Modelling pooling strategies for SARS-Cov-2 RT-PCR testing in a university setting

Authors: UoB COVID modelling group

# Abstract

* Pooling by social cluster could improve pooling performance to some extent, though

# Introduction

In the midst of the COVID-19 pandemic, university students represent a demographic in the population who are likely to have experience low rates of symptomatic infection while being in a high contact social setting. Thus, monitoring and managing potentially high rates of infection and transmission of the disease within this group, both to minimise harm to them and to limit harm to more vulnerable groups, will likely require extensive testing. One strategy to minimise the financial cost of regular mass testing is to pool samples, in which a single test is performed on a group of individuals, with the option that if the group tests positive then follow up tests are performed on the individuals in the group to identify specific infected individuals. Another potential benefit of this approach is that it will reduce test reagent use, the supply of which may be outstripped by demand.

There are a number of decisions that need to be made when designing a test pooling strategy. A primary question is whether the number of samples within a pool may relate to the sensitivity of the test. For example, if prevalence is low and only one infected individual is present per pool, the uninfected individuals will dilute the viral load within the pool which could increase false negative rates. Further, when prevalence is low few pools will contain infected samples, and so the number of pools that require follow-up tests will be low. It may then be beneficial to have relatively large pools in this scenario. By contrast, when prevalence is high we might expect large pools to commonly test positive, and so we would tend towards testing larger numbers. A strategy that starts with large pools and then iteratively create sub pools to minimise testing numbers will have the disadvantage of a longer lead time between sample collection and test result, and informing the infected individual swiftly is of importance in limiting spread.

Another consideration is whether it is important to pool individuals who are likely to be co-infected. In the university context, a method for reducing social contacts is to assign students to living circles (e.g. 10 people per living circle). Under a density dependent transmission model we might expect clustering of cases within living circles. Testing by living circle could improve test sensitivity under the assumption that if one person is infected then others will be also, thus the dilution of the viral load is minimised. It could also reduce costs because fewer pools will be detected as positive than under random pooling, and therefore fewer pools will require follow up tests to identify infected individuals.

A third consideration is whether to follow up positive pooled samples to determine which individuals are specifically infected. If tests are performed by living circle then the follow up tests may be rendered largely moot if everyone within a living circle is expected to self-isolate when any one member of the living circle is tested positive. However, the impact of only returning pool-level test results to individuals on the false positive rate needs to be quantified.

Here we build a pooling model based on the living circles within halls of residence at the University of Bristol. We generate different levels of potential infection clustering by varying levels of social contacts within and between living circles. We also allow for heterogeneity of viral load between individuals based on time course of infection and individual-level infection severity.

# Methods

## RT-PCR dilution calibration

Help…

## Modelling overview

Our objective is to simulate the RT-PCR test for presence of SARS-Cov-2 in samples from all individuals under different disease transmission scenarios. There are four main components to the model – viral load sampling, disease transmission between students, pooling allocation of collected samples, and testing performance. We compare different aspects of testing performance across three strategies – per-individual testing, pooled testing, or pooled testing with per-individual follow-up in positive pools.

## Viral load model

To generate heterogeneity in the viral load of infected individuals, we model the viral load using two components – the maximum viral load over the course of infection, and the point in time at which the sample is taken during the infection. We assume the viral load changes over time following a log normal curve. The viral load for each individual over time is hence given by

Where ranges from 0 to 14 days and represents the per-individual scaling value of the viral load. We select the parameters and , and sample the scaling value as

Finally, to obtain an individual’s viral load at the time of sample collection, we sample from uniformly across the number of days of infection, and denote their sampled viral load as .

We also model a scenario in which there is no heterogeneity between individuals in their viral load.

