Estimating the proportion of SARS-CoV-2 infections reported through diagnostic testing

Ewan Colman[*1], **Jessica Enright**[2], **Gavrila A. Puspitarani**[1], **and Rowland R. Kao**[†1]

[1]Royal (Dick) School of Veterinary Studies and the Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, UK
[2]School of Computing Science, University of Glasgow, Glasgow, UK

The number of positive diagnostic tests for SARS-CoV-2 is a critical metric that is commonly used to assess epidemic severity and the efficacy of current levels of control. However, a proportion of individuals infected with SARS-CoV-2 may never receive a diagnostic test, while many of those who are tested may receive a false negative result. Consequently, cases reported through testing of symptomatic individuals represent only a fraction of the total number of infections, and this proportion is expected to vary depending on changes in natural factors and variability in test-seeking behaviour. Here we combine a number of data sources from England to estimate the proportion of infections that have resulted in a positive diagnosis. Using published estimates of the incubation period distribution and time-dependent test sensitivity, we estimate SARS-CoV-2 incidence from daily reported diagnostic test data. By calibrating this estimate against surveillance data we find that approximately 25% of infections were consistently reported through diagnostic testing before November 2020. This percentage increased through the final months of 2020, predominantly in regions with a large presence of the the UK variant of concern (VOC), before falling rapidly in the last two weeks of January 2021. These changes are not explained by variation in rates of lateral flow device or PCR testing, but are consistent with there being an increased probability for the VOC that infection will result in an eventual positive diagnosis.

Testing for SARS-CoV-2 in the UK aims to accomplish two things - first, to rapidly confirm suspected cases of COVID-19 disease via symptomatic testing in order to contain outbreak clusters, and second, to establish the overall burden of infection by taking a random sample of the population. So long as the number of false positive results is sufficiently low, the number of positive symptomatic tests provides a *lower* estimate of the number of people exposed to the virus. In contrast, random testing can provide an accurate estimate of prevalence, but is an inefficient way to rapidly identify infection clusters, and may also have biases depending on the extent to which a positive test indicates the true infection status of the individual.

Both types of data are available in England: the number of positive tests from people with suspected infection are published daily by the UK government [1], and the Office for National Statistics (ONS) regularly publishes estimates of the population prevalence based on unbiased sampling [2]. The existence of these sources creates an opportunity to answer an important question question: *what proportion of all infections are being reported through diagnostic testing?* Knowing this can help to estimate true incidence rates, a quantity central to modelling the spread of the virus, and determine the infection fatality rate of the disease [3].

Here we describe how diagnostic case numbers can be used to estimate the proportion of the population testing positive, and, by calibrating this against ONS surveillance data, estimate the proportion of infections that were reported through diagnostic testing. Our

[*]ecolman@ed.ac.uk
[†]rowland.kao@ed.ac.uk

method uses published estimates the time-dependent PCR test sensitivity, however, daily case data also include rapid *in situ* tests conducted with lateral flow devices (LFDs), typically used for the screening of non-symptomatic people in high-risk environments (e.g. schools, care homes). There is concern that LFDs have lower sensitivity than PCR tests, particularly when administered by members of the general public [4].

Further complications come from new variants of SARS-CoV-2, which have been shown to result different clinical symptoms [5,6]. New variants might have different pathological characteristics, a difference in incubation period or rate of viral shedding for example, that effect the likelihood of receiving a positive test. One such difference might be a different rate of test-seeking behaviour, which we expect to directly affect the proportion of infections being reported through diagnostic testing. Examination of regional and temporal variation in the proportion of cases reported provides a novel way to consider these developments, and to enrich our understanding of the epidemiology of the virus.

1 Data

We are primarily concerned with daily Pillar 1 and 2 case data [1], hereafter referred to as diagnostic test cases, shown in Fig. S2. We use the following notation:

$$C(t) \;=\; \text{reported Pillar 1 and 2 cases on day } t.$$

These cases counts come from lab-based PCR tests and lateral flow device (LFD) testing, as performed in many community settings [7]. We assume the PCR tests have consistent sensitivity across all uses. Our analysis uses the 7-day rolling average number of people teted via PCR, and the number of LFD tests conducted, shown in Fig. S1. At the time of writing, the number of cases detected by LFDs were not available. Population counts for the 9 regions of England were taken from the ONS [8].

From the ONS infection survey [2], we use the reported percentage of people in England testing positive for coronavirus, and apply it to the population of the area to give

$$M(t) \;=\; \text{number of test-positive people}$$

where t is the 4th day of the 7-day period during which the samples were taken. Additionally, for the 9 regions of England we use estimates that differ from the national percentage due to additional modelling to interpolate between time points and correct for biases [9]. Since the latest release only contains the most recent time point for these data, the time series we use is compiled from the archived data releases. These are plotted in Fig. S2. We note a potential inconsistency in these data due to the selection of different modelling methods at different points in time.

Finally, the infection survey provides an estimate of the proportion of tests that achieve different testing targets using the TaqPath test [10]. We consider tests that are negative for the S target gene and positive for the two other targets to be a proxy for infections that derive from the B.1.1.7 lineage of the virus, hereafter referred to as the variant of concern (VOC). Tests that are positive for only one of the other targets may indicate the new variant or another lineage [11], and so we define

$$V(t) \;=\; \text{proportion of infections that are the VOC}$$

as the estimated proportion of tests that are negative on the S target after removing those that are negative on the S target and one other target. Since the S dropout can occur by chance, this characterisation is not appropriate when the prevalence of the VOC is very low. We assume the proportion to be 0 before November 1st. Additionally, the mean cycle threshold of PCR-positive tests used in the ONS survey are provided, shown in Fig. S3.

2 Estimating the time of exposure from the time of positive test

We define two random variables, x and t^+, representing the time an individuals was exposed and the time they received a positive test, respectively. Assuming discrete time steps, we express the probability that an individual who received a positive test from a sample taken at time T was first exposed to the virus at time X,

$$P(\ x = X \mid t^+ = T\ ) = \frac{P(\ x = X\ \&\ t^+ = T\ )}{\sum_{i \geq 0} P(\ x = T - i\ \&\ t^+ = T\ )}. \tag{1}$$

The joint probability distribution $P(\ x = X\ \&\ t^+ = T\ )$ can be pieced together from various sources by considering the sequence of events that result in an individual testing positive.

