity, income and urban-rural classification.



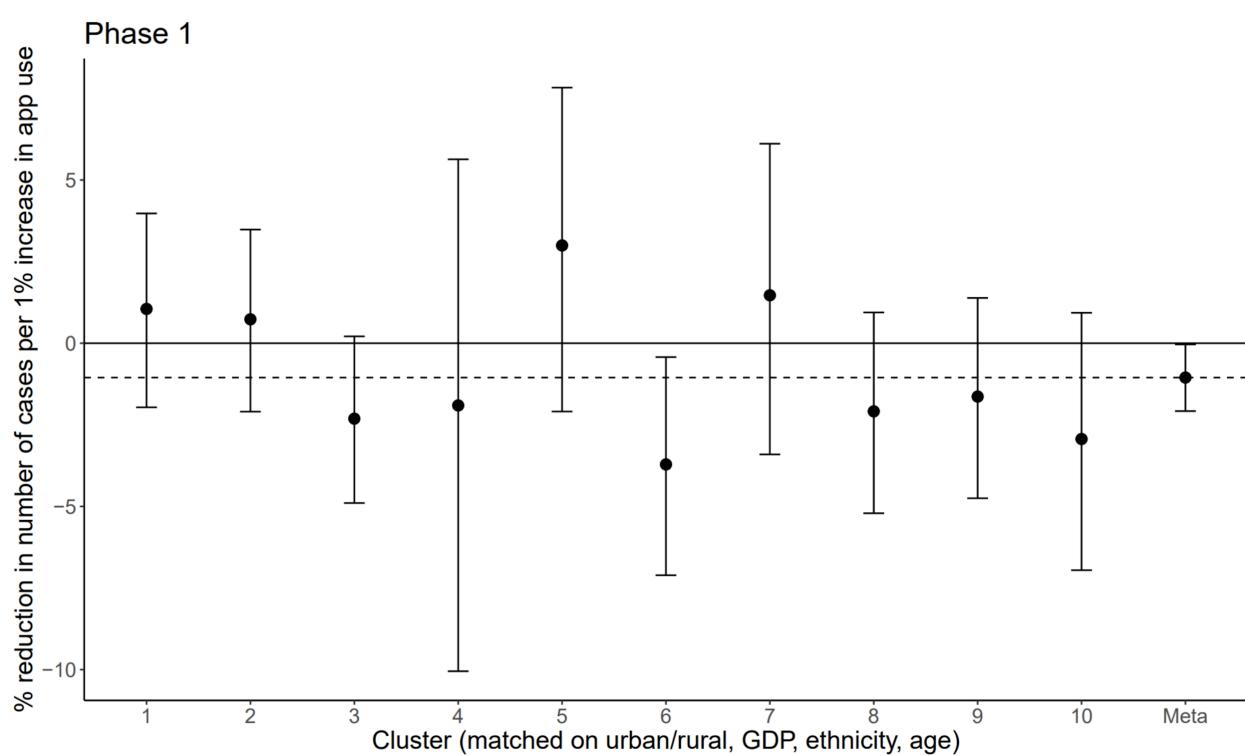
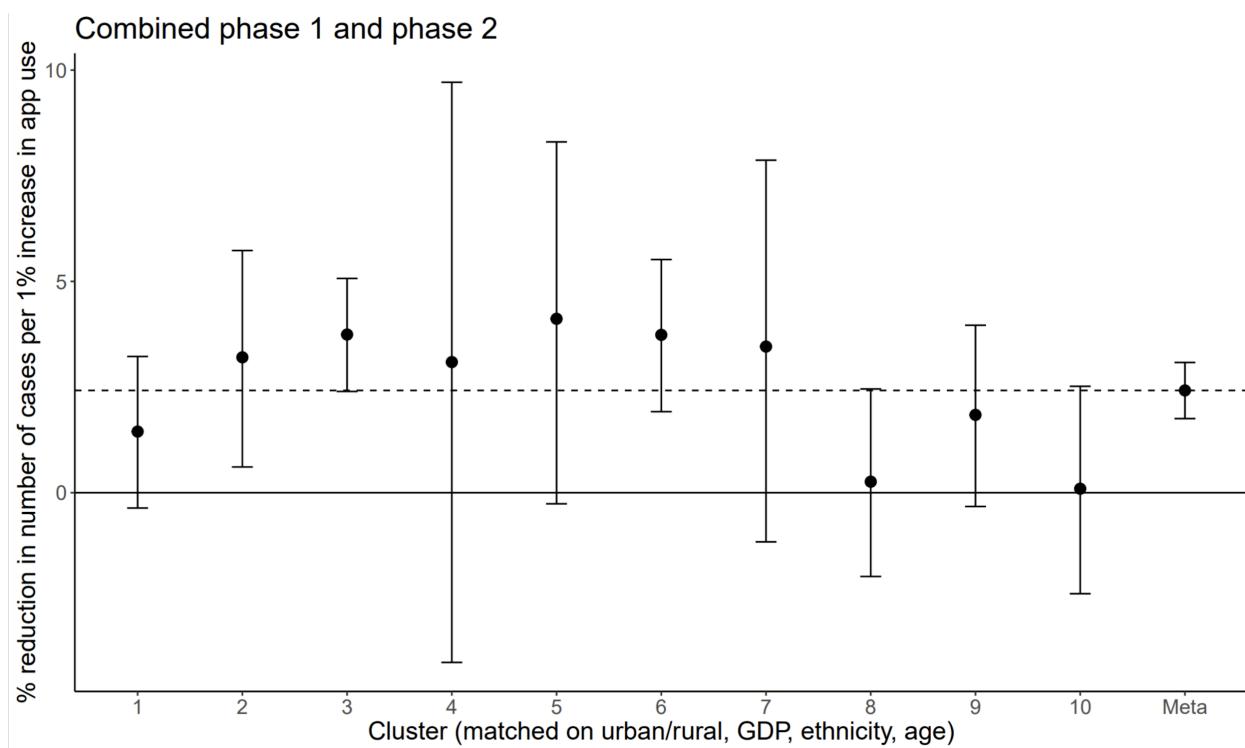
**Supplementary Figure 8.** Map showing local authorities coloured according to the ten clusters of similar demography. The statistical analysis, performed as a robustness check, is a stratified linear regression, adjusted for confirmed case numbers in phase 0.

