FAST: a Feasible, Accurate and Speedy Test Strategy for COVID-19

Linjiajie Fang[1], Bing-Yi Jing[2], Shen Ling[3], and Qing Yang[4]

***Abstract*— As the COVID-19 pandemic continues worldwide, there is an urgent need to detect infected patients as quickly and accurately as possible. Group testing proposed by Technion [2] could improve efficiency greatly. However, the false negative rate (FNR) would be doubled. Using USA as an example, group testing would have over 70,000 false negatives, compared to 35,000 false negatives by individual testing.**

In this paper, we propose a Flexible, Accurate and Speedy Test (FAST), which is faster and more accurate than any existing tests. FAST first forms small close contact subgroups, e.g. families and friends. It then pools subgroups to form larger groups before RT-PCR test is done. FAST needs a similar number of tests to Technion's method, but could sharply reduce the FNR to a negligible level. Again taking USA as example, FAST reduces the number of false negatives to just 2000 while it is seven times faster than individual testing.

I. INTRODUCTION

Reverse Transcription Polymerase Chain Reaction (RT-PCR) is one of the main diagnostic methods for early detection of coronavirus, SARS-CoV-2. However, RT-PCR is a technical intensive testing method, which requires professional technicians to perform in some specialized laboratories [1]. The overload at testing facilities would hinder the process of virus tests for a large population. As a result, there is an urgent need to improve the efficiency of overall testing. Researchers at Technion - Israel Institute of Technology and Rambam Health Care Campus introduced a faster way to test coronavirus by pooling samples (Fig.1). Such idea of applying group testing to reduce the number of necessary tests can be traced back to Dorfman [2]. They also suggested a positive pool up to the size of 32 or even 64 containing a single infected sample could still be confidently identified [3][4].

Despite the high efficiency of Technion's proposal, one major drawback is its false negative rate will double that of individual testing. Taking USA as an example, as in late May, there were about 1.7 million confirmed COVID-19 cases[5]. Assuming the false negative rate of RT-PCR test is 2%, individual test will estimate that around 35,000 infected patients as healthy, while the number would be 70,000 using Technion's method.

To overcome this problem and maintain the high efficiency, we propose a *Flexible, Accurate and Speedy Test* (FAST), an adaptive group testing strategy (Fig.2). We first form small sub-groups of people who are close to each other socially and/or physically, e.g. families, colleagues, passengers from the same flight. These subgroups are then pooled together to form a larger group to be tested with a design of repeated tests. FAST can substantially reduce the number of tests and in the meantime drastically lower the false negatives from over 70,000 to about 2,000.

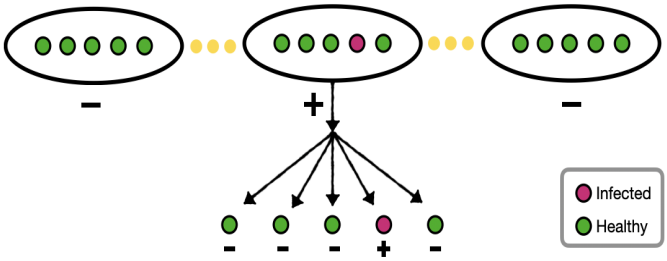


Fig. 1: Group testing suggested by Technion. The pools of samples are tested and positive pools are later tested individually.

Interestingly, we discover a simple relationship between the optimal pool size and the prevalence i.e. $k_{opt} \approx 1/\sqrt{prevalence}$, which can be used to determine the optimal pool size in practice easily. More details will be described in RESULTS.

Unless otherwise stated, all the results below concern the COVID-19 cases in USA The same analysis can be carried out to the other nations.

II. RESULTS

Efficiency comparisons We evaluate the efficiency of testing strategies by calculating persons per test, i.e. the size of population divided by the total number of tests performed. Fig.3 compares the performance of individual test, Technion's strategy and FAST on the simulation set of 10,000 families. We assume the prevalence $\rho = 0.005$ and within-close-contact-subgroup-infection rate $\beta = 0.7$. β is defined by the probability that one individual is infected by a virus carrier in the same subgroup. Fig.3(a) shows that optimal pooling sizes of FAST and Technion's strategy are 15 and 32 respectively. And both FAST and Technion's strategy are over seven times efficient than individual test.

Dramatically lower FNR Fig.3(b) shows the results of false negative rates under different pooling sizes. Technion's strategy doubles the false negative rate from 2% to 4%. FAST dramatically lowers the false negative rate to nearly 0.0012. Such improvement is significant in reality: Taking USA as an example, 1.7 million cases were confirmed in late May.