First, we consider the time the individual was exposed to the virus and acquired the infection. We denote the prior probability that the infection was acquired at time X with $B(X)$. Next, we consider the time between exposure and the time that they received a test. For symptomatic cases we assume that the test occurs shortly after symptom onset, i.e. the time since exposure $\tau = T - X$ is equal to the sum of the incubation period and a delay parameter δ that we assume to be a fixed quantity. The probability of a test on day T, is thus $R(T - \delta - X)$ where $R(i)$ is the probability that the duration of the incubation period is i, which we assume to be Log-normal with a mean of 5.5 days and dispersion parameter 1.52 [12].[1]

Lastly, once the individual has acquired the infection and has had a test, the test must be positive to become a recorded infection. The probability of testing positive varies as a function, $S(\tau)$, of the time since exposure $\tau = T - X$; here we use the function provided by Hellewell et al. [13] and shown in Figure 1. The curve is similar in shape to the shedding profile found in other studies [14–16] and is largely consistent with the literature on viral shedding according to the most current review on the subject [17], which reports viral load peaks at day 3-5 and a mean viral shedding duration of 17 days. Variation is associated with age and severity of illness. Studies that test for a difference between asymptomatic and symptomatic infections do not report consistent results [15, 18], we therefore assume the same sensitivity profile for all infections.

We express the probability that an infected individual was exposed on day X and tested positive on day T by multiplying the probabilities mentioned above,

$$P(\ x = X\ \&\ t^+ = T\ ) = B(X)R(T - X - \delta)S(T - X), \tag{2}$$

which can be substituted into Eq. (1).

3 Estimating the reporting rate

We define the *reporting rate* as the proportion of SARS-CoV-2 infections that result in a PCR-positive test through the government Pillar 1 and 2 testing programmes. To account for the unreported infections we assume a multiplicative scaling factor

$$\theta \ = \ \text{proportion of infections reported through diagnostic testing}$$

where θ accounts for both test sensitivity and the proportion of infected individuals who seek testing.

[1] To get a probability distribution expressing the length of the incubation period in discrete days, we integrate over consecutive intervals of length 1

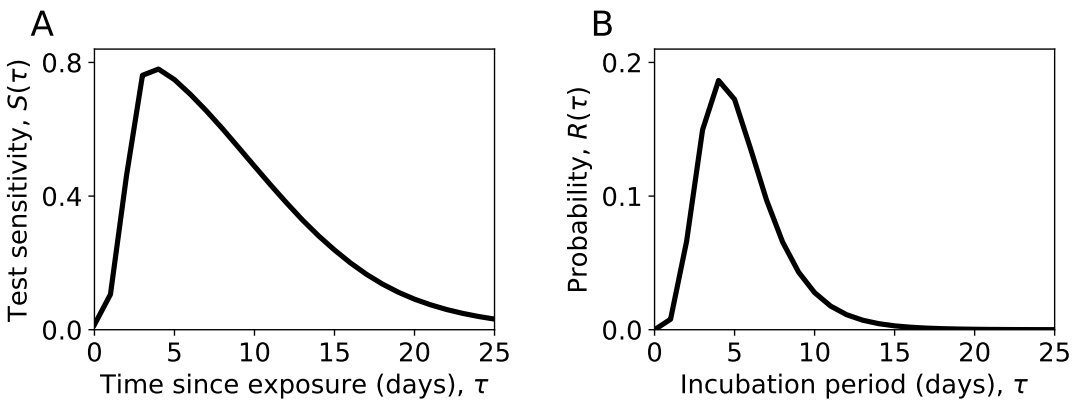

Figure 1: (A) The test sensitivity as a function of time from Hellewel et al. [13]. The function, $S(\tau)$ gives the probability that a PCR test will be positive if performed on an infected person τ days after exposure. (B) The incubation period probability distribution from Lauer et al. [12], shown here are the probability of symptom onset on each day since exposure.

To estimate the incidence, $I(t;\theta)$, defined as the number of newly acquired infections on day t, let us first consider that the number of cases exposed on day t and tested positive on day, $t+j$ is approximately $P(\ x = t \mid t^+ = t+j\ )C(t+j)$. The first part of this expression is given by Eq. (1) which in turn depends on Eq. (2). To resolve this, we assume an uninformative prior $B(t) = \beta$ where β is a constant whose value is inconsequential to the final result,[2] and assume a 1-day delay between symptom onset and receiving a test ($\delta = 1$). Finally, to account for the unreported cases we multiply by $1/\theta$. Summing over all possible days in which the exposure could be reported gives

$$I(t;\theta) = \frac{1}{\theta}\sum_{j\geq 0} C(t+j) \times \frac{R(j-\delta)S(j)}{\sum_{i\geq 0} R(i-\delta)S(i)} \tag{3}$$

Next, we map the individuals exposed on each day to the number of reported cases that would test positive on a given future day t,

$$\hat{M}(t;\theta) = \sum_{k=0} I(t-k;\theta)S(k). \tag{4}$$

For a given point in time, t, solving $\hat{M}(t;\theta) = M(t)$ for θ gives a time-dependent estimate for θ,

$$\theta(t) = \frac{\sum_{k\geq 0} S(k) \sum_{j\geq 0} C(t-k+j)R(j-\delta)S(j)}{M(t)\sum_{i\geq 0} R(i-\delta)S(i)}. \tag{5}$$

This equation combines the daily diagnostic case counts, the population positivity from surveillance, the incubation period distribution, and the time-dependent test sensitivity, to provide an estimate of the proportion of infections being reported at time t.

