A Single Stage Pooling Scheme for Large-Scale Pathogen Detection

Alexander Anderson[1], Samuel Dixon[1], Suresh Pillai[2], and Krishna Narayanan[1]

[1] Department of Electrical & Computer Engineering, Texas A&M University, College Station, USA
[2] Department of Food Science & Technology, Texas A&M University, College Station, USA

Abstract

In the event of a foodborne illness outbreak, testing needs to be done quickly and effectively to identify infected products. This study proposes a method of group testing to reduce the time and cost burden required to correctly identify samples infected with a food pathogen in a population with sparse prevalence of 1-3%. Group testing is the concept that collections of samples can be pooled together in various ways and tested in order to reduce the overall number of tests required to identify each sample as negative or positive, while still retaining a high sensitivity and specificity. This study will focus on a non-adaptive testing scheme with three pooling groups, using the Combinatorial Orthogonal Matching Pursuit (COMP) [1] recovery algorithm. This scheme was found to be as effective as individually testing samples, while reducing the number of tests required by 70%.

Keywords: group testing, pooling, non-adaptive, food pathogen

1. Theoretical Background

Group testing is a strategy to test a set of samples in a way that eliminates the need to test each individual within the set. This is done by pooling groups of samples and testing those groups in an organized, algorithmic way. The strategy intends to maximize sensitivity while reducing the number of tests done on the set.

There are two generalized methods for group testing schemes: adaptive testing, where testing stages must be run sequentially, and non-adaptive testing, where testing stages can be run in parallel. Various adaptive algorithms are explored in detail in Du and Hwang's textbook [3], and some of their specific applications and tradeoffs in food pathogen testing are explored in the article by Furstenau et al. [4].

Several adaptive group testing schemes proposed have used soft information from RT-qPCR cycle counts to infer viral loads in samples [5] [6]. This study will focus on visually observing a binary outcome using a growth medium, Tryptic Soy Broth (TSB). Utilizing this method reduces the cost and complexity of testing while increasing robustness.

This study shows that a single stage non-adaptive group testing scheme can be used to consistently identify positive samples inoculated with Salmonella in a sparsely infected population (a prevalence of 2%).

1.1 Dorfman Style Pooling

The topic of group testing was first explored by Robert Dorfman in World War II to identify soldiers infected with Syphilis [2]. In Dorfman's approach, every sample was pooled together to test. If the pool comes back negative, one would know that no pathogen was present in any sample, thus the entire set is negative. However, if the pool comes back positive, one would only know at least one sample is positive and as a result, the entire set would need to be retested individually to determine which individuals are positive. Dorfman's strategy is extremely effective for cases where both the testing assay is highly sensitive to dilution and the prevalence is relatively low. If these two criteria are met, Dorfman's strategy is much more effective than individual testing.

In the context of food pathogen testing, this scheme can be very challenging to successfully replicate. Not only would a high sensitivity assay for testing be required, but dilution of sample concentrations and losses due to pipette error must be considered. For example, consider a set of 100 5mL samples

with exactly 1 positive sample containing 10 CFU of a bacteria, for convenience, Salmonella Typhi. If we were to pool all 100 samples in their entirety together, the final concentration in the pool would be 10 CFU/500mL or 0.02 CFU/mL. Clearly, this low concentration becomes a problem when using a medium to low sensitivity testing assay. For this reason, we must consider a scheme with a lower number of tests per pool relative to the number of samples.

1.2 Combinatorial Orthogonal Matching Pursuit (COMP)

The COMP algorithm is one of many combinatorial group testing schemes that can be applied two dimensionally. A variation was analyzed in the study by Furstenau et al. [4], but we will consider an algorithm with three main pooling groups rather than two.

In this combinatorial algorithm, any item in a negative test is considered *definitely nondefective* (DND), and the remaining items are considered *possibly defective* (PD). The output of the algorithm is the set of all *possibly defective* items. This procedure is possible for a variety of sensing matrices, but for simplicity and proof of concept, we will consider the 2D *nxn* square case with three pooling groups (i.e. a sensing matrix with *3n* rows and n^2 columns).

Consider a set of n^2 items, arranged in a square array of size *n*. The three pooling (testing) groups would be as follows: along the *n* rows, along the *n* columns, and along the *n* diagonals for a total of *3n* tests.