[1]Mr. Linjiajie Fang is a PhD student in Data Science and Analytics, Hong Kong University of Science and Technology Email: lfangad@connect.ust.hk

[2]Prof. Bing-Yi Jing is a Professor in Mathematics, Hong Kong University of Science and Technology Email: majing@ust.hk

[3]Ms. Shen Ling is a PhD candidate in Mathematics and Statistics, Dalhousie University, NS, Canada Email: lisaling@dal.ca

[4]Ms. Qing Yang is a PhD candidate in Mathematics, Hong Kong University of Science and Technology Email: qyangad@connect.ust.hk

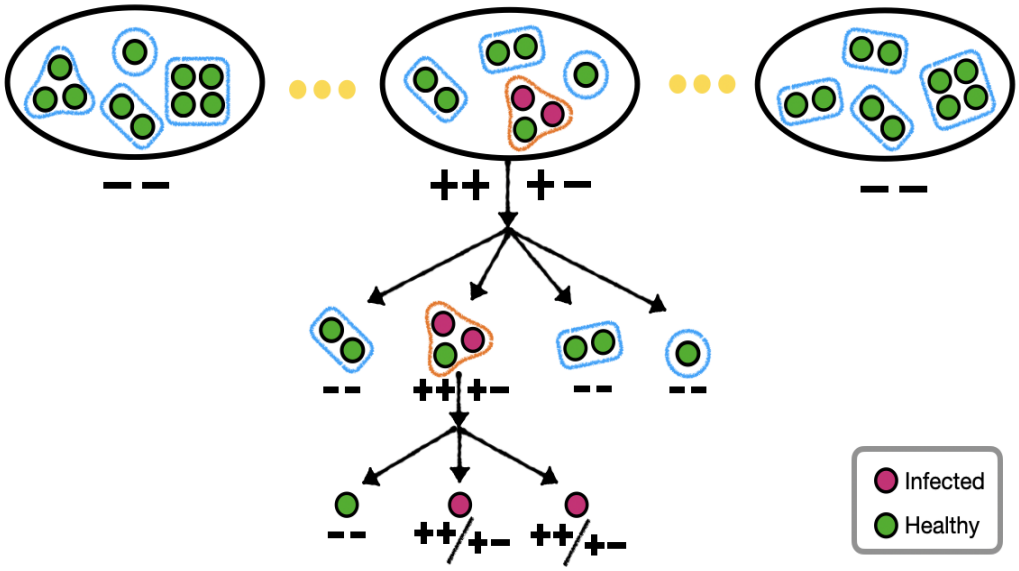


Fig. 2: An example of performing FAST when each group containing four close contact subgroups. FAST performs two RT-PCR tests for each test simultaneously, and either one of the two tests shows positive, such group, subgroup or individual would be labeled positive. FAST first tests groups of samples in the first step, subgroups in a positive group will be tested one-by-one. Finally, samples in a positive subgroup will be re-tested individually.

Individual test would miss 35,000 infected patients while Technion's strategy would double such number to 70,000. If FAST is applied, the number would sharply drop to around 2,000.

Low and stable FPR Fig.3(c) shows the results of false positive rates under different pooling sizes. By assuming the false positive rate to be 0.001 for RT-PCR test, the simulation shows that FAST has stable and almost zero false positive rate. Despite the false positive rate of Technion's strategy is also far less than individual test, it is higher than FAST and nearly linear to the pooling size, which suggests avoiding pooling too much samples. However, FAST can pool much larger number of samples without increasing false positive rate. If the universal virus testing is applied to USA with around 300 million residents, and the optimal pooling sizes are applied, individual tests would wrongly diagnose 300,000 healthy people as infected. Technion's strategy would reduce this number to 22,500, yet FAST can lower this number to 3,600 in total.

How does FAST achieve such an improvement? To answer this question, it is helpful to compare the design of FAST and other group testing methods. Fig.2 illustrates the flow chart of performing FAST. FAST first forms small subgroups of people who have been in close contact, then a fixed number of subgroups are pooled into larger groups to be tested. Like Technion's strategy, FAST gathers samples into pools to perform group testing, and only subgroups or individuals in positive pools will be re-tested one by one. But FAST differs in its procedure of performing each test and its design of groups.

FAST performs two RT-PCR tests for each test simultaneously, and either one of two tests shows positive, such subgroup or individual would be labeled positive. Such procedure can decrease false negative rate of each single test to second order immediately. But, as a cost, this procedure also doubles both the number of tests and the false positive rate.

To alleviate the loss of efficiency and increase the false positive rate by repeated tests, FAST adopts two-layer grouping methods, which pools subgroups to a larger group. By virtue of hierarchical two-layer structure, FAST can largely reduce the number of required tests and at the same time sharply reduce false positive rate.

Unlike Technion's random pooling, FAST pools close contact subgroups to a larger group which contributes to the improvement of efficiency as it can dense the distribution of infected patients. In our simulation, we take families as close contact groups. The concept can be generalized to any groups of people who are in close contact to each other socially and/or physically, such as friends, colleagues, or passengers from the same flights. Since subgroups are in different sizes, compared to uniform sizes of groups in Technion's strategy, FAST does not restrict the number of samples in each group or subgroup to be fixed, which makes FAST even more flexible.