3.1 Time-independent reporting rates To get a measure of the reporting rate that does not change over time we take θ to be a constant. Motivated by the suggestion that the VOC may have a different pathological characteristics to other variants, we in choose to use two reporting rates corresponding to the usual reporting rate, θ_{OLD}, and the reporting rate for cases caused by the VOC, θ_{VOC}. Recalling that $V(t)$ is the proportion of infections caused by the VOC, we consider a revised estimate of $M(t)$ using the convex combination

$$\tilde{M}(t;\theta_{OLD},\theta_{VOC}) = (1-V(t))\hat{M}(t;\theta_{OLD}) - V(t)\hat{M}(t;\theta_{VOC}). \tag{6}$$

[2]This assumes that the probability of exposure is constant over the time period for which the individual could feasibly have been exposed given they tested positive at time t

We can then estimate by taking the value that minimizes the absolute error between $M(t)$ and $\tilde{M}(t)$. Specifically, the time-independent reporting rates are estimated by numerically solving

$$\{\theta_{OLD}, \theta_{VOC}\} = \underset{x,y}{\operatorname{argmin}} \sum_t \left| M(t) - \tilde{M}(t; x, y) \right|. \tag{7}$$

Results

Reporting rates calculated using Eq. (5) are presented for the 9 regions of England in Figure 2. There is an apparent difference when comparing the four southern-most regions (inc. East of England) to the others. Until the beginning of December, these regions show a general increase in the percentage of infections being reported, while the others show a decrease. While the number of PCR tests conducted increased steadily in all regions and LFD testing had not yet begun (Fig. S1), the correlation between testing and the reporting rate over the whole time period is positive in southern regions and negative for those in the north (Table S1).

Reporting rates increased during November and December 2020 in all regions. While there may be various behavioural factors causing this increase, we note that the VOC increased considerably in proportion to other variants during this time period. Since there is some evidence suggesting that the VOC has a higher severity of clinical symptoms [5,6], it is reasonable to suggest it may also have a reporting rate distinct from the average of other variants.[3] Secondly, over the same time period LFD testing, typically used for screening non-symptomatic individuals, became widely accessible. Since asymptomatic infections make up an estimated 20% of all infections [20], we would expect this scheme to have increased the reporting rate.

To test both possibilities, we compute the time-independent rates for the VOC and other variants, calculated using the method described in section 3.1. We include the effect of LFD tests by estimating the number of LFD positives and subtracting these from the

[3] One study of self-reported symptoms did not find support for this [19].

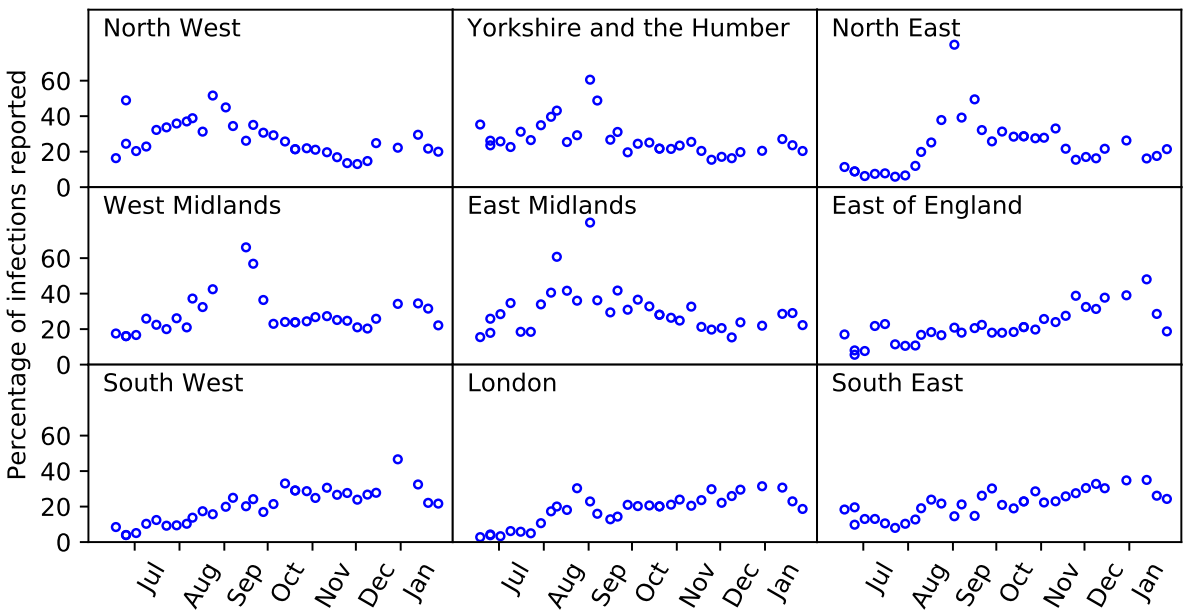

Figure 2: Reporting rate in the 9 regions of England expressed as a percentage. Each point represents the estimate, $\theta(t)$ given by Eq. (5) evaluated at each point in time for which surveillance data have been published.

total Pillar 1 and 2 cases before repeating the analysis. In doing so, the reporting rates θ_{OLD} and θ_{VOC} become an estimate of the proportion of infections that are reported with a positive PCR test (for the old variants and new variant, respectively). In doing so, we are assuming independence between the testing regimes, i.e. ignoring any interaction between the two. Since at the time of writing the number of LFD positives are not published, we explore the impact of having a fixed percentage of all LFD tests being positive (Tables 1 and S2).

We find that larger LFD positivity rates result in lower reporting rates for the VOC, however, the conclusion that reporting rate is higher for the new variant compared to older variants holds for a wide range of assumed LFD positivity rates. The best fit to the surveillance data occurs when 0% of all LFD tests are positive. This does not necessarily imply that all LFD tests are negative, only that the new variant has considerably more power to explain the increases in reporting rate observed since November 2020, unless the proporton of LFD tests that are positive changes substantially over time. In all cases, the estimated reporting rate in England for variants that are not the VOC is 25%, with regional variation shown in Figure 3. In the best fitting case (LFD positivity = 0) the reporting rate for the VOC is 32%. Estimates of θ_{VOC} are smallest in regions where $V(t)$ provides a relatively small amount to the overall infections, having little impact on the outcome compared to other causes of variation.