To visualize this approach, let us apply this scheme to the *n=10* case (*Figure 1)*. To form the tests, we would pool along the 10 rows, 10 columns, and 10 diagonals for a total of 30 tests. Suppose samples 23 and 48 are positive, while the remaining samples are negative.

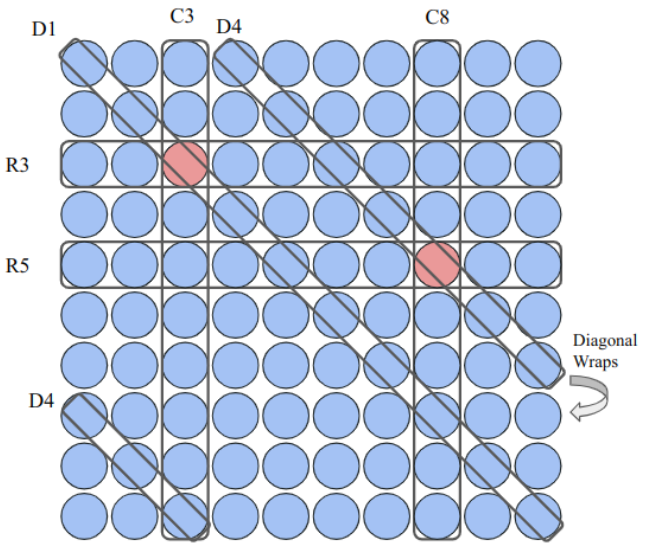

Figure 1: 2D COMP3 Pooling Example

From this set of positive samples, we would obtain the following positive tests: R3, R5, D1, D4, C3, and C8. From the COMP algorithm, we would know that any item in one of the 24 negative tests would be *definitely nondefective*, while the remaining items would be our set of *probable defectives*, in this case, items 23 and 48. The prevalence of 2% illustrated in this example gives a high degree of specificity, that is, the set of *probable defectives* aligns perfectly with what we can consider the 'true positives'. Clearly in practice one will never know the set of positive items or even the true prevalence, so we must consider the tradeoffs of this algorithm *in silico* over a variety of different prevalence, approximately 1-5%.

2. *In Silico* Results

To study the quantitative behavior of the two dimensional COMP3 pooling scheme, we have developed a computational model[1]. Given a prevalence and sample size, the model randomly "inoculates" an exact number of items in the test matrix[2]. It is interesting to note that the number of infected individuals may be modeled with a Binomial distribution or a summation of Bernoulli distributions. However, an adjustment is made in the corresponding model where for a given prevalence, the expected value of infected individuals is guaranteed. That is, there is no variance when inoculating the samples in simulation. The text matrix is a *nxn* binary array, where each index contains either a 0 or 1, denoting the item's status (e.g. negative or positive, respectively). From here, the model forms test "pools" by constructing 3 binary arrays of size *n* (the testing groups), where once again a 0 represents a negative test and a 1 represents a positive test. The pools are formed along the test matrix's rows, columns, and diagonals as shown in (*Figure 1*). These resulting tests are passed through the COMP3 algorithm to determine the set of probable defectives. The result is then compared to the set of "true positives" (inoculated items) to determine the specificity. The specificity is defined as the number of individuals correctly identified divided by the total number of individuals[3].

The above model was applied to a Monte Carlo simulation to determine the asymptotic behavior of the scheme for a variety of parameters over many iterations (ten thousand).

Clearly, it is most natural to first discuss the behavior of the algorithm for increasing prevalence, that is, as the

[1] The source code can be found on the project's Github repository: https://github.com/aanderson60/group-testing

[2] Here the word 'matrix' refers to the mathematical definition, a rectangular table of numbers, rather than the microbiological definition.

[3] It is important to note that defining the specificity in this manner implies that we are taking the set of Probable Defectives (PDs) to be the Definite Defectives (DDs). In reality, the algorithm can be more robust if one assumes the output is uncertain (the set of PDs).

number of positive items in a given size test matrix increases. In *Figure 2* the behavior of specificity vs. prevalence is displayed for a Monte Carlo simulation of size 10000. The test matrix used was of size 10x10 or 100 items. The algorithm performs poorly for higher prevalences, but for the scope of this study we will focus on 0-5% prevalence (*Figure 3*).

