CAGE

Measuring the epidemiological impact of a false negative: Evidence from a natural experiment

CAGE working paper no. 596

November 2021

Thiemo Fetzer

WARWICK
THE UNIVERSITY OF WARWICK

UKRI Economic and Social Research Council

Measuring the Epidemiological Impact of a False Negative: Evidence from a Natural Experiment*

Thiemo Fetzer†

November 14, 2021

Abstract

Reliable COVID-19 testing remains a central pillar to manage the pandemic. Yet, the accuracy and reliability of tests and test equipment has regularly been brought into question. Both false-positive and false-negative test results convey costs. Yet, false negatives are likely more problematic due to the risk of onward transmission and the failure to break infection chains as a result. This paper studies the epidemiological impact of a false negative in the context of a high vaccine uptake country. Between 2 September and 12 October an estimated 43,000 PCR tests in the UK may have produced a false negative test result with individuals infected being told that they tested negative. These instances were particularly pronounced in the South West of England. Using a synthetic control method approach concentrating on the 13 most affected regions, this paper estimates that every false negative COVID-19 case is likely to have caused between 0.6 to 1.6 additional infections in the subsequent weeks.

Keywords: False Negative, Natural Experiment, Test Error, Health, Coronavirus

JEL Classification: I31, Z18

*The author thanks Luis Candelaria and Pedro Souza for helpful comments. He also thanks the UK for its repeated supply of natural experiments and interesting policy variation.

†University of Warwick, CAGE, CESifo and CEPR. t.fetzer@warwick.ac.uk.

1 Introduction

The case in support of non-pharmaceutical interventions to suppress the spread of COVID-19 is paramount. Even after vaccines and treatments have become available, such measures will remain necessary for a considerable amount of time (Ferguson et al., 2020). Effective testing regimes are vital to break transmission chains even in the presence of significant numbers of vaccinated individuals for at least three reasons: to protect those in the population that can not be vaccinated; to lower the risk of the emergence of new virus variants; to reduce pressures on the healthcare system. Naturally, the ability to break chains of infections relies on the accuracy of the tests themselves. While a false positive test result may induce economic costs and inconvenience for the individual involved, a false negative test result may result in further community transmission of the virus, especially if the infected individual is asymptomatic.

The rollout of vaccinations and new COVID-19 variants have changed the dynamic of the pandemic considerably. While vaccinations may be able to reduce onward transmission (Singanayagam et al., 2021; Harris et al., 2021), the different virus variants may have increased the reproduction numbers. Further, many of the effective non-pharmaceutical interventions such as mask mandates (see e.g. Abaluck et al., 2021; Mitze et al., 2020) or restrictions on public or private gatherings have ended, while at the same time in particular vaccinated individuals may have lowered their guard (Bagues and Dimitrova, 2021) and community mobility is far from the lows during the lockdowns. Naturally, how large the transmission is in such a changing context and what role the testing system can play remains an important question.

As part of the efforts to contain the spread of COVID-19, the UK has developed a widespread testing system. The most important component of the test infrastructure is the broad availability of free-of-charge tests through both, a dense network of in-person tests available through national drive- or walk through testing sites or through home test kits. To date, more than 280 million tests have been carried out through this system. The guidance for households and individuals is to regularly perform self-administered tests using lateral flow test kits that are

widely available free of charge. This is particularly aimed at individuals who may be asymptomatic. PCR tests are routinely and freely available for individuals that have COVID-19 symptoms. Further, every positive lateral flow test is confirmed via a PCR tests, many of which are subsequently sequenced to ensure consistent monitoring of new emerging variants.

Much of the testing infrastructure was build rapidly involving a broad set of private contractors that operate the testing system on behalf of the government which has drawn significant criticism from medical professionals (see e.g. Iacobucci, 2020) and parliamentary oversight committees which conclude that the English COVID-19 testing and tracing system has failed to achieve its objectives despite a budget of nearly GBP 37 billion (which is roughly 1/3 of the total budget for the overall healthcare system - see Committee of Public Accounts, 2021).

This paper exploits a consequential error in the COVID-19 testing system. On October 15, the UK Health Security Agency (UKHSA) published a statement that it has suspended testing operations provided by Immensa Health Clinic Ltd at its laboratory in Wolverhampton operations following reports of people receiving negative confirmatory PCR test results after they had tested positive on a lateral flow test. To date, the precise cause of the error is not known but NHS Test and Trace estimate that the error occurred between 2 September and 12 October, 43,000 people may have been given incorrect negative PCR test results. These were individuals who were mostly based in the South West of England. To date, an exact geographic breakdown of the affected COVID-19 tests has not been provided. Nevertheless, thirteen districts that operate test sites in the South West that would predominantly send their PCR swabs to the Wolverhampton lab have been identified.[1] It is not clear how contact tracing was affected by the false negative, given that contact tracing efforts would be initiated already following a positive lateral flow test result (Fetzer and Graeber, 2021). Whether these efforts are subsequently halted with a false negative result is not clear.

To quantify the epidemiological impact of the false negative test results, a synthetic control method approach is adopted. For the thirteen districts that are considered treated, a synthetic control is constructed such that it has a very similar

[1] See `https://mobile.twitter.com/AlastairGrant4/status/1452897308140523520`.

epidemiological profile just prior to the testing error occurring. This approach allows the estimation of the number of missing cases in these thirteen districts during the period of the test error and also subsequently allows the estimation of the number of additional cases following the suspension of the laboratory. Depending on the setup, the estimates suggest that, for every missed infection there may have been between 0.6 to 1.6 additional COVID-19 cases up to Nov 12, 2021. Combined, this implies that the 43,000 individuals that tested positive may have caused between 25,800 to 68,800 additional infections.

This highlights the importance of a reliable testing system as it enables other non-pharmaceutical interventions to function effectively, such as contact tracing (see Fetzer and Graeber, 2021; Wymant et al., 2021; Sun and Viboud, 2020) and self isolation Patel et al. (2021); Bedford et al. (2020). Further, reliable testing data is important to enable the health care sector to plan capacities ahead of time given that infections are a leading indicator for hospitalizations (Moghadas et al., 2020).

The rest of this paper is organized as follows. Section 2 provides background information on the context, data and the measurement approach. Section 3 describes the empirical strategy. Section 4 presents the results, discusses their significance and explores the robustness of our findings. Section 5 concludes.

2 Context and data

2.1 National Testing System and the testing error

In the wake of the first wave of the COVID-19 pandemic in 2020, the UK has rapidly developed its national testing infrastructure through its NHS Test and Trace system. Rather than relying on a network of public laboratories, the testing infrastructure was mostly developed involving a broad set of private contractors that operate the testing system on behalf of the government. The combined budget of NHS Test and Trace, which now is overseen by the newly created UK Health Security Agency (UKHSA) amounts to nearly GBP 37 billion – which contrasted to the last pre-pandemic budget year 2018/2019 makes up 32% of the budget available to the whole UK public health care system NHS. Currently, more than 1,128 private testing providers are listed.

There are significant concerns that many of the private testing providers lack the expertise, equipment and experience to carry out reliable testing; further, many contracts for testing services were awarded not following regular procurement processes, raising concerns about potential conflicts of interest and misappropriation of public funds (Committee of Public Accounts, 2021). Many of the laboratories processing COVID-19 tests were not accredited by the independent UK Accreditation Service, with most providers self-declaring that they meet minimum requirements for testing.[2]

On October 15, 2021, the UKHSA published a statement that "NHS Test and Trace (NHS TT) have suspended testing operations provided by Immensa Health Clinic Ltd at its laboratory in Wolverhampton, following an investigation into reports of people receiving negative PCR test results after they have previously tested positive on a Lateral Flow Device (LFD)." Operations were suspended on October 12. The precise cause of the test error is, as of yet, unknown, but an initial analysis of NHS TT suggested that around 400,000 samples have been processed through the lab. It was estimated that 43,000 people may have been given incorrect negative PCR test results between 2 September and 12 October, mostly in the South West of England. Further, it was indicated that NHS TT would be contacting the people that could still be infectious to advise them to take another test. Further, close contacts who are symptomatic will also be advised to take a test. Despite rises in infections in the weeks following the incident, the UK government denied that the testing failure is responsible for the surge in infections.[3]

The issues were reported for weeks and became visible to experts in official test statistics. For example, the NHS TT provide weekly aggregated test statistics that provide a breakdown of the total number of positive lateral flow test results.[4] According to the official guidance every positive lateral flow test result should be confirmed with a PCR test which is used to confirm the lateral flow test result but

[2]In fact, only 13 labs are currently listed on the UKAS website as accredited COVID-19 testing providers, see https://www.ukas.com/find-an-organisation/browse-by-category/?cat=3731.

[3]See e.g. https://www.theguardian.com/world/2021/oct/15/uk-ministers-face-questions-firm-linked-suspected-covid-test-errors.

[4]The data is available on https://www.gov.uk/government/collections/nhs-test-and-trace-statistics-england-weekly-reports.

also serves to monitor new emerging virus variants. The data provide the number of positive lateral flow tests collected either via an asymptomatic test site or via a test carried out at home. The majority, around 73% of all positive lateral flow tests since calendar week 24 onwards can be matched to confirmatory PCR test. Yet, as is illustrated in Figure 1, from 2 September until early October, the total number of positive lateral flow tests that produced a negative PCR test result more then doubled vis-a-vis the weekly average for the 8 weeks prior to the test error indicated by the dashed grey line in Panel A. Similarly, the share of positive lateral flow tests producing a negative PCR test result increased significantly doubling from around 9% to 18%. Unfortunately, this data is not available with a more granular geographic breakdown, yet, community reports suggest that the phenomena of positive asymptomatic tests being not confirmed with a positive PCR test were geographically concentrated in the South West.[5]

While investigations are underway into the precise cause, the community reports suggest that the test error mostly affected districts in the South West of England. Out of the English 307 local authority districts, the South West is made up of 29 districts. For the purpose of the analysis in this paper, the focus will be on community reported (likely) districts that were affected. The author has launched a Freedom of Information (FOI) requests to the UK Health Security Agency (UKHSA) that oversees and collects data pertaining to the national testing program. This FOI aims to pry available the exact geographic distribution of the tests that were carried out by the Immensa Lab on behalf of the UK government.[6]

2.2 Data on COVID-19 in England

Our baseline analyses leverage three sources of publicly available data.

Reporting dashboard The primary dataset used for the analysis is constructed using the UK's COVID-19 dashboard.[7] This dashboard provides granular data on

[5]See for example `https://neighbournetworks.uk/2021/09/18/18th-sept-2021-data-update/` and `https://www.opendemocracy.net/en/ournhs/i-raised-an-early-alarm-on-pcr-test-scandal-but-authorities-ignored-me/`.

[6]This FOI and all associated data and communication is available on `https://www.whatdotheyknow.com/request/immensa_dante_labs_missing_tests`.