In addition to a change in reporting rate, it is possible that the test-sensitivity for the VOC is substantially different from the assumed $S(\tau)$. For example, infections could be more likely to have been detected more than 3 weeks after exposure; a higher proportion of these "historical"infections would then be included in the ONS population positivity estimates. If we assume that a higher mean cycle threshold implies a higher probability of being a historical infection, it follows that the mean cycle threshold for the PCR tests used in the ONS infection survey indicate the abundance of historical infections in the data. We observe that the times for which our estimate is lower than the ONS figure are times when ONS cycle threshold values are highest, and reporting rates are negatively correlated with the ONS cycle threshold values in several regions, adding support to this hypothesis.

Table 1: Reporting rates for old variants and the VOC under different LFD positivity assumptions for England. The residual sum is the sum of absolute differences between the estimated percentage and the percentage given by the ONS infection survey.

LFD positivity	θ_{OLD}	θ_{VOC}	Residual sum
0%	0.25	0.32	2.28
1%	0.25	0.31	2.44
2%	0.25	0.31	2.59
3%	0.25	0.3	2.76
4%	0.25	0.3	2.91
5%	0.25	0.29	3.1
6%	0.25	0.28	3.37
7%	0.25	0.27	3.64
8%	0.25	0.25	3.93
9%	0.25	0.27	4.24

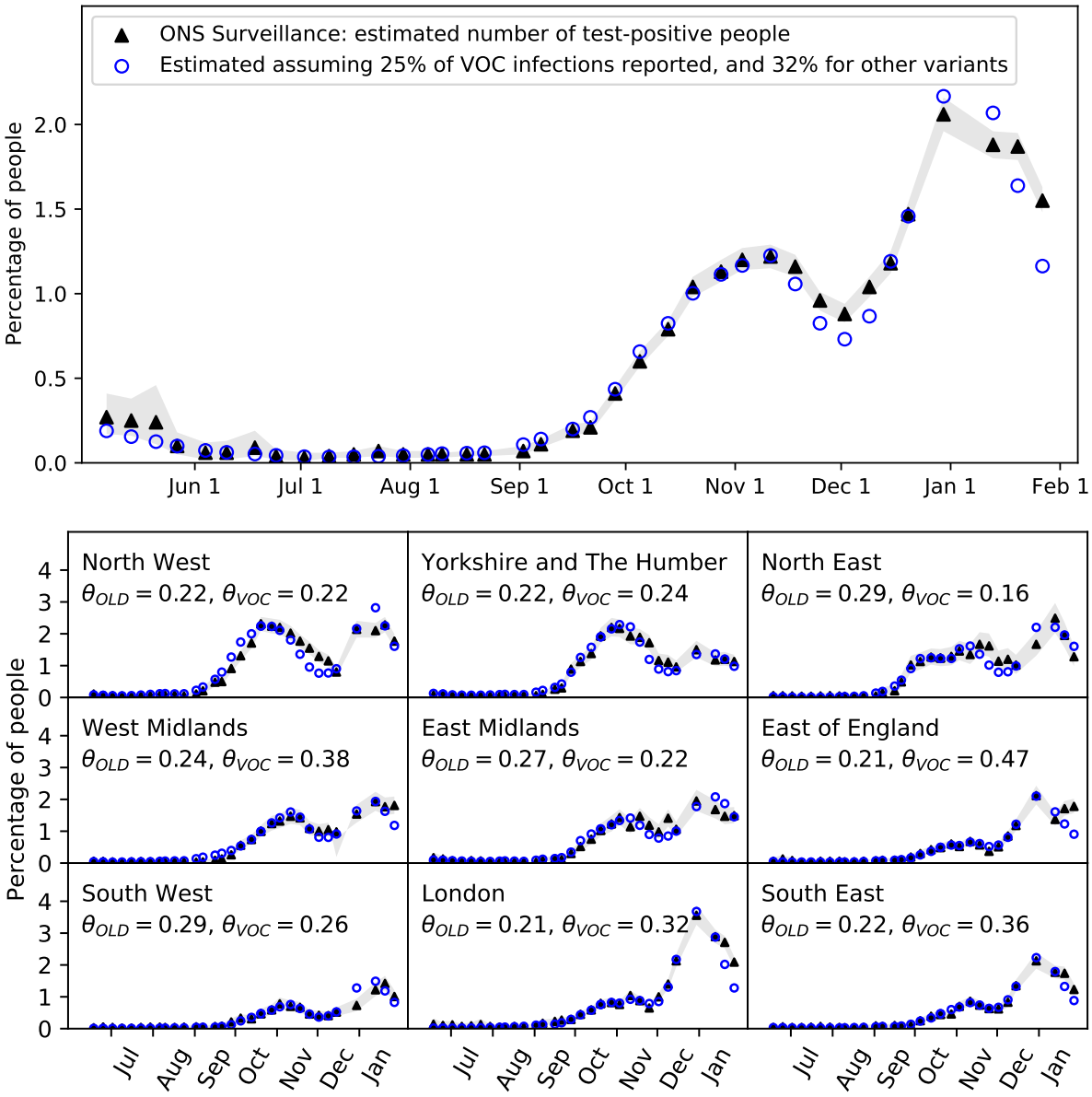

Figure 3: The number of SARS-CoV-2 test-positive people in England (top) and in each of the 9 regions of England (bottom). Blue circles show the estimate given by Eq.(6), black triangles show the ONS infection survey data.

Conclusion

In this analysis, we have used a combination of two important data streams to estimate the true incidence of SARS-CoV-2 infections in England over time. This approach would allow us to directly estimate other quantities of epidemiological importance for COVID-19, such as the infection fatality rate, and, as a methodology, could be used in future to provide early estimates for novel emerging pathogens provided estimates of incubation periods and test sensitivity over time are available.