Another useful tradeoff to model is that of number of tests vs. specificity (*Figure 4*). For the two dimensional test matrix we have considered so far, the ratio of number to tests to number of items to be tested does not scale linearly and, in fact, decreases. For this reason, we also see a decreasing specificity as we increase the matrix size (*Figure 5*). For larger sample sizes, perhaps a different scheme with more testing groups should be used. However, for the 10x10 case with 2% prevalence that we will consider experimentally, the specificity is quite high.

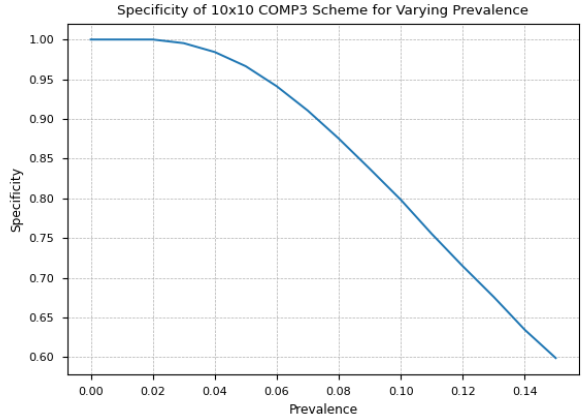

Figure 2: 10x10 COMP3 Specificity vs. Prevalence (P<0.14)

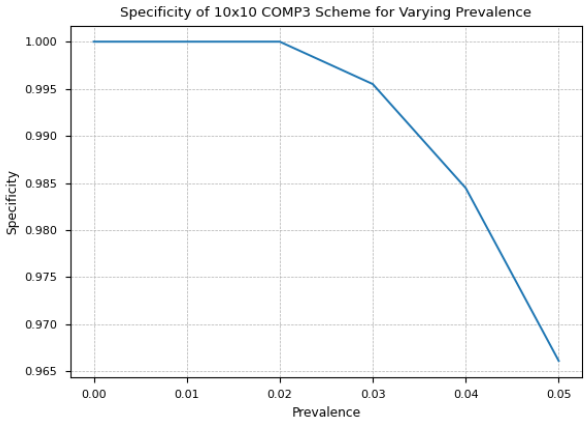

Figure 3: 10x10 COMP3 Specificity vs. Prevalence (P<0.05)

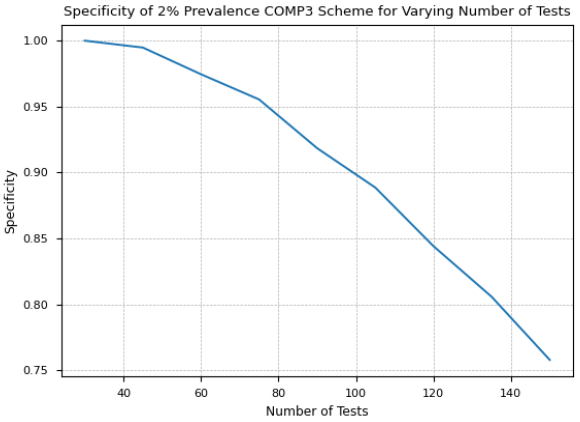

Figure 4: COMP3 (P=0.02) Specificity vs. Number of Tests

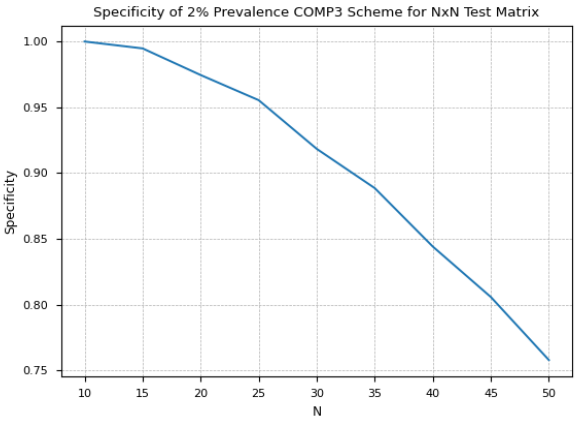

Figure 5: COMP3 (P=0.02) Specificity vs. Test Matrix Size

One more interesting quantity to consider is the number of probable defectives returned by the algorithm for different prevalences (*Figures 6,7,8*). We will consider the *n=10* case for simplicity. Clearly, the average number of probable defectives increases with the prevalence, as does the deviation from the mean. For a prevalence of 3% (*Figure 6*), the algorithm displays perfect specificity (3 PDs) about 60% of the time, returns 4 PDs about 30% of the time, and returns between 5 and 7 PDs the remaining 10% of times.