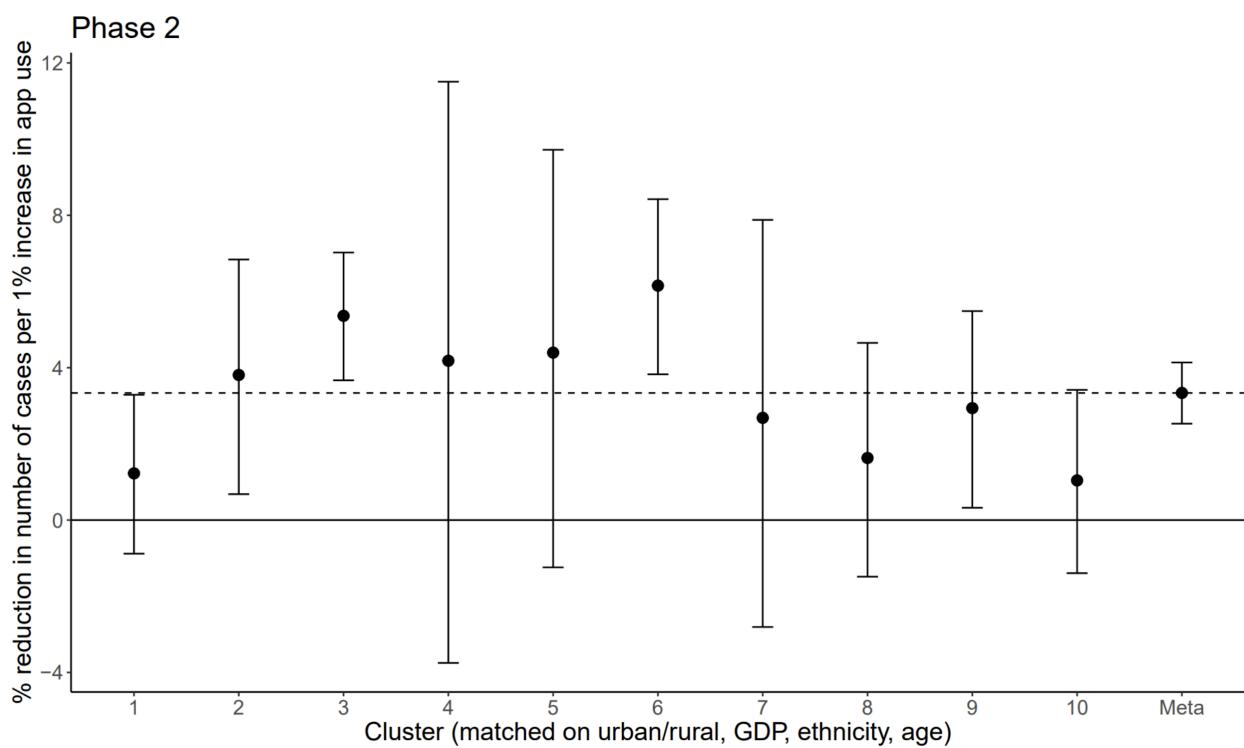
Due to the availability of statistics, we performed the analysis for England only. For each cluster, we performed a linear regression