Effects of within-close-contact-subgroup-infection rate Fig.4 compares the performance of Technion's strategy and FAST with optimal pooling size under different within-close-contact-subgroup-infection rate $\beta \in [0, 1]$. β is defined as the probability that one individual is infected by a virus carrier in the same subgroup. Here we fix the prevalence $\rho = 0.005$. Fig.4(a) shows that FAST always pools more samples than Technion's strategy and the pooling size is linear to β. Fig.4(b) shows that when $\beta \geq 0.72$,

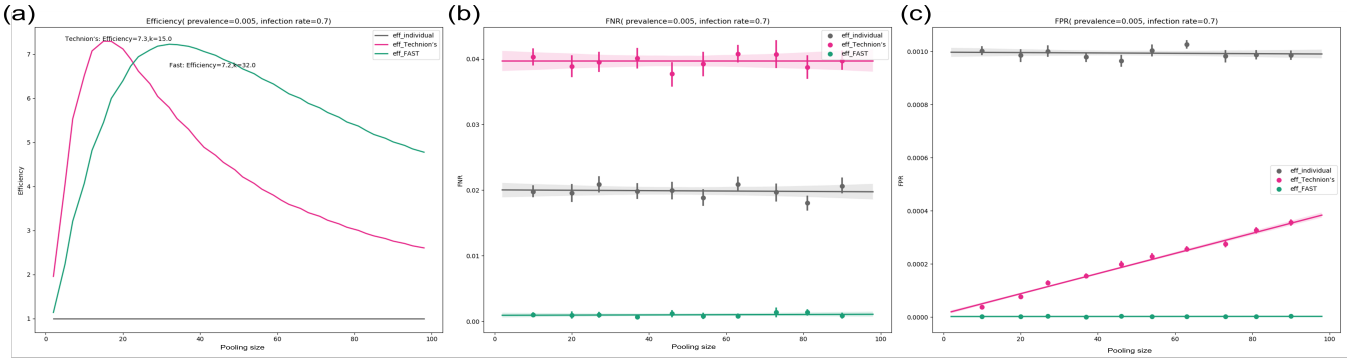


Fig. 3: The performance of FAST, Technion's strategy and individual tests on 50 simulation sets. Each simulation set contains 10,000 families with distribution of household sizes following data in USA[6]. We fixed individual prevalence $\rho = 0.005$ and within-close-contact-subgroup-infection rate $\beta = 0.7$. We evaluate the efficiency and accuracy along different pooling sizes. (a): The efficiency versus pool size. (b): The false negative rate versus pool size. (c): The false positive rate changing pool size.

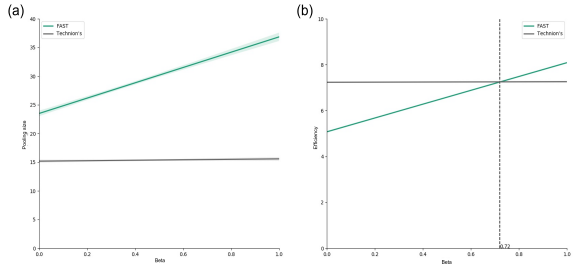


Fig. 4: The best performance of FAST and Technion's strategy under different within-close-contact-subgroup-infection rate. The prevalence and infection rate is fixed to be the same as Fig.2. (a): The optimal pooling size of FAST and Technion's strategy with respect to changing β. (b): The best efficiency and relative efficiency of FAST and Technion's strategy with respect to changing β.

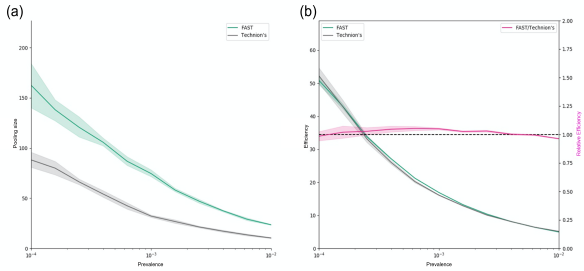


Fig. 5: The best performance of FAST and Technion's strategy under different prevalence. The prevalence and infection rate is fixed to be the same as Fig.2.(a): The optimal pooling size of two methods with respect to changing ρ. (b): The efficiency and relative efficiency of two methods with respect to changing ρ.

FAST beats Technion's strategy in terms of efficiency. Since SARS-CoV-2 spreads very easily and sustainably between people[7], it is reasonable to speculate that FAST would be more efficient in reality.

Different prevalence levels Fig.5 compares the performance of two methods under different prevalence. We fix the within-close-contact-subgroup-infection rate $\beta = 0.7$ and change the prevalence from 0.01% to 1%. Fig.5(a) and Fig.5(b) show that the number of optimal pooling sizes and efficiency of both two group testing strategies shrink almost exponentially when the prevalence increases. Fig.5(b) shows that the relative efficiencies of both methods do not change a lot by prevalence. If FAST performs better than Technion's strategy when prevalence is high, it would also be favorable for universal testing of areas with lower prevalence.

Estimation of the optimal pooling size. Surprisingly, we discover a simple relationship between the optional pooling size and prevalence level. The significance of these simple estimations is that they can be used in reality to guide the implementation of grouping testing without tedious computation. Here we introduce the results without proofs:

1. Given the prevalence ρ, the optimal pooling size k_{opt} of individuals could be estimated by $k_{opt} \approx 1/\sqrt{\rho}$ for small ρ.

For example, when the prevalence is 0.005, it is estimated that the optimal pooling size of Technion's strategy is $1/\sqrt{0.005} \approx 14$. It is very close to the simulation result which optimally pools 15 samples. This can also be used to estimate the optimal pooling size of sub-pools:

2. Suppose the probability of a sub-pool containing infections is ρ^*, the optimal pooling size k^*_{opt} of sub-pools could be estimated by $k^*_{opt} \approx 1/\sqrt{\rho^*}$ for small ρ^*.

