The epidemiological impact of the NHS COVID-19 App

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

*equal contribution
†correspondence to christophe.fraser@bdi.ox.ac.uk
[1]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department for Medicine, University of Oxford, Oxford, UK
[2]Zühlke Engineering Ltd., London, UK
[3]The Alan Turing Institute, London, UK
[4]Department of Statistics, University of Warwick, Coventry, UK
[5]Department of Economics, University of Oxford, Oxford, UK
[6]Department of Statistics, University of Oxford, Oxford, UK

Last updated 9 Feb 2021 - For public release

Supplementary Information

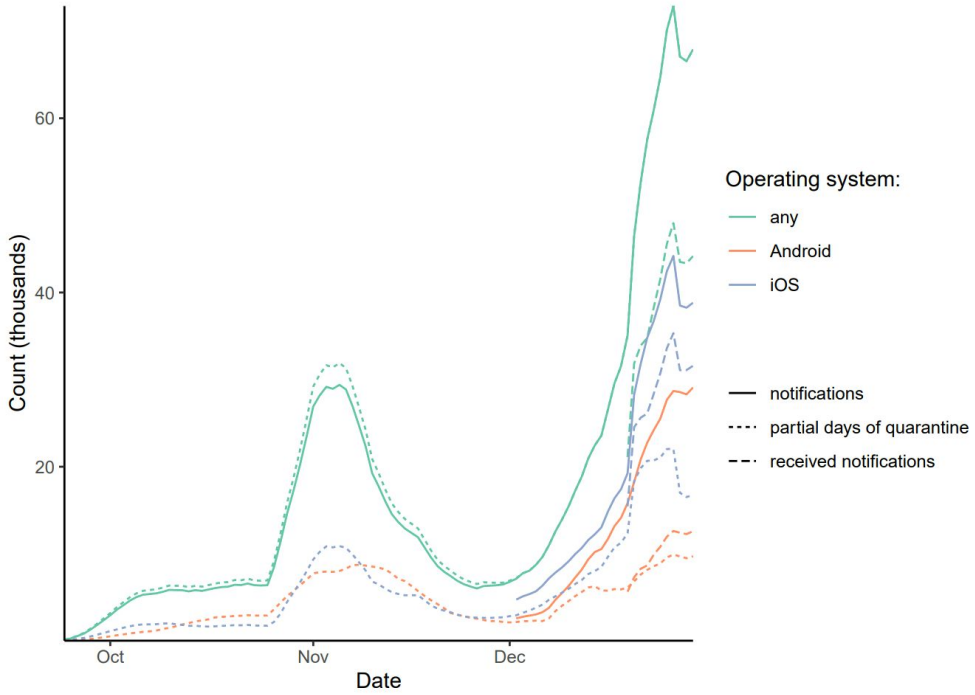
Estimate of cases averted from modelling