$\log(\text{cases per capita in phase X}) =$

$$\begin{aligned} & \text{beta\_phase\_0} * \log(\text{cases per capita in phase 0}) + \\ & \text{beta\_uptake} * (\text{percentage of population that use the app}) + \\ & \text{epsilon\_residual} \end{aligned}$$

The analysis is performed separately for each cluster, and then aggregated, weighting by variance of each sub-analysis.





**Supplementary Figure 9.** Regression coefficients for app effect for each cluster, and aggregate variance-weighted estimate (labelled ‘meta’). Panels are labelled for Phase 1 and 2 combined, and Phase 1 and Phase 2 separately. Aggregate estimates are reported in Table 2.

### Matched pairs

As a check for robustness, we use a different approach to account for the number of cases in Phase 0 and for other confounders. Given the likely non-linear impact of the number of cases in Phase 0 and of geographical aspects, we stratify by both features. Namely, we consider all pairs of neighbouring LTLAs in England, restricting the selection to pairs that have almost the same number of cases per capita in Phase 0 (i.e. do not differ by more than 2.5 percentiles in Phase 0 cases). We then treat each pair as a distinct comparison, consider the difference between the logarithm of the number of cases per capita or other statistics between the two LTLAs in each pair, and we run the regressions for these differences in statistics across all pairs. With this choice, the placebo regression comes out non-significant for all predictors (rural/urban score, GDP band, poverty before housing costs, fraction of app users), showing that the stratification is effective in removing multiple confounders at once.

### Phase 0 (placebo regression)

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	-0.0033	-0.014 - 0.07	0.51
beta_gdp_band	-0.002	-0.0059 - 0.002	0.32
beta_poverty	0.0011	-0.039 - 0.006	0.65
beta_users*	0.0002	-0.0045 - 0.0049	0.93

\*when negative, beta\_users is the decrease in log(cases) per 1% increase in app users.

On the other hand, regressions for Phase 1 and Phase 2 show that this approach removes the effect of the other confounders, which are never significant, while emphasising a clear effect of app usage on the number of cases in Phase 1 and 2.

### Phase 1 and 2 combined

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.016	-0.043 - 0.075	0.59
beta_gdp_band	0.002	-0.021 - 0.025	0.86
beta_poverty	-0.0047	-0.033 - 0.024	0.74
beta_users*	-0.044	-0.071 - -0.017	0.0023

\*when negative, beta\_users is the decrease in log(cases) per 1% increase in app users.