For FAST, $\rho^* = \frac{m}{1+(m-1)\beta}\rho$ with known average close contact subgroup size m and infection rate β. In our simulation $m = 2.5$[6], $\beta = 0.7$, $\rho = 0.005$, thus $\rho^* \approx 0.0061$, we can quickly compute optimal pooling size to be about $1/\sqrt{0.0061} \approx 13$ families or equivalently 32 people, which is the same as our simulation result. In our experiments, such estimation is effective with an error of at most one sample for $\rho \leq 0.4$.

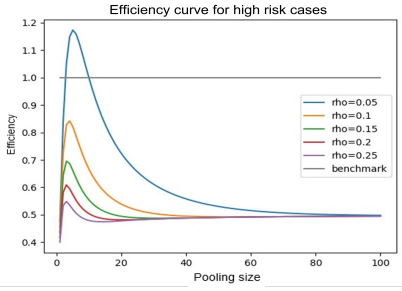


Fig. 6: The theoretical efficiency of one-layer group testing method with repeated tests when the prevalence level is high.

III. DISCUSSION

How accurate can group testing achieve? If repetition is not adopted, two-layer grouping, which applies to the groups of subgroups, is much more efficient than one-layer grouping. However, compared with individual test, one-layer grouping (like Technion's method) doubles FNR, and its FPR is nearly linear to the group size. Two-layer grouping triples FNR, and their FPR is nearly linear to the subgroup size. These facts direct us to avoid overmuch number of layers and prevent pooling too many individual samples in the last layer of group testing strategy.

To alleviate the problem of multiplying FNR without loss of much efficiency, FAST adopts two-layer grouping structure and performs repeated tests, whose FNR is controlled to be at the level of second order and FPR is stable along different pooling sizes. Although repetition intuitively increases the total number of tests, by virtue of two-layer structure and infection information, FAST maintains comparable efficiency as Technion's strategy.

Necessity of group testing When prevalence becomes very high, both efficiency and accuracy of group testing will be greatly reduced. Thus we have to consider if it is necessary to implement group testing when level of prevalence is severe. Fig.6 provides theoretical efficiency of naive grouping testing with repeated tests, which suggests that when prevalence is over 10%, the number of tests of grouping patients would outnumber the individual tests. It could be a loss of the original intention for designing group testing. For areas with very severe epidemics, small pooling size will be appreciated or just remain as simple as individual testing. However, to the best of our knowledge, prevalence in most areas for COVID-19 is about 1% [8] or even lower, so the necessity of group testing is still evident for COVID-19 pandemic.

Determine Close contact group The concept of close contact group can be freely generalized to any clique of close contacts, not only families. For example, the residential relations, roommates staying in the same dormitory unit. Furthermore, when applying group testing strategies in airport to diagnose the virus status of commuters, passengers sitting in front and back rows on the same plane can be pooled in the same close contact group.

Unlike Technion's strategy, which pools fixed number of individuals, FAST pools samples by fixed number of close contact subgroups. Since the size of close contact subgroups is flexible, FAST pools different number of samples from group to group, which concerns more about contact information than restrictions on specific pooling size. In this paper, the contact between groups or subgroups is ignored. However, in reality, the contact between people is much more complicated. The contact information of patients plays an important role to distinguish high risk susceptible patients from low risk ones. A flexible grouping strategy based on contact network might help cluster the infected patients to certain pools and largely reduce the number of tests.

Flexibility of FAST strategy Not limited to USA, FAST can be flexibly adopted to different nations in different disease stage. With appropriate settings including prevalence and within-close-contact-subgroup-infected rate, FAST can generate the optimal pooling size with the corresponding best efficiency to achieve. Under the guidance of FAST, the false negative rate and the false positive rate can be maintained at a dramatically low level with good performance in efficiency in some range of parameters.

"Predictive" group testing In some areas, government would propose universal testing among all the population, where millions of individuals could be involved. In order to perform a more efficient testing, it is prioritized to test people who have high probability of getting infected. Before testing, we can make prediction of the risk level for someone based on their contact information, symptoms and residential areas. Such risk prediction is usually non-uniform among different groups of people. In the future, we hope to develop methods to predict the risk of people and build testing strategies based on different predictions.

IV. METHODS

Simulation Setup. To fairly evaluate the performance of FAST, we set up our simulations with practical assumptions. We assume 0.005 prevalence according to the number of infections in USA up to 29th May 2020 [5], and the distribution of family sizes follows data of the household in America [6]. That is, we choose number of family members from $\{1, 2, 3, 4, 5, 6, 7\}$ with probabilities equal to $\{0.28, 0.35, 0.15, 0.13, 0.06, 0.02, 0.01\}$ respectively. As for the accuracy of each RT-PCR test, we assume FNR=0.02 and FPR=0.001, following the same configuration proposed in Hanel etc.(2020)[9].

In Fig.3 we assume a within-close-contact-subgroup infection rate i.e. $\beta = 0.7$, in order to model the realistic case for COVID-19. In our data preparation, we first generate 10,000 families with fraction of infection equal to 0.0061, then we generate the number of members and infections of each family by the household distribution and the infection rate to make sure the overall individual prevalence is 0.005. For each pooling size, we simulate 50 replicates and evaluate the average performance.