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

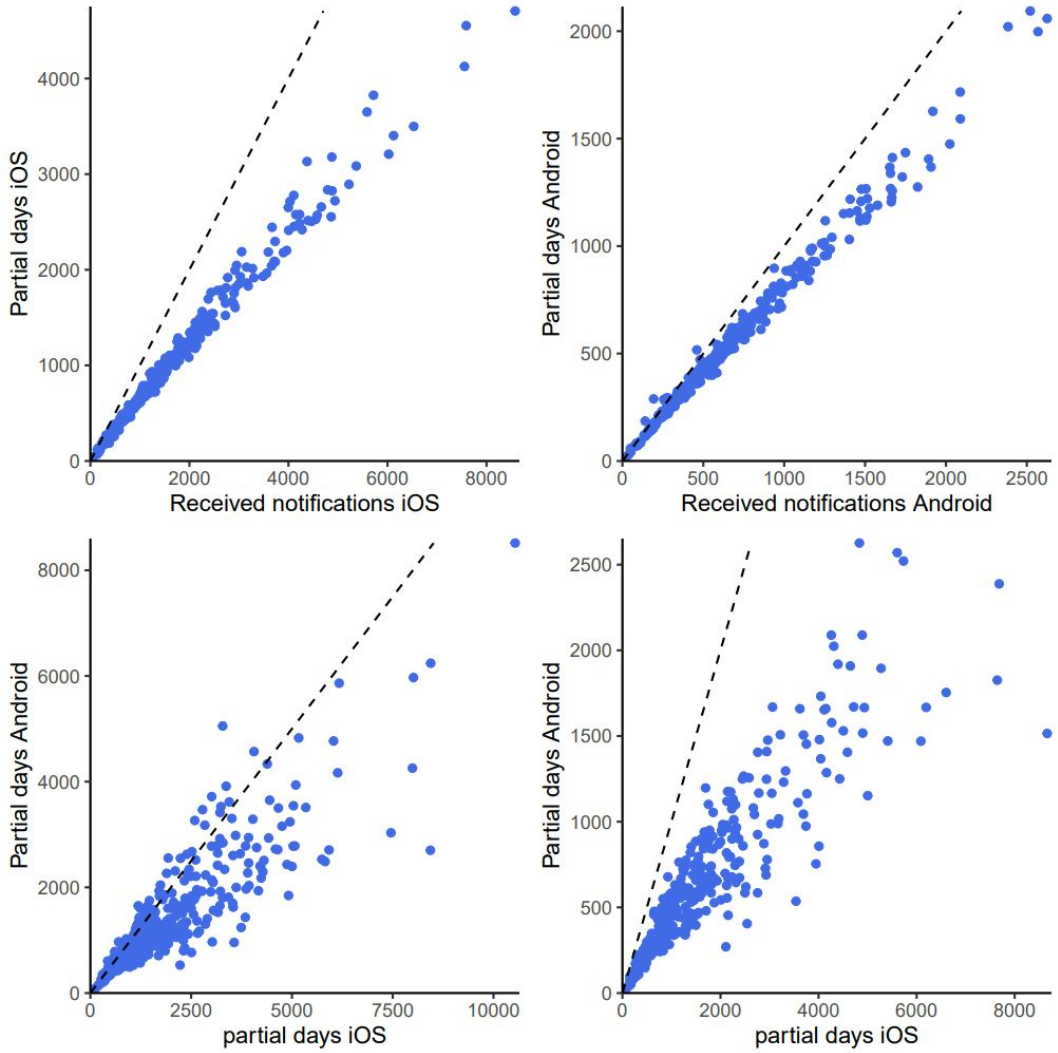
$$c_{iOS}(t) = \frac{\sum_x P_{iOS}(x,t)}{N_{iOS}(t)} \quad , \quad c_{Android}(t) = \frac{\sum_x P_{Android}(x,t)}{N_{Android}(t)}$$

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

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

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

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



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, McAloon 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^2.

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 δ 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 μ_i, σ_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 μ_i, σ_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.

