Pooled Testing in the Presence of Congestion: The Dorfman Model Revisited

Bingnan Lu[1], Benjamin Legros[2], Oualid Jouini[3], Saif Benjaafar[4]

[1] University of Minnesota, Department of Industrial & Systems Engineering, Minneapolis, MN, USA
luxx0389@umn.edu
[2] EM Normandie, Laboratoire Métis, Paris, France
blegros@em-normandie.fr
[3] Université Paris-Saclay, CentraleSupélec, Laboratoire Genie Industriel, Gif-sur-Yvette, France
oualid.jouini@centralesupelec.fr
[4] University of Michigan, Department of Industrial & Operations Engineering, Ann Arbor, MI, USA
saifb@umich.edu

Keywords*: Pooled testing, healthcare, dynamic models, queueing delays, optimal batch size, matrix analytic approach.*

1 Extended abstract

During periods of high demand for diagnostic testing for infectious diseases, such as during the COVID-19 pandemic, testing facilities may use an approach called pooled testing to increase testing efficiency. With pooled testing samples from multiple individuals are combined and tested together for the presence of infection instead of being tested individually. A negative test implies that none of the individual samples in the group has markers for infection, while a positive result implies that at least one sample in the group has. The individual samples within a sample group that returns a positive test are then retested to identify which ones are positive.

The extent to which this approach can help reduce the time and resources needed for testing clearly depends on the prevalence of the infection in the population being tested. In particular, the benefit of pooling is greater when the prevalence rate is low. In that case, most group tests will return a negative result and testing them individually, all else being the same, would have been wasteful of resources and caused unnecessary delay. On the other hand, if the prevalence rate is high, most group tests will return a positive result, requiring a testing of samples individually. Hence, no efficiency is gained in this case and the system incurs the additional load of an unnecessary group test. Moreover, in settings where samples arrive to the testing facility over time, potential gains in the speed of screening and in dampening the arrival variability to the testing facility because of the batch arrival process must be traded-off against the delay incurred waiting for enough samples to arrive. Clearly there is tension between savings on individual tests (when the results of group tests end up being negative) and the unnecessary burden of group testing (when group tests end up being positive), and between potential speed up and variability reduction at the testing stage and delays at the sample batch forming stage.
This tension can be resolved by choosing optimally, depending on the objective, the number of samples to include in the pool.

Pooled testing has a rich history, in both practice and academic research, across various fields, including medicine, statistics, operations management, and operations research. [2] appears to be the first to have suggested pooled testing as a strategy for improving the efficiency of screening for infectious diseases. The literature that followed is extensive (see for instance [1, 3] and the references therein). However, as with the original Dorfman paper, the focus of this literature has been on maximizing the throughput of the testing facility. This means maximizing the number of samples that

can be screened per unit of time or, in the case of a specified number of samples, minimizing the amount of time it takes to complete the screening of all the samples. While this approach maximizes the efficiency of the facility, it may not necessarily minimize the delay in obtaining test results experienced by the patients who provide the samples. Producing test results quickly took on particular urgency during the COVID-19 pandemic as test results determined, among others, whether an individual needed to quarantine, is able to resume work, or is allowed to travel. This urgency was compounded, at the height of the pandemic, by scarcity in testing kits, testing equipment, and qualified medical staff.

In this paper, we revisit the pooled testing problem with the objective of minimizing delay, namely the expected time to produce a test result for each submitted sample in settings where samples arrive over time with stochastic inter-arrival times. Central to our investigation is the examination of whether the prevalent and ostensibly simpler practice of selecting sample pool sizes to maximize throughput is compatible with minimizing delay. More generally, we are interested in characterizing settings for which pooled testing is particularly helpful in reducing delays. We ground our initial analysis in a queueing model of a single testing facility where samples arrive over time according to a Poisson process. Samples go through two stages: a batching stage and a testing stage. In the batching stage, samples wait for enough other samples to arrive in order to form a batch before proceeding to the testing stage. In the testing stage, batches wait for the testing facility to become available where they are processed on a first come-first served basis. Batch testing times are independent and identically distributed with the exponential distribution. Samples from a batch that tests positive must be tested individually, with each test taking an amount of time that follows the same distribution as a batch test. Delay experienced by the samples (and presumably the associated patients) is the sum of the delay at the batching stage and delay at the testing stage.

We develop a matrix-analytic approach to compute total delay for a given batch size, which we then use to obtain the batch size that minimizes this delay. In extensions, we consider the case where sample arrivals and testing times follow general distributions. We also consider an adaptive batching procedure that does not necessarily involve the testing of all samples from a positive batch and a policy that does not employ a fixed batch size and instead dynamically decides on how many samples to test based on the state of the system.

Managerially, the paper makes the following contributions. We confirm that pooled testing (relative to no pooling) can significantly reduce expected delay. This is particularly the case when either the prevalence rate is low or testing times are long. While choosing batch sizes to maximize throughput, per the Dorfman approach, can also in some cases reduce expected delay relative to no pooling, we show that a delay-minimizing batch can result in even greater reduction in delay. This is particularly the case when either the prevalence rate is low or testing times are long. We also observe that choosing batch sizes to maximize throughput can lead to higher delay than no pooling at all. This is the case when the prevalence rate is either too high or too low. Another finding is that a dynamic policy that adjusts the batch size based on the current number of samples that are yet to be screened does not significantly improve delay relative to a policy that keeps the batch size fixed. However, we find that a policy that determines whether samples are tested as a group or individually based on testing outcomes can significantly improve performance.

References

[1] Matthew Aldridge, Oliver Johnson, and Jonathan Scarlett. Group testing: an information theory perspective. arXiv preprint: 1902.06002, 2019.

[2] Robert Dorfman. The detection of defective members of large populations. *The Annals of Mathematical Statistics*, 14(4): 436–440, 1943.

[3] Dingzhu Du, Frank K Hwang, and Frank Hwang. Combinatorial group testing and its applications. *World Scientific*, Volume 12, 1983.