In Fig.4, we follow the same configuration in Fig.3 except for changing infection rate β from 0 to 1 with a gap equal

to 0.1. For each β, we search for the optimal pooling size by evaluating strategies with pooling sizes less than 500 and choose the size with the best efficiency. Similarly, in Fig.5 we fix $\beta = 0.7$ and change prevalence from 10^{-4} to 10^{-2} with a gap of log-scale equal to 0.2 in order to evaluate the best performance on different prevalence levels. To avoid the variance when searching for the best efficiency, we simulate 10 replicates of simulation set containing 50,000 families and the average performance is compared.

V. CONCLUSION

In this paper, we introduce FAST, an *Flexible, Accurate and Speedy Test*, based on the design of group testing, adoption of close contact information and necessary repetition. Simulations concerning the cases in USA demonstrate that FAST can be over 7 times efficient as the individual test, and greatly reduce the false negative rate and the false positive rate to be far below the level of naive group testing. We also introduce a simple and very close estimation of the optimal pooling size $k_{opt} \approx 1/\sqrt{\text{prevalence}}$, which can be flexibly implemented in reality.

VI. CODE AVAILABILITY

Our code is available on https://github.com/Fang-Lin93/FAST_group_testing

REFERENCES

[1] T. Nolan, R. E. Hands, and S. A. Bustin, "Quantification of mrna using real-time rt-pcr," *Nature protocols*, vol. 1, no. 3, p. 1559, 2006.

[2] R. Dorfman, "The detection of defective members of large populations," *The Annals of Mathematical Statistics*, vol. 14, no. 4, pp. 436–440, 1943.

[3] N. Jeffay, "To ease global virus test bottleneck israeli scientists suggest pooling samples pooling method for accelerated testing of covid-19," 2020.

[4] "Pooling method for accelerated testing of covid-19," 2020.

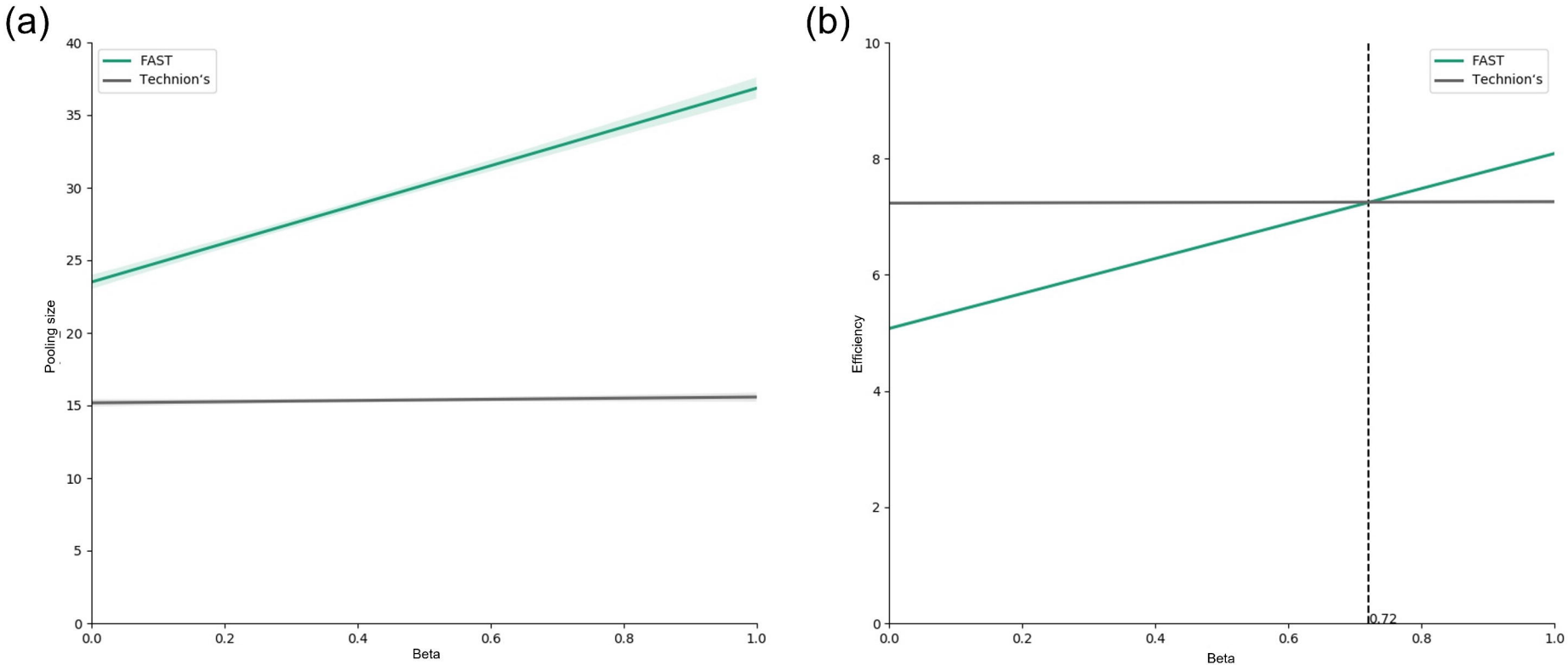
[5] "Coronavirus disease 2019 (covid-19): Cases in the u.s.," 2020.