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

COVID-19 infections and deaths at different spatial resolutions. Our geographical focus is on England, as England is said to be most affected by the test error and due to the consistency of data published by English authorities. The number of COVID-19 cases is reported by the date on which the test swab was taken (the so-called specimen date). Cases data include all positive lab confirmed virus test results plus, in England, positive rapid lateral flow tests that do not have negative confirmatory lab-based polymerase chain reaction (PCR) tests taken within 72 hours. Further, the data provide the cumulative total number of double vaccinated individuals resident in an area. Further, other cumulative data on the share of infected individuals prior to the test error occurring is obtained to get a baseline measure of (likely) community immunity. Measures on deaths are also counted, yet, the preferred measure capturing whether a death has occurred within 28 days of a positive COVID-19 test may be inaccurate, given the test error may have resulted in positive lateral flow tests being overturned by a (wrong) negative test result. The analysis is conducted at the level of the 307 English local authority districts (as per their 2021 boundary definitions). The resulting core dataset is a balanced daily panel. Our estimation window focuses on the period starting in calendar week 24 (starting July 6, 2020) all the way to the most recent day data is available.

Test and Trace statistics We also draw on data on testing and tracing statistics provided by NHS Test and Trace (GOV.UK, 2020). These data are published weekly and provide additional data on the testing system. Specifically it provides a daily dataset measuring the total number of all COVID-19 tests taken per day (lateral flow and PCR tests under the national testing program). This data is available at the local authority district level and is updated weekly. Further, weekly views of the data is also provided, in addition to some further breakdowns of the positive test results which measure whether individual tests are matched or confirmed with a PCR test as was illustrated in the data presented in Figure 1. The daily number of all tests carried out is combined with the infection data reported in the reporting dashboard to arrive at a measure of the overall test positivity rate.

Additional data Additional data on area characteristics are obtained from the UK's Office of National Statistics. These capture the 2019 age make-up of the local population, further measures from the 2011 census. The data can be leveraged to construct the synthetic control but is primarily used to normalize outcome measures.

2.3 A preview of the data

Figure 2 presents a time-series plot comparing the positive test rate and the number of new COVID-19 cases over time across the 13 districts (likely) most affected by the testing error with the rest of the average positivity rate of tests across districts outside of the South West in Panel A, and, the total number of confirmed cases in Panel B. We note that prior to the first vertical line marking the (likely) onset of the testing errors at the Immensa laboratory, the share of tests returning a positive results significantly declined from above 3% to less than 1% in the following weeks. This is followed by a sharp increase in the share of tests returning a positive result from October 12 onwards with the share peaking at around 8% of all tests. Subsequently, the positivity rate remains above average compared to the rest of the country for a few more weeks. It is noteworthy that the positivity figure closely tracks the average positivity rate across all districts outside the South West for the weeks prior to the test error.

In Panel B we observe what happens to case figures. The total number of COVID-19 cases across the most affected districts are slightly higher, on average, when compared to districts in the rest of the country outside the South West. During the weeks affected by the testing error, the case figures are quite consistently lower compared to the rest of the country, only to subsequently shoot up. The below average test positivity rate in the districts in the South West most exposed to the testing error is consistent with the notable jump in test positivity rate from the point onwards, as some individuals that would have tested negative via the PCR test were retested. The gap between the number of positive cases that appear "missing" during the period of the error and the subsequent increase in cases may provide for an estimate of ranges of the number of additional infections that may have been caused due to the uncontrolled community transmission that the false

negative tests may have facilitated.

A second visualization that highlights the geographic concentration of the patterns in the share of tests returning a positive result that is illustrated in Panel A of Figure 2 is provided in Figure 3. This map presents the spatial distribution of the share of COVID-19 tests that returned a positive test result over three different time periods. Panel A provides the distribution of the share of positive test results in the four week window prior to the test error. Panel B presents the share of positive tests over the four week window during which tests may have been producing negative test results. While there was a decline in test positivity rate, many districts in the South West stands out as virtually disappearing from the epidemiological map. Panel C displays the distribution in the weeks following the test error highlighting a strong resurgence of the test positivity rate in the South West. Panel D lastly plots the distribution of the changes highlighting that, in particular districts in the South West, stand out with increases in the share of tests being positive vis-a-vis the period during which the false negatives were produced, with many districts seeing increases 1 SD higher compared to the average across districts.

2.4 Identifying treated districts

This paper adopts a synthetic control method approach to estimate the epidemiological impact of the false-negative test results. We classify districts as treated- or untreated based based on community reports which have identified 13 districts most affected by the test error due to a mix of contextual information from testing sites in those districts sending test swabs to the Immensa lab along with identified irregularities in the testing data during the period when the false negative test errors were being produced. [8]

We next present details of the empirical strategy.

[8] See for example `https://neighbournetworks.uk/2021/09/18/18th-sept-2021-data-update/` and `https://www.opendemocracy.net/en/ournhs/i-raised-an-early-alarm-on-pcr-test-scandal-but-authorities-ignored-me/`.

3 Empirical strategy

3.1 Synthetic control approach

The paper adopts a synthetic control method (SCM) approach to estimate the impact of the testing error on subsequent infection dynamics (Abadie and Gardeazabal, 2003; Abadie et al., 2010, 2015).[9] The synthetic control method approach can be leveraged as the geographic concentration of the testing error in the South West of England provide a relatively well-defined treatment unit. The approach offers an alternative to difference-in-differences designs whereby a control district is constructed as a "synthetic" unit that represents a weighted combination of many untreated districts. Weights are calculated in order to maximize the similarity between the synthetic control and the treatment unit in terms of specified "matching" variables. In the context of the COVID-19 pandemic, such approaches have been used to explore the efficacy of some non-pharmaceutical interventions, such as mask mandates (see Mitze et al., 2020).

We assume there are data on J total districts. The first J_0 are untreated, while the remaining $J - J_0$ are treated. Further, there are I time-varying outcomes measured across T total time periods where we denote the first T_0 as the pre-treatment time periods, while $T - T_0$ are the post-treatment time windows. Let Y_{itj} denote the value of outcome i at time t for district j. Further, there are L cross-sectional time-invariant characteristics with R_{lj} denoting the value of the cross sectional characteristic l for district j. A synthetic control group is calculated by assigning a weight to each non-treated case. These weights are denoted w_j for $j \in (1, ..., J_0)$. There are three sets of constraints that the synthetic control should meet.

First, the sum of the weights should equal the number of treated districts. That is

[9] A generalized difference-in-difference approach as e.g. leveraged in Fetzer and Graeber (2021) may become feasible as an alternative if data on the exact geographic distribution of the false negative test results become available. The author has launched a FOI to the UKHSA. All associated data and communication is available on `https://www.whatdotheyknow.com/request/immensa_dante_labs_missing_tests`.

$$\sum_{j=1}^{J_0} w_j = J - J_0$$

Second, the weighted synthetic control matches the treatment area across covariates for all $l \in (1, ...L)$.

$$\sum_{j=1}^{J_0} w_j R_{lj} = \sum_{j=J_0+1}^{J} R_{lj}$$

And third, the synthetic control and treatment also match across pre-intervention time points of each outcome in that

$$\sum_{j=1}^{J_0} w_j Y_{itj} = \sum_{j=J_0+1}^{J} Y_{itj}$$

for all $i \in (1, ...I)$ and $t \in (1, ..., T_0)$.

This amounts to a total of $1 + L + IT_0$ constraints that the w_j must satisfy.

3.2 Implementation

For the most preferred estimate we implement the above using the R package `microsynth` (Robbins and Davenport, 2021). In the simplest case, we leverage a single outcome measure, and four set of time-invariant baseline characteristics measured in the four weeks prior to the error occurring: the average % of residents that have received a second COVID-19 vaccination dose, the cumulative COVID-19 mortality rate per capita, the average cumulative number of COVID-19 cases in a district since the start of the pandemic, the average new number of COVID-19 cases per day, the average number of COVID-19 tests performed as well as the average positivity rate of lateral flow tests. We also use the average number of new cases per capita across 19 age groups in the pre-treatment period as well as the average cumulative number of cases per capita in an age group as additional time-invariant controls. This implies that there are 47 covariates or, $L = 47$. The pre-intervention period from calendar week 26 to the 1. September 2021 includes 70 days. In the above terminology, using the sharp definition of treated- versus control group areas, this amounts to there being 118 constraints. Using only sub-

sets of the features produces very similar results.

3.3 Estimating surplus infections

In order to estimate the combined treatment effect measuring the total number of infections that are (likely) to have been caused by the testing error, we separate the post-treatment time period $T - T_0$ into two sup periods. Let T_1 denote the time point when the Immensa lab ceased operating from October 12 onwards.

This allows us to measure the cumulative number of *missing cases* across the treated areas during the period when the lab was operating producing false negatives as

$$\widehat{\alpha_{\text{miss}}} = \sum_{t=T_0+1}^{T_1} \Big(\sum_{j=J_0+1}^{J} Y_{itj} - \sum_{j=1}^{J} w_j Y_{itj} \Big)$$

Similarly, we can compute the number of *surplus cases* that resulted endogenously due to the testing error as

$$\widehat{\alpha_{\text{surplus}}} = \sum_{t=T_1+1}^{T} \Big(\sum_{j=J_0+1}^{J} Y_{itj} - \sum_{j=1}^{J} w_j Y_{itj} \Big)$$

Appendix Figure A1 illustrates the idea behind the two measures. The red shaded area captures captures the number of cases missed vis-a-vis a credible counterfactual during the period that the test error occurred in areas (most) affected by the error. The blue shaded area captures the number of surplus cases after the error was rectified.

The absolute numbers are naturally hard to interpret, given that the missing infections during the period of the test error may have produced further missing infections that also went on to be undetected while the laboratory producing false negatives was still operating. Nevertheless, the ratio of the two measures is likely a good proxy for the epidemiological impact of the undetected cases in the subsequent weeks, proxying for the onward transmissions that are likely to have been caused. That is, we can compute the additional infections per undetected COVID-19 case in treated areas vis-a-vis the synthetic control as

$$\hat{\Delta} = \frac{\widehat{\alpha_{\text{surplus}}}}{\widehat{\alpha_{\text{miss}}}}$$

Specifically, we can define an upper- and a lower-range estimate, given that $\widehat{\alpha_{\text{surplus}}}$ may contain some cases that were previously missed that got re-tested after the error was detected. A potential range of the multiplier thus may be

$$[\hat{\Delta}_{lower}, \hat{\Delta}_{upper}] = [\frac{\widehat{\alpha_{\text{surplus}}}}{\widehat{\alpha_{\text{miss}}}} - 1, \frac{\widehat{\alpha_{\text{surplus}}}}{\widehat{\alpha_{\text{miss}}}}]$$

This measure provides the relative impact of one missed case on subsequent differential infections occurring in the regions most treated. This scaling factor or multiplier can be used to quantify the extent to which false-negative test results produce onward infections. This is an interesting metric given that with significant immunity conveyed by the vaccination, the presence of COVID-19 variants and given changes in individual behavior may have significantly altered the dynamic of the pandemic and the implied reproduction numbers.

3.4 Identification and potential biases in measurement

To date, the results of an investigation by NHS TT into the issues that caused the false negative test results are still pending. Yet, it is quite clear that the error constitutes a natural experiment that was not anticipated or expected to happen. By chance, the extent to which false negative test results were issued was much more severe for some areas of England than others. This source of random variation allows us to investigate to which extent districts treated by such false negatives subsequently experienced a different pandemic progression.

