PNAS
www.pnas.org

Supplementary Information for

Measuring the Scientific Effectiveness of Contact Tracing: Evidence from a Natural Experiment

Thiemo Fetzer, Thomas Graeber

*Corresponding author: Thomas Graeber
Email: tgraeber@hbs.edu

This PDF file includes:

Supplementary Information Text

We detail the findings outlined in the main text in five steps. First, we provide additional detail on Fig. 3. Second, we shed light on the mechanisms associated with the main effects on infections and deaths by looking at data on testing activity. Third, we zero in on the impact on the performance of the Test and Trace system. Fourth, we report a battery of robustness analyses outlined in the main text. In the fifth and final step, we discuss the quantification of the effects estimated across our set of estimation strategies.

Additional information on Fig. 3

In Table S2, we report estimation results underlying the coefficient plot in Fig. 3. We estimate the effect of late referrals to contact tracing per capita on new infections per capita (columns 1-3) and new COVID-19 deaths per capita (columns 4-5). We re-run our baseline analyses (columns (1) and (3)) on new infections and deaths while including additional controls. In columns (2) and (4), we add a large vector of 55 additional area characteristics and interact them fully with a set of time fixed effects. The area characteristics are: employment shares in 1-digit industries; educational attainment; socio-economic status of the resident population, which also captures shares in full time education or in university; and regular in-, and out commuting flows. These characteristics come from the 2011 Census. We also leverage the detailed demographic makeup of an area's population by expressing population demographics as shares in ten-year age intervals. We further control for death rates in the first wave of the pandemic in spring 2020; population density and its variability across small geographies within an area. Throughout, despite these empirically highly demanding specifications that control for non-linear case growth that may be induced by, e.g., school- or university reopenings, the results remain virtually unchanged.

Note that the data glitch simultaneously led to lower publicly announced new case numbers. Local variation in the share of cases that was missing from these announcements might have affected people's behavior. More specifically, lower local case counts may be associated with less social distancing, more public activities etc. It is unclear to which extent the population attends to and internalizes case numbers at the local level. The most salient figures are arguably those at the national level. The aggregate national growth in cases due to the data glitch, however, is orthogonal to the regional variation that our analyses exploit. To assess to which extent this endogenous response to reported local cases numbers affects our results, we employ a method of approximating people's activity based on mobility data (*1*) as has been successfully done in other work related to COVID-19 (*2*). Previous research showed that mobility measures pick up people's response to COVID-19 such as staying at home and predict the progression of the pandemic.

As shown in columns (3) and (6) of Table S2, however, controlling for a mobility metric does not affect estimates of the effect of late referrals, indicating that the behavioral response to the local number of reported cases does not play a significant role here.

To complement these results, we estimate an extended regression specification that provides these treatment effects separately for each calendar week. Fig. S7 displays coefficient estimates for the weekly effects on infections (Panel A) and COVID-19 deaths (Panel B). Reassuringly, there is no systematic relationship between late referrals and infections or deaths in the pre-treatment period. We observe highly significant and quantitatively large positive effects of late referrals on new infections in calendar weeks 39 to 41. The effect is largest in calendar week 40, roughly 2 weeks following the first delays in contact tracing and subsides around calendar week 44. In Panel B, we observe positive effects starting in week 40 and peaking in calendar week 43. The effect on the death toll lags behind the increase in infections due to the normal lag between infections and COVID-19-related deaths.

Effect on COVID-19 testing

Contacts of infected persons are encouraged by contact tracers to self-quarantine for 14 days. Note that without symptoms, a contact is neither required nor advised to take a test

themselves (*3, 4*). A negative test does not rule out an infection and the procedures required to conduct tests can by themselves contribute to the spread of the disease if a person is already infectious. By contrast, a contact who is not reached by the contact tracing system and never learns of their potential infection will not self-quarantine or take other precautionary measures, especially if they are asymptomatic or pre-symptomatic. A failure of contact tracing means that non-contacted individuals cannot respond to their potential infection, increasing the likelihood both of infecting others and of getting infected by a third infectious person if they are not already infected.

In Panel A of Fig. S5, we show the effect of late referrals due to the data glitch on the total number of tests taken in a given district, based on a regression specification analogous to the ones above (equation (2)). We document a sizeable increase in the number of tests conducted. The COVID-19 testing data are available at the weekly level. In order to be able to directly compare the magnitudes with the previous results, we divide the weekly testing figures by seven to obtain an estimate of the daily testing rate. Our main difference-in-differences estimate suggests that each late referral led to, on average, 2.7 additional tests taken per day between calendar week 39 and 44 (Table S3, Panel A, column (1)).

At the same time, we report a strongly positive effect on the number of positive tests per capita as well as the test positivity rate, see Panels B and C of Fig. S5, respectively. We estimate that each additional late referral led to a significant increase in a district's test positivity rate by 0.1 percentage points, given an average positivity rate of 3.6% (Table S3, Panel C, column (1)). The share of positive tests reverted back to pre-treatment levels in calendar week 43.

Our estimate of the effect on weekly positive tests per capita data lends credence to our above estimate for the effect on new infections which uses a different data source. We obtain a baseline estimate of 0.67 on positive tests (Table S3, Panel B, column (1)), which is closely in line and statistically indistinguishable to our estimate of 0.61 for new cases (Table S7, Panel A, column (1)).

Effects on the performance of contact tracing

Next, we analyze the repercussions on the effectiveness of the contact tracing system. To contain the further spread of the pandemic, a timely referral of cases to the contact tracing is essential. Unfortunately, the publicly available data on the Test and Trace system, especially on contact tracing performance, are far from exhaustive. As described above, contact tracing begins after positive cases are reported by laboratories to PHE, which in turn transfers case information to NHS Test and Trace. Contact tracers contracted by NHS Test and Trace then first contact individuals who tested positive. At this stage already, not all individuals that tested positive may be successfully reached. Even if an individual is reached and asked to provide contact details of recent close contacts, they may not properly recall or they may not be willing to disclose all relevant information. The actual contact tracing only sets in after contact information is obtained either through the contact tracer or the secure website. This implies that there are multiple margins through which contact tracing – even under normal circumstances – may fail: (a) not all COVID-19 positive individuals may be successfully reached; (b) those individual may imperfectly recall or incompletely disclose recent contacts; (c) and the contact tracing system may fail to reach all identified contacts.

Fig. S2 studies aggregate performance data capturing the fraction of close contacts that were advised to self-quarantine by the time taken to reach them. It demonstrates the possible effect that the data glitch had on the time taken to reach contacts. Note that this figure zooms in on the number of contacts that were actually reached, i.e., it focuses on step (c) above conditional on success in steps (a) and (b). While the fraction of those who were reached *within the first 24 hours* hovered above 80% in the weeks preceding the data glitch, the fraction plummeted to just above 60% in calendar week 40. Strikingly, we find that the tracing system's performance remains low even in the three weeks following the correction of the data glitch. The share of contacts reached within 24 hours only appears to revert back to pre-treatment levels by week 44. This suggests that the tracing system was jammed by the late referrals from late September, adversely affecting the tracing performance for cases referred

after the data glitch was corrected on October 3. Put differently, the tracing system may not have been well adapted to handle both a sudden influx of thousands of COVID-19 positive tested individuals referred to contact tracing with a delay and the subsequent higher infection levels that arose due to the preceding failure of a timely referral to Test and Trace.

In Fig. S6 we present findings on the Test and Trace performance at the Upper Tier Local Authority level, analyzed using the same baseline specifications as above. Areas that experienced a larger impact on late referrals saw a deluge of referrals to contact tracing from calendar week 40. Note that a part of this increase may be mechanical as the contact tracing statistics for the week from October 1 to October 7 – that straddles calendar weeks 40 and 41 – is matched to calendar week 40. We find that the impact on referrals persists throughout the subsequent weeks, similar to our estimates of the effects on infections and testing activity. This prolonged impact likely captures the fact that many of the individuals that were referred late to the contact tracing system further spread the disease, resulting in an overall worsening of the local pandemic situation.

The only subnational performance measure available at the UTLA level captures the share of contacts reached out of all contacts recorded from those positively tested individuals who were both referred to the contact tracing system and successfully reached. These data do not include the time it took to reach individuals. In Panel C of Fig. S6, we document some (more noisily estimated) evidence suggesting that the performance of contact tracing declined more drastically in parts of England that experienced a stronger impact on late referrals due to the data glitch. These estimates imply that the performance deteriorated with fewer close contacts being successfully reached. Late referrals are associated with a prolonged negative effect on the performance of the Test and Trace system that extends well beyond the correction on October 3. The corresponding difference-in-differences estimates are presented in Table S4.

Robustness exercises

We conduct a number of additional analyses to shed light on the robustness of our findings.

Refined difference-in-differences estimation using matched pairs We construct a refined difference-in-differences estimator based on a procedure of matching areas which are highly similar in terms of their pre-treatment exposure to the pandemic. This approach aims at creating even more accurate treatment-control comparisons. In columns (4) to (6) of Table S7, we report results that correspond to those in columns (1) to (3) except for the different construction of control groups. We reliably estimate treatment effects that are statistically indistinguishable from the plain difference-in-differences approach. Similar robustness exercises are reported for the other outcome measures that we study, see columns (4) to (6) in Tables 3 and 4.

We point out that the matched-pairs design is empirically exceptionally demanding. By creating matched pairs and controlling for time-fixed effects specific to each pair, we conduct like-for-like comparisons by studying pandemic outcomes within pairs of districts that have been on a highly similar trajectory just prior to the data glitch.

Regional heterogeneity Before exploring to which extent our results are driven by individual regions, we examine the regional heterogeneity of the estimated treatment effects. To obtain these estimates, we refer back to our main difference-in-differences model and interact the main treatment measure with a set of region dummies, plotting out the coefficients along with 90% confidence bands. These are presented in Fig. S14. The results suggest that the impact of delayed referrals on subsequent infections is most pronounced in the East Midlands, the North West, the South East as well as in Yorkshire. The effects on COVID-19-related deaths are more noisily estimated. This analysis suggests that the positive impact on deaths is most severe in the East, London, the North West and Yorkshire.

Sensitivity to geographic regions and spatial disaggregation So far, we reported our analyses at the level of the Lower Tier Local Authority (315 units, Table S7). In Tables S9 and S10 we replicate our findings at the Upper Tier Local Authority Level (149 units) as well as the NUTS3 region level (Nomenclature of Territorial Units for Statistics, 93 units). We obtain very similar results irrespective of the spatial resolution of the data, which further shows that the results are not an artefact of or significantly impacted by inter-regional spillovers of the treatment effect. Moreover, in Fig. S15 we examine the sensitivity of our findings to excluding individual areas from the estimation. We show the distribution of the leave-one-out-estimator of the effect of late referrals on new cases and deaths, separately for analyses conducted at the LTLA, UTLA and NUTS3 levels. The observed sensitivity of the treatment effects to excluding individual regions is small.

Alternative functional forms for the relationship between late referrals and COVID-19 spread Our main regression specifications estimate the effect of the per capita level of late referrals on the per capita level of measures of COVID19 spread, controlling for the level as well as non-linear trends in pre-treatment exposure to the pandemic. The non-linear nature of infection dynamics suggests specifications with logarithms as an alternative. In Table S11, we additionally estimate the same type of regressions using different combinations, such as a log-log as well as a log-levels specification, replicating our main findings.

Alternative measures of late referrals We made conservative assumptions to construct our baseline measure of late referrals, but there is some degree of flexibility in the calculation of the treatment measure. We explore the sensitivity of our findings to the use of alternative approaches in Table S12. Our main measure of late referrals aggregates all cases with a specimen date between September 20 and September 26 that were not referred to Test and Trace as of October 2. This measure is conservative in terms of the number of late referrals it predicts: it relies on 7,242 late referrals that can most clearly be identified as such, which is less than half of the officially reported figure of 15,841 late referrals. We report regression results analogous to those in Table S7 for three alternative ways of inferring of late referrals that are due to the data glitch.

To this end, we non-parametrically estimate the time path of the *typical* reporting lag from the time immediately preceding the data glitch, i.e., for specimen dates between September 1 and September 20. This allows us to predict the fraction of cases with a given specimen date that should be reported a given number of days after the test was taken.

We use this prediction exercise, first, to create an even more conservative measure than our baseline by subtracting the number of cases that we would expect to not have been reported by October 2 under the typical reporting lag. As argued above, this barely affects our measure. Even for the latest date in the specimen date range considered, September 27, we would expect 95.9% of cases to have been reported by October 2 under normal circumstances (see Table S1). This more conservative measure reduces our predicted number of late referrals from 7,242 to
6,044.

Second, we create a more comprehensive, yet less conservative measure by including specimen dates of up to October 1, and by accounting for the typical reporting lag using the same non-parametric estimation as above. Note that due to the potential divergence between the estimated typical reporting lag from pretreatment data and the actual reporting lag, this measure is noisy, especially for lower spatial aggregation. This measure leads to a total number of 9,755 late referrals, still below the officially reported figure of 15,841 missed cases.

Third, we re-run our analyses using our baseline measure of late referrals as a fraction of the total number of cases between September 20 and September 27 in a given area. This measure suffers from statistical bias: because a fraction measure is noisy in areas with low case counts, it has an artificial upward bias. We present estimation results for the fraction measure as well as a version that exclude areas with a total case count below 50 during the time of September 20 and 27.

All of these results are presented in Table S12. We replicate our main results for each of these three alternative measures, and more compellingly, find that the estimated effect magnitude varies little across specifications.

Placebo tests In Fig. S16, we present a series of placebo tests. To do so, we construct a simple estimator of late referrals by specimen date as follows. First, for each specimen date, we retrieve the number of cases that was reported as of seven days following that specimen date. Second, we subtract this number from the final, "true" case count for that specimen date, which is the number of cases known for this specimen date as of the most recent version of the data. This allows us to construct, for each specimen date, a measure of the number of cases that are were not yet reported as of one week following the specimen date. We construct this measure for each specimen before, during and after the period that was affected by the Excel error.

The hypothesis of this placebo exercise is that judging from this measure of late referrals for a specific specimen date, only specimen dates between September 20 and September 25 should be predictive of future case growth. For tests taken on September 26, the case count seven days was already subject to the correction of the data glitch that occurred on October 3.

We test this hypothesis by running our main specification (equation (2)) using these measures. The results are presented for both new infections and new COVID-19-related deaths in Fig. S16. We document that only the missing case figures constructed in this fashion from around September 20 to September 26 strongly predict subsequent case growth and deaths.

Alternative inference Inference in the paper is conducted using clustering of standard errors at the spatial level at which the outcome data is measured. An alternative is to conduct a type of randomization inference. To do so, we draw repeated random samples of the main missing cases measure, redistributing the missing cases randomly across districts. We do this in three ways: reshuffling district exposure measures M_i across all districts in England; across all districts within the 9 NUTS1 regions; and across all districts within the 33 NUTS2 regions. For each exercise, we create 100 reshuffled treatment exposure measures using these three approaches. This allows us to estimate the treatment effects for these placebo treatment assignments. We would expect that the point estimate that are obtained based on the true spatial distribution of the missed cases to be sharply different from the null effects we would expect for the reshuffled distribution.

The latter may not be the case, especially for the reshuffling exercises at the region or NUTS2 region level: due to potential spatial autocorrelation, our treatment effect estimates may spuriously pick up treatment effects due to such spatial correlation. We present these results for new COVID-19 cases and deaths in Fig. S17 as a set of kernel density plots of the distribution of the 100 point estimates that are obtained from these placebo exercises. We indicate with a vertical line the point estimate obtained from using the true distribution of the district exposure measures M_i. Throughout the exercises and the outcomes, we can reject the null hypothesis that the effect we observe is spurious with implied p-values that are below 0.01%.