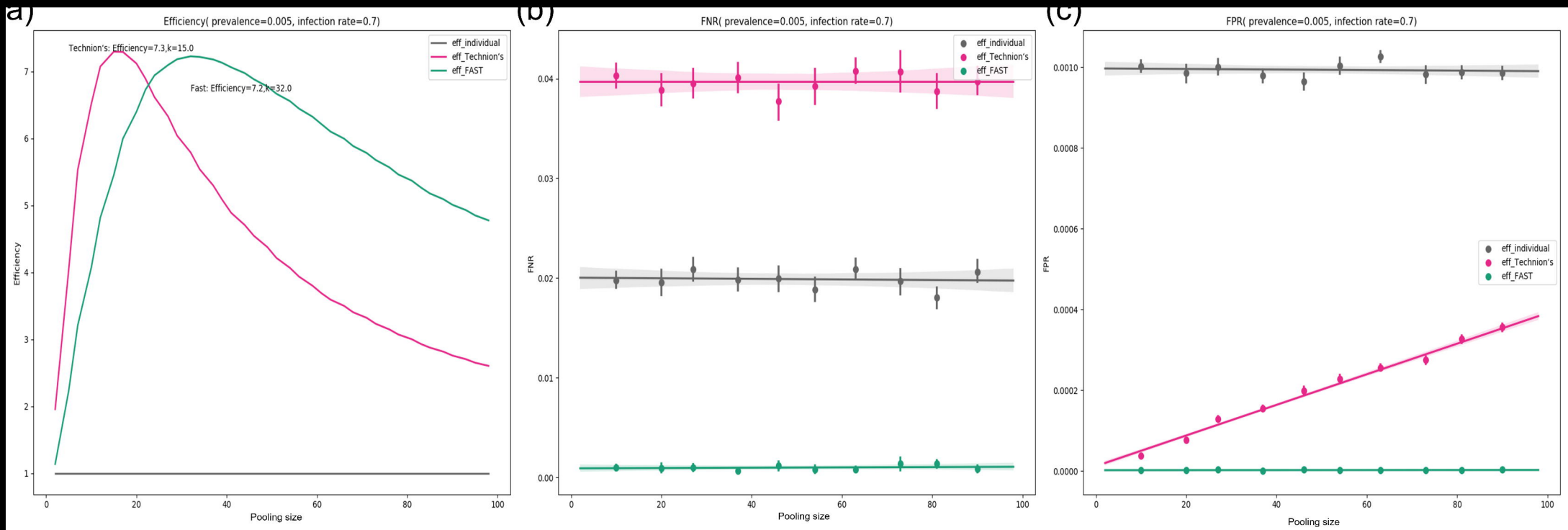
[6] E. Duffin, "Distribution of households in the united states from 1970 to 2019, by household size," Nov. 2019.

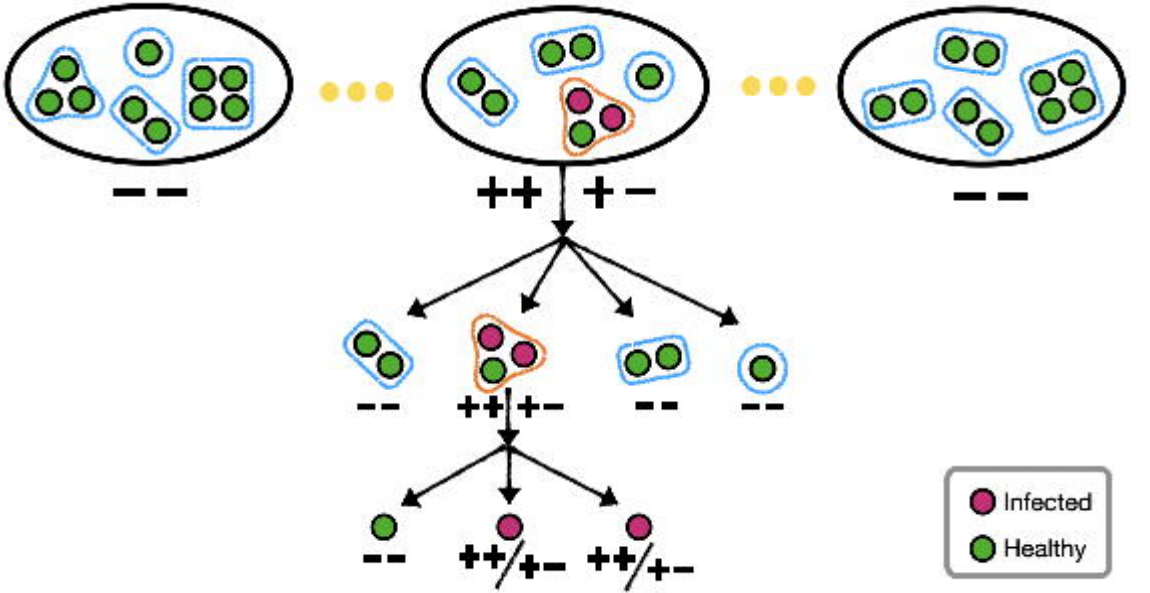
[7] "Coronavirus disease 2019," 2020.