Naturally, the above measures $\widehat{\alpha_{\text{miss}}}$, and $\widehat{\alpha_{\text{surplus}}}$, and $\hat{\Delta}$ are noisy, as it is impossibly to identify the full chain of contacts and potential interactions that may have resulted in onward transmission. Nevertheless, we can discuss to what extent the measures will be biased upwards or downwards. Since at the time of writing this paper detailed data on the exact geographic distribution of the false negatives is still pending, we have to rely on the expert analysis and contextual information to identify the treated districts for the purpose of the analysis. While the synthetic control construction draws its donor pool from outside of the South West through-

out the analysis, there is likely going to be contamination of the donor pool that arises from the fact that some false negatives may have been sent to individuals residing outside of the South West. Mechanically, this should imply that the estimates of the treatment effects may be downward biased as the synthetic control includes some areas that may be partially treated, resulting in a smaller estimated treatment effect.

Further, when it comes to $\widehat{\alpha_{\text{miss}}}$, this measure has three components. First, the genuine number of missing cases due to the false negatives, along with the endogenous response consisting of additional false negatives that may be generated due to community transmission during the period when PCR tests were producing false negatives. Similarly, $\widehat{\alpha_{\text{surplus}}}$ may be biased for three reasons: first, some of the individuals that received a false negative where invited to subsequently get tested again, after the error was detected. This mechanically implies more cases being measured, which may already be included in the measure $\widehat{\alpha_{\text{miss}}}$. Further, if the testing error increased community testing in the affected areas in general, this may imply a detection of more cases, that would, counterfactually, go undetected. This would induce an upward bias in $\widehat{\alpha_{\text{surplus}}}$. Lastly, $\widehat{\alpha_{\text{surplus}}}$ includes the infections that endogenously arise due to the false negatives inducing chains of infections. This naturally is the relevant parameter of interest.

As we will see, while in treated districts, there is an increase in testing, this increase is relatively small and can not account for the overall increase in positive cases. Further, we can mechanically bound the increase in surplus infections that arises mechanically due to retesting by simply subtracting $\widehat{\alpha_{\text{surplus}}} - \widehat{\alpha_{\text{miss}}}$, providing us with a lower bound on onward infections.

We next present the results from the analysis.

4 Results and discussion

Figure 5 presents the results from the synthetic control construction as outlined above. The left panel provides the time series of the number of cases in the treatment group (red line) and the synthetic control (black dashed line). The two lines track each other very well, with the two notably diverging with the onset of the

testing error: areas likely most affected by the testing error saw a notable decrease in reported infections, suggesting that the false negatives artificially depressed the true case count during the period when the error went undetected. Subsequently, there is a sharp increase in infections following the operations in the Immensa lab being shut down.

The light grey lines are *placebo* placebo treatment units using random permutations of the control units. That is, a total 100 alternative treated districts from the set of control group districts are selected, at random, to serve as treated units and subsequently, synthetic control estimates are constructed. This allows for an alternative way of conducting inference as naturally, we would not expect to see notable treatment effects in areas that are unaffected. Quite reassuringly, we note that during the period of the testing error, the treated units appear to have a notably lower level of infections throughout compared to the placebo synthetic control estimates. What is noteworthy is that the missing cases appear to be particularly concentrated on specific days.

While the investigation is ongoing this may be of particular interest to assess what may have been the cause of the errors. The lab processes test swabs that are placed in tubes containing a solution of chemicals that make copies of the virus genetic material if any virus is present using the polymerase chain reaction (PCR) process. The tubes are placed heating block that is the main part of the PCR machine. After that, the solution is tested for any genetic material. The machines that are said to be in use are manufactured by Perkin Elmer PCR 9700 and can hold and process up to 96 test tubes simultaneously, which can be purchased used or new for around GBP 800 - 3,000. The lumpiness of daily figures of missing cases could indicate processing errors that were particularly concentrated around individual batches, specific days or in specific shifts, which could indicate that environmental conditions or issues with the equipment may be particularly problematic.[10] The COVID-19 case data is based on the specimen dates capturing the

[10]Whistleblower evidence suggests that failures may be due to environmental conditions in the lab being erratic due to poor air conditioning or air regulation. The specific concentration of false negatives on specific days would suggest that possibly poor maintenance of equipment or other technical factors may be an even more important source compared to `https://www.independent.co.uk/news/health/covid-lab-wolverhampton-immensa-tests-result-b1945785.html`.

date that stand out with a particularly large number of missing cases relative to the synthetic control are test swabs that were taken from October 5 - October 11.

Quantification We can arrive at an estimate of the implied measures $\widehat{\alpha_{\text{miss}}}$, and $\widehat{\alpha_{\text{surplus}}}$, and $\hat{\Delta}$. The results presented in Figure 5 suggest that $\widehat{\alpha_{\text{miss}}} = 12,971$, i.e. in the thirteen districts, relative to the synthetic control during September 2 to October 11 inclusive, there are at least 12,971 COVID-19 cases that are *missing*. During the period from October 12 onwards, the difference in case numbers between the synthetic control and the reported case count by specimen date suggests that $\widehat{\alpha_{\text{surplus}}} = 20853$. This results in the estimate $\hat{\Delta} = 1.6$, suggesting that the best estimate with this methodology would suggest that every missing case that may have arisen out of the Immensa lab testing error produced an additional 1.6 infections.

This represents an upper bound of the differential effect given that some individual cases included in $\widehat{\alpha_{\text{miss}}}$ may have been retested. A conservative way of accounting for this would suggest that all missing cases are included in $\widehat{\alpha_{\text{surplus}}}$, which would render

$$[\hat{\Delta}_{lower}, \hat{\Delta}_{upper}] = [0.6, 1.6]$$

Estimates by district We can conduct the above analysis district-by-district. In this case, rather than pooling the 13 treated districts, we construct a synthetic control for each treated district individually. This allows us to arrive at estimates of $\widehat{\alpha_{\text{miss}}}$, and $\widehat{\alpha_{\text{surplus}}}$, and $\hat{\Delta}$ for each district i. The results are visually presented in Appendix Figure A2. The pattern looks very similar throughout across the 13 districts with below expected case numbers vis-a-vis the synthetic control during the period that the lab was active and likely producing false negative, followed by a rapid increase in case figures that capture the community spread and, partially, the response due to retesting. Table 1 provides the estimated impacts by district when fitting a synthetic control for each treated district separately.

The column pre is a measure of the goodness of fit, capturing the deviations between synthetic control and treated units for the period that was unaffected by

the lab processing errors. The subsequent columns capture the estimated number of missing cases by districts and the subsequent surplus cases. We note that the estimates obtained when fitting a synthetic control separately for each district compares very well with the estimates obtained from pooling the treated units with a similar estimated $\hat{\Delta}_{upper} = 1.59$.

As indicated, it is likely that the synthetic control based measure is unlikely to capture the full impact, given that it is quite likely that some areas in the control group would have been affected by the testing error as well. This should induce the estimated treatment effects to be biased downwards as part of the control group may be treated itself given that the treatment unit is an individual test and not a region. With this caveat in mind, we can quantify the likely overall impact based on the governments own first impact assessment.

If the effect of a false-negative is similar in districts that are not identified as particularly treated given the volume of false negatives through community reports, we can arrive at an overall estimate by applying the range of estimates $[\hat{\Delta}_{lower}, \hat{\Delta}_{upper}] = [0.6, 1.6]$ to the estimated total number of missed cases from the UKHSA first assessment of the testing error. This suggested that there may have been up to 43,000 false negatives across the UK. Using the above, we could arrive at the combined indirect effect of these false negatives using the estimated multipliers to suggest that the 43,000 false negatives may have lead to an additional 25,800 to 68,800 additional infections.

As indicated, more granular data is needed for an even more refined assessment and such data has been requested through a freedom of information request.

Mechanisms The increase in case numbers, especially after the the testing error was noticed could be due to more extensive testing efforts being targeted at the regions most affected. This may lead to more positive cases being detected that would otherwise go unnoticed. We also construct synthetic controls to measure the evolution of testing activity and the test positivity rate in treated versus control group districts. These are presented in Figure 6.

Panel A documents what is happening to the number of tests. We observe that during the period of the testing error, the 13 most affected districts record

nearly 108,718 fewer tests than what would be expected vis-a-vis the respective synthetic control. This suggests that the likely failure to detect positive cases resulted in fewer subsequent tests of potential contacts that counterfactually, would have taken place. After the testing error, there is an increase in the number of tests cumulatively of around 44,467. This is far fewer than the tests that counterfactually have not taken place during the period when infections may have gone undetected due to the false negatives producing further community spread. It is worth nothing that the endogenous response on testing mechanically should have an impact on the number of positive cases (and hence, missing cases) and the test positivity rate. During the period when the lab was operating producing false negative PCR tests, lower testing prevalence should imply an upward bias in the test positivity rate as fewer asymptomatic infections may not be recorded.

Panel B presents the evolution of the positive test rate across all tests conducted under the pillar 2 community testing program. We note that during the period affected by the lab error, the share of tests returning a positive result declines significantly. Following the suspension of lab operations, the share increases notably across the 13 most affected districts, increasing by nearly 100% relative to the baseline.

More granular data is needed, ideally at the individual level, in order to tease out the exact mechanisms. While naturally, it is impossible to detect the full transmission dynamics that the false negative test results may have provided, the estimates are quite clear suggesting that they led to a net increase in infections that is not negligible, with the ensuing impact on adding to pressures on the health care system and causing anxiety and distress among those affected.

5 Conclusion

This research notes shows that a significant and large scale processing problem of confirmatory PCR tests in a laboratory in England producing an estimated 43,000 false negative test results has had a causal impact on increasing pandemic progressing with an estimated 25,800 to 68,800 additional infections having been due to the testing error. The assertion by the government that the testing error

is unlikely to have had an impact on infection dynamics can not be sustained, though more granular data is needed in order to further refine the analysis using alternative inference methods.

An independent analysis of the failures in the laboratory is needed in order to improve practices and to identify common pitfalls and issues that may undermine the quality of tests carried out in England. This is particularly important as such large scale testing errors undermine the public trust in the governments response to handling the COVID-19 pandemic. Given the significant fiscal resources that have been deployed to develop a national testing programme involving private sector players, accountability and transparency are imperative in order to ensure that faith in the system and compliance with COVID-19 testing guidelines remains high.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller, "Comparative Politics and the Synthetic Control Method," *American Journal of Political Science*, feb 2015, *59* (2), 495–510.

_ , _ , Hainmueller, and Jens, "Synthetic control methods for comparative case studies: Estimating the effect of California's Tobacco control program," *Journal of the American Statistical Association*, 2010, *105* (490), 493–505.

_ and Javier Gardeazabal, "The economic costs of conflict: A case study of the Basque Country," *American Economic Review*, 2003, *93* (1), 113–132.

Abaluck, Jason, Laura H Kwong, Ashley Styczynski, Ashraful Haque, Md. Alamgir Kabir, Ellen Bates-Jefferys, Emily Crawford, Jade Benjamin-Chung, Salim Benhachmi, Shabib Raihan, Shadman Rahman, Neeti Zaman, Peter J. Winch, Md. Maqsud Hossain, Hasan Mahmud Re, and Ahmed Mushfiq Mobarak, "Normalizing Community Mask-Wearing: A Cluster Randomized Trial in Bangladesh," *NBER Working Paper*, 2021.