Distribution of estimated effect sizes We estimate a distribution of effect sizes across the universe of our robustness analyses laid out above. Regarding our key outcomes of new infections and new COVID19-related deaths, we estimate the following effect ranges: in our collection of point estimates, we find that each late referral that we identify as being due to the data glitch was related to between 17.5 and 19 additional cases, and to between 0.21 and 0.29 additional COVID19-related deaths during the six-week post-treatment period.

Quantification of effects

We offer a tentative quantification of the effects across the whole of England and the English regions in Table S8. We anchor these point estimates on the main point estimates

presented in Table S7 as well as the most conservative point estimate obtained from our most saturated specification in Table S2.

To arrive at the presented figures, we leverage the point estimate and simulate the full distribution of effects for the post-treatment period that ranges from calendar week 39 to including calendar week 44. For the cumulative new infections, our point estimates suggest that with 90% confidence, between 13% to 40% of the nearly 600,000 new *detected* COVID-19 infections may be attributable to the failure to contact tracing. This calibration implies that 127,018 infections, or around 21% of all detected infections may be due to the contact tracing failure.

The numbers of additional COVID-19-related deaths linked to the error are estimated less precisely. Our central conservative point estimate would suggest that, out of the total of 7,196 COVID-19 deaths during the time window, a similar share of around 21% are due to the contact tracing error.

The table provides a range of further upper- and lower-bound estimates as implied by the 90% confidence intervals spanning around the point estimates. It also highlights that, not surprisingly, the effect is quite homogenous across the English regions in relative terms.

We advise caution, however, against taking these effect sizes at face value: due to the complex structure of a pandemic, such as externalities across areas and the non-linear nature of infectious developments, effect magnitudes are inherently difficult to interpret.

Materials and Methods

Empirical context: contact tracing in England

In England, laboratories report positive COVID-19 test results to *Public Health England* (PHE) on a daily basis. The PHE aggregates all nation-wide test results using an automated reporting dashboard, which forms the basis for the official reporting of case numbers as well as contact tracing (*5*). Specifically, data on positive cases are passed on to the *NHS Test and Trace* (Test and Trace) system, a government-funded service that was established in 2020 to organize all contact tracing at the national level (*4*). For all cases that do not come from a high exposure setting such as a school or a prison, the infected person is contacted via a text, email alert or phone call and asked to shared details of their recent close contacts and places they have visited. They can respond online via a secure website or by telephone with a contact tracer.

Data on COVID-19 in England

Our baseline analyses leverage three sources of publicly available data.

Reporting dashboard Our primary dataset is constructed using the UK's COVID19 dashboard.[1] This dashboard provides granular data on COVID-19 infections and deaths at different spatial resolutions. Our geographical focus is on England, because other countries in the UK were not affected by the Excel error. The data include daily lab-confirmed positive test results and deaths. Data on positive cases are characterized by two dates: the specimen date, i.e., the date when the sample is taken from the person being tested, and the reporting date, i.e., the date when a positive case is first included in the published totals and referred to Test and Trace, so that contact tracing can begin. In order to reconstruct the time line of case reporting for each specimen date, we collect "vintage datasets" published on past reporting dates. The distinction between specimen and reporting date forms the basis for our analysis of delays in contact tracing due to the Excel error.

We conduct analyses at different levels of spatial disaggregation. England has 315 lower tier local authority districts (LTLA). While most COVID-19 data are published at this level, some data are only available at the upper tier authority district level (UTLA) – of which there are 149 in England. Our baseline analyses exploit variation at the LTLA level but we replicate our results at the UTLA level as well as NUTS3 region level, of which there are 93 units.

The resulting core dataset is a balanced daily panel. Our estimation window focuses on the period starting in calendar week 28 (starting July 6, 2020) all the way to calendar week 44 (starting October 26), covering a total of 37,485 observations.

Test and Trace statistics We also draw on data on testing and tracing statistics provided by NHS Test and Trace (*6*). These data are published weekly and provide some statistics on the effectiveness of the contact tracing efforts such as the fraction of contacts reached, delays as well as the total number of tests taken and test positivity rates. These data are available at different geographical and temporal resolutions than the daily case data. Specifically, while the COVID-19 test statistics are provided for the most granular lower tier local authority district level (LTLA), the contact-tracing data are more patchy and only available at the coarser UTLA level. The data are provided at the weekly level for weeks starting on Thursday and ending on Wednesday. This implies that calendar weeks are not cleanly separated in this dataset. We matched reporting windows to calendar week based on the largest overlap. For example, calendar week 39 ranges from September 21 to September 28. The nearest reporting window for the Test and Trace statistics is the week starting on September 24 and ending on September 30, which straddles four days of calendar week 39 and three days of calendar week 40. We match this week to calendar week 39. This implies, however, that the identification of the exact timing of effects is more challenging in the weekly data.

[1] Available at https://coronavirus.data.gov.uk/.

Figure S13 provides an overview of the process flow and highlights the sources of delays and potential caveats to bear in mind when studying and interpreting the data, especially relating to contact tracing and its performance. The performance of the system, for example, is undermined if, e.g., a high fraction of COVID19 positive individuals cannot be contacted or reached. This naturally implies that potential close recent contacts may not be identified. Similarly, even if an individual that tested positive is successfully contacted, they may not remember the individuals they spent notable time together during the time they may have been infectious. And even if individuals provide details of close contacts, these may not be reached in a timely fashion or may not be reached at all.

There appears to be room for improvement in the comprehensiveness of public reporting and the statistical presentation of the data.[2]

Additional weekly death statistics In addition to the daily death statistics, we also leverage weekly death statistics at the local authority level as published by the Office for National Statistics (7). These data report on new COVID-19-related deaths by the type of location where the death occurred, e.g., at home, in hospitals or in care homes.

Identifying delayed referrals to contact tracing

We rely on granular data on positive COVID-19 tests to construct a measure capturing the extent to which positive COVID-19 cases have been affected by the delayed referral to contact tracing across different parts of England. The official PHE announcement only specified the total number of late referrals but provided no information about the geographical distribution and the specimen dates of these cases. A Freedom of Information request has been raised by the authors to obtain a detailed geographic and temporal break down of all cases that were referred to contact tracing with a delay – so far, these data have not been made available.[3]

Baseline measure of late referrals Despite the lack of official data, we can infer which individual cases have been affected by a delayed referral. To do so, we study the reported case figures at different points in time. The logic of our approach follows from Table S6 and Fig. S8. Table S6 shows the COVID-19 case counts as they were reported on three different dates: November 15, October 4 and October 2. The case counts are broken down by the date on which the test sample was taken (specimen date). For all tests taken on September 24, the most recent figures from November 15 imply a total of 6199 positive cases known as of November 15. Because more than 1.5 months have passed between the specimen date of September 24 and the reporting date of November 15, all tests should have been processed and entered the statistics. We can interpret the number from November 15 as the final case count for this specimen date. In fact, the typical time lag between the specimen date and the reporting date is much shorter. Table S1 and Fig. S1 highlight that usually, between 94% to 96% of all positive cases are identified and reported within the five days following the specimen date. For our baseline measure, we therefore restrict our attention to the earliest specimen dates that where likely impacted by the Excel error, September 20 to September 27. By the time the error was discovered on October 3, tests taken during this specimen date range should have almost fully entered the statistics under normal circumstances. Panel B of Fig. S8 visualizes the striking discontinuity caused by the data glitch in the otherwise smooth increase of the fraction of cases reported in the days following a given specimen date. Taking the example of the specimen date September 24, we observe that the fraction reported had converged to a steady level by October 2, but then a sudden upward revision occurred on October 3. This stands in contrast to the overall smooth evolution of the fraction reported for

[2] The authors have launched a public FOIA request to request more granular data. The FOIA request can be accessed here: https://www.whatdotheyknow.com/request/nhs_test_and_trace_statistics_re.

[3] The FOIA is in the public domain on https://www.whatdotheyknow.com/request/regional_breakdown_of_cases_not.

specimen dates not affected by the data glitch, as shown in Appendix Fig. S9. These figures capturing the over-time conversion to the final case count on a given specimen date leverage data from different historically published versions of the COVID-19 dataset.

Judging from the typical reporting lag as observed between September 1 and September 19, we would expect that at least 95.9% of the positive tests taken on September 27 and at least 99.3% of the positive tests taken on September 20 have been reported before October 3. In reality, however, this share turned out to be much lower as a result of the late referrals. Fig. S1 visualizes the share of cases reported with different reporting delays – the fraction reported by day five following a specimen date dropped to roughly 60% during the period affected by the data glitch.

For reasons of parsimony, our baseline measure is constructed assuming that *all* cases taken between September 20 and 27 would have been reported by October 2 in the absence of the data glitch. Formally, this means we define the number of late referrals in district *i* that were likely due to the Excel error as

$$M_i = \sum_{\omega=20\ \text{Sep}\ 2020}^{27\ \text{Sep}\ 2020} \text{True Case}_{i,\omega}^{11\ \text{Nov}\ 2020} - \text{Case}_{i,\omega}^{02\ \text{Oct}\ 2020}.$$

Specifically, across target specimen dates we sum up the difference between the final, "true" case count approximated by the most recent dataset version (November 15) and the case count known as of October 2. Note that, first, this baseline measure is transparent and does not impose auxiliary assumptions about the structure of the counterfactual reporting lag. Second, missed cases from earlier specimen dates are likely to have had the most pronounced effect on the development of the pandemic. A contact who contracted the disease from a person who tested positive on September 20, for example, could in turn infect others before the contact was finally traced by Test and Trace on October 3 or thereafter. This implies that the adverse effect of delayed contact tracing is stronger for cases with earlier specimen dates. In total, we calculate a number of 7,242 late referrals to contact tracing with specimen dates between September 20 and 27. This figure broken down to the LTLA level forms the basis for our measure of the local impact on late referrals due to the Excel error. We thus have a time-invariant scalar measure of treatment intensity in terms of late referrals.

Our baseline measure captures substantial variation in the extent to which different areas were affected by the data glitch. To illustrate, Panel D of Fig. S8 shows a distinct geographic signature – there is substantial heterogeneity in the fraction of cases that we categorize as late referrals in each area. We confirm below that this heterogeneity is *as-if* random: it appears to be unrelated to all area-specific characteristics that are relevant for the local development of the pandemic, and we can therefore exploit it to evaluate the quasi-causal effects of the intensity of contact tracing.

Alternative measures of late referrals While our baseline measure relies on just 7,242 late referrals out of the total of 15,841 cases that were officially acknowledged, our subset of late referrals from early specimen dates are likely to have had the strongest impact on the progression of COVID-19, and we can most cleanly identify these from the available sources of data. We construct a set of alternative measures of late referrals covering shorter or longer windows of specimen dates, e.g., from September 20 to 25, or from September 20 to 30. We further complement these analyses with a more parametric approach that statistically approximates the time path of the "typical reporting lag", i.e., the distribution of delays in the absence of a processing error. Specifically, we proceed in two steps. First, for a given window of specimen dates, e.g., September 20 to October 1, we determine the reported case numbers that should be expected under the typical reporting lag, which we obtain by statistically approximating the usual evolution of reported fractions as shown in Fig. S1. To this end, we estimate the fraction of cases that would be reported *d* days after the test was taken by fitting the following function using non-linear least squares:

$$f_{i,d} = 1 - (1 - \left(\frac{r}{1 + e^{-c(d-t)}}\right)^{d}). \quad (1)$$

The above functional form is often invoked to approximate converging processes in variety of domains and can be estimated using non-linear least squares (see (*8*) for an implementation in R). The fit of this estimation for the pre-treatment period of September 1 to September 19 is illustrated in Fig. S11.

In a second step, we compare the predicted number of reported cases to the actually reported number of cases by October 2 to construct our measure of late referrals:

$$M_i = \sum_{\omega=20\ \text{Sep}\ 2020}^{1\ \text{Oct}\ 2020} \text{True Case}_{i,\omega}^{11\ \text{Nov}\ 2020}(1 - \widehat{f_{\iota,d(\omega)}})$$

The above model 1 can be estimated at the country-level, but can also be trained at the region level to allow for region-specific variation in the typical reporting lag.

Our baseline specification relies on the number of late referrals normalized by the population size, while flexibly controlling for the local level and dynamics of the evolution of the pandemic. As an alternative measure, we can express the number of late referrals as a fraction of the final number of positive cases reported for September 20 to 27, by computing

$$m_i = \frac{\sum_{\omega=20\ \text{Sep}\ 2020}^{27\ \text{Sep}\ 2020} \text{True Case}_{i,\omega}^{11\ \text{Nov}\ 2020} - \text{Case}_{i,\omega}^{02\ \text{Oct}\ 2020}}{\sum_{\omega=20\ \text{Sep}\ 2020}^{27\ \text{Sep}\ 2020} \text{True Case}_{i,\omega}^{11\ \text{Nov}\ 2020}}$$

This measure is conceptually appealing in that it accounts for the local severity of the pandemic but it is statistically problematic due to a small sample issue. A fraction measure is noisy for areas with a low true case count, which creates a positive bias in our application. We use the above as an auxiliary measure imposing some sample restrictions, i.e., by focusing on places with at least a minimum number of cases. In the Appendix, we explore a variety of measures and show that our findings are robust to those.

Empirical strategy

Our empirical strategy exploits cross-sectional variation in the extent to which different parts of England were affected by the delayed referral of COVID-19 positive cases to contact tracing efforts. This cross-area variation in exposure is quasirandom as a result of the Excel data entry error, allowing us to study the causal effect of contact tracing on measures of subsequent COVID-19 spread. We follow a simple and a refined difference-in-differences estimation approach at different geographic resolutions of the data.

Difference-in-differences specification The basic difference-in-differences estimate is obtained from estimating

$$y_{i,t} = \eta_i + \gamma_i + \eta \times Post_t \times M_i + \beta' X_{i,t} + \epsilon \quad (2)$$

where $y_{i,t}$ denotes a measure of COVID-19 spread in area i at time t (either a specific date or a week). The regression controls for district fixed effects, η_i, as well as a set of time fixed effects γ_t. To account for the non-linear nature of case growth we add a host of additional measures $X_{i,t}$ of the disease progression across areas and control flexibly for these.

Specifically, we measure an area's average number of new COVID-19 cases per capita, the number of tests per capita as well as the positivity rate of the tests during calendar weeks 37 to 38 (September 7 to September 20), directly preceding the data glitch. For each of these measures, we categorize districts into deciles according to its empirical distribution across

districts. We successively control for non-linear time trends in these variables by decile. This ensures that we are not confounding or wrongly attributing differences in the outcome variables to the fact that different parts of England had been at different stages in the pandemic. Instead, this specification aims at identifying the differential effect that late referrals to contact tracing had on the subsequent spread of COVID-19, comparing areas that have been on a very similar trajectory in the pandemic in the weeks just prior to the data glitch.

Naturally, the above exercise can be extended to flexibly estimate treatment effects over time. This will further allow us to shed light on the common-trends identification assumption implicit in the above difference-in-differences approach.
Specifically, we estimate

$$y_{i,t} = \eta_i + \gamma_i + \sum_t \eta_t \times \mathbb{I}(Week_t = t) \times M_i + \beta' X_{i,t} + \epsilon \quad (3)$$

This allows us to plot the estimated coefficients $\widehat{\eta_t}$ and explore to what extent differences in the outcome emerge around the time that the contact tracing shock happened and to what extent this affected the pandemic development going forward.