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

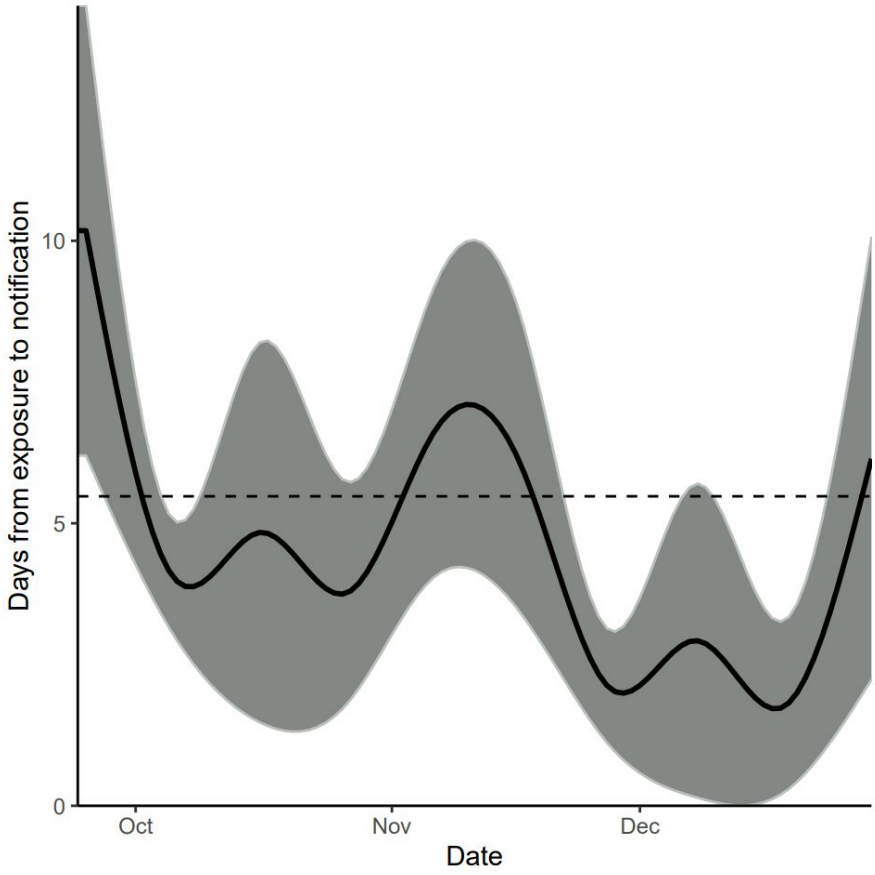
In the ideal scenario, these quantities are expected to satisfy the equation

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

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 τ 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 τ), summed over τ (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 $\widehat{P}(x,t), \widehat{F}(x,t)$ we solved for this distribution by Least Squares, assuming the delays to be independent of LTLAs. This corresponds to the minimisation of $\sum_t \left[\sum_x \left(\widehat{F}(x,t) - \sum_{\tau=1}^{14} \widehat{P}(x, t-\tau) \sum_{l=\tau}^{q(t-\tau)} p(l|t-\tau) \right) \right]^2$ with respect to the parameters μ_i, σ_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).

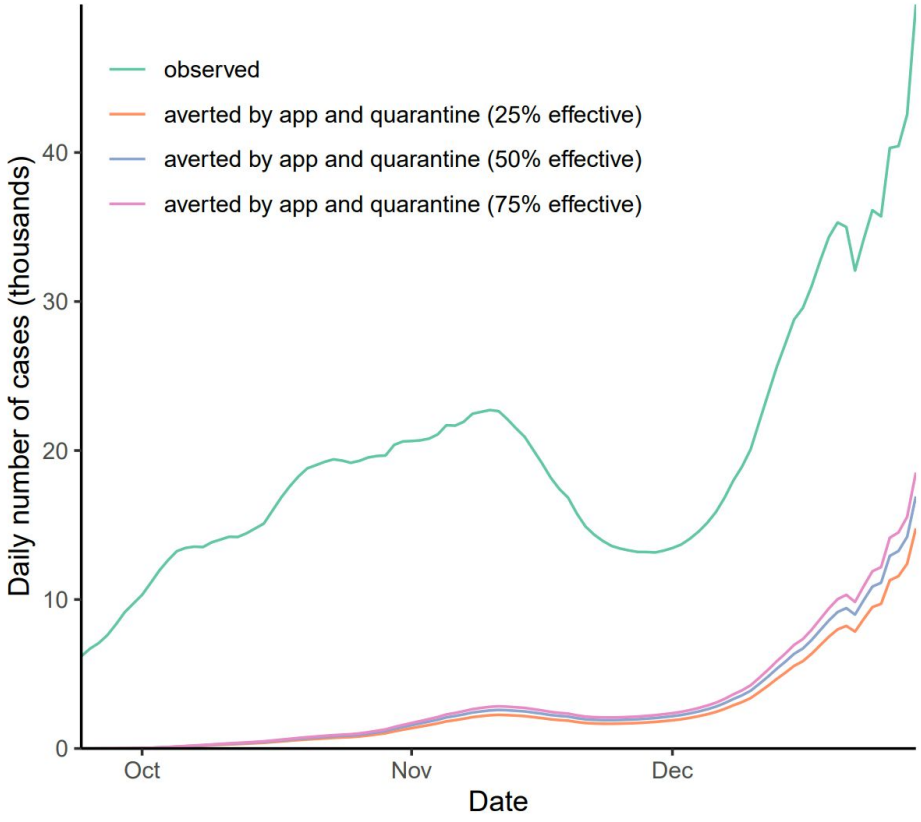
Final estimate for the total number of cases averted