It is interesting to note that this behavior largely follows a normal distribution, especially as the prevalence increases. It is important to note that this normal distribution is not centered at a perfect specificity (i.e. when the number of PDs is equal to the number of true positives), but rather a number of PDs larger than that. This is intuitive when we consider that the COMP algorithm cannot generate a “false negative” (not considering any lab or human error). Since for this reason the number of PDs cannot be below the number of true positives, the distribution is cut off at this amount while still retaining normal behavior above this threshold. We can generalize the behavior of this distribution as one that shifts right and “stretches out” with respect to the number of PDs.

In other words, the standard deviation and mean of the distribution increase as the prevalence increases.

Practically, this means that the testing scheme is less robust for uncertain prevalence. Some ideas to combat this problem will be discussed in the following sections.

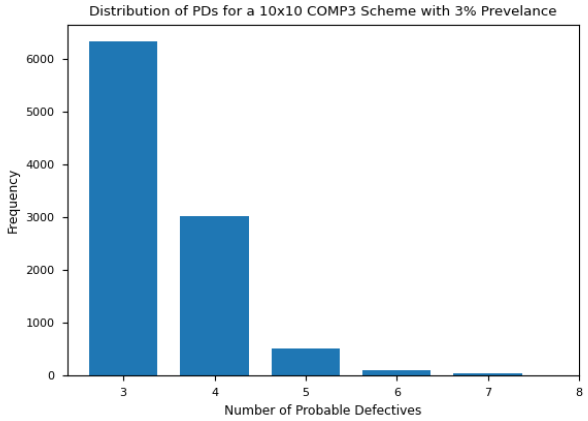

Figure 6: COMP3 (P=0.03) Distribution of Probable Defectives

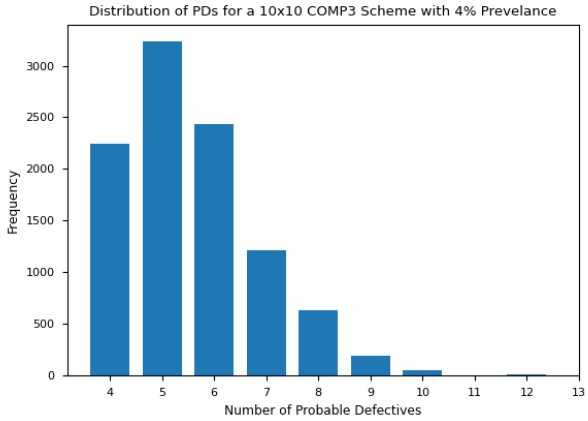

Figure 7: COMP3 (P=0.04) Distribution of Probable Defectives

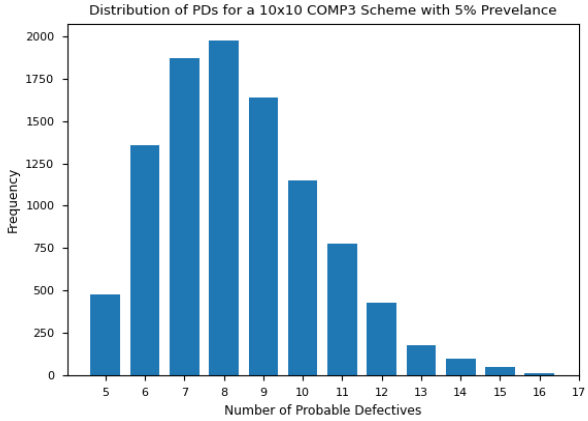

Figure 8: COMP3 (P=0.05) Distribution of Probable Defectives

3. Experimental Methods

The field of group testing has been substantially explored on a theoretical basis, but rarely have its methods been verified experimentally, especially in the context of pathogen detection. With the rise of the COVID-19 pandemic, some have explored group testing methods as a solution to the cost and time burden associated with testing [5] [6] [7]. For simplicity and applicability to food safety applications, we will perform experiments using a Salmonella Typhi bacteria culture. The purpose of the experiments in this study was to determine where the limitations lie with this theory and serve as a proof-of-concept for further pooling schemes. Three main groups of experiments were developed. First, a Dorfman style pooling test to determine the sensitivity of the test to multiple dilutions. Next, a simple matrix pooling scheme using a buffer to test a larger-scale and more effective scheme. Finally, the same matrix pooling scheme was repeated using apple juice as a medium to introduce background noise in the samples.