Zooming in on districts with similar pandemic evolution We supplement our baseline difference-in-difference exercise with an additional exercise that aims to tackle potential concerns about the non-linear growth in cases. To do so, we refine the control group for our difference-in-differences design. For each district *i*, we compute the similarity between that district *i* and every other district *j* in terms of their disease progression just prior to the Excel error. To measure distance, we use the cosine similarity metric, applied to the following vector of seventeen characteristics $\mathbf{X_i}$ capturing the disease progression in a district *i*: new COVID-19 cases and deaths per capita reported on October 1, 2020; the number of COVID19 tests along with the positivity rate in calendar week 40; the average number of new COVID-19 cases and deaths per capita in calendar weeks 37 and 38; the average number of COVID-19 tests per capita and positivity rate during weeks 37 and 38; and the growth in new COVID-19 cases, tests, deaths and the positivity rate between week 37 and 38. We also add measures of the pandemic progression in the first wave, such as the death rates in March to June, as well as other area characteristics, such as population density. The similarity measure is computed as:

$$\text{similarity}_{ij} = \cos(\theta) = X_i \cdot \frac{X_j}{\| X_i \| \| X_j \|} = \frac{\sum_{p=1}^{n} x_{i,p} x_{j,p}}{\sqrt{\sum_{p=1}^{n} x_{i,p}^2} \sqrt{\sum_{p=1}^{n} x_{j,p}^2}} \quad (4)$$

To illustrate this exercise, Fig. S12 displays the cosine similarity measure between Adur and Watford (red diamonds), as well as their similarity to all other districts (blue dots) on the horizontal axis, along with a set of measures capturing the pandemic situation prior to the data glitch. Based on cosine similarity, Adur and Watford are closest to each other and are used to form a matched pair. Across all individual measures included in the similarity measure, the two districts are highly similar, showcasing that cosine similarity allows to identify districts with similar disease progression statistics. Even though the two districts share a cosine similarity score of 0.97, they differ substantially in terms of their number of late referrals, owing to the idiosyncratic effect of the data glitch. While Adur experienced 6.3 late referrals per 100k, Watford only saw 3.44 late referrals in that same time period. This highlights that our measure captures heterogeneity in exposure to the data glitch even in this matching approach that zooms in on otherwise highly similar districts.

For each district *i*, we identified a “best match” *j* using matching without replacement. We obtain 157 matched pairs from 314 districts, omitting the last district. We estimate a version of the above specification,

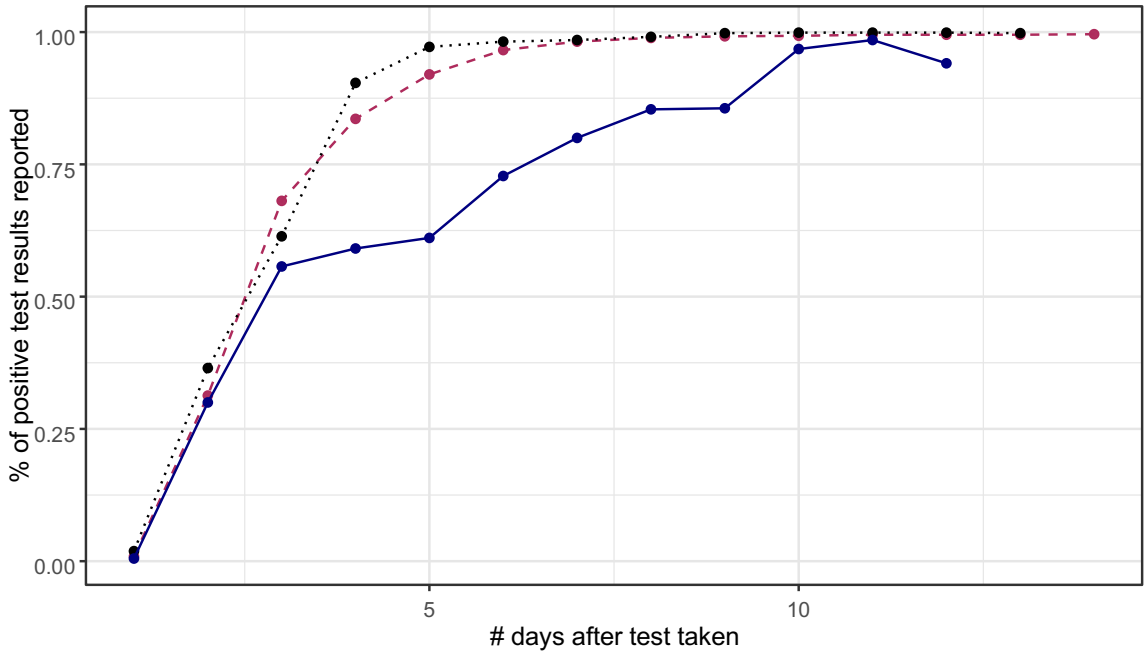
$$y_{i,t} = \eta_i + \gamma_{p,t} + \eta \times Post_t \times M_i + \beta' X_{i,t} + \epsilon, \quad (5)$$

where we now control for matched-pair-by-time fixed effects $\gamma_{p,t}$. In our most demanding specification, we control for 157 different sets of time fixed effects, allowing individual non-linear time trends for places that have been on a similar pandemic trajectory in the pre-treatment period. Practically, we estimate nearly 18,683 (157 x 119 days) separate time effects. This addresses the fact that infection dynamics in a pandemic may produce non-linear growth in cases. By virtue of zooming in on matched pairs of districts that look very similar in terms of the pandemic just around the Excel error occurred and affected these districts quite differentially, this will further strengthen the identification offered by this natural experiment.

References

1. T. Fetzer, *et al.,* Global Behaviors and Perceptions in the COVID-19 Pandemic. Harvard Business School Working Paper 20-111 (2020).
2. Google, "Google COVID-19 Community Mobility Reports.," 2020, (https://www.google.com/covid19/mobility/). [the easiest access to this source is via the URL]
3. NHS, "If you're told to self-isolate by NHS Test and Trace or the NHS COVID-19 app" (https://www.nhs.uk/conditions/coronavirus-covid-19/testing- and-tracing/nhs-test-and-trace-if-youve-been-in-contact-with-a-person-who-has- coronavirus/). [the easiest access to this source is via the URL]
4. Department of Health & Social Care, "NHS Test and Trace statistics (England): methodology" (https://www.gov.uk/government/publications/nhs-test-and-trace-statistics-england-methodology/nhs-test-and-trace-statistics-england-methodology). [the easiest access to this source is via the URL]
5. Department of Health & Social Care, "COVID-19 testing data: methodology note" (https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note). [the easiest access to this source is via the URL]
6. Department of Health and Social Care, "NHS Test and Trace (England) and coronavirus testing (UK) statistics" (https://www.gov.uk/government/publications/nhs-test-and-trace-england-and-coronavirus-testing-uk-statistics-22-october-to-28-october). [the easiest access to this source is via the URL]
7. ONS, "Death registrations and occurrences by local authority and health board" (https://www.ons.gov.uk/datasets/weekly-deaths-local-authority/editions/time-series/versions/8). [the easiest access to this source is via the URL]
8. N. C. Steer, P. M. Ramsay, M. Franco, nlstimedist: An R package for the biologically meaningful quantification of unimodal phenology distributions. Methods in Ecology and Evolution. 10, 1934–1940 (2019).

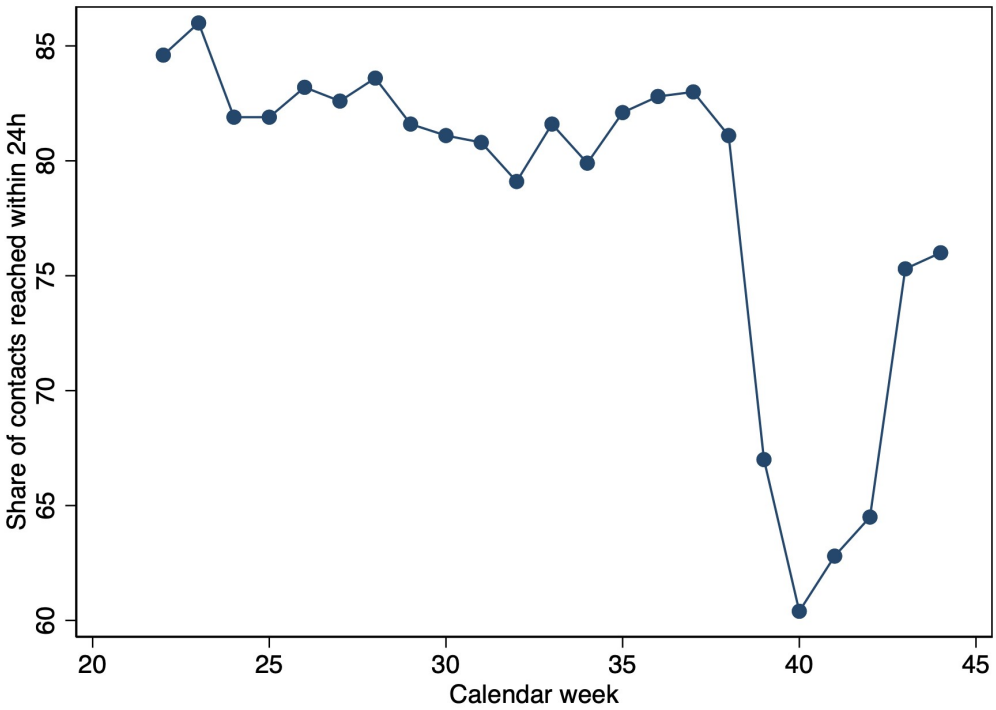
Fig. S1. Fraction of positive COVID-19 cases tested on a specific date by reporting delay



Notes: Figure plots the share of positive COVID-19 test results that are reported, published and referred to contact tracing as a function of the number of days since the test was taken. The maroon dashed line represents case data from Sept 1 to Sept 20, 2020. On day 5 after the test was taken, on average, 92% of all test results have been published and individuals have been referred to contact tracing. The blue line represents the same curve but for tests performed from Sept 20 to Oct 1st. There are notably fewer positive cases reported and referred to contact tracing as a result of the spreadsheet error. Up to five days after the specimen for a test was taken only 61% of positive test results have been published. The black dotted line presents the same data but for the period from Oct 5 to Oct 15 highlighting this was a temporary glitch.

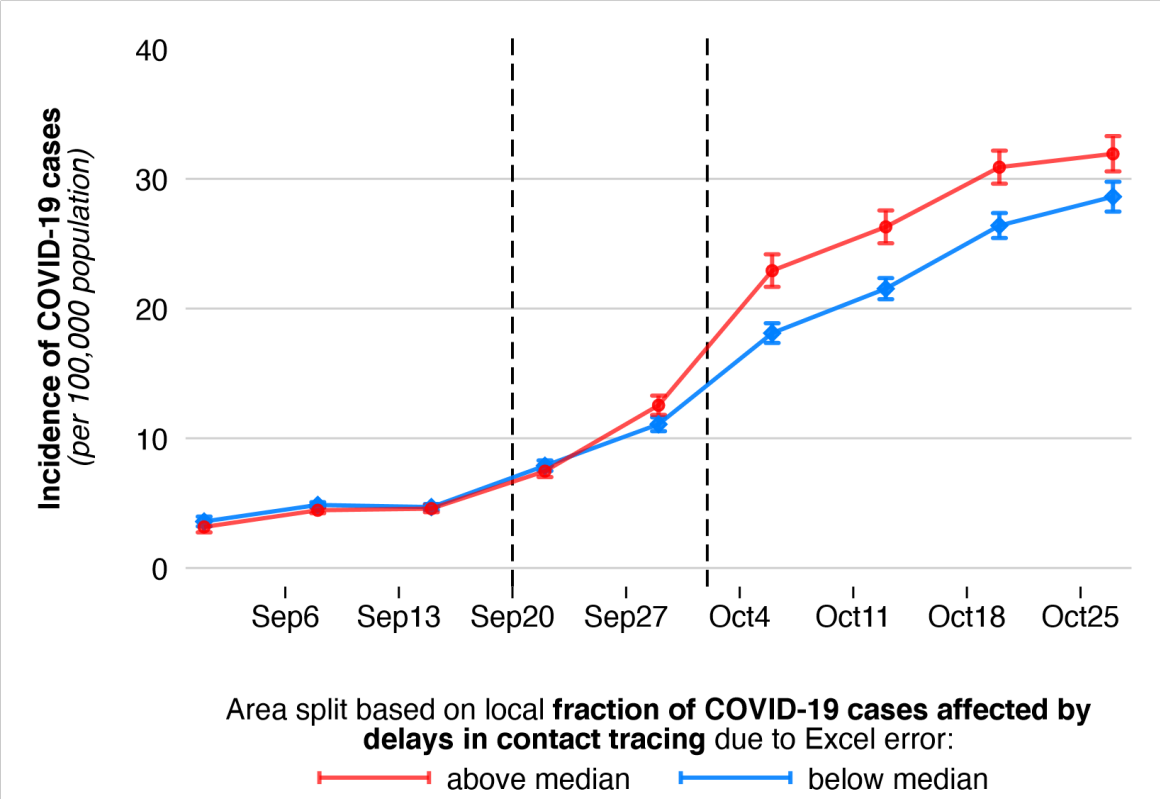
.

Fig. S2. Evolution of performance of centrally managed contact tracing effort over time



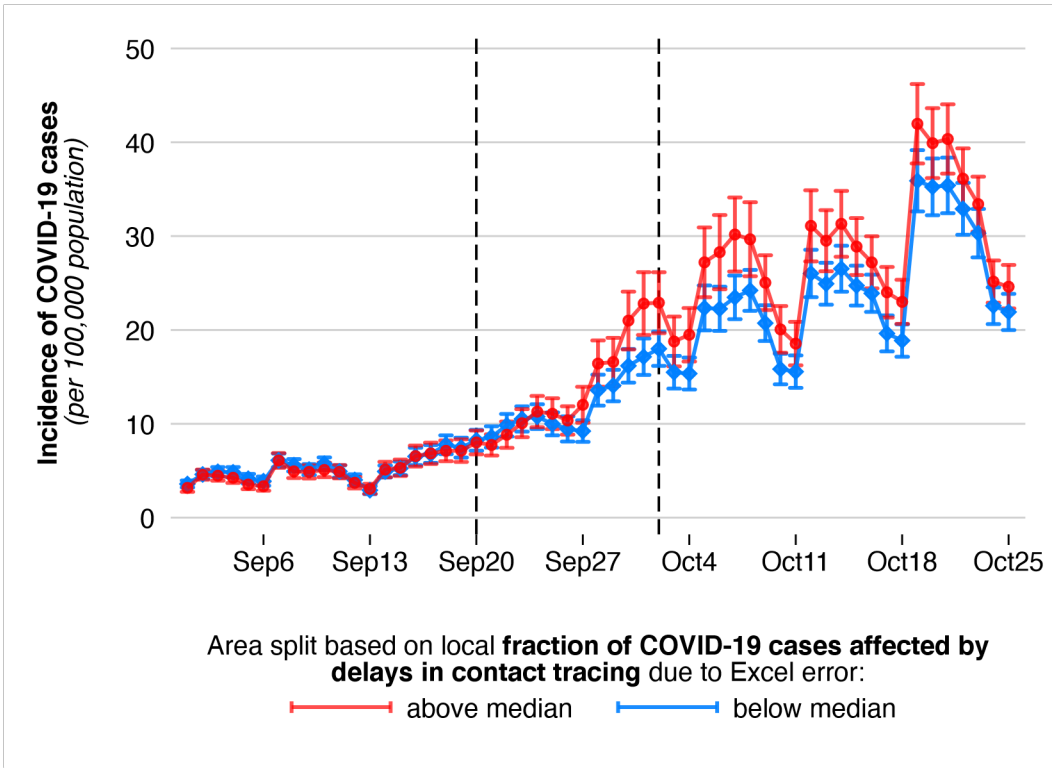
Notes: Figure plots the share of contacts of individuals who were advised to self-quarantine by time taken to reach them. The vertical axis presents the share of all contacts of individuals that were asked to self-quarantine that have been reached within 24h. This excludes data pertaining to cases where the individuals that are supposed to self-quarantine have not been contacted and may also exclude individuals who have not provided any details of close contacts. Individuals that were asked to self-quarantine in response to a positive test in weeks 39 and 40 were affected by the Excel error. The untimely referral caused the contact tracing performance to decline.