Our analysis reveals considerable variation across regions of England in the proportion of infected individuals testing positive. Since this variation is unlikely to be the result of biological differences, we have reason to suspect differences in test-seeking behaviour, with regions in the south of England and those in the North showing markedly different patterns. A particular concern is the evidence of a decline in test-seeking behaviour in the Northern regions from the autumn onwards, that might compromise the ability to control the spread of the virus in these regions. The reasons why may be a combination

of differences in attitudes, messaging, accessibility of tests, and natural epidemiological phenomena. Confirming this observation and, if confirmed, understanding why will be important for assessing future directions of control.

There is some evidence to suggest that the new variant has a higher reporting rate than other variants, however other explanations for the increase in $\theta(t)$ should also be considered. If we assume that it is only the VOC causing the observed change, then approximately 32% of new variant infections are reported with a positive test compared to 25% for other variants. These results may be change over time as new data continues to become available. Nonetheless, the sudden deviation from the reliable pattern raises questions about other possible changes that have occurred that have not been observed until now. The impact of the VOC on testing data is perhaps a warning that what we think we know about SARS-CoV-2 may become redundant over time as the virus continues to evolve.

References

[1] Uk coronavirus dashboard: cases. https://coronavirus.data.gov.uk/details/cases.

[2] Office for national statistics infection survey. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/latest.

[3] Gideon Meyerowitz-Katz and Lea Merone. A systematic review and meta-analysis of published research data on covid-19 infection-fatality rates. *International Journal of Infectious Diseases*, 2020.

[4] Jacqui Wise. Covid-19: Lateral flow tests miss over half of cases, liverpool pilot data show, 2020.

[5] Pengfei Wang, Lihong Liu, Sho Iketani, Yang Luo, Yicheng Guo, Maple Wang, Jian Yu, Baoshan Zhang, Peter D. Kwong, Barney S. Graham, John R. Mascola, Jennifer Y. Chang, Michael T. Yin, Magdalena Sobieszczyk, Christos A. Kyratsous, Lawrence Shapiro, Zizhang Sheng, Manoj S. Nair, Yaoxing Huang, and David D. Ho. Increased resistance of sars-cov-2 variants b.1.351 and b.1.1.7 to antibody neutralization. *bioRxiv*, 2021.

[6] Coronavirus (covid-19) infection survey: characteristics of people testing positive for covid-19 in england, 27 january 2021. https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/articles/coronaviruscovid19infectionsinthecommunityinengland/characteristicsofpeopletestingpositiveforcovid19inengland27january2021.

[7] Uk coronavirus dashboard: tests. https://coronavirus.data.gov.uk/details/tests.

[8] Office for national statistics population estimates. https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland.

[9] Koen B Pouwels, Thomas House, Julie V Robotham, Paul J Birrell, Andrew Gelman, Nikola Bowers, Ian Boreham, Heledd Thomas, James Lewis, Iain Bell, John I Bell, John N Newton, Jeremy Farrar, Ian Diamond, Pete Benton, Ann Sarah Walker, and COVID-19 Infection Survey team. Community prevalence of sars-cov-2 in england: Results from the ons coronavirus infection survey pilot. *medRxiv*, 2020.

[10] Investigation of novel sars-cov-2 variant. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/950823/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3_-_England.pdf.

[11] New-variant compatibility in the ons infection survey. https://theo.io/post/2021-01-22-ons-data/. Accessed on 01/02/2021.

[12] Stephen A Lauer, Kyra H Grantz, Qifang Bi, Forrest K Jones, Qulu Zheng, Hannah R Meredith, Andrew S Azman, Nicholas G Reich, and Justin Lessler. The incubation period of coronavirus disease 2019 (covid-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine*, 172(9):577–582, 2020.

[13] Joel Hellewell, Timothy W. Russell, Rupert Beale, Gavin Kelly, Catherine Houlihan, Eleni Nastouli, and Adam J. Kucharski. Estimating the effectiveness of routine asymptomatic pcr testing at different frequencies for the detection of sars-cov-2 infections. *medRxiv*, 2020.

[14] Xi He, Eric HY Lau, Peng Wu, Xilong Deng, Jian Wang, Xinxin Hao, Yiu Chung Lau, Jessica Y Wong, Yujuan Guan, Xinghua Tan, et al. Temporal dynamics in viral shedding and transmissibility of covid-19. *Nature medicine*, 26(5):672–675, 2020.

[15] Quan-Xin Long, Xiao-Jun Tang, Qiu-Lin Shi, Qin Li, Hai-Jun Deng, Jun Yuan, Jie-Li Hu, Wei Xu, Yong Zhang, Fa-Jin Lv, et al. Clinical and immunological assessment of asymptomatic sars-cov-2 infections. *Nature medicine*, 26(8):1200–1204, 2020.

[16] Roman Wölfel, Victor M Corman, Wolfgang Guggemos, Michael Seilmaier, Sabine Zange, Marcel A Müller, Daniela Niemeyer, Terry C Jones, Patrick Vollmar, Camilla Rothe, et al. Virological assessment of hospitalized patients with covid-2019. *Nature*, 581(7809):465–469, 2020.

[17] Muge Cevik, Matthew Tate, Ollie Lloyd, Alberto Enrico Maraolo, Jenna Schafers, and Antonia Ho. Sars-cov-2, sars-cov-1 and mers-cov viral load dynamics, duration of viral shedding and infectiousness: a living systematic review and meta-analysis. *SARS-CoV-1 and MERS-CoV Viral Load Dynamics, Duration of Viral Shedding and Infectiousness: A Living Systematic Review and Meta-Analysis*, 2020.

[18] Stephen M. Kissler, Joseph R. Fauver, Christina Mack, Scott W. Olesen, Caroline Tai, Kristin Y. Shiue, Chaney C. Kalinich, Sarah Jednak, Isabel M. Ott, Chantal B.F. Vogels, Jay Wohlgemuth, James Weisberger, John DiFiori, Deverick J. Anderson, Jimmie Mancell, David D. Ho, Nathan D. Grubaugh, and Yonatan H. Grad. Sars-cov-2 viral dynamics in acute infections. *medRxiv*, 2020.

