What proportion of all SARS-CoV-2 infections in England are reported through diagnostic tests?

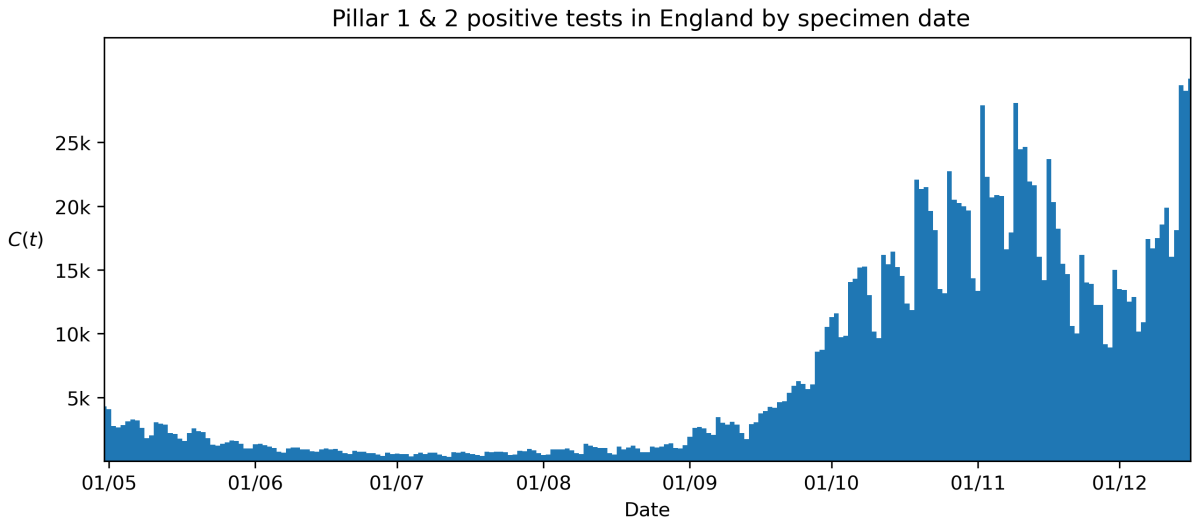
Ewan Colman*[1], Jessica Enright[2], and Rowland Kao[1]

[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

A number of sources of data exist that allow assessment of the SARS-CoV-2 situation in England. The Office for National Statistics provide weekly surveillance reports based on a random unbiased sampling, however, these are usually published a number of days after the date of collection. As such, daily case counts from diagnostic tests are often preferred, but these only include infections that result in test-seeking behaviour and miss most asymptomatic cases. Here we propose a method to map diagnostic case numbers to the total number of people who would test positive on a given day. By comparing this estimate to the surveillance data we find that for every case reported through diagnostic testing there are approximately three infections that go unreported.

Data. There are four sources of data used in this analysis. We are primarily concerned with the Pillar 1 & 2 case data provided daily on the website of the UK government [1], hereafter referred to as diagnostic test cases, and use the following notation:

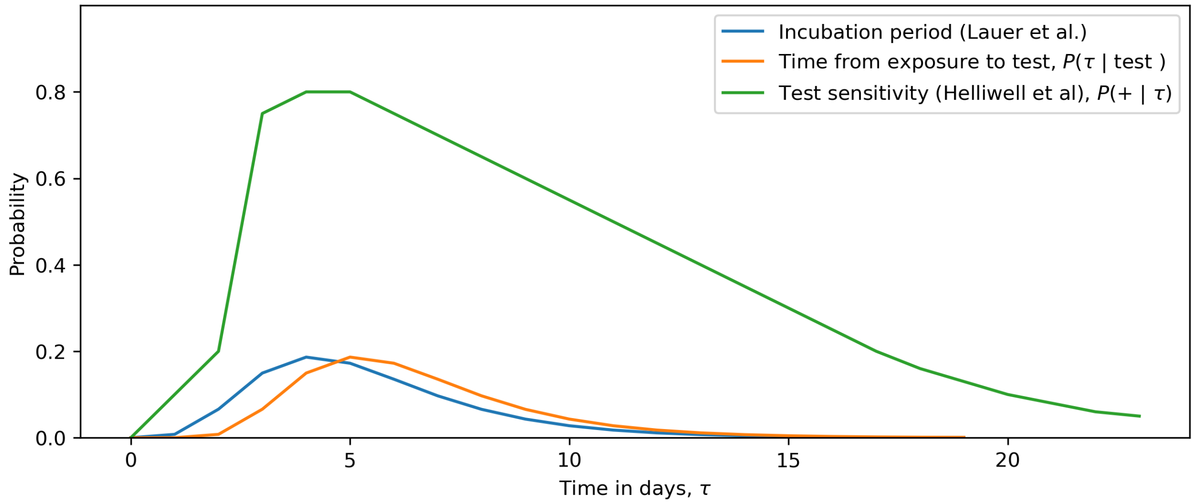
$$C(t) \;=\; \text{reported pillar 1 \& 2 cases on day } t$$



We aim to estimate the incidence (number of new infections) and the percentage of the population testing positive on a given day. We make some assumptions about the timeline of infection; for the incubation period we use the Lognormal distribution found in [2] integrated over consecutive intervals of length 1 give a discrete probability distribution pertaining to the length of the incubation period in days. We further assume

*ecolman@ed.ac.uk

that all reported cases were symptomatic and received the test one day after the onset of symptoms. Thus, the probability that someone will get a diagnostic test $\tau+1$ days after exposure is the same as the probability that the incubation period is τ days, denoted by $P(\ j \mid \text{test}\ )$. For the test sensitivity we use the result of [3] that gives the probability that a test performed on someone exposed τ days earlier will return a positive result, denoted by $P(+ \mid j)$.