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 $\widehat{N}(x,t)$, 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 τ with respect to the time of exposure, and the distribution $p(\delta|t)$ of the delay δ 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.

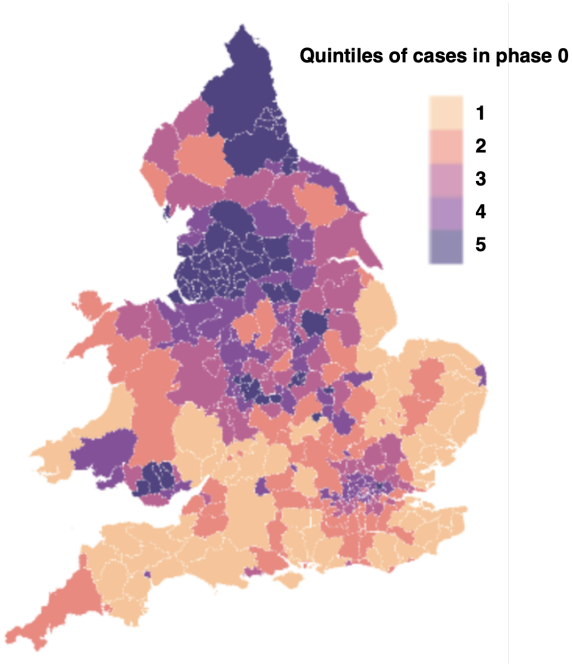


Supplementary Figure 4: *rolling 7-day average of the number of cases observed and predicted number of cases averted thanks to the notifications sent by the app, for different values of adherence to imperfect quarantine (25%,50%,75%) for the 31% of notified individuals who we assumed adhere imperfectly,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

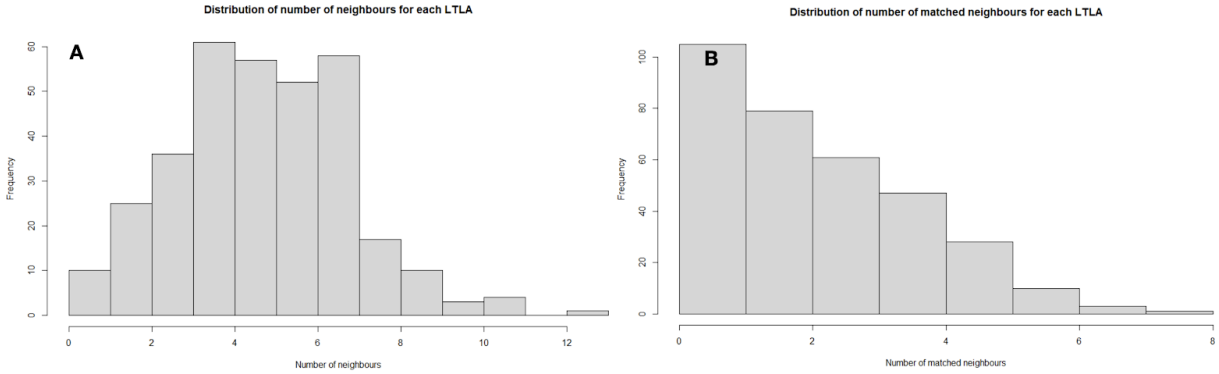


Supplementary Figure 6. *Map showing local authorities coloured according to the number of cases during phase 0. The statistical analysis focuses on differences between neighbouring local areas that are in the same quintile.*