## Transmission model

To generate clustering of cases according to social structures amongst the students, such as living circles and halls of residence, we use a single time-step transmission model. At time step 0, a random set of individuals is selected to be infected according to some initial prevalence . The number of people each individual goes on to infect is set as

We allow the value to vary across simulation scenarios. The individuals who are infected are determined by contact patterns and the level of containment assumed for a particular simulation scenario. The following table lays out the probability that an uninfected individual is infected by individual given their social contact:

|  |  |  |  |
| --- | --- | --- | --- |
| Containment | Same living circle | Same building | Anywhere else |
| High | 0.9 | 0.09 | 0.01 |
| Medium | 0.6 | 0.3 | 0.1 |
| Low | 0.34 | 0.33 | 0.33 |

Hence, at time step 1, a new set of individuals will be infected, leading to an updated prevalence .

## Testing pool allocation

We adopted two strategies to assign individuals to testing pools. First, we allowed assignment to be random. Second, we attempted to maximise grouping within test pools given living circles. The living circle sizes vary from 1 to 44, with a median of 7. However, within each simulation scenario the pool size is fixed (though we try different pool sizes between simulation scenarios). To maximise grouping of living circle members within a testing pool we used a simple bin packing algorithm, *binPack* from the *BBmisc* R package (version 1.11), which uses a greedy algorithm to maximise pool occupancy whilst minimising distribution of living circle members across multiple pools.

## Testing model

The RT-PCR testing process involves an exponential growth phase of viral RNA, such that after a number of replication cycles , the viral load

where is the efficiency of the reaction, is the starting viral load in the sample, and is the number samples in the testing pool. In the case of the individual test, D=1 and the is the viral load of the individual. We assume that the efficiency lies between 0.65 and 0.9, such as

In order to test positive, the RT-PCR test must detect a viral load of some fluorescence threshold corresponding to a some number of RNA copies after some number of cycles We can write down the value given , distribution of values, dilution factor , parameter values for the distribution of efficiencies , and the expected test sensitivity as a quantile function

A distribution of values can be obtained from the viral load model. Through experimental calibration we expect that the average sensitivity of the test is when , falling to 0.8 when using cycles. Hence, we must identify a set of and parameters that determine the test efficiency distribution that satisfies the calibrated test sensitivities given the assumed distribution of viral loads in the samples. To achieve this we use a general optimisation function, where we want to find an threshold that is estimated to be identical for the two dilution scenarios:

We used the *optim* function in R 4.0.2 to solve this optimisation function, obtaining parameter values of and .

We assume the cost of a sample kit is £2 and the cost of a PCR test is £25.

## Simulation setup

We explore the performance of pooling strategies under the following simulation scenarios

Viral load: between-individual heterogeneity or no heterogeneity

Initial prevalence: 0.001, 0.01, 0.05

R value: 0.5, 1, 3

Containment: ‘high’, ‘medium’, ‘low’

Pool sizes are chosen to be 2, 3, 4, 5, 10, 15, 20, 25 or 30, and individuals are allocated to pools according to their living circles or at random. This results in 486 simulation scenarios, each of which is repeated 100 times.

# Results

## Cost and reagent use savings vary according to prevalence

Pooling by social contacts improves testing performance

Prevalence estimation using pooling strategies

Sensitivity by pooling strategies

Specificity by pooling strategies

Sensitivity to cost ratio by pooling strategies

# Discussion

A regular testing regimen across the student body may demand test pooling as the only viable strategy that financial and reagent resource limits permit. This study illustrates that there are trade-offs to be made in using this approach, which may require ethical arguments to inform decision making. In particular, reducing testing costs will incur a modest reduction in sensitivity.

Per-individual testing within positive pools improves the specificity of test results but is very costly when the prevalence is high.

An adaptive strategy, whereby different pooling schemes are used depending on the prevalence, could be optimal.

Pooling samples in a way that is informed by social contacts appears to be an effective strategy to improve pooling performance, but does come at a cost of higher management and organisational cost at the levels of sample collection and laboratory systems. Though, clustered samples may fall into the same pools without a management layer required, for example if samples from the same classroom or hall of residence are collected together.

Limitations

* We don’t know the true heterogeneity
* We don’t know the true dilution sensitivity beyond 10x
* We only used a simple 2-time-step model to simulate clustering, but the true extent of clustering is not known