Bagues, Manuel and Velichka Dimitrova, "The Psychological Gains from COVID-19 Vaccination : Who Benefits the Most ?," *CEPR Working Paper*, 2021.

Bedford, Juliet, Delia Enria, Johan Giesecke, David L. Heymann, Chikwe Ihekweazu, Gary Kobinger, H. Clifford Lane, Ziad Memish, Myoung don Oh, Amadou Alpha Sall, Anne Schuchat, Kumnuan Ungchusak, and Lothar H. Wieler, "COVID-19: towards controlling of a pandemic," *The Lancet*, 2020, *395* (10229), 1015–1018.

Committee of Public Accounts, "House of Commons Committee of Public Accounts Test and Trace update," Technical Report October 2021.

Ferguson, Neil, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, Zulma Cucunubá, Gina Cuomo-Dannenburg et al., "Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand," *Imperial College London*, 2020, *10*, 77482.

Fetzer, Thiemo and Thomas Graeber, "Measuring the scientific effectiveness of

contact tracing: Evidence from a natural experiment," *Proceedings of the National Academy of Sciences*, 2021, *118* (33), e2100814118.

GOV.UK, "NHS Test and Trace (England) and coronavirus testing (UK) statistics," 2020, *https://www.gov.uk/government/publications/nhs-test-and-trace-england-and-coronavirus-testing-uk-statistics-22-october-to-28-october.*

Harris, Ross J., Jennifer A. Hall, Asad Zaidi, Nick J. Andrews, J. Kevin Dunbar, and Gavin Dabrera, "Effect of Vaccination on Household Transmission of SARS-CoV-2 in England," *New England Journal of Medicine*, 2021, *385* (8), 759–760.

Iacobucci, Gareth, "Covid-19: Government is criticised for âscandalousâ £10bn spent on test and trace programme," *BMJ*, jul 2020, *370* (July), m2805.

Mitze, Timo, Reinhold Kosfeld, Johannes Rode, and Klaus Walde, "Face masks considerably reduce COVID-19 cases in Germany," *Proceedings of the National Academy of Sciences of the United States of America*, 2020, *117* (51), 32293–32301.

Moghadas, Seyed M., Affan Shoukat, Meagan C. Fitzpatrick, Chad R. Wells, Pratha Sah, Abhishek Pandey, Jeffrey D. Sachs, Zheng Wang, Lauren A. Meyers, Burton H. Singer, and Alison P. Galvani, "Projecting hospital utilization during the COVID-19 outbreaks in the United States," *Proceedings of the National Academy of Sciences of the United States of America*, 2020, *117* (16), 9122–9126.

Patel, Jay, Genevie Fernandes, and Devi Sridhar, "How can we improve self-isolation and quarantine for covid-19," *The BMJ*, 2021, *372* (table 1), 1–6.

Robbins, Michael W. and Steven Davenport, "Microsynth: Synthetic control methods for disaggregated and micro-level data in R," *Journal of Statistical Software*, 2021, *97* (2), 1–31.

Singanayagam, Anika, Seran Hakki, Jake Dunning, Kieran J Madon, Michael A Crone, Aleksandra Koycheva, Nieves Derqui-Fernandez, Jack L Barnett, Michael G Whitfield, Robert Varro, Andre Charlett, Rhia Kundu, Joe Fenn, Jessica Cutajar, Valerie Quinn, Emily Conibear, Wendy Barclay, Paul S Freemont, Graham P Taylor, Shazaad Ahmad, Maria Zambon, Neil M Ferguson, Ajit Lalvani, Anjna Badhan, Simon Dustan, Chitra Tejpal, Anjeli V Ketkar, Janakan Sam Narean, Sarah Hammett, Eimear McDermott, Timesh Pillay, Hamish Houston, Constanta Luca, Jada Samuel, Samuel Bremang,

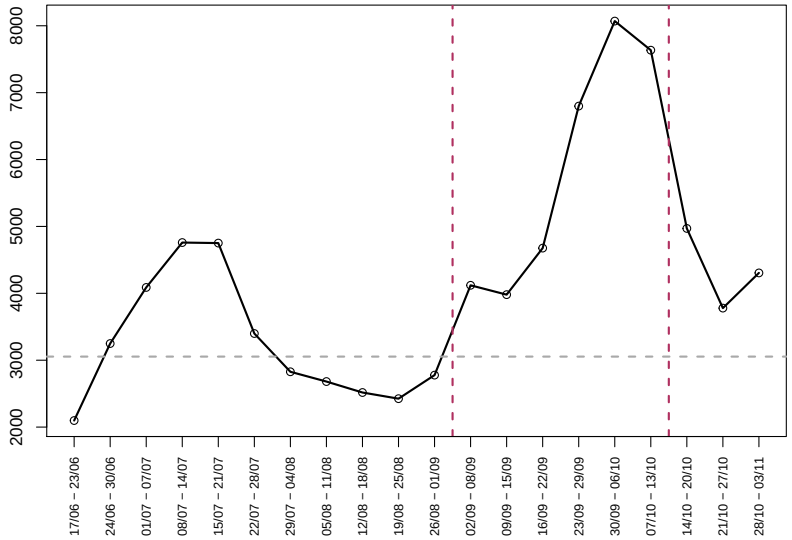
Samuel Evetts, John Poh, Charlotte Anderson, David Jackson, Shahjahan Miah, Joanna Ellis, and Angie Lackenby, "Community transmission and viral load kinetics of the SARS-CoV-2 delta (B.1.617.2) variant in vaccinated and unvaccinated individuals in the UK: a prospective, longitudinal, cohort study," *The Lancet Infectious Diseases*, oct 2021, *3099* (21).

Sun, Kaiyuan and Cécile Viboud, "Impact of contact tracing on SARS-CoV-2 transmission," *The Lancet Infectious Diseases*, 2020, *20* (8), 876–877.

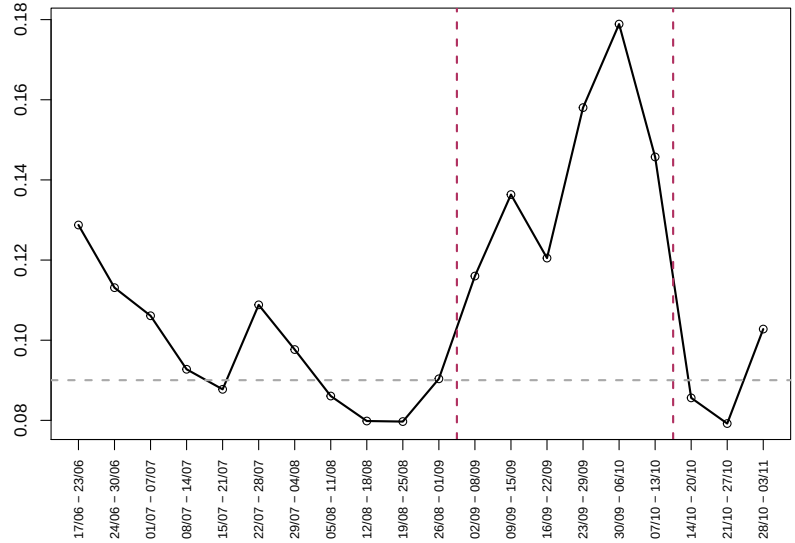
Wymant, Chris, Luca Ferretti, Daphne Tsallis, Marcos Charalambides, Lucie Abeler-Dörner, David Bonsall, Robert Hinch, Michelle Kendall, Luke Milsom, Matthew Ayres, Chris Holmes, Mark Briers, and Christophe Fraser, "The epidemiological impact of the NHS COVID-19 app," *Nature*, 2021, *594* (7863), 408–412.

Figure 1: Evolution of confirmatory PCR tests producing a negative result following a positive lateral flow test result in the whole of England