3.1 Dorfman Style Pooling Experiments

The first experiment performed was one involving a single pool of 5 samples with 1 positive sample. The 5 samples were prepared in 15ml sterile tubes and were filled each with 10mL of Phosphate-Buffered Saline (PBS) as a background medium. The 10 growth tubes and positive/negative controls were prepared with 5mL of Tryptic Soy Broth (TSB) in 15mL sterile tubes. A sample of approximately 10^9 CFU/mL of Salmonella Typhi culture was then serially diluted in PBS to a concentration of 100 CFU/mL. 1mL of this concentration was then added to a single sample out of the 5. Additionally, 100μL of a higher concentration was added to the positive control growth tube. The negative control was left as is, with only TSB. From here, the 5 samples were pooled together into one 50mL volume using a 50mL sterile tube. Finally, 1mL was extracted from the pool 10 times to be added to each of the 10 growth tubes (*Figure 9*). The tubes were incubated at 35°C for approximately 2 days.

The second experiment performed was a variation on the first, with the same setup. However, in the pooling stage, rather than taking the entire 10mL volume of each sample, 2mL was taken out of each sample to be pooled, for a final volume of 10mL. From here, 1mL was extracted from the pool 10 times to be added to each of the 10 growth tubes (*Figure 10*). The tubes were also incubated for the same time and temperature as the first experiment.

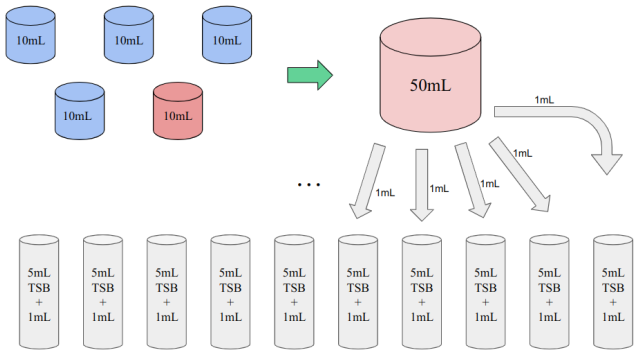

Figure 9: Dorfman Pooling Schematic - First Variation

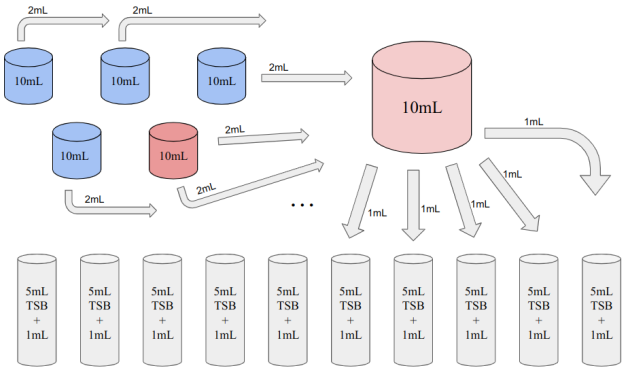

Figure 10: Dorfman Pooling Experiment - Second Variation

3.2 Matrix Pooling Experiments

For the initial experiments, 100 15mL sterile tubes were arranged in a 10x10 grid containing 5 mL of a background medium, Phosphate-Buffered Saline (PBS). Additionally, 32 50mL sterile tubes for the test pools and controls were filled with 10mL of 2X strength Tryptic Soy Broth (TSB) to encourage bacterial growth. The positive control was additionally filled with 9mL of PBS. The remaining 1mL was taken from a mock inoculated sample, using the same procedure as the positive samples. The negative control was additionally filled with 10mL of PBS. The experiments were performed in a single-blind manner to reduce bias, with the researchers performing the pooling and observing the results unaware of which samples were inoculated.