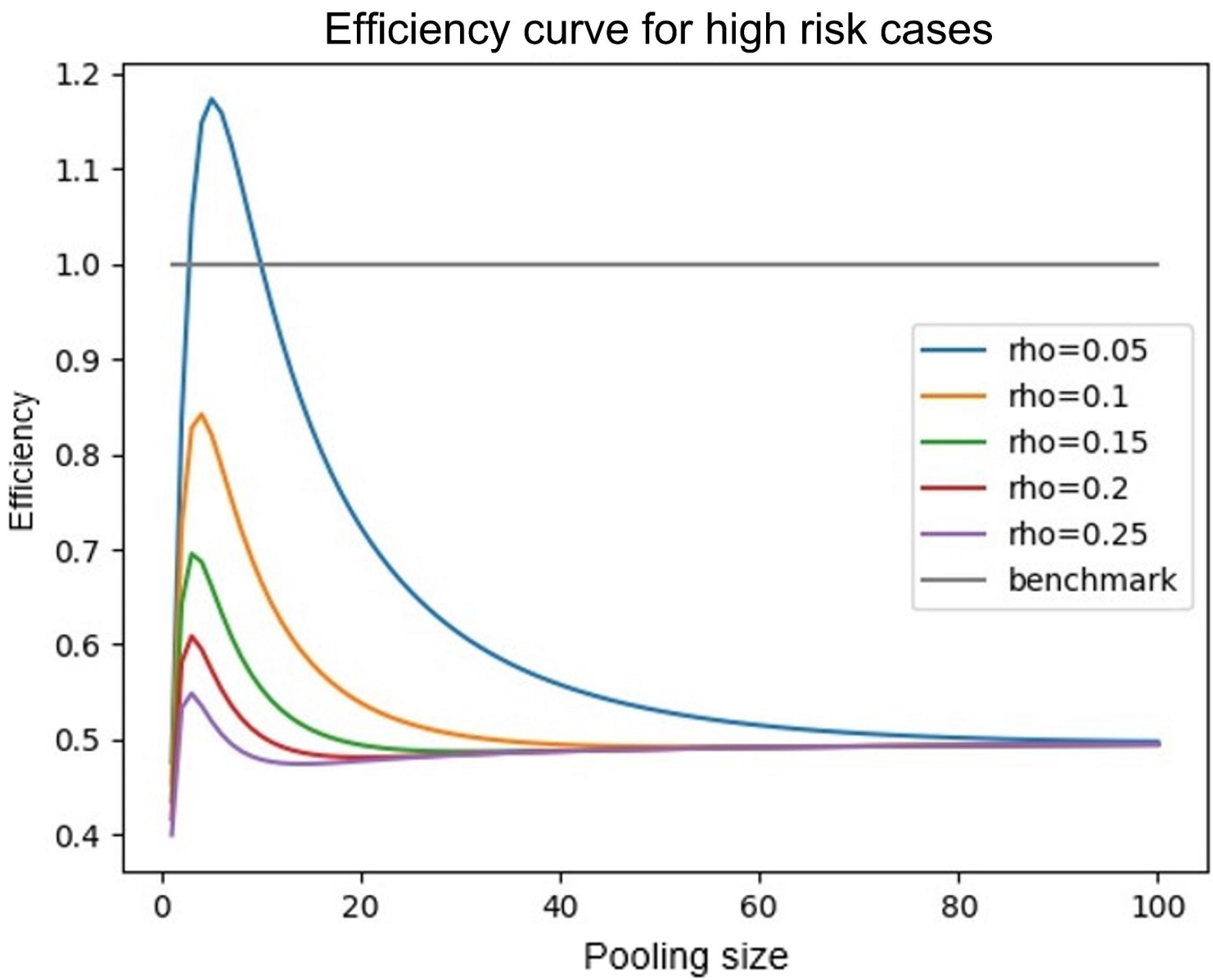
[8] S. Lohse, T. Pfuhl, B. Berkó-Göttel, J. Rissland, T. Geißler, B. Gärtner, S. L. Becker, S. Schneitler, and S. Smola, "Pooling of samples for testing for sars-cov-2 in asymptomatic people," *The Lancet Infectious Diseases*, 2020.

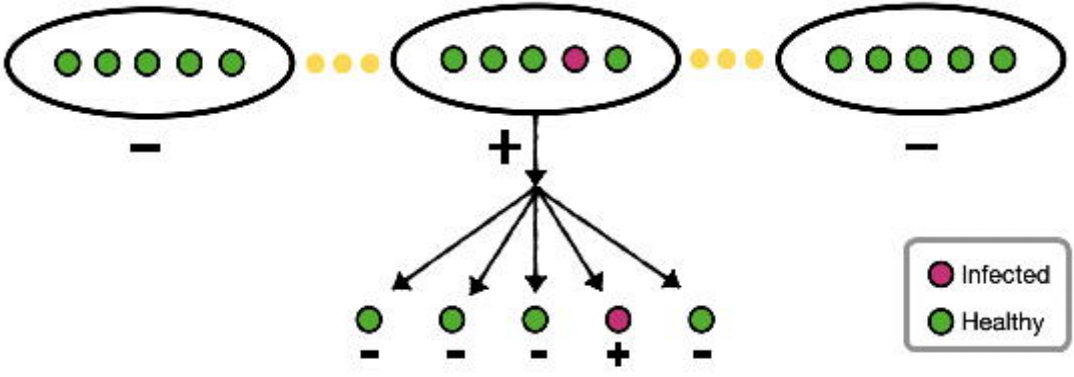
[9] R. Hanel and S. Thurner, "Boosting test-efficiency by pooled testing strategies for sars-cov-2," *arXiv preprint arXiv:2003.09944*, 2020.