## Phase 1

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.014	-0.059 - 0.087	0.7
beta_gdp_band	0.002	-0.026 - 0.03	0.89
beta_poverty	-0.013	-0.049 - 0.022	0.46
beta_users*	-0.051	-0.084 - -0.018	0.004

\*when negative, beta\_users is the decrease in log(cases) per 1% increase in app users.

## Phase 2

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.016	-0.046 - 0.078	0.61
beta_gdp_band	0.004	-0.02 - 0.028	0.73
beta_poverty	-0.005	-0.036 - 0.025	0.72
beta_users*	-0.039	-0.067 - -0.01	0.009

\*when negative, beta\_users is the decrease in log(cases) per 1% increase in app users.

## Regression adjusted for quality of manual contact tracing

The same regression on matched pairs was run including the fraction of contacts reached by the National NHS Test & Trace program in the regression. For this purpose, we considered only pairs belonging to two different English UTLAs, since the fraction of contacts reached for each case is reported by UTLA. The aim of this regression is to correct for the potential impact of the manual contact tracing programme as well.

The regression for cases in Phase 0 is not a placebo regression, because manual contact tracing was active in that phase, and indeed we observe a weakly significant effect of manual contact tracing in Phase 0 (-0.3% cases for each 1% increase in contacts reached, p<0.1).

## Phase 0

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	-0.0096	-0.025 - 0.056	0.20
beta_gdp_band	-0.0024	-0.0068 - 0.002	0.28
beta_poverty	0.0032	-0.0016 - 0.008	0.18
beta_manual_tracing*	-0.003	-0.0067 - 0.00063	0.10
beta_users*	0.0005	-0.0043 - 0.0052	0.84

\*when negative, beta\_users and beta\_manual\_tracing are the decrease in log(cases) per 1% increase in app users and per 1% increase in contacts reached, respectively.

The regressions for Phase 1 and 2 confirm an effect of app usage per 1% users, and a significant comparable effect of manual tracing per 1% contacts reached in Phase 1. Both effects are non-significant in Phase 2.

## Phase 1 and 2 combined

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.023	-0.084 - 0.13	0.66
beta_gdp_band	0.0037	-0.027 - 0.035	0.8
beta_poverty	0.0024	-0.031 - 0.036	0.88
beta_manual_tracing*	-0.0057	-0.031 - 0.02	0.65
beta_users*	-0.035	-0.068 - -0.0016	0.041

\*when negative, beta\_users and beta\_manual\_tracing are the decrease in log(cases) per 1% increase in app users and per 1% increase in contacts reached, respectively.

## Phase 1

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.036	-0.088 - 0.16	0.55
beta_gdp_band	0.0037	-0.032 - 0.04	0.83
beta_poverty	0.0018	-0.037 - 0.04	0.92
beta_manual_tracing*	-0.037	-0.067 - -0.0072	0.018
beta_users*	-0.047	-0.085 - -0.0079	0.021

\*when negative, beta\_users and beta\_manual\_tracing are the decrease in log(cases) per 1% increase in app users and per 1% increase in contacts reached, respectively.

## Phase 2

Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.018	-0.09 - 0.13	0.73
beta_gdp_band	0.0057	-0.026 - 0.037	0.71
beta_poverty	0.0017	-0.036 - 0.032	0.92
beta_manual_tracing*	0.008	-0.018 - 0.034	0.53
beta_users*	-0.028	-0.062 - 0.0064	0.11

\*when negative, beta\_users and beta\_manual\_tracing are the decrease in log(cases) per 1% increase in app users and per 1% increase in contacts reached, respectively.

## Estimation of the case fatality rate

The case fatality rate is estimated as 1.47% as the ratio of total deaths (27,922) to cases (1,891,777) for Phases 1 and 2 combined. To test for heterogeneity, it was also estimated as the regression of local deaths to cases, but no substantial heterogeneity was observed. It is a lower estimate due to right censoring of the time series of deaths.