Fig. S3. Evolution of local COVID-19 incidence in areas with above median vs. areas with below median exposure to delays in contact tracing due to the Excel error: Median split by deciles of pre-treatment infection intensity



Notes: Figure is similar to Fig. 2, but defining above/below median share of local delays in contact tracing based on deciles of pre-treatment infection intensity compared to quintilesin Fig 2. For each of the 315 Lower Tier Local Authorities in England, we calculate the share of positive COVID-19 tests taken between September 20 and September 27 that were referred to contact tracing with an unusual delay of 6 to 14 days due to the Excel error. We create two equally sized groups of areas based on whether they – by chance – experienced above median or below median exposure to unusual delays in contract tracing. We plot the average incidence of COVID-19 for each group by test date. We observe virtually identical pre-treatment trends across groups but a substantive divergence in COVID-19 spread at the onset of the period during which the Excel error occurred, which is highlighted by the dashed lines. 90% confidence intervals displayed.

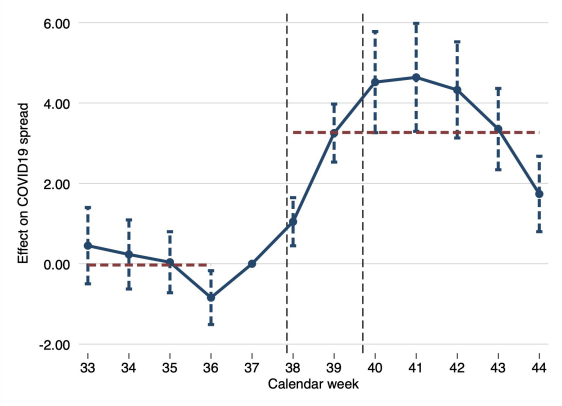
Fig. S4. Evolution of local COVID-19 incidence in areas with above median vs. areas with below median exposure to delays in contact tracing due to the Excel error: Daily data



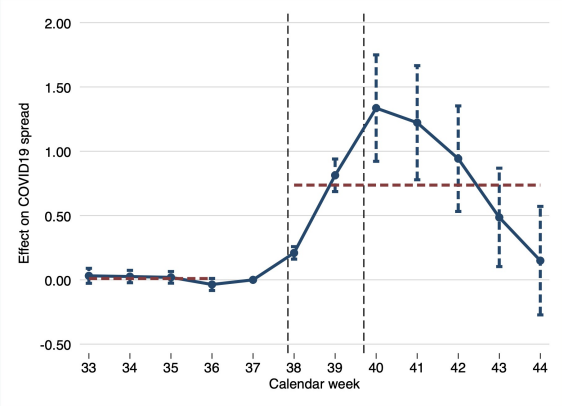
Notes: Figure is similar to Fig. 2, but using daily instead weekly data. For each of the 315 Lower Tier Local Authorities in England, we calculate the share of positive COVID-19 tests taken between September 20 and September 27 that were referred to contact tracing with an unusual delay of 6 to 14 days due to the Excel error. We create two equally sized groups of areas based on whether they – by chance – experienced above median or below median exposure to unusual delays in contract tracing. We plot the average incidence of COVID-19 for each group by test date. We observe virtually identical pre-treatment trends across groups but a substantive divergence in COVID-19 spread at the onset of the period during which the Excel error occurred, which is highlighted by the dashed lines. 90% confidence intervals displayed.

Fig. S5. Impact of delayed referral of COVID-19 positive cases to Test and Trace on subsequent COVID-19 testing performance

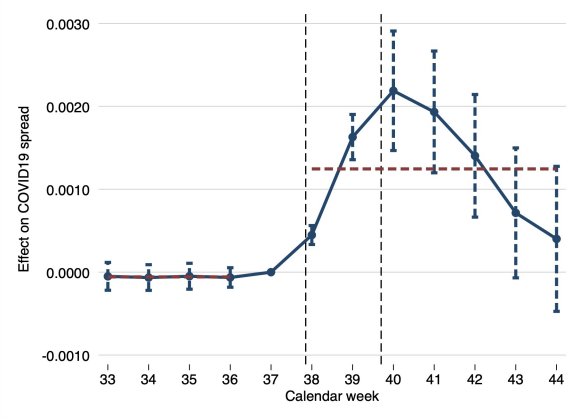
Panel A: Tests per capita



Panel B: Positive Tests per capita



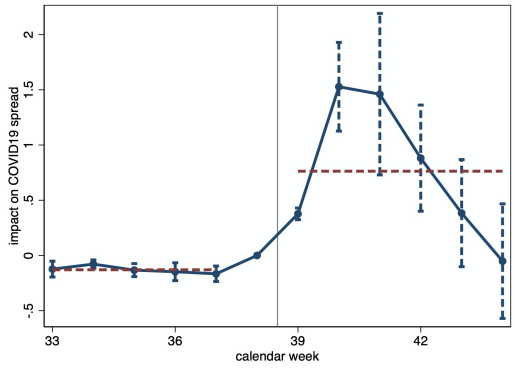
Panel C: Share of positive tests



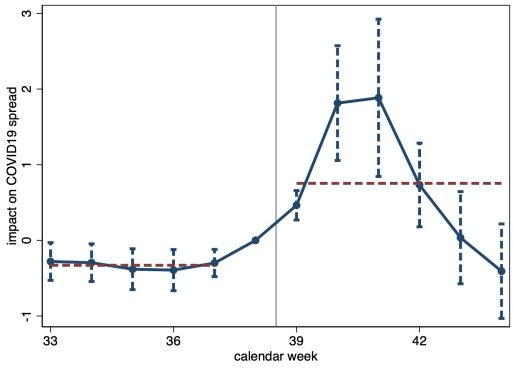
Notes: Figure presents regression estimates capturing the impact of cases that tested positive between Sept 20 to Sept 27 but were not referred to contact tracing until the earliest October 3, 2020 on the outcome variables indicated in the figure panel heads. Note that the subnational Test & Trace statistics are made available lack a lot of detail and reporting is not following conventional calendar week definitions. Rather, a week refers to a time window ranging from Thursday to Wednesday of the subsequent week. That implies that the week 39 label, covering to the period from 24 Sep 2020 to 30 Sep 2020, straddles four days of calendar week 39 and three days of calendar week 40. All regressions control for district fixed effects and date fixed effects, along with non-linear time trends in the extent of true infections measured as of today during calendar weeks 37 and 38. Standard errors are clustered at the district level with 90% confidence intervals shown.

Fig. S6. Impact of delayed referral to contact tracing on Test and Trace

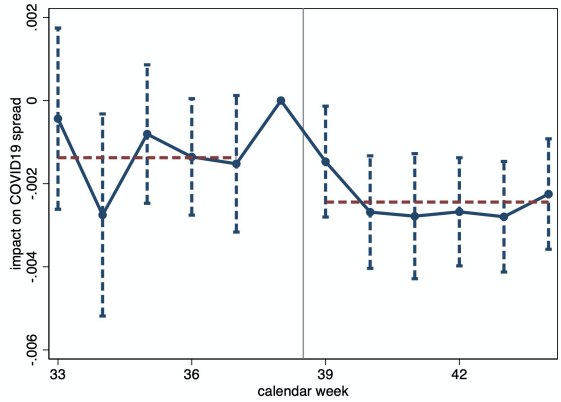
Panel A: Referrals to Test and Trace per capita



Panel B: Number of contacts reached

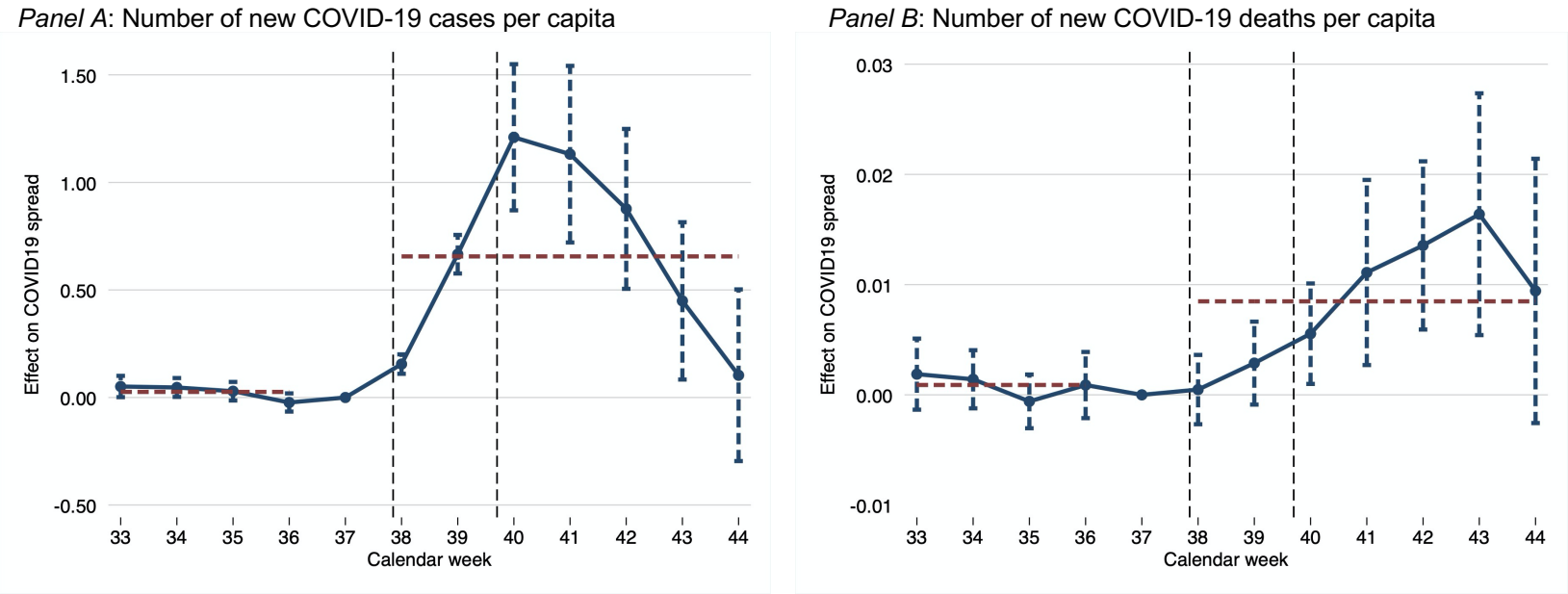


Panel C: Share of contacts reached



Notes: Figure presents regression estimates capturing the impact of cases that tested positive between Sept 20 to Sept 27 but were not referred to contact tracing until the earliest October 3, 2020 on the outcome variables indicated in the figure panel heads. Note that the subnational Test & Trace statistics are made available lack a lot of detail and reporting is not following conventional calendar week definitions. Rather, a week refers to a time window ranging from Thursday to Wednesday of the subsequent week. That implies that the week 39 label, covering to the period from 24 Sep 2020 to 30 Sep 2020, straddles four days of calendar week 39 and three days of calendar week 40. All regressions control for district fixed effects and date fixed effects, along with non-linear time trends in the extent of true infections measured as of today during calendar weeks 37 and 38. Standard errors are clustered at the district level with 90% confidence intervals shown.

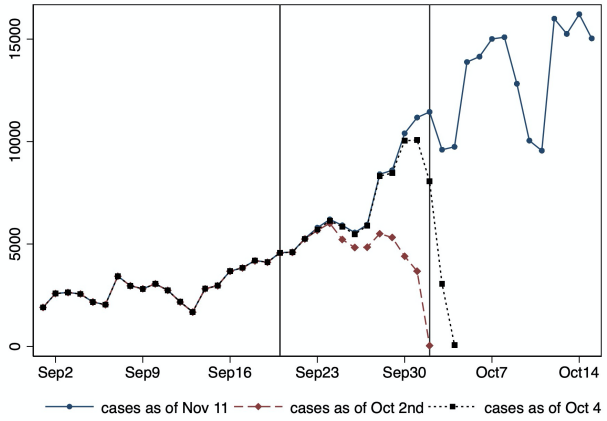
Fig. S7. Impact of delayed referral to contact tracing, new COVID-19 infections and deaths



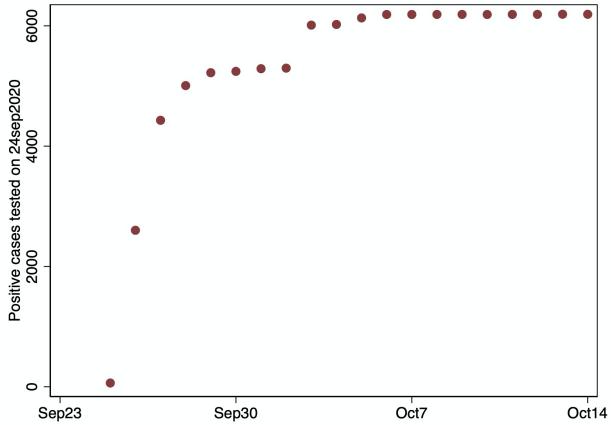
Notes: Figure presents regression estimates capturing the impact of cases that tested positive between Sept 20 to Sept 27 but were not referred to contact tracing until the earliest October 3, 2020 on the outcome variables indicated in the figure panel heads. All regressions control for district fixed effects and date fixed effects, along with non-linear time trends in the extent of true infections measured as of today during calendar weeks 37 and 38. Standard errors are clustered at the district level with 90% confidence intervals shown.