In the experiments, a desired concentration of 10-100 CFU/mL was obtained by performing a multi-stage serial dilution by a factor of 10^{-7} of a turbid Salmonella bacteria culture with an approximate concentration 10^{8} - 10^{9} CFU/mL. After the serial dilution, 2 out of the 100 samples were inoculated arbitrarily. From the diluted sample of Salmonella Typhi used for inoculation, a 100μL volume was plated on Tryptic Soy Agar (TSA) with a swab in order to quantify the amount of bacteria being injected into each sample. This was done three times for a given bacteria sample used for inoculation to gather a more accurate sample mean, minimizing variance.

From this point 100 samples were pooled into row, column, and diagonal tests[4]. The pooling was performed by taking a 1mL volume out of each sample to be placed in its respective pool using a pipette. This was done 3 times for each sample (3 pools per sample). The pooled tests contained a total of 10mL of Tryptic Soy Broth (TSB), which was the growth medium for the bacteria. The ratio by volume of pooled sample to growth medium was maintained to be 1. Once the samples were collected into their respective pools, the pools were placed in an incubator at 35°C for 16 hours. After 16 hours, the pooled tests were removed and positive test results were determined qualitatively. Tests observed as turbid were marked as positive, while non-turbid tests were marked as negative. This binary method for testing samples was remarkably robust, because it took advantage of the innate growth behavior of bacteria and relied on natural processes that occur independent from a lab setting. From this point, the COMP algorithm for recovery, implemented in software, was used to identify the set of possible infected samples. This scheme was repeated to produce three successful iterations.

The next round of experiments involved changing the background medium of PBS to apple juice to gather evidence that the process was still reliable with an alternative, food-specific sample containing considerable background noise. Three iterations were done, in the same manner as described above.

4. Experimental Results

The outcomes of the experiments described in the previous section were highly successful, with some problems that can be attributed to human lab error. Some potential solutions to increase the robustness of group testing schemes to this error will also be discussed.

4.1 Dorfman Style Pooling Experiment Results

In both variations of the experiment, the results were successful, with each of the 10 growth tubes turning visually turbid after some time, indicating a positive pool as expected.

This experiment served as a baseline and proof-of-concept to demonstrate that group testing in the simplest case, a Dorman style pool, could be used to recover an infected individual amongst a sample size of 5. The results provided evidence that group testing could be applied even where the physical phenomena and complication of dilution was present when mixing multiple samples into a pooled group. Furthermore, by testing two different pooling volume schemes, the results indicated that recovery was feasible both

[4] Note that these pools are not specific to the COMP algorithm, which is what is used for recovery.

by pooling the whole sample and by extracting just a portion of the sample.

3.2 Matrix Pooling Experiment Results

The matrix pooling experiments were used to explore a possible pooling scheme using the Combinatorial Orthogonal Matching Pursuit (COMP) [1] in order to accomplish the objectives and positive attributes group testing allows for (i.e. reduction of tests with minimal loss of sensitivity).

Among the five iterations using PBS as the background medium (*Table 1*), there were three results where the exact samples were identified, one result where one of the exact samples was identified, and one result where no samples were recovered. The primary cause of these failures can be attributed to lab error and contamination of both the negative samples and pooling tubes. Lab error, specifically in pipetting, can create false negative results. Such a low concentration of bacteria is in reality not evenly distributed and can accumulate on the pipette tips, never making it into the growth tubes. For this reason, the samples were vortexed prior to extraction to mix the sample and more evenly distribute the bacteria if present. One potential solution to this would be to use the entire volume of each sample in creating the pools, but this can be unreasonable when dealing with samples of very large volume. This idea should be explored in further research.

For the three iterations using apple juice as the background medium (*Table 2*), both positive samples were successfully identified in each of the three cases.

Table 1: Matrix Pooling with PBS Summary Results

Iteration	Inoculated Samples	Average Infected Sample Concentration (CFU/mL)	COMP3 Identified PDs
1	62, 67	1.69-16.95	62, 67
2	17, 30	1.69-16.95	17, 30
3	1, 100	13.79	1[5]
4	27, 83	73.6	None[6]
5	28, 54	35	28, 54

[5] During this iteration, sample 100 was contaminated, as was evident from its plate growth. Additionally, one of the pooling tubes took 72 hours to turn positive. The algorithm was able to correctly identify the individuals after this time.