For the statistical analysis, each area is compared to its matching neighbours. So consider an area labelled x with matching neighbours y1, y2, y3 … 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 (difference in cases in phase X) =
beta_rural_urban * (difference in local rural/urban score) +
beta_gdp_band * (difference in local GDP band) +
beta_poverty * (difference in percent of population living in poverty) +
beta_users * (difference in percent of population using the app) +
epsilon_residual

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 (difference in cases in phase 0) =

beta_rural_urban * (difference in local rural/urban score) +
beta_gdp_band * (difference in local GDP band) +
beta_poverty * (difference in percent of population living in poverty) +
beta_users * (difference in percent of population using the app) +
epsilon_residual

Phase 0 placebo regression

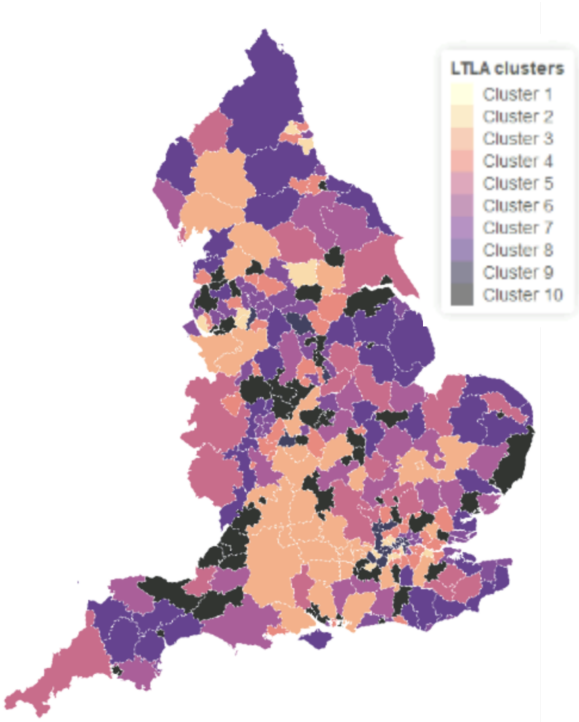
Coefficient	Estimate	95% confidence interval	P value
beta_rural_urban	0.054	0.030 - 0.079	2e-5
beta_gdp_band	0.0097	0.0005 - 0.019	0.038
beta_poverty	0.039	0.026 - 0.052	1e-8
beta_users*	-0.0062	-0.016 - 0.0041	0.24