Fig. S8. Delayed contact tracing referral: Identification of delayed referral to contact tracing

Panel A: Identifying missing cases across vintages

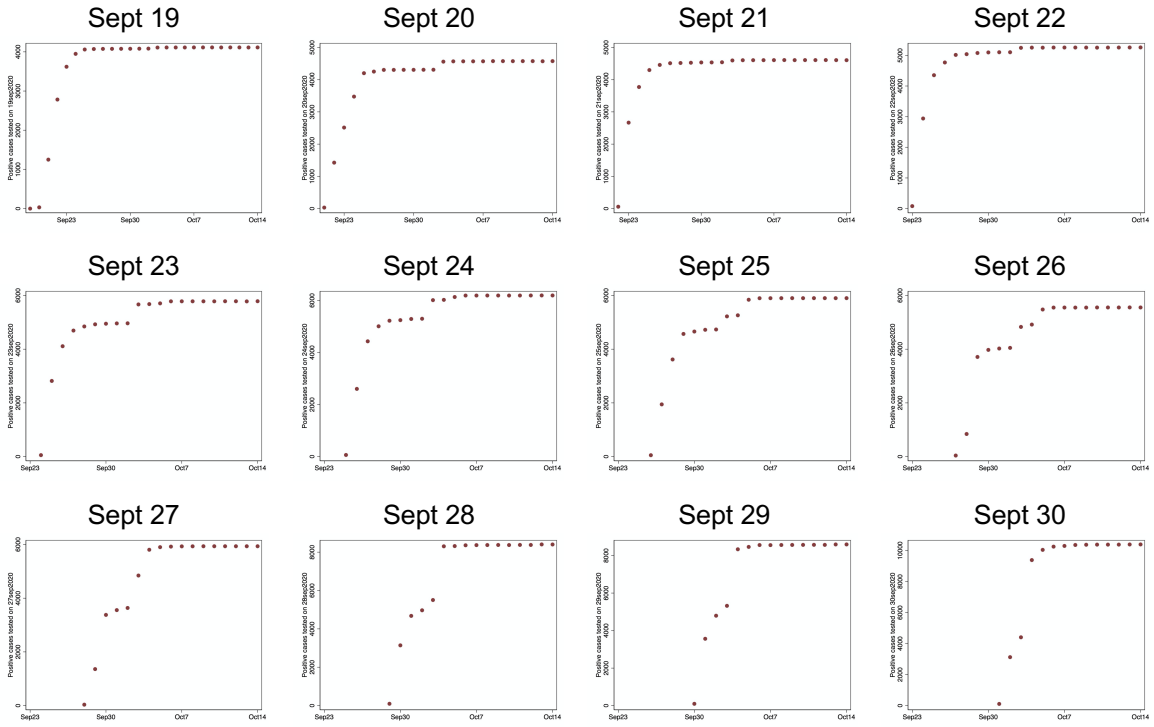


Panel B: Cumulative positive test results for tests taken Sept 24



Notes: Panel A documents the number of cases by date on which a test was taken for three different versions of the dataset: Nov 11, Oct 4 and Oct 2nd. The data for Oct 4 includes a large set of the missing positive cases that were not reported in the Oct 2 data version resulting in large upward revisions. These revisions capture cases that were not referred to contact tracing until Oct 3 or 4th the earliest. Panel B illustrates this using data for all tests taken on Sept 24. Over time the reported cumulative value of positive COVID-19 cases converges to the true value as all test results get processed. Usually, 5 days after a test is taken at least 95% of all test results have been published. Between October 2 and October 3, the case count for Sept 24 jumps by around 715 cases or 12% of all cases due to the Excel glitch.

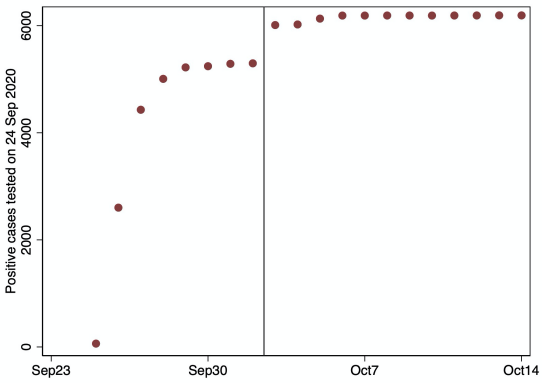
Fig. S9. Delayed referral to contact tracing affecting individual cases with positive test result from Sept 20 - Sept 30



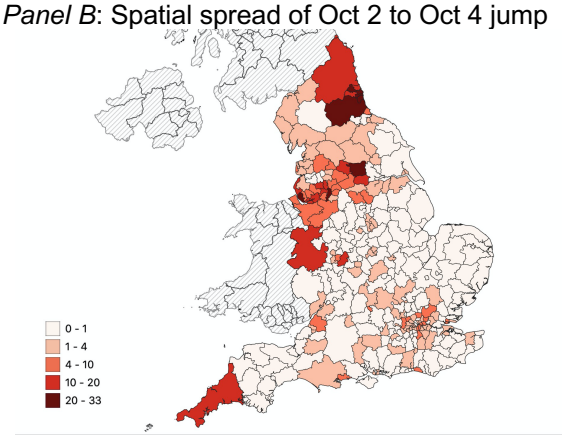
Notes: Figure plots the cumulative number of positive tests on the date indicated in the column head. The vertical axis presents the number of positive cases while the horizontal axis presents the date on which a case count was published. There are notable jumps in the case counts starting Sept 20 due to positive cases not being reported and submitted to contact tracing due to an excel spreadsheet error.

Fig. S10. Geographic signature of Oct 4 upward revision of Sept 24 COVID-19 positive cases across districts

Panel A: Sept 24 cumul. cases across vintages

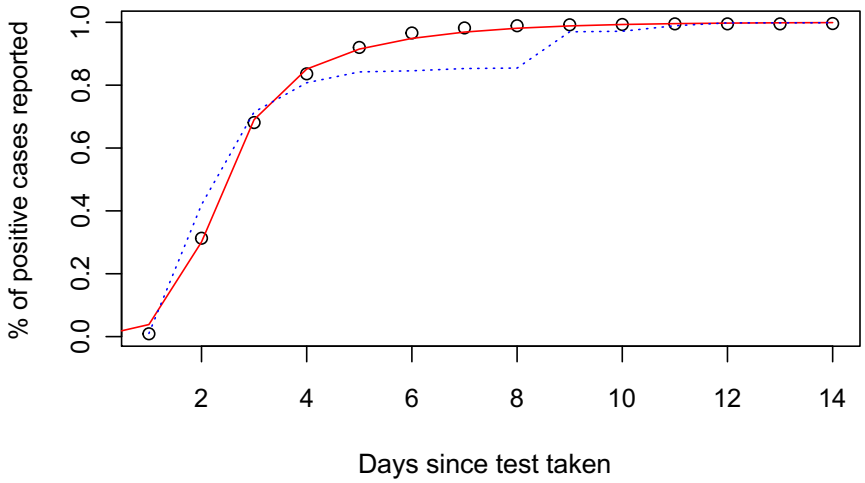


Panel B: Spatial spread of Oct 2 to Oct 4 jump



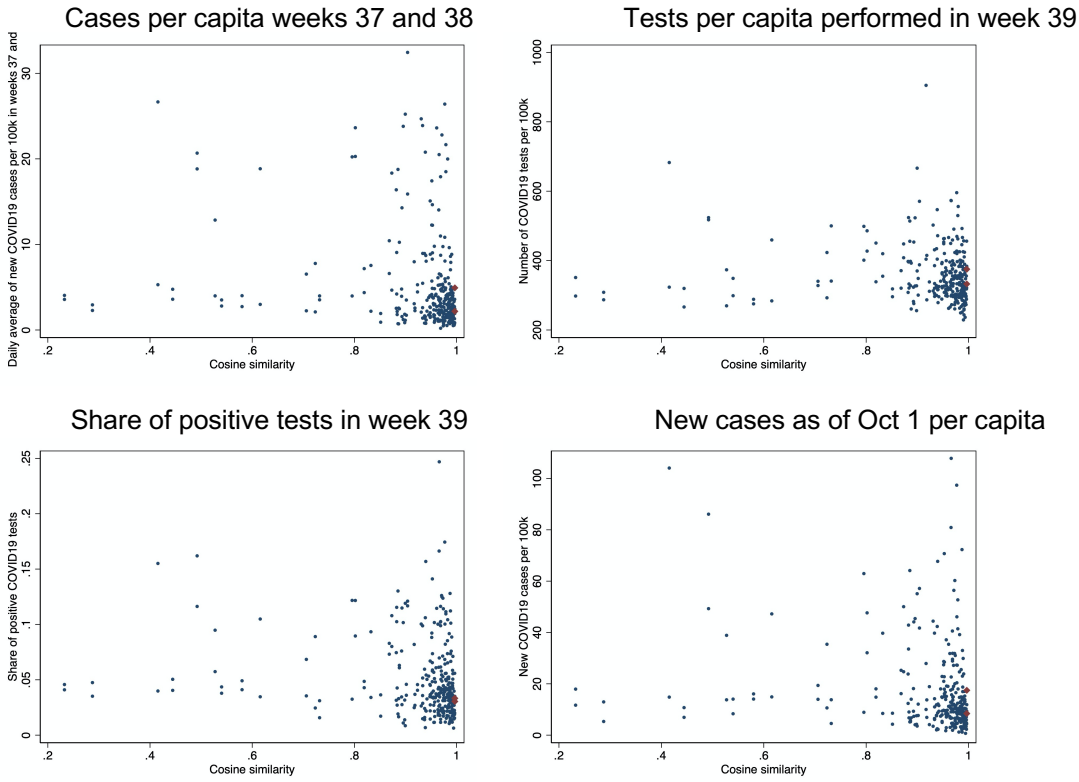
Notes: Figure plots the COVID-19 cumulative case figures as reported for Sept 24 across different reporting dates in Panel A. Panel B presents the spatial distribution of the absolute number of cases that were added between Oct 2 and Oct 4 (for the specimen date Sept 24). The color shades represent different quintiles of the distribution of this absolute number of added cases.

Fig. S11. Fitting evolution of reported cases since test date



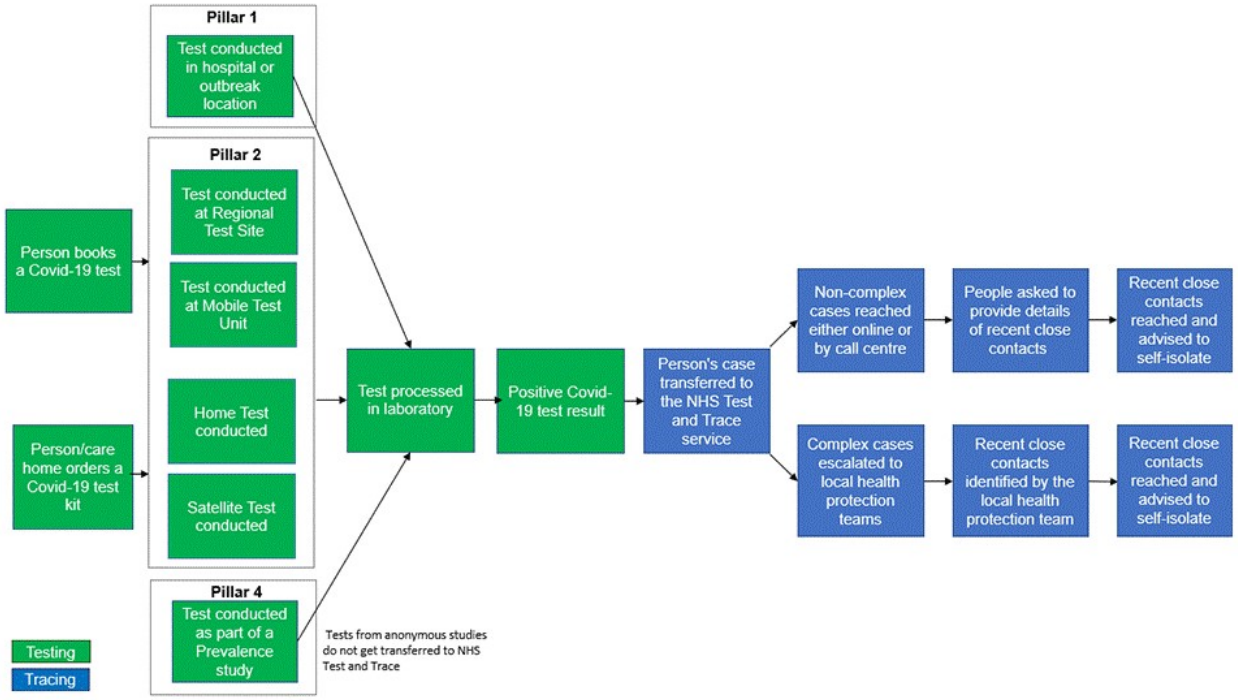
Notes: Figure plots data capturing the share of all positive COVID-19 tests that have been processed, reported and referred to contact tracing as a function of the number of days that have passed since the COVID-19 test was taken on the horizontal axis. The hallow circles refers to the average pattern in the data for Sept 1 to Sept 19. The red line is the one obtained from fitting non-linear least squares of equation 1. The dashed line presents the evolution of the fraction of COVID-19 cases reported and referred to contact tracing for tests taken on Sept 24. The fraction of reported cases jumps nine days after the test was taken which coincides with the upward revision of October 3.

Fig. S12. Example visualization of cosine similarity measure for a pair of districts



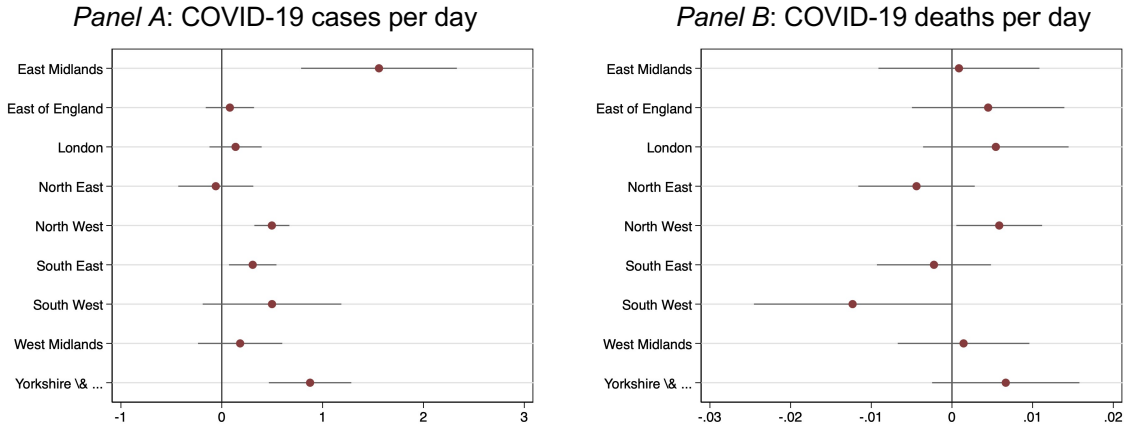
Notes: Figure plots example of the similarity measure used to construct matched pairs. The cosine similarity measure is plotted along the horizontal axis. The vertical axis presents a subset of features that are included in the cosine similarity measure. The matched pairs are indicated as red diamonds representing two districts that are closest in terms of cosine similarity and form a matched pair. Throughout, the two districts are very similar not just in terms of cosine similarity but also in terms of similarity regarding each individual uni-dimensional measure.

Fig. S13. Contact tracing flowchart



Notes: Figure presents the process activating contact tracing as presented on (*18*).