Panel A: Absolute Numbers

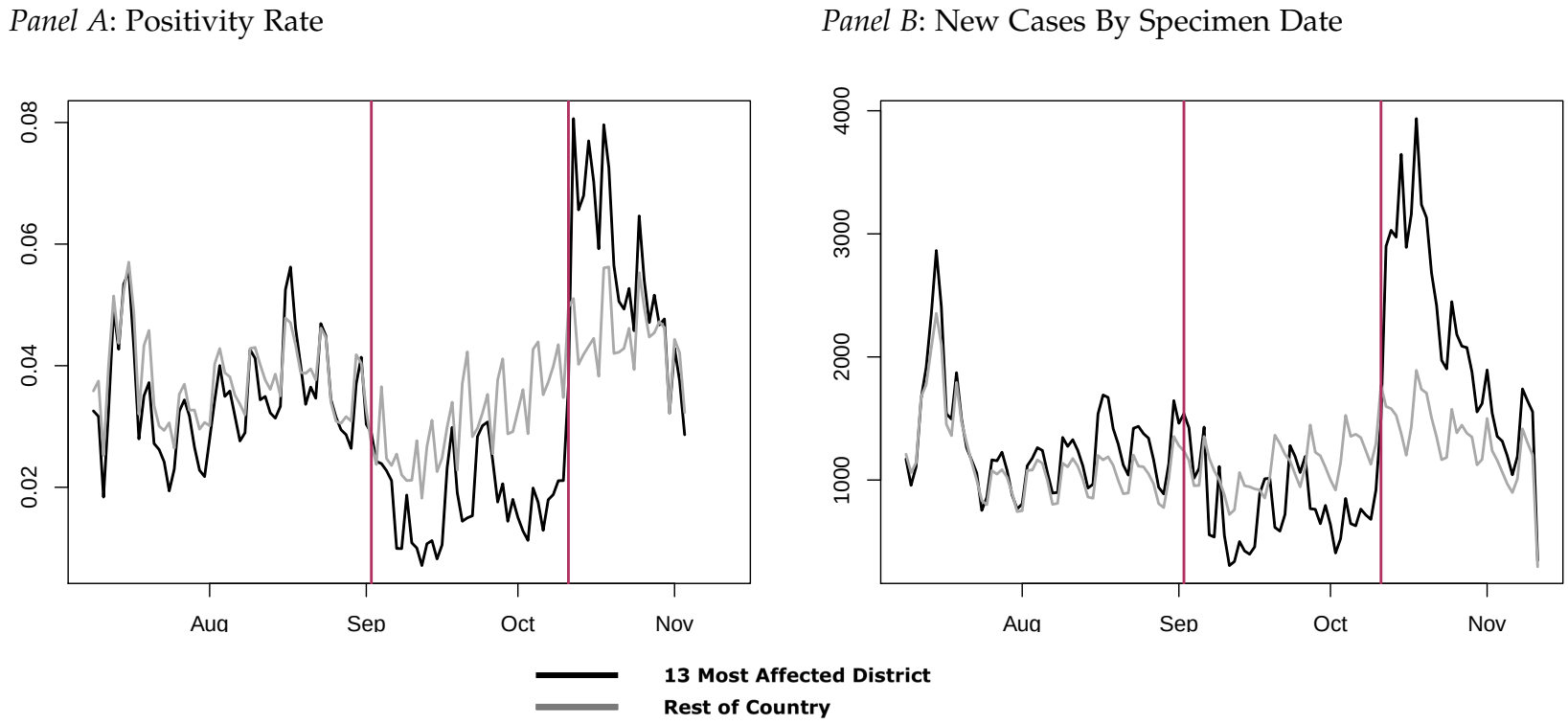


Panel B: as % of all tests



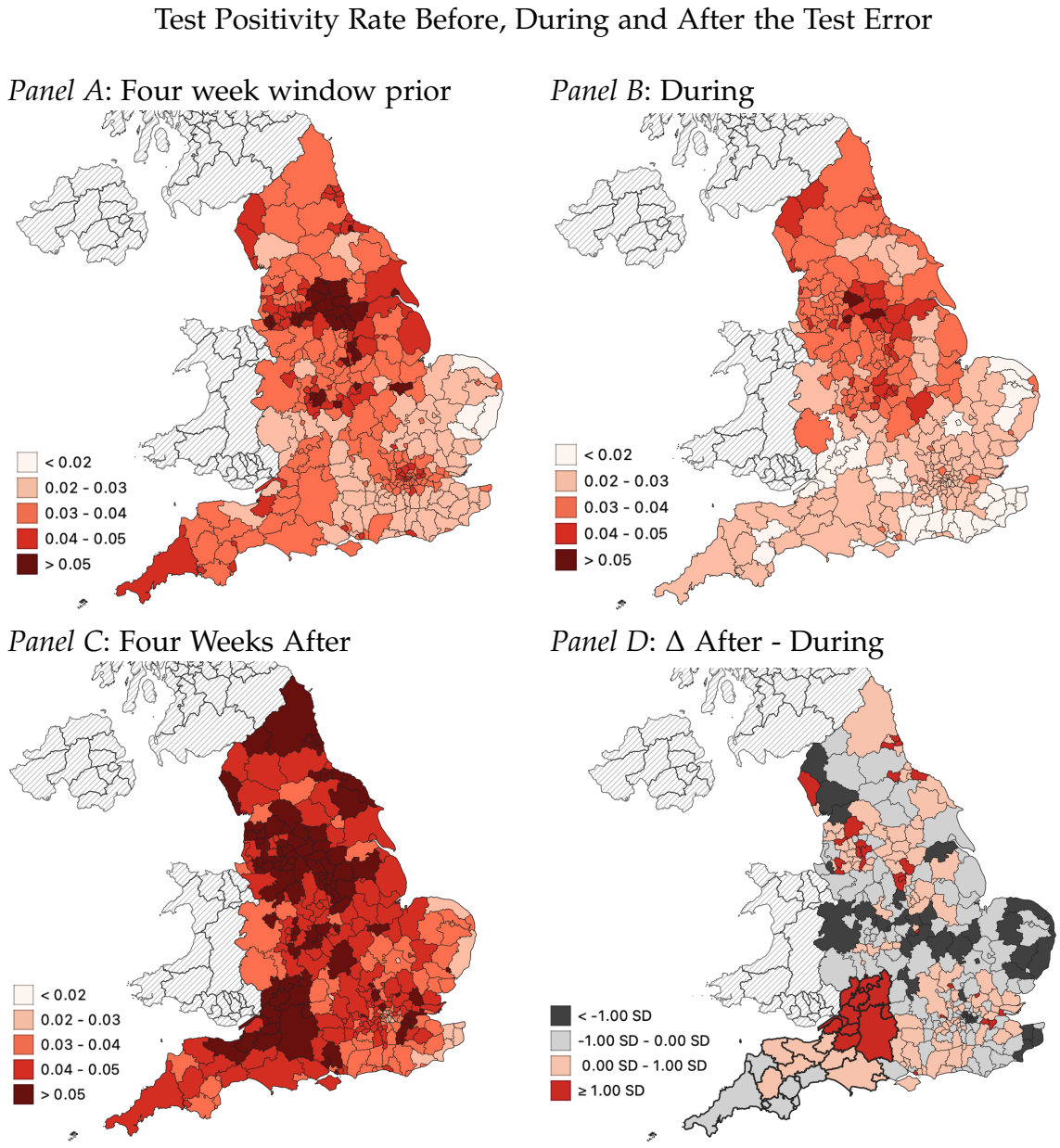
Notes: Figure plots the absolute number and the share of lateral flow tests that returned a positive COVID-19 test result among all that were followed by a subsequent PCR test where the subsequent confirmatory PCR test result was negative. Panel A presents the absolute number highlighting that from the start of week 35 there is a notable increase in the absolute number of positive lateral flow test matched with a PCR test that produce a negative PCR test result. Panel B presents the % of positive lateral flow tests that are matched to a confirmatory PCR test producing a negative result starting in the whole population of confirmatory PCR tests that were conducted. We observe a notable increase during weeks 35 to 40 when the Immensa lab operations were suspended.

Figure 2: Comparing Positive Test Rate and Case Incidence in (Most) Affected Districts in South West vis-a-vis the rest of the country



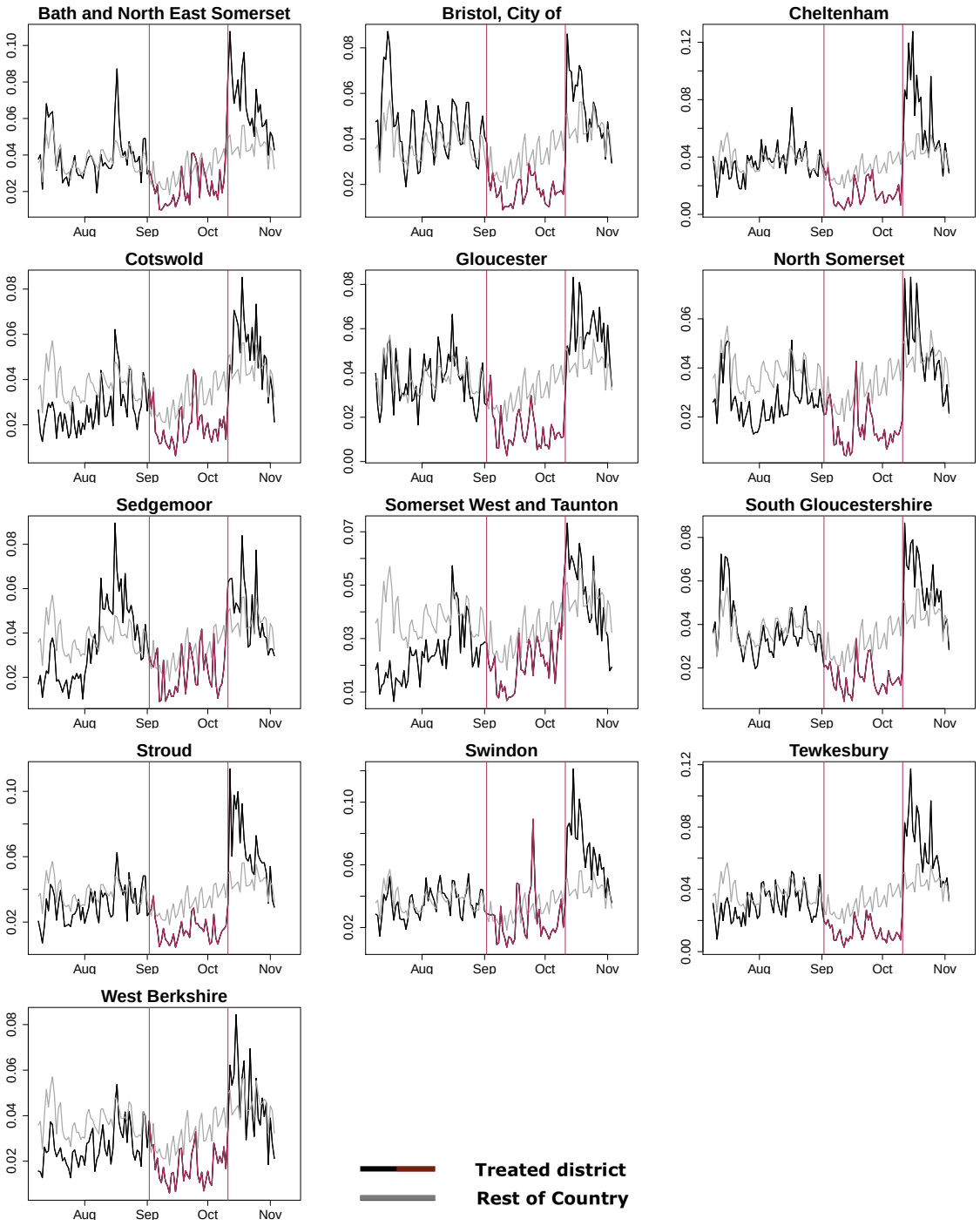
Notes: Figure documents how in the 13 districts (likely) most affected by the testing error the positivity rate and the number of new COVID-19 cases diverge vis-a-vis the rest of the country. Panel A focuses on the distribution over time of the share of COVID-19 tests that returned a positive results, while Panel B focuses on the number of cases. Dark grey is the combined statistic for the 13 districts most affected by the tracing error, while the gray indicates the statistic relating to the rest of the country outside the South West.

Figure 3: Distribution of Test Positivity Rate Across English Local Authority Districts Over Three Time Windows