The final source of data comes from the Office for National Statistics (ONS) surveillance survey [4] that reports the number of people in England testing positive for coronavirus. The data is estimated from samples taken over 7-day periods. The reported figure applies most accurately to the mid point of the sampling week. We treat this data as correct and unbiased and our goal is to approximate these data as accurately as possible by scaling up the diagnostic test data.

Method. The process for estimating the the test-positive population from diagnostic case data is made of three parts. First, we map the reported cases to their date of exposure. Using Bayes' rule we can express the probability that a reported case was exposed j days earlier as

$$\begin{aligned} P(\ j \mid +) &= \text{probability of exposure } j \text{ days before positive test} \\ &= \frac{P(+ \mid j)P(\ j \mid \text{test}\ )}{\sum P(+ \mid i)P(\ i \mid \text{test}\ )} \end{aligned} \tag{1}$$

Second, we account for the unreported infections by multiplying by a factor

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

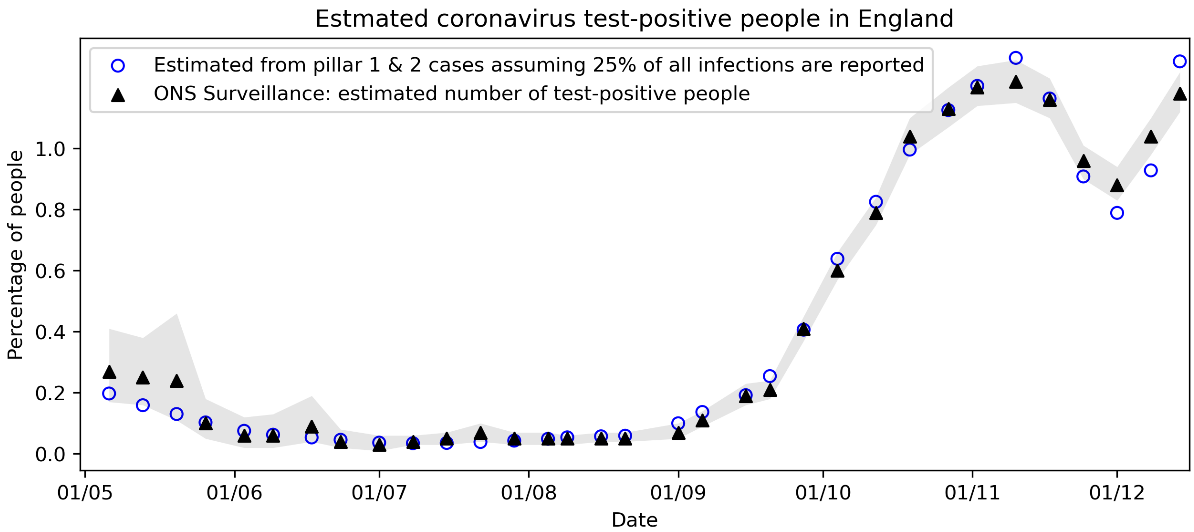
This parameter is not known but will be estimated later. The number of cases that were exposed on day t and tested on a future day, $t+j$, is approximately $P(\ j \mid +)\ C(t+j)$. Summing over all possible future days $t+j$ and scaling up by $1/\theta$ to account for undetected cases gives an estimate of the daily incidence,

$$\begin{aligned} X(t) &= \text{number of exposures on day } t \\ &= \frac{1}{\theta}\sum_{j=0} C(t+j)P(\ j \mid +) \end{aligned} \tag{2}$$

In the final step, we map the individuals exposed on each day to an estimate of the test-positivity rate (the percentage of people who would test positive in a random sample of the population) on each future day.

$$\begin{aligned} S(t) &= \text{number of test-positive people} \\ &= \sum_{k=0} X(t-k)P(+ \mid k) \end{aligned} \tag{3}$$

Results We evaluated equation (3) over a range of values of θ. To measure the error we used the absolute difference between $S(t)$ and the corresponding value given by the ONS surveillance data summed over all days for which such data is available. The best fit was found at $\theta = 4.0$ (to 1 decimal place). The quality of the fit is apparent from the figure below and we have chosen not to present any further validation at this stage. In sum, this implies that there are 4 times as many cases compared to the number reported in England - or, in other words, 25% of infections are reported with a positive diagnostic test.



Final remark. Our analysis assumes that all cases identified through diagnostic testing are symptomatic individuals who receive a test the day after symptom onset. This is unlikely to be true for all cases, but the apparent accuracy of our result suggests that the majority of cases do conform to this pattern. The result is contrary to the idea that increased testing capacity increases the number of reported cases, or it at the least shows that the effect is small in comparison. For example, if 1% are testing positive, testing an additional 10^5 asymptomatic people will only yield an additional 10^3 cases whereas diagnostic testing is currently identifying around 20 times that number.

References

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

[2] `https://www.acpjournals.org/doi/full/10.7326/M20-0504`

[3] `https://www.medrxiv.org/content/10.1101/2020.11.24.20229948v1.full.pdf`

[4] `https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveypilot/latest`