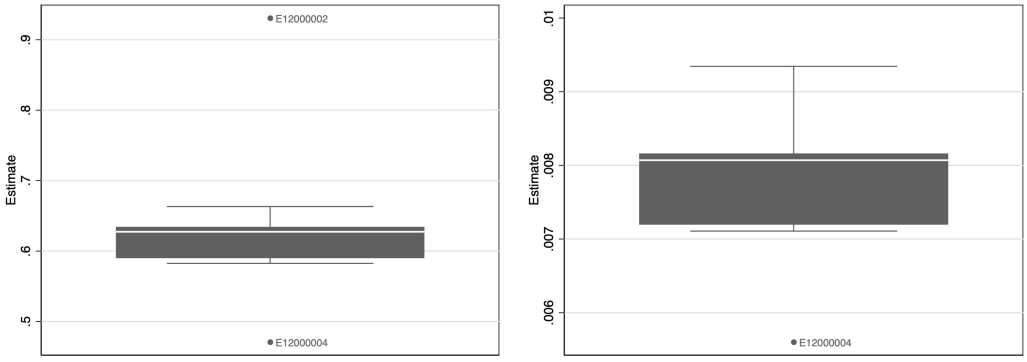
Fig. S14. Regional heterogeneity in impact of delayed contact tracing referral on new cases and deaths



Notes: Figure plots the impact of delayed referrals to test and trace on subsequent new COVID-19 cases (left panel) and new COVID-19 deaths (right panel). All regressions correspond to the specifications presented in column (1) of Table S7, but allowing the effect to be heterogenous across regions. 90% confidence intervals obtained from clustering standard errors at the district level are indicated.

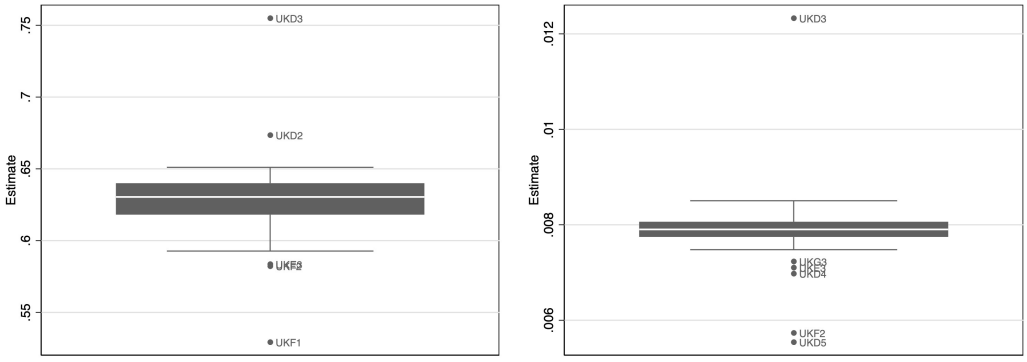
Fig. S15. Distribution of point estimates when dropping one geographical unit at a time (NUTS = Nomenclature of Territorial Units for Statistics)

Panel A: Dropping each of the 9 NUTS1 regions in turn



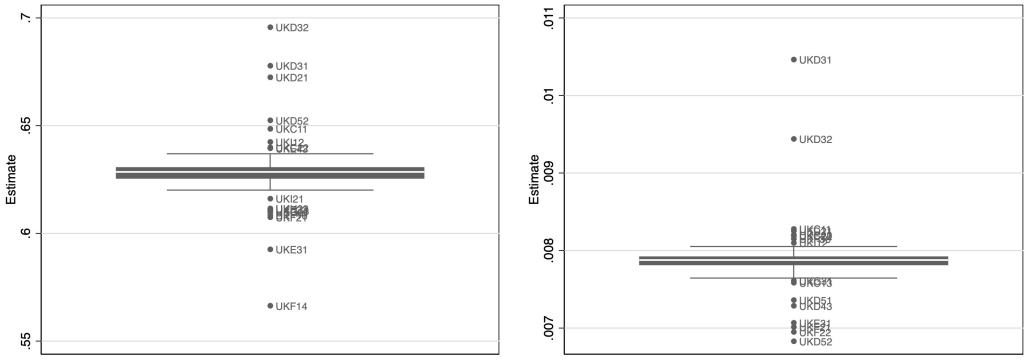
(a) New Cases per capita (b) New Deaths per capita

Panel B: Dropping each of the 30 NUTS2 regions in turn



(c) New Cases per capita (d) New Deaths per capita

Panel C: Dropping each of the 93 NUTS3 regions in turn

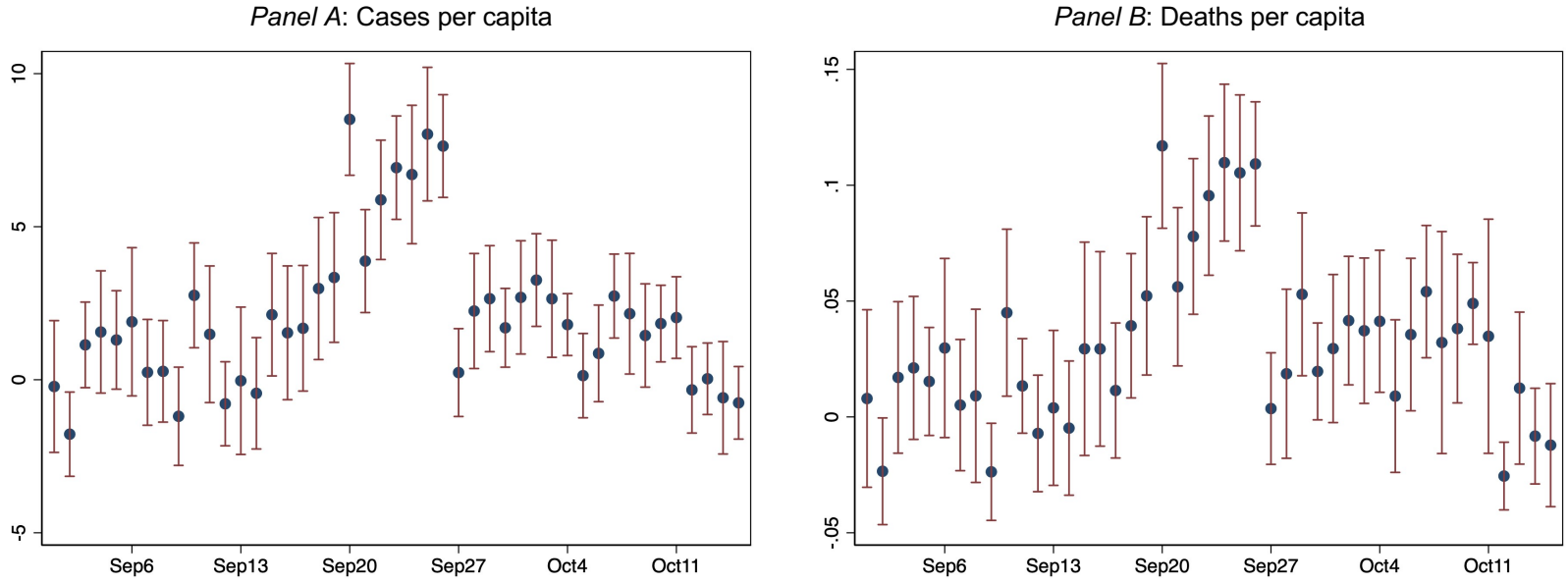


(e) New Cases per capita (f) New Deaths per capita

Notes: Figures present the distribution of the point estimates obtained when dropping all observations pertaining to one region a time. The estimating regression has as dependent variable either the number of new COVID-19 infections or the number of new COVID-19 deaths after week 40 as recorded in the most recent data version. All regressions control for area fixed effects, time

fixed effects and a non-linear time trend in the extent of the local COVID-19 spread measured per capita during weeks 37 and 38. Standard errors are clustered at the district level.

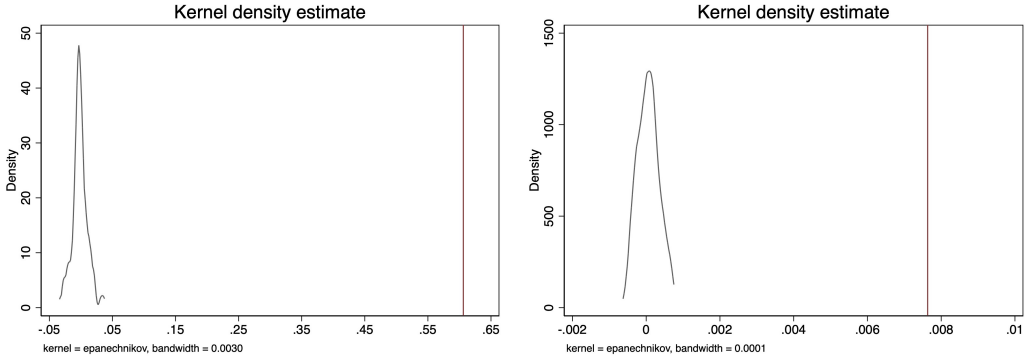
Fig. S16. Estimates of the effects of missing cases on COVID-19 case growth and deaths for the treatment period (Sept 20 to Sept 27) and the placebo periods (before Sept 20 and after Sept 27).



Notes: Figure presents regression estimates of the number of missing cases on subsequent new COVID-19 case growth post calendar week 39 in Panel A and new COVID-19 deaths in Panel B. The number of missing cases on a specific date is computed by measuring, for each date, the difference between the case count reported in the most recent data version from November 15, 2020 and the case count published *seven days after* the actual test was taken. That is, the Sept 23 figures represent the gap in reported cases between the Sept 30 version of the case count and the Nov 10, 2020 version of the case count for Sept 23. This implies missing cases affected by the Excel glitch would appear in all data from Sep 20 to Sep 26 as the Excel error was only starting to be rectified from October 3. The point estimates obtained for the "missing cases measure" for dates before Sept 20 and after Sept 30 serves as a placebo estimate. Standard errors are clustered at the district level with 90% confidence intervals shown.

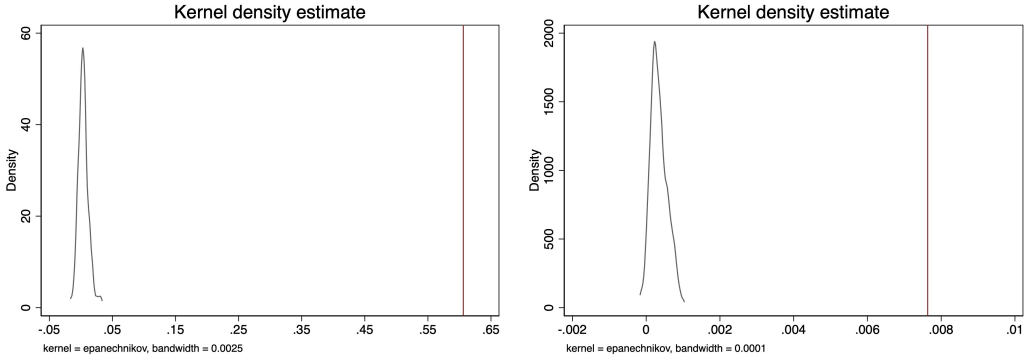
Fig. S17. Randomization inference: reshuffling the district-level exposure measure randomly (NUTS = Nomenclature of Territorial Units for Statistics)

Panel A: Reshuffling M_i across whole of England



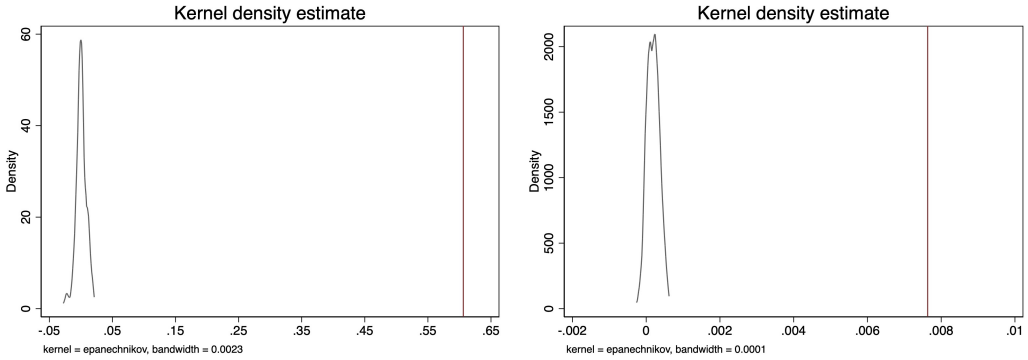
(a) New Cases per capita (b) New Deaths per capita

Panel B: Reshuffling M_i across districts within each of the 9 NUTS1 regions



(c) New Cases per capita (d) New Deaths per capita

Panel C: Reshuffling M_i across districts within each of the 33 NUTS2 regions



(e) New Cases per capita (f) New Deaths per capita

Notes: Figures present the distribution of point estimates obtained from estimating the main difference-in-difference specification in column (1) of Table S7 when using 100 different reshuffled treatment exposure measure. Reshuffling is either across all districts in England in Panel A; across all districts within each of the 9 NUTS1 regions; across all districts within each of the 33 NUTS2

regions. The kernel density plots the distribution of the point estimates. The vertical line indicates the point estimate obtained when using the actual M_i estimate which corresponds to the point estimates presented in column (1) of Table S7.

Table S1. Comparison of usual case count share reported at least five days after a test was taken across different data windows

Published date	Sept 24	Oct 2	Oct 14
Test date	Sept 12 - Sept 19	Sept 20 - Sept 27	Oct 2 - Oct 9
Number of days since test			
5	95.9%	61.0%	94.0%
6	97.1%	72.7%	97.4%
7	98.8%	79.9%	99.2%
8	98.9%	85.3%	99.4%
9	98.9%	85.5%	99.3%
10	99.2%	96.6%	99.6%
11	99.0%	98.3%	99.6%
12	99.3%	94.1%	99.5%

Notes: Table presents the share of all positive tested cases published by the date indicated in the column head covering different time windows up to a specific number of days after the test was conducted. During the period affected by the Excel error, only 61% of all cases occurring between Sept 20 and Sept 27 have been published and referred to contact tracing five days after the test was done. This compares with 95.9% of positive cases immediately prior to the Excel glitch and 94% of cases immediately after the Excel glitch.