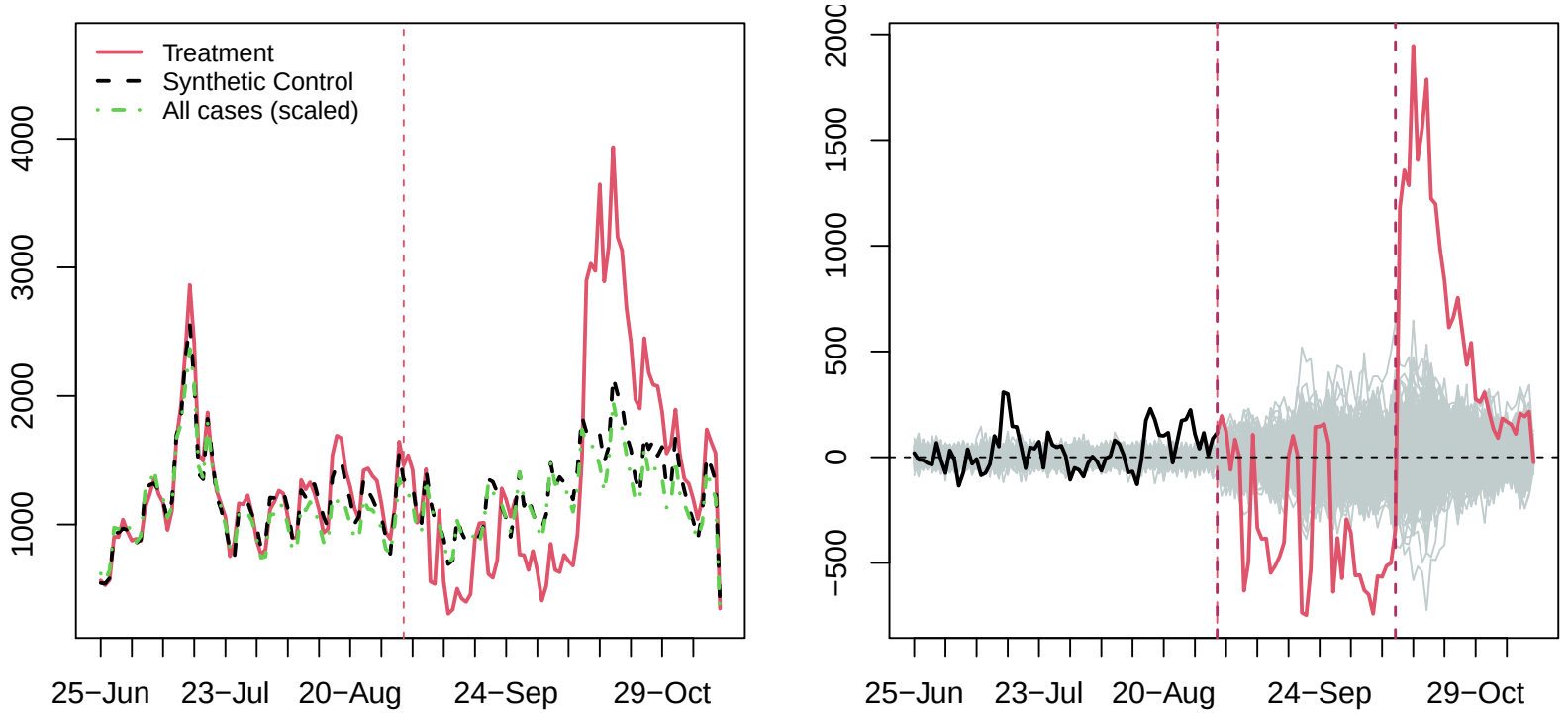
Notes: Panel A, B and C plots the spatial distribution of the share of COVID-19 tests that returned a positive result across English Local Authority Districts at three different time points. Panel D plots the empirical distribution of the difference between Panel C and Panel B documenting which districts saw a notably larger increase in the positive test rate rate. This is most sharply concentrated in the South West. The population of tests includes all undertaken under the official UK Government Testing Program (also known as Pillar 2 tests). These make up around 85% of the around 340 million tests carried out to date. COVID-19 cases are coded based on the date on which the test was taken. Panel A plots the share of all COVID-19 positive tests in the four week window prior to the time period affected by the testing error. Panel B focuses on the period during which the testing error is said to have occurred from 2 September to 12 October. Panel C focuses on the time period since then. Panel D plots the distribution of the changes. The sharp decline in test positivity rate between Panel A and Panel B and subsequent sharp increase in Panel C is spatially concentrated in the South West that was most affected by the error as is also illustrated in Panel D.

Figure 4: Test Positivity Rate Across 13 Regions (Likely) Most Affected



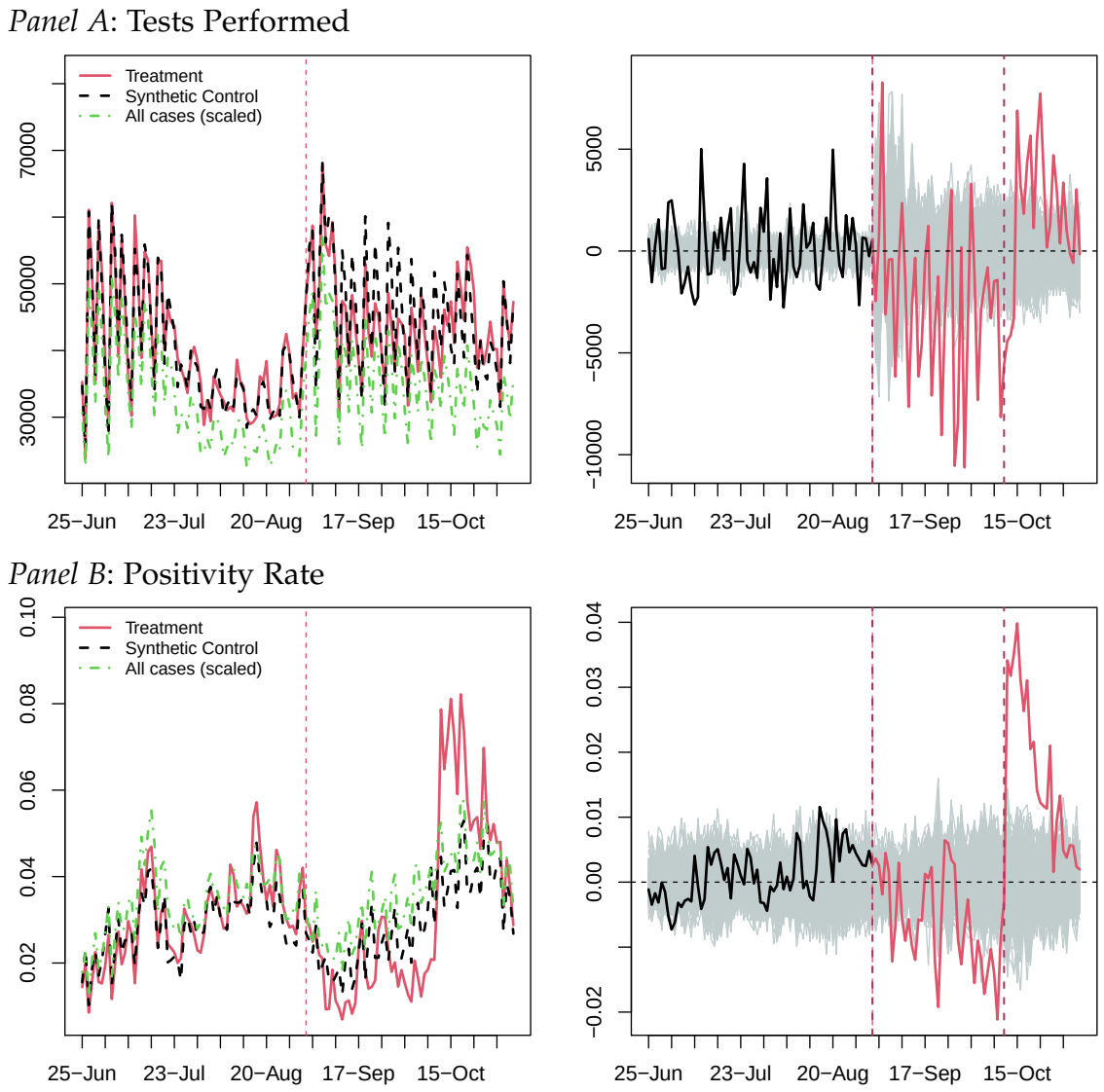
Notes: Figure plots the share of COVID-19 tests that are carried out as part of the pillar 2 community testing program across the thirteen most affected districts. The light grey line indicates the share of positive tests in the rest of England. we note that in most instances the share of positive tests declines significantly during the period that the lab was producing false negatives and subsequently results in a sharp increase in the test positivity after the lab operations were suspended.

Figure 5: Impact of False Positives in South West on COVID-19 Cases Detected Measured as New Cases By Specimen Date



Notes: Figure plots the results from the synthetic control estimates studying the number new COVID-19 reported cases by specimen date. The left figure plots the constructed synthetic control group and the treated districts documenting that prior to the testing error the two move closely together. In the right panel the difference between the number new cases by specimen date between the treated areas and the synthetic control is plotted out. In addition permutation placebo tests that assign placebo treatments to the set of control group districts are presented as light grey lines. We note that the number of new cases by specimen date in treated districts is significantly lower from September 2 to October 11 when the laboratory was producing false negatives. From October 12 onwards there is a notable increase in the number of positive cases being reported. This is used to arrive at an estimate of the overall likely impact that the testing error had on COVID-19 cases.

Figure 6: Impact of False Positives in South West on Epidemiological Dynamic



Notes: Figure plots the results from the synthetic control estimates studying the number COVID-19 tests carried out by district and day in Panel A and the share of these tests that return a positive result in Panel B. The left column plots the time series of the constructed synthetic control group against the time series for the treated districts. In addition permutation placebo tests that assign placebo treatments to the set of control group districts are presented as light grey lines. We note that both the number of tests performed and the positivity rate of those tests in treated districts is significantly lower from September 2 to October 11 when the laboratory was producing false negatives. From October 12 onwards there is a notable increase in the number of tests and the positivity rate.

Table 1: Estimated Impact By District

	Pre	α_{miss}	α_{surplus}	Δ_{lower}	Δ_{upper}
Gloucester	40.92	-1078.47	1218.65	0.13	1.13
South Gloucestershire	517.43	-2317.92	3160.34	0.36	1.36
West Berkshire	43.95	-1033.83	768.34	-0.26	0.74
North Somerset	89.56	-1622.48	1533.62	-0.05	0.95
Swindon	69.14	-599.14	2373.68	2.96	3.96
Somerset West and Taunton	155.76	-604.78	1055.84	0.75	1.75
Tewkesbury	13.23	-608.91	1562.84	1.57	2.57
Bristol, City of	2963.94	-3017.00	3833.97	0.27	1.27
Cheltenham	127.96	-671.11	1748.07	1.60	2.60
Bath and North East Somerset	270.42	-564.52	2814.02	3.98	4.98
Cotswold	47.27	-245.75	658.04	1.68	2.68
Sedgemoor	406.06	-881.62	277.68	-0.69	0.31
Stroud	164.87	-988.21	1597.18	0.62	1.62
Combined		-14233.74	22602.27	0.59	1.59

Notes: Table illustrates the estimates of the testing error across the individual districts when constructing a separate synthetic control group for each of the 13 districts that have been particularly impacted by the testing errors based on community and press reporting. The column Pre indicates the quality of the fit of the synthetic control, the closer this is to zero the better is the goodness of fit.