*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

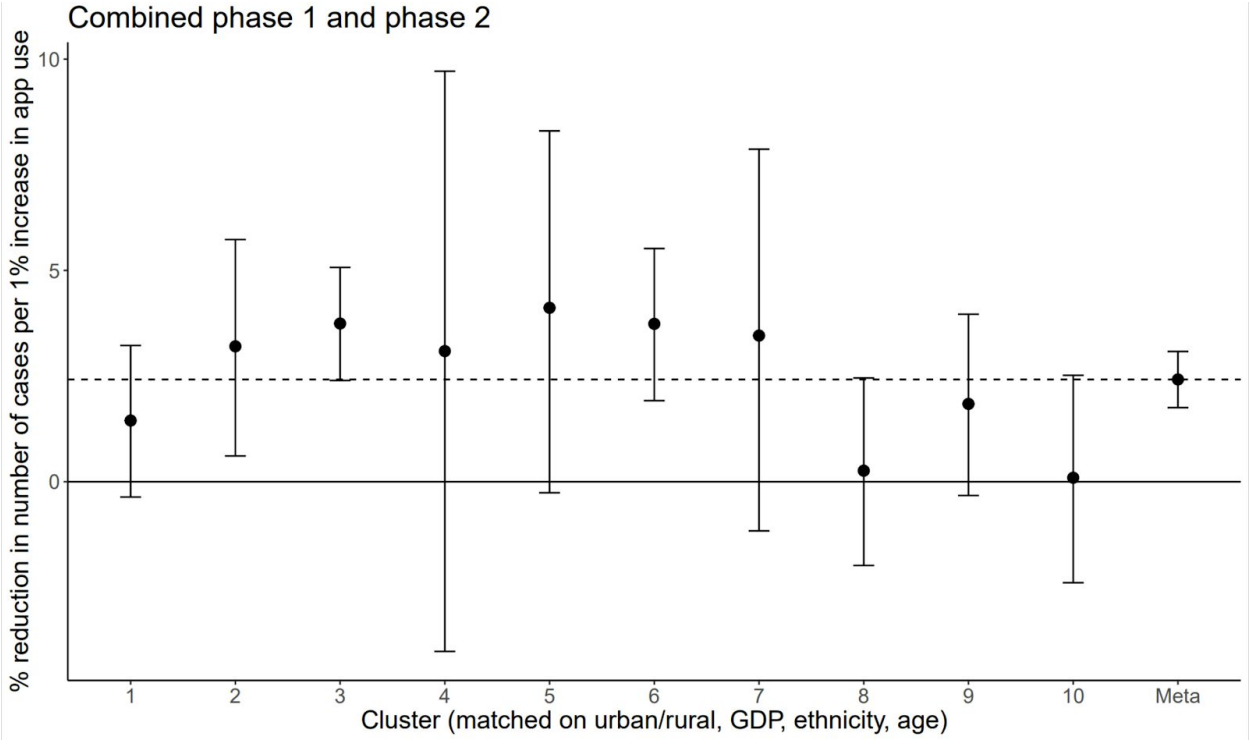
log(cases per capita in phase X) =

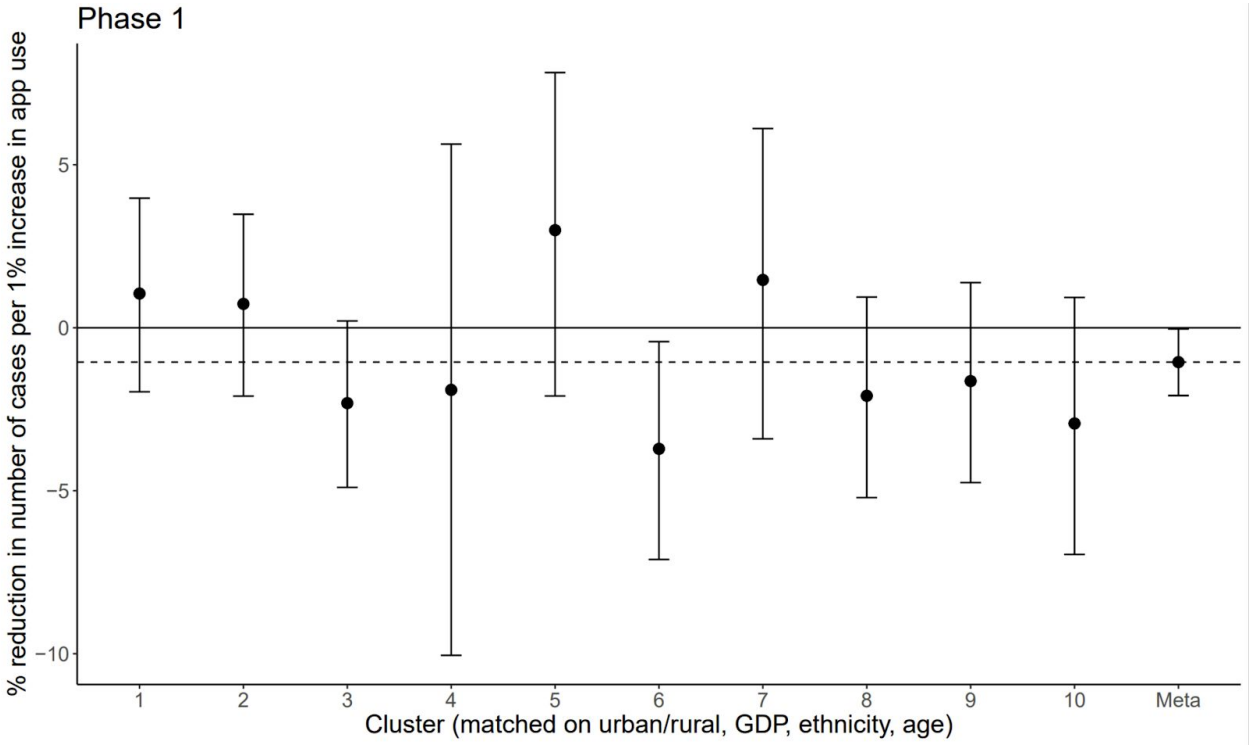
beta_phase_0 * log(cases per capita in phase 0) +