[19] Mark S Graham, Carole H Sudre, Anna May, Michela Antonelli, Benjamin Murray, Thomas Varsavsky, Kerstin Klaser, Liane Dos Santos Canas, Erika Molteni, Marc Modat, David Alden Drew, Long Alden Nguyen, Lorenzo Polidori, Somesh Selvachandran, Christina Hu, Joan Capdevila Pujol, Alexander Hammers, Andrew T Chan, Jonathan Wolf, Timothy Spector, Claire Steves, and Sebastien Ourselin. The effect of sars-cov-2 variant b.1.1.7 on symptomatology, re-infection and transmissibility. *medRxiv*, 2021.

[20] Diana Buitrago-Garcia, Dianne Egli-Gany, Michel J Counotte, Stefanie Hossmann, Hira Imeri, Aziz Mert Ipekci, Georgia Salanti, and Nicola Low. Occurrence and transmission potential of asymptomatic and presymptomatic sars-cov-2 infections: A living systematic review and meta-analysis. *PLoS medicine*, 17(9):e1003346, 2020.

Table S1: Coorrelation coefficint (Pearson r) for the association between each the estimated $\theta(t)$ an the other variable at the same time. These are evaluated only on days for which ONS surveillance data are available. For PCR tests the value is the rolling 7-day average corresponding to the 7-day period for which t is the 4th day, and the correlation is evaluated over all available data points. For LFD tests it is the sum over the 7-day period for which t is the 4th day, only time points after November 1st are used. For VOC proportion and cycle threshold values are found by linearly interpolating between the time points given in the source data and evaluating on day t. For VOC only points after November 1st are used, for the cycle threshold all available data points are used starting at September 21st.

	PCR tests	LFD Tests	VOC proportion	Cycle threshold
North West	−0.39*	−0.10	0.26	−0.80**
Yorkshire and the Humber	−0.45**	−0.51	−0.06	−0.37
North East	0.34*	−0.39	−0.17	−0.72**
West Midlands	−0.00	0.05	0.62	−0.59*
East Midlands	−0.15	−0.55	−0.28	−0.21
East of England	0.84**	0.61	0.79*	−0.23
South West	0.83**	0.12	0.73*	−0.38
London	0.80**	0.51	0.70	−0.41
South East	0.80**	0.94**	0.97**	−0.21

Table S2: Reporting rates for old variants and the VOC under different LFD positivity assumptions for each region of England. The residual sum is the sum of absolute differences between the estimated percentage and the percentage given by the ONS infection survey.

North West

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.22	0.22	5.46
1%	0.22	0.2	5.7
2%	0.22	0.19	5.99
3%	0.22	0.17	6.28
4%	0.22	0.15	6.62
5%	0.22	0.14	7.06
6%	0.22	0.16	7.54
7%	0.22	0.15	8.01
8%	0.22	0.14	8.52
9%	0.22	0.14	9.02

Yorkshire and The Humber

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.22	0.24	3.46
1%	0.22	0.22	3.52
2%	0.22	0.19	3.59
3%	0.22	0.17	3.68
4%	0.22	0.16	3.85
5%	0.22	0.16	4.08
6%	0.22	0.16	4.32
7%	0.22	0.15	4.58
8%	0.22	0.15	4.83
9%	0.22	0.15	5.11

North East

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.29	0.16	4.19
1%	0.29	0.15	4.07
2%	0.29	0.14	4.01
3%	0.29	0.12	4.04
4%	0.29	0.12	4.32
5%	0.29	0.11	4.61
6%	0.29	0.1	5.15
7%	0.29	0.1	5.58
8%	0.29	0.1	6.2
9%	0.29	0.09	6.74

West Midlands

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.24	0.38	2.59
1%	0.24	0.36	2.75
2%	0.24	0.34	2.92
3%	0.24	0.33	3.09
4%	0.24	0.31	3.29
5%	0.24	0.3	3.49
6%	0.24	0.28	3.72
7%	0.24	0.27	3.94
8%	0.24	0.26	4.21
9%	0.25	0.24	4.47

East Midlands

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.27	0.22	3.71
1%	0.27	0.19	3.69
2%	0.28	0.18	3.7
3%	0.28	0.18	3.7
4%	0.27	0.18	3.7
5%	0.27	0.18	3.7
6%	0.28	0.17	3.78
7%	0.28	0.16	4.05
8%	0.28	0.16	4.32
9%	0.28	0.16	4.6

East of England

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.21	0.47	2.51
1%	0.21	0.46	2.61
2%	0.21	0.46	2.73
3%	0.21	0.45	2.85
4%	0.21	0.45	2.97
5%	0.21	0.44	3.09
6%	0.21	0.44	3.2
7%	0.21	0.43	3.34
8%	0.21	0.43	3.52
9%	0.21	0.42	3.72

South West

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.29	0.26	2.11
1%	0.29	0.26	2.3
2%	0.28	0.27	2.5
3%	0.28	0.27	2.71
4%	0.29	0.25	2.93
5%	0.29	0.23	3.2
6%	0.29	0.22	3.54
7%	0.29	0.21	3.93
8%	0.29	0.21	4.33
9%	0.29	0.23	4.75

London

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.21	0.32	3.21
1%	0.21	0.31	3.39
2%	0.21	0.31	3.58
3%	0.21	0.3	3.76
4%	0.21	0.29	3.94
5%	0.21	0.29	4.16
6%	0.21	0.28	4.37
7%	0.21	0.27	4.58
8%	0.21	0.27	4.82
9%	0.21	0.31	5.05

South East

LFD Pos.	θ_{OLD}	θ_{VOC}	$\sum$ Res.
0%	0.22	0.36	1.66
1%	0.22	0.35	1.83
2%	0.23	0.33	2.05
3%	0.23	0.33	2.29
4%	0.23	0.32	2.56
5%	0.23	0.32	2.82
6%	0.23	0.32	3.1
7%	0.23	0.32	3.39
8%	0.23	0.31	3.68
9%	0.22	0.33	3.99