Appendix to "Measuring the Epidemiological Impact of a False Negative: Evidence from a Natural Experiment"

For Online Publication

Figure A1: Illustration of Measure of Pandemic Impact of False Negative Error

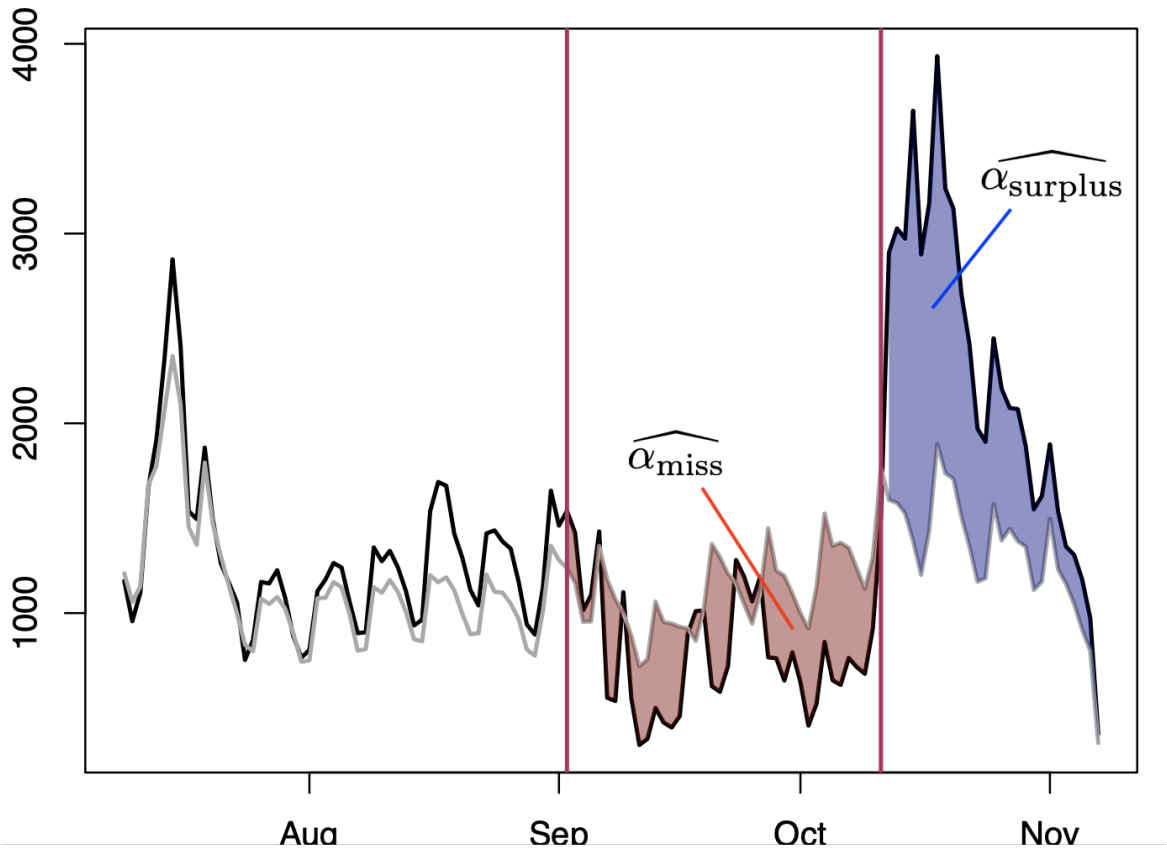
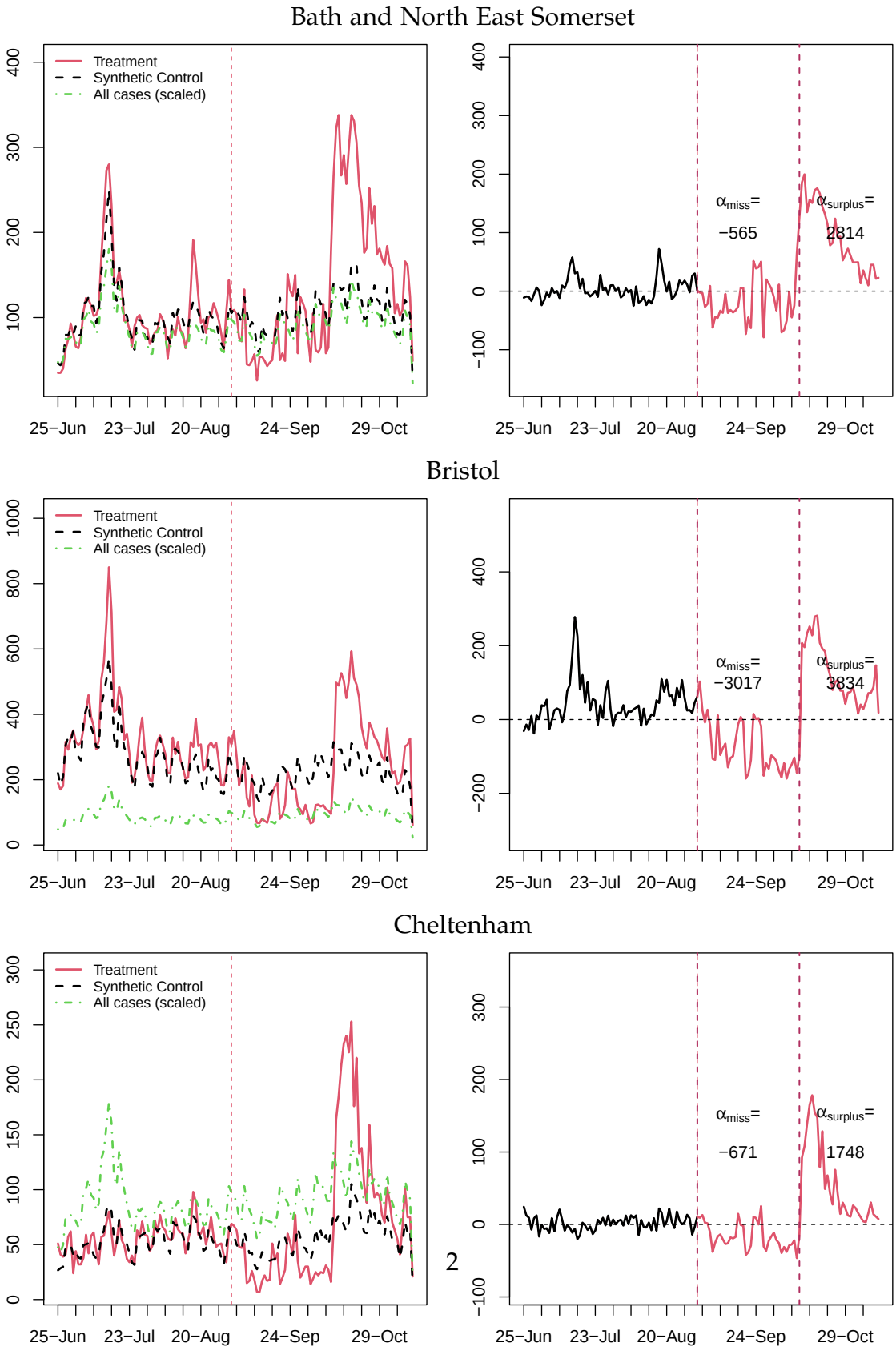
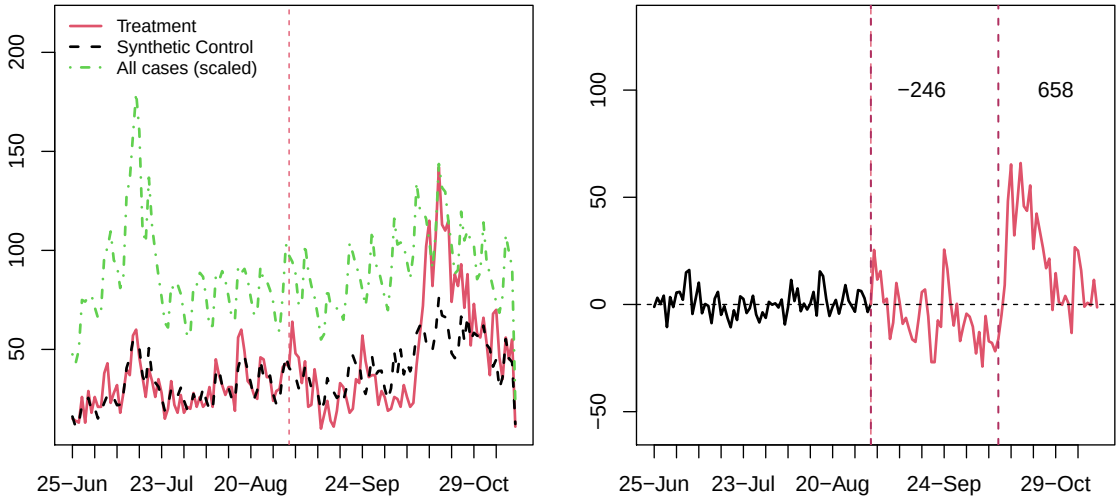


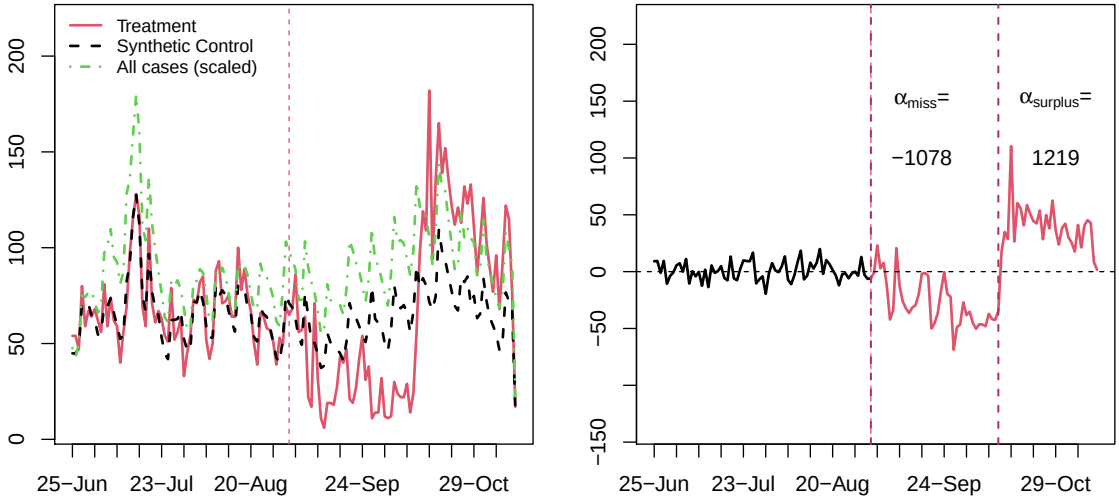
Figure A2: Impact of False Negatives on COVID-19 cases across most treated districts