Table S2. Robustness of results to additional controls

	New COVID-19 cases			*New COVID-19 deaths*		
	(1)	(2)	(3)	(4)	(5)	(5)
Post week 39 × Missing cases 20-27 Sept per capita	0.606***	0.419***	0.395***	0.007**	0.006**	0.005**
	(0.119)	(0.108)	(0.102)	(0.003)	(0.003)	(0.002)
Mean DV	11.906	11.906	11.991	0.185	0.185	0.185
Observations	28119	28119	27755	28119	28119	27755
Spatial units	309	309	305	309	309	305
Additional controls	2730	7462	7462	2730	7462	7462
Non-linear time trends in COVID-19 intensity weeks 36-38	X	X	X	X	X	X
Non-linear time trends in vector of area controls		X	X		X	X
Google Daily Mobility Scores			X			X

Notes: Impact of delays in contact tracing on new COVID-19 infections and deaths. Difference-in-differences regression estimates (at level of Lower Tier Local Authority) for the effect of the number of delayed referrals to contact tracing per capita on new infections per capita (columns 1-3) and new COVID-19 deaths per capita (columns 3-6). All regressions control for district fixed effects and date fixed effects. Columns 1 and 4 control for non-linear time trends in cases per capita just prior to the error. Columns 2 and 5 additionally control for non-linear time trends in a vector of 55 area-specific characteristic. Columns 3 and 6 further control for mobility data that accounts for the potential behavioral effect of lower locally reported case numbers. Standard errors are clustered at the district level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S3. Impact of non-timely contact tracing on weekly COVID-19 test data

	plain difference-in-difference			*matched pairs*		
	(1)	(2)	(3)	(4)	(5)	(5)
Panel A: Weekly Tests Per Capita						
Post week 39 × Missing cases 20-27 Sept per capita	2.665***	2.894***	2.746***	2.459***	2.792***	2.656***
	(0.512)	(0.501)	(0.516)	(0.767)	(0.796)	(0.860)
Mean DV	304.007	304.007	304.070	304.322	304.322	304.387
Observations	28301	28301	28119	28210	28210	28028
Spatial units	311	311	309	310	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Panel B: Weekly Positive Tests Per Capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.674***	0.685***	0.676***	0.792***	0.760***	0.790***
	(0.122)	(0.122)	(0.130)	(0.166)	(0.168)	(0.173)
Mean DV	13.468	13.468	13.496	13.475	13.475	13.502
Observations	27762	27762	27664	27671	27671	27573
Spatial units	311	311	309	310	310	308
Additional controls	1001	1911	2821	15274	16184	17094
Panel C: Share of tests returning a positive result						
Post week 39 × Missing cases 20-27 Sept per capita	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean DV	0.036	0.036	0.036	0.036	0.036	0.036
Observations	27762	27762	27664	27671	27671	27573
Spatial units	311	311	309	310	310	308
Additional controls	1001	1911	2821	15274	16184	17094
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) present the main difference-in-differences results. Columns (4) - (6) control for matched pair by time fixed effects. Matched pairs are constructed by identifying for each district one that is closest in terms of the Cosine distance between the following variables: new COVID-19 cases as of Oct 1, new COVID-19 deaths on Oct 1, the positive test rate in week 39, the number of tests per capita in week 39, the average new cases and deaths per capita, the average positive test share, the average number of tests performed during weeks 36-38, along with the growth rates in new cases, new deaths, positive test rate and tests performed in between weeks 36-38. Standard errors are clustered at the district level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S4. Impact of non-timely contact tracing on the weekly performance of contact tracing

	plain difference-in-difference			*matched pairs*		
	(1)	(2)	(3)	(4)	(5)	(5)
Panel A: Referrals to Test and Trace Per Capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.801***	0.790***	0.748***	0.275	0.287	0.358*
	(0.171)	(0.170)	(0.169)	(0.253)	(0.292)	(0.212)
Mean DV	14.924	14.977	15.102	14.964	15.017	15.144
Observations	14378	14287	14105	14287	14196	14014
Spatial units	149	148	146	148	147	145
Additional controls	1001	1911	2821	7735	8645	9555
Panel B: Contacts Reached Per Capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.897***	0.880***	0.836***	0.312	0.346	0.379
	(0.232)	(0.232)	(0.234)	(0.330)	(0.387)	(0.312)
Mean DV	20.232	20.295	20.442	20.270	20.333	20.482
Observations	14378	14287	14105	14287	14196	14014
Spatial units	149	148	146	148	147	145
Additional controls	1001	1911	2821	7735	8645	9555
Panel C: Share of contacts reached						
Post week 39 × Missing cases 20-27 Sept per capita	-0.000	-0.000	-0.001	-0.002*	-0.001	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean DV	0.633	0.633	0.633	0.633	0.633	0.633
Observations	14378	14287	14105	14014	13832	13468
Spatial units	149	148	146	145	143	139
Additional controls	1001	1911	2821	7735	8554	9282
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) present the main difference-in-differences results. Columns (4) - (6) control for matched pair by time fixed effects. Matched pairs are constructed by identifying for each district one that is closest in terms of the Cosine distance between the following variables: new COVID-19 cases as of Oct 1, new COVID-19 deaths on Oct 1, the positive test rate in week 39, the number of tests per capita in week 39, the average new cases and deaths per capita, the average positive test share, the average number of tests performed during weeks 36-38, along with the growth rates in new cases, new deaths, positive test rate and tests performed in between weeks 36-38. Standard errors are clustered at the district level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S5. Impact of non-timely contact tracing on the pandemic progression

	plain difference-in-difference			*matched pairs*		
	(1)	(2)	(3)	(4)	(5)	(5)
Panel A: Daily new COVID-19 cases per capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.606***	0.619***	0.606***	0.725***	0.692***	0.713***
	(0.111)	(0.111)	(0.119)	(0.146)	(0.148)	(0.156)
Mean DV	11.790	11.848	11.906	11.796	11.854	11.912
Observations	28665	28301	28119	28574	28210	28028
Spatial units	315	311	309	314	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Panel B: Daily new COVID-19 deaths per capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.008***	0.008***	0.007**	0.007***	0.007***	0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean DV	0.183	0.184	0.185	0.184	0.184	0.185
Observations	28665	28301	28119	28574	28210	28028
Spatial units	315	311	309	314	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) present the main difference-in-differences results. Columns (4) - (6) control for matched pair by time fixed effects. Matched pairs are constructed by identifying for each district one that is closest in terms of the Cosine distance between the following variables: new COVID-19 cases as of Oct 1, new COVID-19 deaths on Oct 1, the positive test rate in week 39, the number of tests per capita in week 39, the average new cases and deaths per capita, the average positive test share, the average number of tests performed during weeks 36-38, along with the growth rates in new cases, new deaths, positive test rate and tests performed in between weeks 36-38. Standard errors are clustered at the district level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S6. Measuring the number of missing cases across data set vintages

Date	PHE	*Cases by test date and vintage*			*Δ Missing case measure*		
		Nov 11	Oct 2	Oct 4	Oct 4 - Oct 2	Nov 11 - Oct 2	Nov 11 - Oct 4
20 September 2020	-	4578	4308	4569	261	270	9
21 September 2020	-	4609	4540	4604	64	69	5
22 September 2020	-	5269	5102	5251	149	167	18
23 September 2020	-	5795	4962	5704	742	833	91
24 September 2020	957	6199	5297	6132	835	902	67
25 September 2020	744	5912	4732	5840	1108	1180	72
26 September 2020	757	5560	4048	5476	1428	1512	84
27 September 2020	-	5944	3635	5898	2263	2309	46
28 September 2020	1415	8405	4969	8317	3348	3436	88
29 September 2020	3049	8598	4793	8461	3668	3805	137
30 September 2020	4133	10400	3123	10040	6917	7277	360
01 October 2020	4786	11172	51	10073	10022	11121	1099

Notes: Table illustrates how the number of missing cases is identified contrasting different versions of case data published on the official English Coronavirus dashboard. We focus on our shock measure computing the missing cases between Sept 20 and Sept 26. The bulk of the increase in cases for tests taken between Sept 20 - Sept 27 is corrected by the data update between Oct 4 to Oct 2. Revisions of figures for Sept 20 - Sept 27 are marginal after Oct 4th.

Table S7. Impact of non-timely contact tracing on the pandemic progression

	plain difference-in-difference			*matched pairs*		
	(1)	(2)	(3)	(4)	(5)	(5)
Panel A: Daily new COVID-19 cases per capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.606***	0.619***	0.606***	0.725***	0.692***	0.713***
	(0.111)	(0.111)	(0.119)	(0.146)	(0.148)	(0.156)
Mean DV	11.790	11.848	11.906	11.796	11.854	11.912
Observations	28665	28301	28119	28574	28210	28028
Spatial units	315	311	309	314	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Panel B: Daily new COVID-19 deaths per capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.008***	0.008***	0.007**	0.007***	0.007***	0.006**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Mean DV	0.183	0.184	0.185	0.184	0.184	0.185
Observations	28665	28301	28119	28574	28210	28028
Spatial units	315	311	309	314	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) present the main difference-in-differences results. Columns (4) - (6) control for matched pair by time fixed effects. Matched pairs are constructed by identifying for each district one that is closest in terms of the Cosine distance between the following variables: new COVID-19 cases as of Oct 1, new COVID-19 deaths on Oct 1, the positive test rate in week 39, the number of tests per capita in week 39, the average new cases and deaths per capita, the average positive test share, the average number of tests performed during weeks 36-38, along with the growth rates in new cases, new deaths, positive test rate and tests performed in between weeks 36-38. Standard errors are clustered at the district level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S8. Quantification of impact of delayed or missing contact tracing on pandemic spread between calendar weeks 39 to 44

Panel A: Cases		**Implied estimates**						**Ranges**			
Region	Cases	Col (1) Table S7			Col (3) Table S2			Levels		% of all cases	
		Estimate	90% CI		Estimate	90% CI		Low	High	Low	High
North East	44878	16573	11559	21587	11404	6828	15979	6828	21587	15%	48%
North West	161232	66903	46661	87145	46036	27565	64506	27565	87145	17%	54%
Yorkshire & ...	101873	27164	18945	35382	18691	11192	26190	11192	35382	11%	35%
East Midlands	59919	12703	8860	16547	8741	5234	12248	5234	16547	9%	28%
West Midlands	63546	17922	12500	23345	12332	7384	17280	7384	23345	12%	37%
East of England	29047	6339	4421	8257	4362	2612	6112	2612	8257	9%	28%
London	61371	19017	13263	24771	13085	7835	18336	7835	24771	13%	40%
South East	43307	10845	7564	14126	7462	4468	10456	4468	14126	10%	33%
South West	32208	7128	4972	9285	4905	2937	6873	2937	9285	9%	29%
England	597381	184595	128745	240445	127018	76056	177981	76056	240445	13%	40%

Panel B: Deaths		**Implied estimates**						**Ranges**			
Region	Deaths	Col (4) Table S7			Col (6) Table S2			Levels		% of all deaths	
		Estimate	90% CI		Estimate	90% CI		Low	High	Low	High
North East	694	209	86	331	136	26	246	26	331	4%	48%
North West	2195	842	348	1337	550	107	992	107	1337	5%	61%
Yorkshire & ...	1085	342	141	543	223	43	403	43	543	4%	50%
East Midlands	653	160	66	254	104	20	188	20	254	3%	39%
West Midlands	747	226	93	358	147	29	266	29	358	4%	48%
East of England	474	80	33	127	52	10	94	10	127	2%	27%
London	529	239	99	380	156	30	282	30	380	6%	72%
South East	547	137	56	217	89	17	161	17	217	3%	40%
South West	272	90	37	142	59	11	106	11	142	4%	52%
England	7196	2324	960	3689	1516	295	2738	295	3689	4%	51%

Notes: Table provides a quantification exercise of the implied effects of the delayed contact tracing of COVID19 positive cases on subsequent infections and deaths across English regions from calendar week 39 to 44 inclusive. Cases and Deaths refers to the cumulative total of new COVID19 infections and deaths since week 39 up to week 44 inclusive. The subsequent columns provide the estimate of the number of cases and deaths that appear econometrically linked to the cases that have not been referred to contact tracing. The table provides on the figures implied by the central point estimate as well as the most conservative estimate. It further provides ranges associated with 90% confidence intervals for the individual point estimates. The column head makes a reference to the specific point estimates leveraged.

Table S9. Robustness of impact of non-timely contact tracing on the pandemic progression: Analysis at the Upper Tier Local Authority level

	plain difference-in-difference			*matched pairs*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: New Cases						
Post week 39 × Missing cases 20-27 Sept per capita	0.619***	0.621***	0.635***	0.680***	0.688***	0.776***
	(0.160)	(0.157)	(0.161)	(0.168)	(0.176)	(0.202)
Mean DV	14.069	14.115	14.228	14.054	14.043	14.083
Observations	14378	14287	14105	14014	13832	13468
Spatial units	149	148	146	145	143	139
Additional controls	1001	1911	2821	7735	8554	9282
Panel B: Deaths						
Post week 39 × Missing cases 20-27 Sept per capita	0.010***	0.009**	0.009**	0.011***	0.011***	0.014***
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Mean DV	0.202	0.202	0.204	0.201	0.201	0.202
Observations	14378	14287	14105	14014	13832	13468
Spatial units	149	148	146	145	143	139
Additional controls	1001	1911	2821	7735	8554	9282
Panel C: Tests						
Post week 39 × Missing cases 20-27 Sept per capita	2.934***	2.842***	2.878***	3.599***	3.441***	4.166***
	(0.574)	(0.602)	(0.652)	(0.678)	(0.665)	(0.894)
Mean DV	311.369	311.369	311.714	311.617	311.617	311.740
Observations	14287	14287	14105	13832	13832	13468
Spatial units	148	148	146	143	143	139
Additional controls	1001	1911	2821	7644	8554	9282
Panel D: Positivity rate						
Post week 39 × Missing cases 20-27 Sept per capita	0.001***	0.001***	0.001***	0.001***	0.001***	0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean DV	0.044	0.044	0.044	0.045	0.045	0.045
Observations	13860	13860	13769	13118	13118	12936
Spatial units	148	148	146	143	143	139
Additional controls	1001	1911	2821	7322	8232	9051
Panel E: Contact tracing referral						
Post week 39 × Missing cases 20-27 Sept per capita	-0.000	-0.000	-0.001	-0.002*	-0.001	-0.002*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean DV	0.633	0.633	0.633	0.633	0.633	0.633
Observations	14378	14287	14105	14014	13832	13468
Spatial units	149	148	146	145	143	139
Additional controls	1001	1911	2821	7735	8554	9282
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) present the main difference-in-differences results. Columns (4) - (6) control for matched pair by time fixed effects. Matched pairs are constructed by identifying for each district one that is closest

in terms of the Cosine distance between the following variables: new COVID-19 cases as of Oct 1, new COVID-19 deaths on Oct 1, the positive test rate in week 39, the number of tests per capita in week 39, the average new cases and deaths per capita, the average positive test share, the average number of tests performed during weeks 36-38, along with the growth rates in new cases, new deaths, positive test rate and tests performed in between weeks 36-38. Standard errors are clustered at the UTLA level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S10. Robustness of impact of non-timely contact tracing on the pandemic progression: Analysis at the NUTS3 level (Nomenclature of Territorial Units for Statistics)

	plain difference-in-difference			*matched pairs*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: New Cases						
post × Missing cases 20-26 Sept per capita	1.210***	1.275***	1.039***	0.899**	0.894**	0.704*
	(0.256)	(0.293)	(0.276)	(0.393)	(0.414)	(0.403)
Mean DV	11.051	11.113	11.267	10.986	10.959	11.253
Observations	11067	10948	10710	10948	10710	10234
Spatial units	93	92	90	92	90	86
Additional controls	1309	2499	3689	6783	7854	8806
Panel B: Deaths						
post × Missing cases 20-26 Sept per capita	0.025***	0.023***	0.024***	0.022***	0.021***	0.024***
	(0.006)	(0.007)	(0.007)	(0.006)	(0.007)	(0.008)
Mean DV	0.186	0.187	0.190	0.184	0.184	0.189
Observations	11067	10948	10710	10948	10710	10234
Spatial units	93	92	90	92	90	86
Additional controls	1309	2499	3689	6783	7854	8806
Panel C: Tests						
post × Missing cases 20-26 Sept per capita	6.258***	6.360***	6.183***	5.948***	5.336**	5.292*
	(1.577)	(1.828)	(1.918)	(2.077)	(2.193)	(2.879)
Mean DV	281.180	281.180	281.833	280.602	280.602	282.266
Observations	10948	10948	10710	10710	10710	10234
Spatial units	92	92	90	90	90	86
Additional controls	1309	2499	3689	6664	7854	8806
Panel D: Positivity rate						
post × Missing cases 20-26 Sept per capita	0.002***	0.002***	0.001***	0.001*	0.001*	0.001
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Mean DV	0.036	0.036	0.036	0.038	0.038	0.038
Observations	10017	10017	9926	8946	8946	8764
Spatial units	92	92	90	90	90	86
Additional controls	1309	2499	3689	5740	6930	8029
Panel E: Contact tracing referral post × Missing cases 20-26 Sept per capita	-0.002*	-0.002*	-0.002*	-0.003***	-0.003**	-0.003*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
Mean DV	0.634	0.634	0.633	0.633	0.634	0.634
Observations	8463	8372	8190	8372	8190	7826
Spatial units	93	92	90	92	90	86
Additional controls	1001	1911	2821	5187	6006	6734
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) present the main difference-in-differences results. Columns (4) - (6) control for matched pair by time fixed effects. Matched pairs are constructed by identifying for each district one that is closest

in terms of the Cosine distance between the following variables: new COVID19 cases as of Oct 1, new COVID-19 deaths on Oct 1, the positive test rate in week 39, the number of tests per capita in week 39, the average new cases and deaths per capita, the average positive test share, the average number of tests performed during weeks 36-38, along with the growth rates in new cases, new deaths, positive test rate and tests performed in between weeks 36-38. Standard errors are clustered at the NUTS3 level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S11. Robustness of results to alternative functional forms