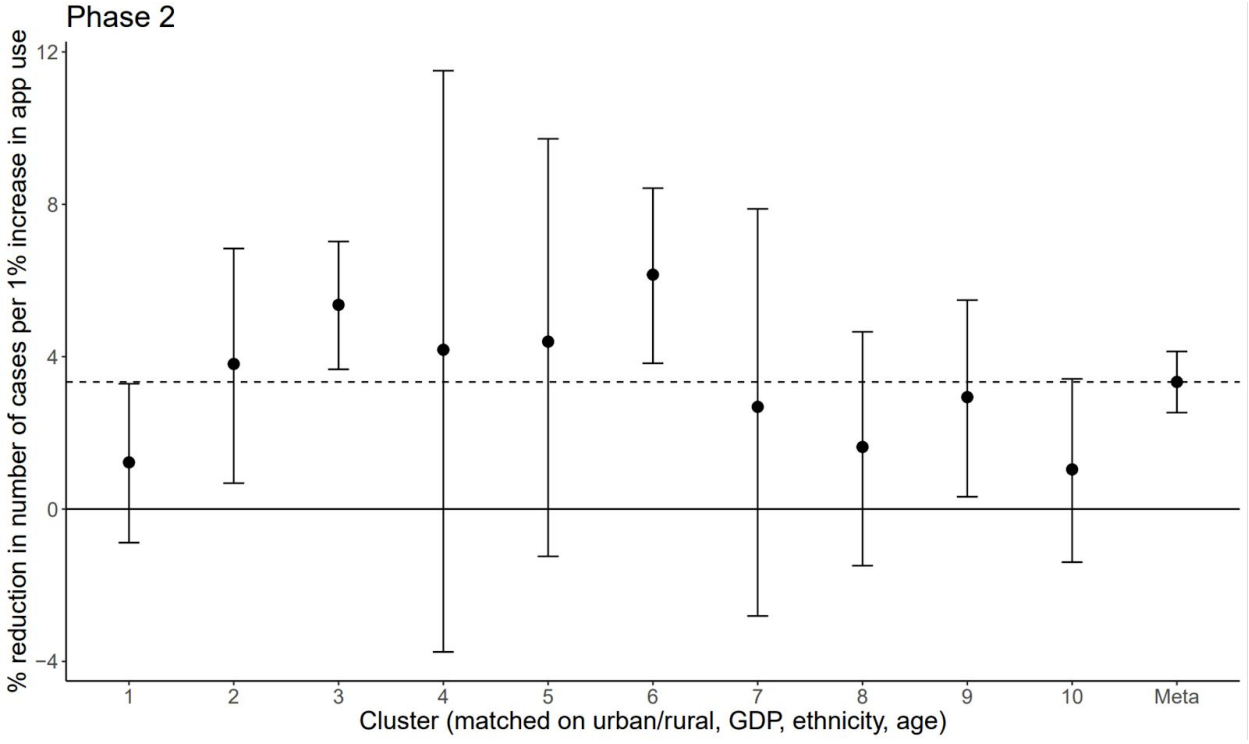
beta_uptake * (percentage of population that use the app) +

epsilon_residual

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


Combined phase 1 and phase 2
% reduction in number of cases per 1% increase in app use
10
5
0
1
2
3
4
5
6
7
8
9
10
Meta
Cluster (matched on urban/rural, GDP, ethnicity, age)


Phase 1
% reduction in number of cases per 1% increase in app use
5
0
−5
−10
1
2
3
4
5
6
7
8
9
10
Meta
Cluster (matched on urban/rural, GDP, ethnicity, age)



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.