[6] During this iteration, contamination due to lab error was also observed, as was evident in a false positive pooled test.

Table 2: Matrix Pooling with Apple Juice Summary Results

Iteration	Inoculated Samples	Average Infected Sample Concentration (CFU/mL)	COMP3 Identified PDs
1	38, 49	132.8	38, 49
2	1, 51	28.727	1, 51
3	7. 93	30.93	7, 93

5. Conclusion and Further Steps

This study provides evidence that group testing can be used in the context of food safety to reduce the cost of testing with low prevalences. The simple COMP3 scheme used serves as a proof of concept that dilution of sample from pooling can be overcome for a bacterial concentration of 10-100 CFU/mL.

The physical constraint of concentration and distribution of bacteria in the sample was one key limiting factor in the specificity. A divide-and-conquer pooling method may be used to combat this limitation when taking a group testing scheme like this to scale. This would be done by subdividing a larger test matrix into smaller matrices, and applying the pooling methods observed in COMP3 to these smaller matrices. A recombining approach may look like taking all the sets of the potential defectives from the smaller matrices and recombining them in such a way that increases specificity, while still maintaining the objective of reducing total tests. This improvement would be similar and emulate the idea of an efficient recursive sorting algorithm, commonly known as merge sort.

The tradeoff between specificity and prevalence was another important limitation of group testing observed throughout the duration of this study. Implementing a second adaptive stage is a potential improvement to COMP3 in order to make group testing feasible for populations where the prevalence of infected individuals is unknown. With this change, the scheme would consist of two stages. The first stage would remain as COMP3 and return a set of possible defective samples. From this point, a second stage would be implemented to test the set of possible defectives in order to improve specificity. In constructing this second stage, soft information from the first stage could be used to obtain an estimate of the prevalence. This soft information could be obtained by observing the tests at varying times up to 16 hours to observe the time until the tests become turbid. A test maturing to be turbid rapidly is indicative of either greater prevalence or concentration of infected individuals. Testing

would need to be done to obtain a characterizing curve for the relationship between concentration and time to turbidity.

From these experiments, group testing theory and its underlying limitations in the context of food sampling have been qualified and investigated in a lab setting. A characterization for the relationships between specificity, prevalence, and samples under test have been developed and utilized to further understand the disconnect between idealities of group testing theory and physical complications introduced by the context of the qualities of the items under test. For the experiments detailed, some specific qualities noted are concentration, distribution within individual samples, and prevalence of the bacteria present in the test matrix. From these preliminary tests, it has been shown that group testing is not only feasible in the context of food testing, but more efficient and effective than the testing schemes currently used in the industry.

References

[1] M. Aldridge, O. Johnson, and J. Scarlet. *Group Testing: An Information Theory Perspective.* Foundations and Trends in Communications and Information Theory, 15(3-4):196-392, 2019.

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

[3] D.-Z. Du and F. Hwang. *Combinatorial Group Testing and Its Applications*. World Scientific.

[4] T. Furstenau, J. Cocking, C. Hepp, and V. Fofanov. *Sample pooling methods for efficient pathogen screening: Practical implications.* PLoS ONE 15(11):e02*36849, 2020.*

[5] A. Heidarzadeh and K. Narayanan. *Two-Stage Adaptive Pooling with RT-QPCR for Covid-19 Screening*. ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021, pp. 8148-8152, doi: 10.1109/ICASSP39728.2021.9413685.

[6] S. Ghosh, A. Rajwade, S. Krishna, N. Gopalkrishnan, T. E. Schaus, A. Chakravarthy, S. Varahan, V. Appu, R. Ramakrishnan, S. Ch et al. *Tapestry: A single-round smart pooling technique for COVID-19 testing*. medRxiv, 2020.

[7] J. Eberhardt, N. Breuckmann, C. Eberhardt. *Multi-Stage Group Testing Improves Efficiency of Large-Scale COVID-19 Screening.* Journal of Clinical Virology (128), 2020.