	New COVID-19 cases			*New COVID-19 deaths*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main specification						
Post week 39 × Missing cases 20-27 Sept per capita	0.606***	0.619***	0.606***	0.008***	0.008***	0.007**
	(0.111)	(0.111)	(0.119)	(0.003)	(0.003)	(0.003)
Mean DV	11.790	11.848	11.906	0.183	0.184	0.185
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Panel B: Log-Log per capita						
Post week 39 × log(Missing cases 20-27 Sept +1 per capita)	0.239***	0.248***	0.224***	0.045**	0.050**	0.045**
	(0.045)	(0.045)	(0.048)	(0.021)	(0.021)	(0.023)
Mean DV	1.746	1.748	1.753	-0.228	-0.229	-0.233
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Panel C: Levels-Levels						
Post week 39 × Missing cases 20-27 Sept	0.977***	0.967***	0.970***	0.009***	0.009***	0.009***
	(0.139)	(0.131)	(0.135)	(0.003)	(0.003)	(0.003)
Mean DV	24.169	24.361	24.505	0.346	0.348	0.350
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Panel D: Log-Levels per capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.020***	0.022***	0.022***	0.006**	0.007**	0.007**
	(0.004)	(0.004)	(0.004)	(0.003)	(0.003)	(0.003)
Mean DV	1.746	1.748	1.753	-0.228	-0.229	-0.233
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Panel E: Log-Log						
Post week 39 × log(Missing cases 20-27 Sept +1)	0.220***	0.228***	0.222***	0.097***	0.101***	0.104***
	(0.029)	(0.029)	(0.031)	(0.015)	(0.016)	(0.017)
Mean DV	2.167	2.172	2.181	0.193	0.194	0.195
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) use new daily COVID-19 cases as dependent variable while columns (4) - (6) explore new daily COVID-

19 deaths as dependent variable. Standard errors are clustered at the district level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S12. Robustness of results to alternative treatment exposure measures

	New COVID-19 cases			*New COVID-19 deaths*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Main specification						
Post week 39 × Missing cases 20-27 Sept per capita	0.606***	0.619***	0.606***	0.008***	0.008***	0.007**
	(0.111)	(0.111)	(0.119)	(0.003)	(0.003)	(0.003)
Mean DV	11.790	11.848	11.906	0.183	0.184	0.185
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Panel B: Curve fit Sep 20 - Sep 27						
Post week 39 × Missing cases Sept/Oct per capita	0.654***	0.664***	0.641***	0.008***	0.009***	0.008**
	(0.125)	(0.126)	(0.133)	(0.003)	(0.003)	(0.003)
Mean DV	11.790	11.848	11.906	0.183	0.184	0.185
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Panel B: Curve fit Sep 20 - Oct 1						
Post week 39 × Missing cases Sept/Oct per capita	0.442***	0.452***	0.430***	0.005***	0.005***	0.005**
	(0.077)	(0.076)	(0.076)	(0.002)	(0.002)	(0.002)
Mean DV	11.790	11.848	11.906	0.183	0.184	0.185
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Panel D: Fraction Missing cases Sept 20-27						
Post week 39 × fracmissed2027sh	12.705**	12.232*	11.369*	0.037	0.053	0.015
	(5.650)	(6.246)	(6.656)	(0.098)	(0.104)	(0.111)
Mean DV	11.790	11.848	11.906	0.183	0.184	0.185
Observations	28665	28301	28119	28665	28301	28119
Spatial units	315	311	309	315	311	309
Additional controls	1001	1911	2821	1001	1911	2821
Panel E: Missing cases Sept 20-27 high incidence places						
Post week 39 × fracmissed2027sh cens2	39.734***	41.164***	45.858***	0.245	0.311	0.324
	(13.521)	(14.468)	(15.262)	(0.216)	(0.223)	(0.241)
Mean DV	14.130	14.212	14.212	0.210	0.211	0.211
Observations	21385	21112	21112	21385	21112	21112
Spatial units	235	232	232	235	232	232
Additional controls	1001	1911	2821	1001	1911	2821
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) use new daily COVID-19 cases as dependent variable while columns (4) - (6) explore new daily COVID-19 deaths as dependent variable. Standard errors are clustered at the district level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S13. Impact of non-timely contact tracing on the weekly death statistics as reported by the Office of National Statistics by place of death

	plain difference-in-difference			*matched pairs*		
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Weekly COVID-19 Deaths Per Capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.006***	0.007***	0.006***	0.005**	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mean DV	0.080	0.081	0.081	0.080	0.081	0.081
Observations	28665	28301	28119	28574	28210	28028
Spatial units	315	311	309	314	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Panel B: Weekly COVID-19 Deaths Recorded in Hospitals Per Capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.006***	0.006***	0.005***	0.006***	0.006***	0.006***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Mean DV	0.060	0.060	0.060	0.060	0.060	0.060
Observations	28301	28301	28119	28210	28210	28028
Spatial units	311	311	309	310	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Panel C: Weekly COVID-19 Deaths Recorded in Care Homes Per Capita						
Post week 39 × Missing cases 20-27 Sept per capita	0.000	0.001	0.001	-0.001	-0.001	-0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Mean DV	0.014	0.014	0.014	0.014	0.014	0.015
Observations	28301	28301	28119	28210	28210	28028
Spatial units	311	311	309	310	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Panel D: Weekly COVID-19 Deaths Recorded in Other Communal Establishments Per Capita						
Post week 39 × Missing cases 20-27 Sept per capita	-0.000	-0.000	-0.000	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Mean DV	0.000	0.000	0.000	0.000	0.000	0.000
Observations	28301	28301	28119	28210	28210	28028
Spatial units	311	311	309	310	310	308
Additional controls	1001	1911	2821	15288	16198	17108
Non-linear time trends in COVID-19 intensity weeks 36-38						
Cases per capita deciles x Date	X	X	X	X	X	X
Tests per capita deciles x Date		X	X		X	X
Positive test rate deciles x Date			X			X

Notes: All regressions control for district fixed effects and date fixed effects. Columns (1) - (3) present the main difference-in-differences results. Columns (4) - (6) control for matched pair by time fixed effects. Matched pairs are constructed by identifying for each district one that is closest in terms of the Cosine distance between the following variables: new COVID-19 cases as of Oct 1, new COVID-19 deaths on Oct 1, the positive test rate in week 39, the number of tests per capita in week 39, the average new cases and deaths per capita, the average positive test share, the average number of tests performed during weeks 36-38, along with the growth rates in new cases, new deaths, positive test rate and tests performed in between weeks 36-38. Standard errors are clustered at the district level with starts indicating *** p< 0.01, ** p< 0.05, * p< 0.1.

Table S14. Covariate Balance: Correlation between area characteristics and the exposure to the Excel error

	missing cases per capita			*% of missed cases*	
	main	curve fit 20-27th	curve fit 20-1st	all	high
	(1)	(2)	(3)	(4)	(5)
Panel A: Employment makeup (2011 census) % working in Agriculture, ...	13.921	13.609	51.797*	-0.163	-0.602
	(18.626)	(16.769)	(29.517)	(0.517)	(0.495)
% working in Mining, ...	107.501	83.134	382.062**	-0.123	-0.702
	(142.681)	(123.317)	(181.288)	(1.553)	(1.418)
% working in Manufacturing	-11.830	-9.082	-25.753	-0.056	-0.071
	(12.449)	(10.442)	(16.466)	(0.125)	(0.096)
% working in Utilities	-62.517	-43.732	-55.491	0.119	-0.157
	(79.738)	(69.080)	(129.500)	(0.888)	(0.702)
% working in Construction	-12.668	-6.335	-43.116	0.037	-0.057
	(28.884)	(24.644)	(35.701)	(0.328)	(0.275)
% working in Wholesale, Retail Trade & Hospitality accommodation	-0.375	-1.383	-2.602	-0.137	-0.061
	(10.762)	(9.449)	(13.803)	(0.192)	(0.191)
% working in Transportation, storage	-13.804	-11.244	-53.799**	-0.293	0.047
	(14.424)	(12.551)	(21.473)	(0.241)	(0.194)
% working in Information, Communication	-4.920	-6.519	-11.191	-0.003	-0.167
	(13.198)	(11.263)	(19.243)	(0.253)	(0.229)
% working in Finance, Insurance	-17.184**	-13.779*	-12.996	0.031	0.003
	(8.524)	(7.672)	(13.034)	(0.181)	(0.150)
% working in Admin and support services	40.455	29.162	-20.569	0.839	0.554
	(56.994)	(48.635)	(77.872)	(0.610)	(0.618)
% working in Public administration	24.292*	21.442*	33.895	0.152	0.176
	(14.260)	(12.242)	(26.334)	(0.142)	(0.138)
% working in Education	7.087	3.663	17.968	-0.117	0.105
	(22.319)	(19.996)	(28.502)	(0.297)	(0.189)
% working in Human Health and other services	18.966	14.577	49.964***	0.166	0.139
	(12.830)	(11.003)	(16.222)	(0.156)	(0.131)
Panel B: Educational attainment of resident population (2011 census) % with No formal qualifications	2.606	3.094	-0.512	-0.069	0.013
	(9.881)	(8.485)	(12.821)	(0.121)	(0.086)
% with Level 1 qualifications	-19.116	-11.724	-48.234*	-0.102	-0.053
	(20.765)	(17.800)	(24.473)	(0.237)	(0.166)
% with Level 2qualifications	-5.466	1.528	-31.952	0.078	-0.031
	(23.940)	(20.350)	(30.308)	(0.208)	(0.171)

% with Level 3+ qualifications	1.243	1.156	5.123	0.059	-0.024
	(5.280)	(4.688)	(6.653)	(0.086)	(0.060)
% School children and full-time students	-132.340	-77.938	-388.788*	0.569	-0.517
	(138.281)	(119.182)	(211.005)	(0.968)	(1.105)
% Full Time Students	71.705	39.993	127.459	0.001	0.524
	(95.119)	(80.542)	(126.980)	(0.729)	(0.630)
Panel C: National Statistics Socio-economic classification of occupation (2011 census)					
% Higher Managerial	-6.171	-4.730	-6.055	0.029	-0.088
	(9.000)	(8.016)	(11.008)	(0.151)	(0.112)
% Lower Managerial	-4.522	-1.924	-5.839	0.145	-0.018
	(10.273)	(9.096)	(13.939)	(0.136)	(0.120)
% Intermediate Occupations	-10.142	-4.146	-39.639	0.077	0.118
	(18.548)	(15.855)	(24.085)	(0.293)	(0.191)
% Small employers	-11.871	-5.188	-20.790	0.190	-0.174
	(16.867)	(14.943)	(24.704)	(0.272)	(0.249)
% Lower supervisory and technical occupations	2.076	4.271	-12.856	-0.155	0.029
	(33.185)	(28.470)	(37.914)	(0.333)	(0.276)
% Semi routine occupations	7.899	6.881	7.773	-0.104	-0.007
	(15.079)	(13.051)	(17.598)	(0.200)	(0.162)
% Routine occupations	-2.930	-3.151	-7.165	-0.153	-0.007
	(12.061)	(10.319)	(15.758)	(0.148)	(0.119)
% Never worked	2.675	-5.451	11.566	-0.047	0.084
	(26.990)	(23.092)	(41.910)	(0.257)	(0.228)
% Long Term Unemployed	35.568	18.464	42.039	-0.690	0.472
	(76.859)	(67.842)	(117.709)	(1.035)	(0.863)
Panel D: First COVID19 wave exposure COVID19 Deaths March-July 2020 per capita	-1307.415	-977.313	-1489.710	-18.609	-5.458
	(1137.140)	(960.435)	(1596.095)	(14.678)	(12.267)
Other Deaths March-July 2020 per capita	520.024	561.006	585.432	-1.220	-2.117
	(665.428)	(565.685)	(888.565)	(8.248)	(6.555)
Panel E: Demographic makeup % aged 0-20	-38.251**	-35.355**	-79.422***	-0.188	-0.126
	(16.383)	(13.943)	(24.337)	(0.210)	(0.202)
% aged 20-30	13.905	8.604	35.363	-0.029	0.122
	(15.739)	(13.434)	(21.746)	(0.119)	(0.099)
% aged 30-40	-8.723	-10.836	-10.682	0.072	0.230
	(16.586)	(14.493)	(25.265)	(0.268)	(0.230)
% aged 40-50	-69.953**	-57.152**	-153.503***	0.116	-0.139
	(29.809)	(25.573)	(43.062)	(0.418)	(0.412)
% aged 50-60	5.711	14.371	-14.076	0.098	-0.128
	(36.314)	(30.926)	(50.124)	(0.300)	(0.286)

% aged 60-70	19.170	21.103	27.455	-0.010	-0.214
	(25.326)	(21.642)	(34.466)	(0.296)	(0.271)
% aged 70+	5.274	7.671	9.847	0.061	-0.102
	(12.927)	(10.957)	(18.227)	(0.180)	(0.156)
Panel F: Area and Population density					
Population	0.001	-0.000	0.006	0.000	0.000
	(0.005)	(0.004)	(0.008)	(0.000)	(0.000)
Area size of district in km2	0.070	-0.041	0.574	0.002	0.004
	(0.525)	(0.436)	(0.838)	(0.003)	(0.003)
Population density	0.014	-0.002	0.113	0.002	0.002
	(0.155)	(0.135)	(0.234)	(0.002)	(0.002)
Variability of population density within district	-0.163	-0.234	0.207	0.002	0.004
	(0.385)	(0.334)	(0.508)	(0.004)	(0.003)
Panel G: Commuting exposure (2011 census) All commuters per resident population	-16.126	-13.050	-24.668	-0.179	-0.194
	(12.976)	(11.235)	(20.061)	(0.157)	(0.128)
Out commuters per resident population	-18.952*	-16.688*	-33.868**	-0.135	-0.077
	46(10.383)	(9.026)	(15.138)	(0.168)	(0.130)
In commuters per resident population	-0.806	-1.067	1.120	-0.044	-0.028*
	(2.463)	(2.022)	(3.240)	(0.029)	(0.015)

Notes: Table presents correlation between area characteristics indicated in the panel head and the treatment intensity measures. Column (1) is the main treatment intensity measure M_i capturing the missing cases (in per capita terms) that were referred to contract tracing with a delay. Column (2) and (3) uses two alternative measures constructed using the curve fitting exercise. Columns (4) and (5) use the fraction-based measure covering the share of infections in an area affected by the error. Column (4) uses the full set of areas while column (5) focuses on areas with significant overall infection levels. All regressions partial out the deciles in the pre-treatment positive test rate, the deciles in the pre-treatment number of cases per capita and the pre-treatment deciles in test performed per capita. Standard errors are clustered at the district level with stars indicating *** p< 0.01, ** p< 0.05, * p< 0.1.