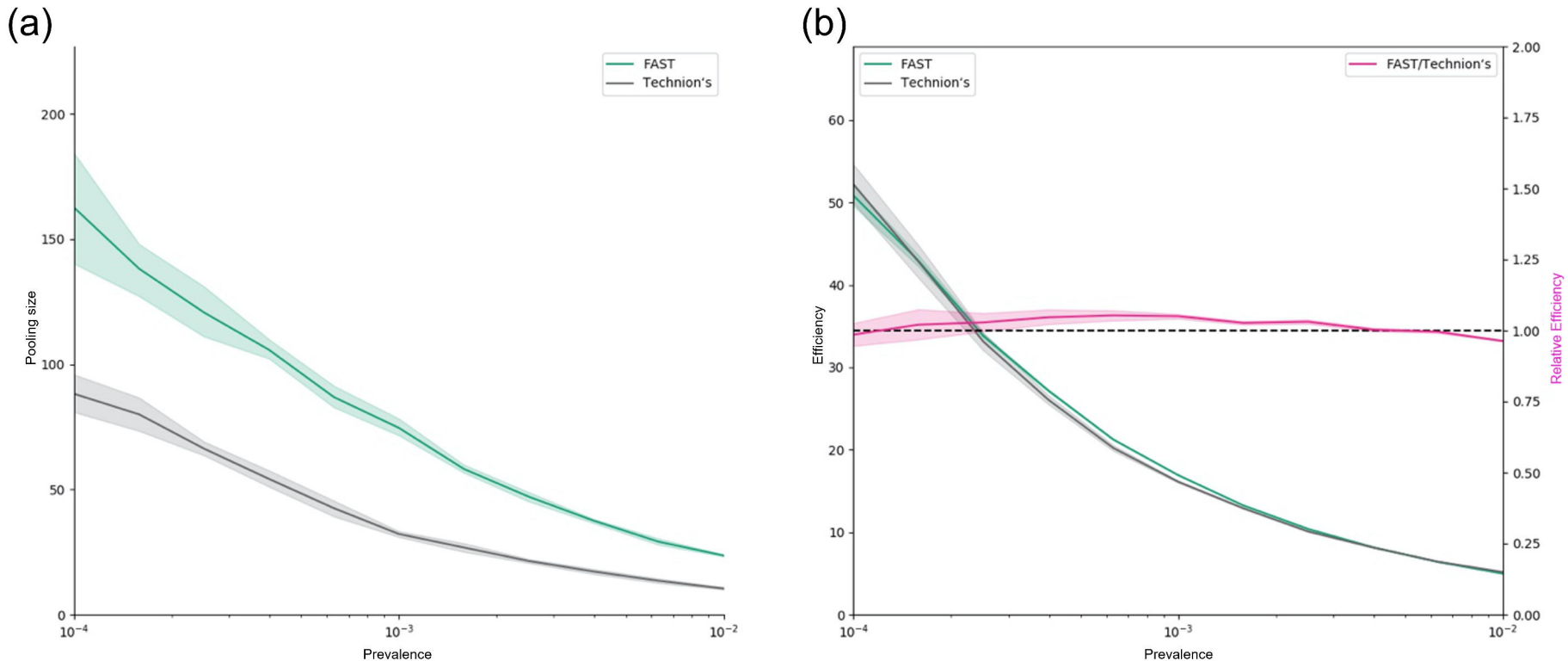

(a)
FAST
Technion's
Pooling size
Beta
0
5
10
15
20
25
30
35
40
0.0
0.2
0.4
0.6
0.8
1.0
(b)
FAST
Technion's
Efficiency
Beta
0.72
0
2
4
6
8
10
0.0
0.2
0.4
0.6
0.8
1.0


a)
Efficiency(prevalence=0.005, infection rate=0.7)
Technion's: Efficiency=7.3,k=15.0
Fast: Efficiency=7.2,k=32.0
eff_individual
eff_Technion's
eff_FAST
Efficiency
Pooling size
b)
FNR(prevalence=0.005, infection rate=0.7)
eff_individual
eff_Technion's
eff_FAST
FNR
Pooling size
c)
FPR(prevalence=0.005, infection rate=0.7)
eff_individual
eff_Technion's
eff_FAST
FPR
Pooling size


−−
++
+−
−−
−−
++
+−
−−
−−
−−
++/+−
++/+−
Infected
Healthy


Efficiency curve for high risk cases
Efficiency
Pooling size
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
20
40
60
80
100
rho=0.05
rho=0.1
rho=0.15
rho=0.2
rho=0.25
benchmark


-
+
-
-
-
-
+
-
Infected
Healthy


(a)
FAST
Technion's
Pooling size
Prevalence
(b)
FAST
Technion's
FAST/Technion's
Efficiency
Relative Efficiency
Prevalence