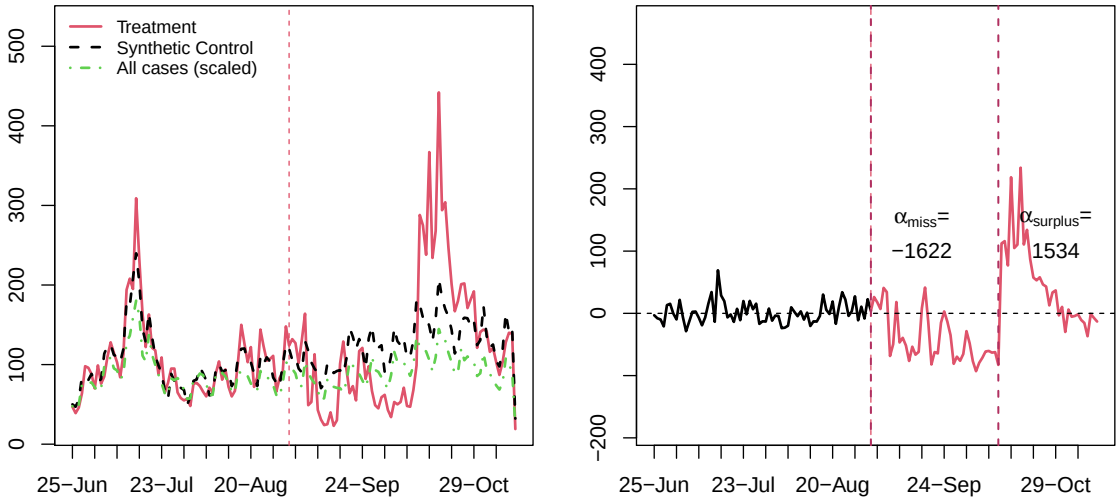
Cotswold



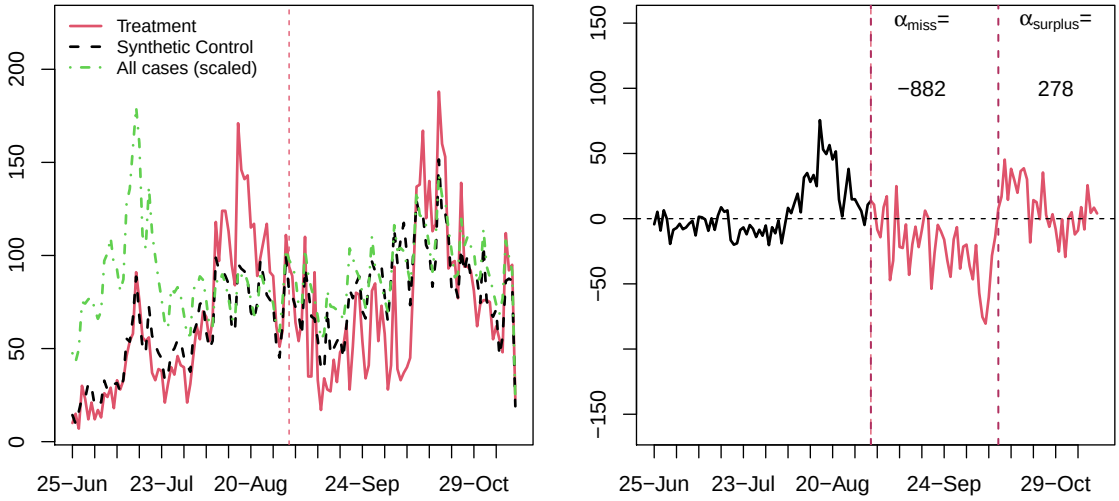
Gloucester



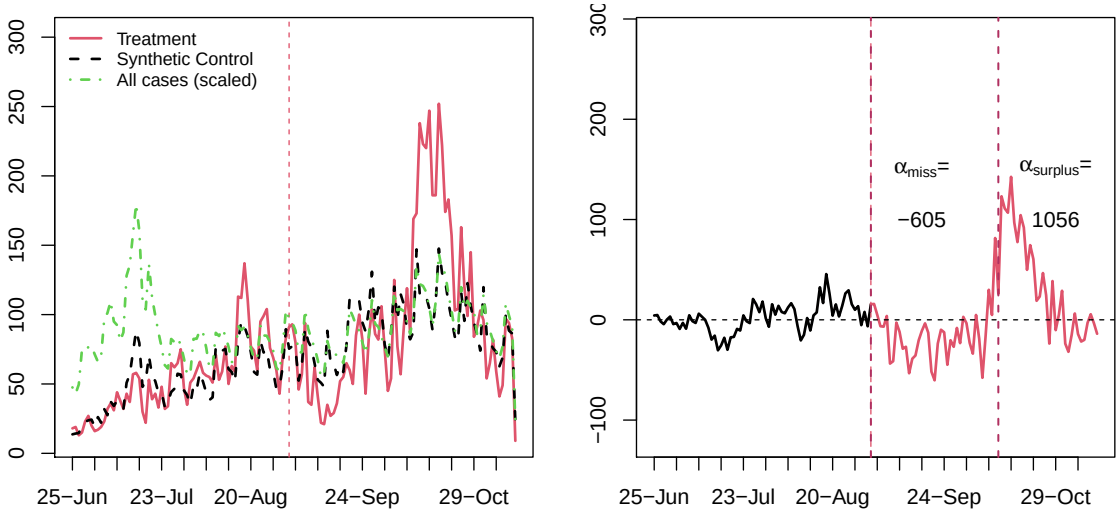
North Somerset



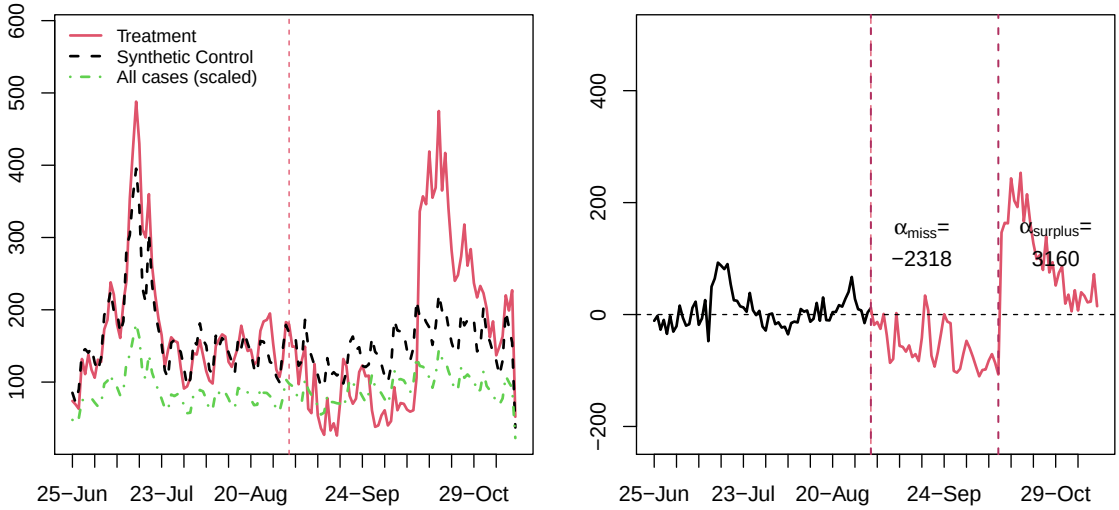
Sedgemoor



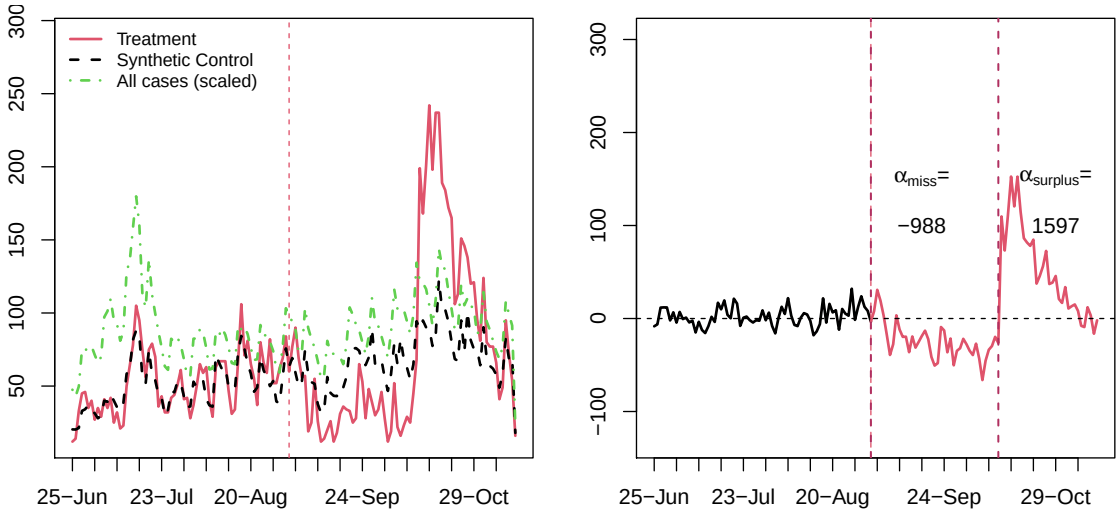
Somerset West and Taunton



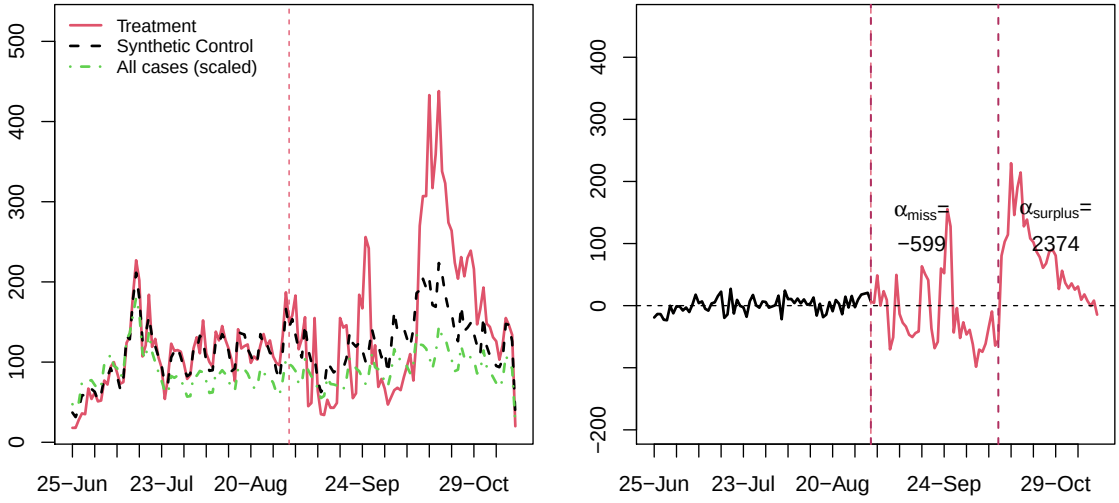
South Gloucestershire



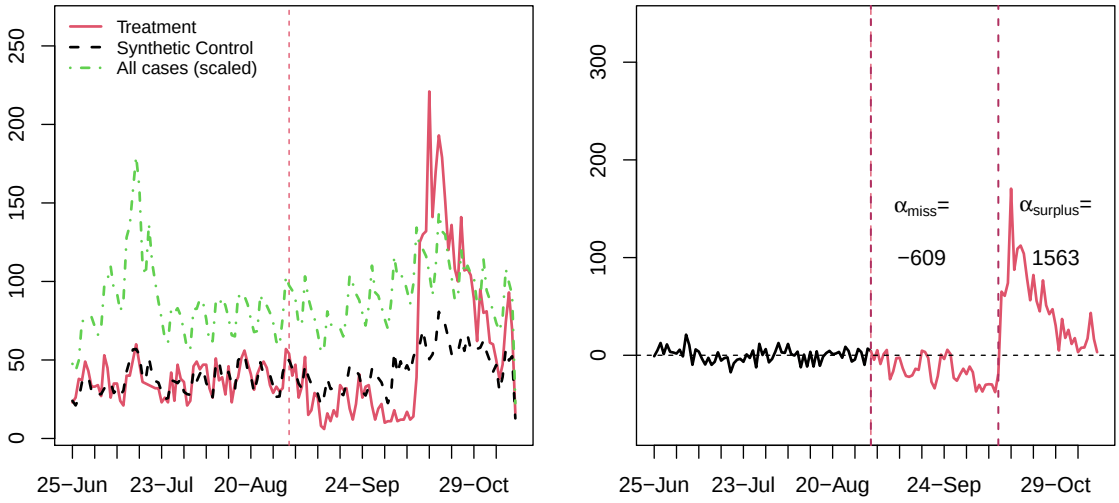
Stroud



Swindon



Tewkesbury



West Berkshire