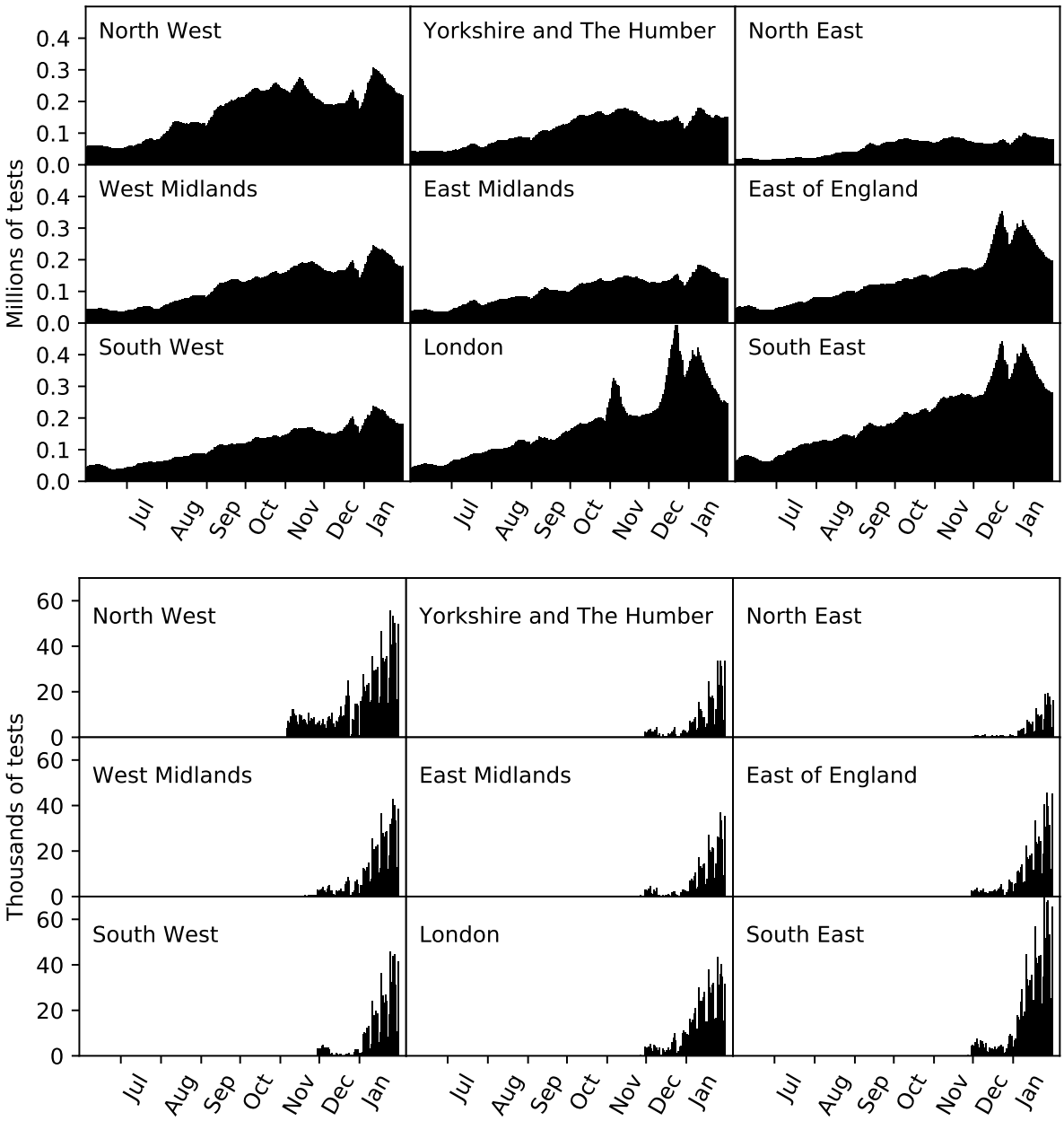

Figure S1: Top: Rolling 7-day sum of the number of Pillar 1 and 2 PCR tests performed. Bottom: daily number of Pillar 2 LFD tests performed.

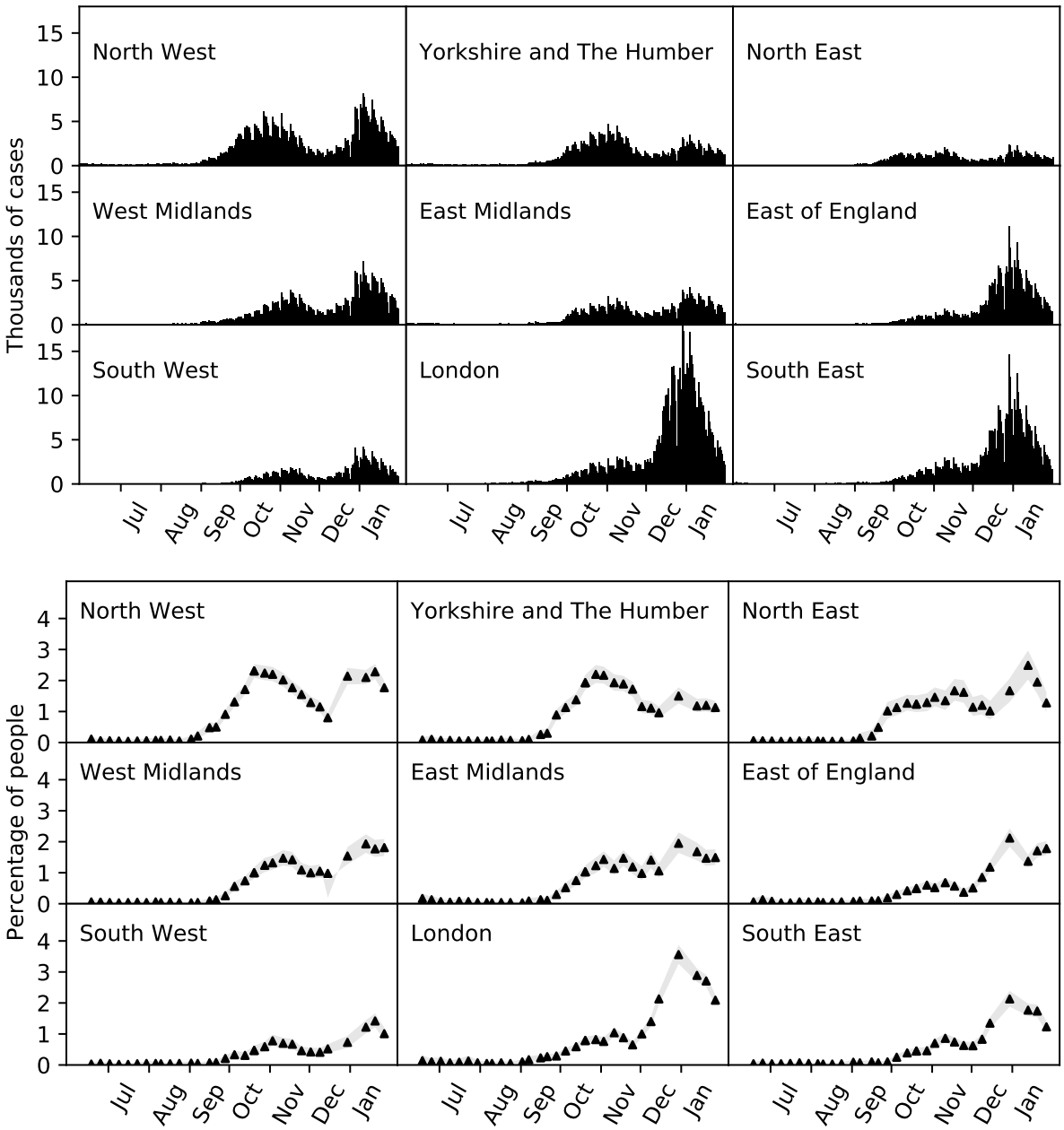

Figure S2: Top: Daily number of positive Pillar 1 an 2 cases. Bottom: percentage of population testing-positive for SARS-CoV-2 as presented in the ONS infection survey, grey region shows the 95% confidence interval.

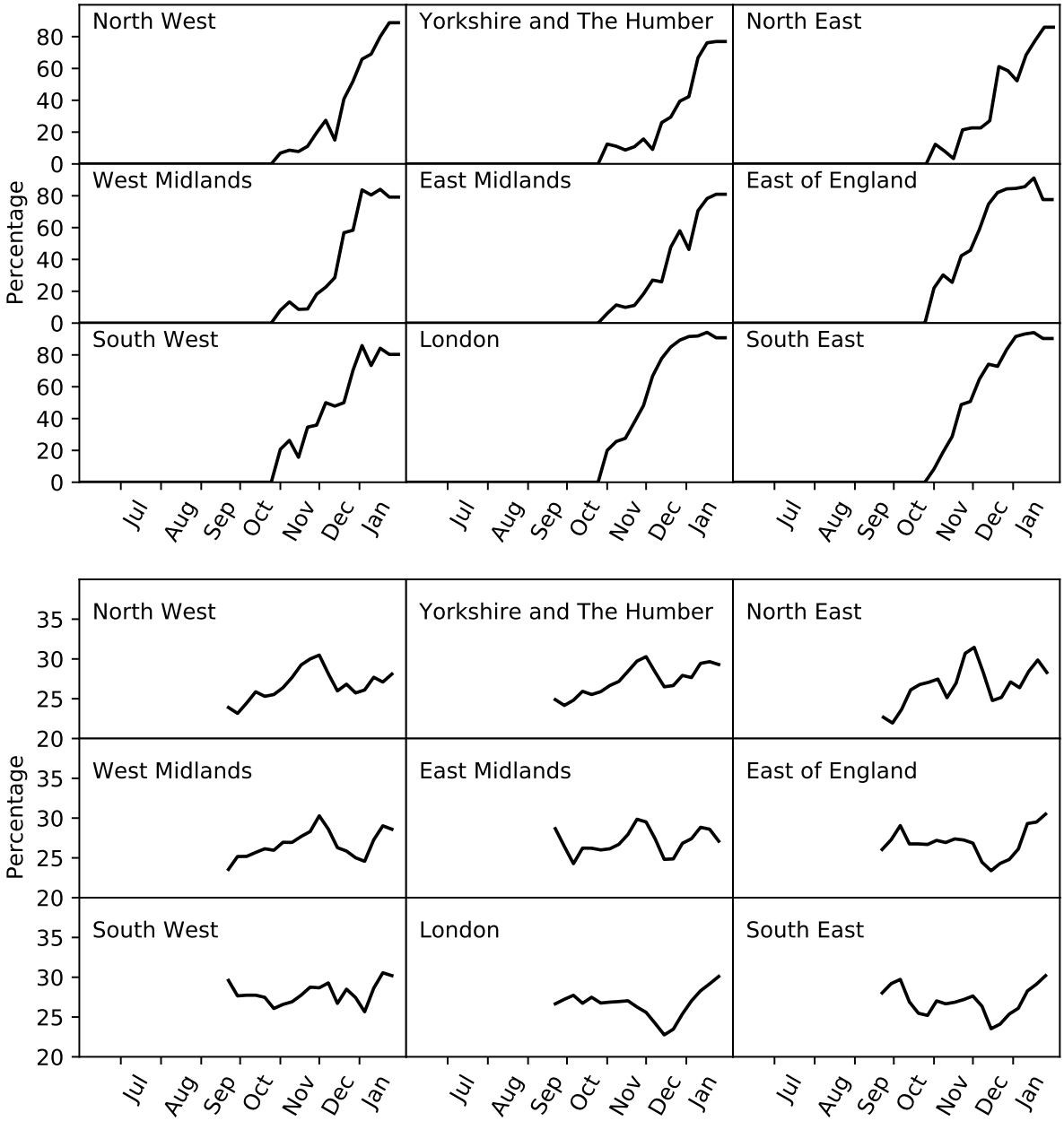

Figure S3: Top: estimated proportion of cases that are the VOC. Bottom: mean cycle threshold value of PCR tests used in the ONS infection survey.