CBMM Memo No. 104 **March 26, 2020**

Can we Contain Covid-19 without Locking-down the Economy?

Shai Shalev-Shwartz and Amnon Shashua

Abstract

We present an analysis of a risk-based selective quarantine model where the population is divided into low and high-risk groups. The high-risk group is quarantined until the low-risk group achieves herd-immunity. We tackle the question of whether this model is safe, in the sense that the health system can contain the number of low-risk people that require severe ICU care (such as life support systems).

This work was supported by the Center for Brains, Minds and Machines (CBMM), funded by NSF STC award CCF-1231216.

Can we Contain Covid-19 without Locking-down the Economy?

Shai Shalev-Shwartz and Amnon Shashua*

March 26, 2020

Abstract

We present an analysis of a risk-based selective quarantine model where the population is divided into low and high-risk groups. The high-risk group is quarantined until the low-risk group achieves herd-immunity. We tackle the question of whether this model is safe, in the sense that the health system can contain the number of low-risk people that require severe ICU care (such as life support systems).

1 Introduction

One could consider three models for handling the spread of Covid-19.

1. Risk-based selective Quarantine: Divide the population into two groups, low-risk and high-risk. Quarantine the high-risk and gradually release the low-risk population to achieve a managed herd immunity of that population. The managed phase is designed to allow the health system to cope with the expected number of Severe cases. Given the herd immunity of the low-risk group, we can gradually release the high-risk population. The question is how to manage the release from quarantine of the low and high risk populations in a way that will not overwhelm the health system.

2. Containment-based selective quarantine: Find all the positive cases and put them in quarantine. This requires an estimation of $[t_0, t_1]$ the "contagious time interval" per age group, then given this time interval one could recursively isolate all the individuals at risk from a person that is carrying the virus using "contact tracing". Another tool is predictive testing using contact-tracing to identify people with many contacts with other people and perform tests on them.

3. Country wide (or region wide) lock-down until the spread of the virus is under control. The lock-down could take anywhere from weeks to months. This is the safest route but does not prevent a "second wave" from occurring.

Models 2,3 could work in tandem and have been tried in China and Singapore. Model 3 is currently the default model around the globe and naturally has a tremendous crippling impact on the economy. In the remainder of this white-paper we derive some tools for analyzing the viability of the risk-based model. Specifically, what level of sampling and confidence level can be obtained to make sure that the health system can contain the model?

2 The Risk-based Selective Quarantine Model: How do we Know Whether it is Safe?

Consider a plausible definition of a high-risk group based on a cut-off age and certain pre-existing conditions. For the sake of concreteness, assume the cut-off age is $67+$ which represents the retired segment of society. The low-risk group is the remainder of society which are released to their daily routine while following certain distancing protocols

*Both authors are affiliated with The Hebrew University of Jerusalem, Israel

that is aimed at slowing the spread, while keeping the economy un-disrupted to a large degree, but ultimately reaching a herd immunity level. At that point it is safe to gradually release the high-risk group from quarantine. The question is how do we guarantee that the health system will not be overwhelmed during the spread of the virus in the low-risk group?

Let b be the number of *severe* cases, e.g., those that require a life support system, that the health system can handle, say $b = 600$ for a country of the size of Israel. Let m_s be the number of low-risk people that will develop severe symptoms assuming we will adopt the risk-based selective model. Then, the model is "safe" if $b > m_s$. Our goal is to derive an upper bound on m_s, so that we'll be able to ensure that $b > m_s$.

Let m be the size of the low-risk group and let p^* be the *current*, unknown, percentage of positive cases among the low-risk population. Let k be the *current* number of severe cases among the low-risk population[1] For example, at the time of writing this paper, in Israel $k = 1$ (out of 39 Severe cases over the entire population). Based on this notation, we have that $k/(p^*m)$ is the probability for a low-risk person to be in the Severe category[2]. Thus, for the health system to contain the spread we need to have:

$$b > \frac{k}{p^*m}m,$$

from which we obtain:

$$p^* > \frac{k}{b}\,.$$

Since k and b are known, it is left to derive a method for ensuring that p^* is large enough. We rely on the following lemma.

Lemma 1 *Fix $\delta \in (0,1)$, assume $b > 2k$ an denote $\tilde{p} = \frac{k}{b}$. Suppose we sample $n = \frac{\frac{4}{3}\log(1/\delta)}{\tilde{p}}$ people, uniformly at random from the low-risk population, test all of them for the virus, and find out that there are at least $\frac{8}{3}\log(1/\delta)$ positive cases. Then, with probability of at least $1-\delta$ we have that $p^* \geq \tilde{p}$.*

Proof If $p^* \geq \tilde{p}$ there is nothing to prove, so from now on we assume that $p^* < \tilde{p} < 0.5$. Denote by S_n the number of positive cases. Observe that S_n is the sum of n Bernoulli variables, each of which has probability of p^* to be positive. Then, Bernstein's inequality [1] tells us that for every $t > 0$,

$$\mathbb{P}[S_n - np^* > t] \leq e^{-\frac{t^2}{np^*(1-p^*)+t/3}}$$

Using $p^* < \tilde{p}$ we get that

$$\mathbb{P}[S_n - np^* > t] \leq e^{-\frac{t^2}{n\tilde{p}(1-\tilde{p})+t/3}}\,.$$

Set $t = n\tilde{p}$ we get

$$\begin{aligned}\mathbb{P}[S_n > np^* + n\tilde{p}] &\leq e^{-\frac{(n\tilde{p})^2}{n\tilde{p}(1-\tilde{p})+n\tilde{p}/3}} \\ &= e^{-\frac{n\tilde{p}}{(1-\tilde{p})+1/3}} \\ &\leq e^{-\frac{n\tilde{p}}{4/3}}\end{aligned}$$

Using again $p^* < \tilde{p}$ we also have that

$$\mathbb{P}[S_n > 2n\tilde{p}] \leq \mathbb{P}[S_n > np^* + n\tilde{p}]\,.$$

[1] It is very likely that when the pandemic grows at an exponential rate, the number of current Severe cases does not accurately reflect the total number of Severe cases that corresponds to the current p^*. This is addressable by adding to k the number of Severe cases taken over a future period, say two weeks.

[2] Note that even though p^*, k are based on current data, the probability of a low-risk person to be in the Severe category is fixed.

Combining all of the above, and setting the value of n, yields

$$\mathbb{P}[S_n > 2n\tilde{p}] \leq e^{-\frac{n\tilde{p}}{4/3}} = \delta \ .$$

This means that the probability that S_n will be at least $2n\tilde{p} = \frac{8}{3}\log(1/\delta)$ is at most δ, which concludes our proof. ■

Corollary 1 *If we set $\delta = 0.05$, then if we test $n = 4\frac{b}{k}$ random people for the virus* [3] *and find out that there are at least 8 positive cases, then with probability of at least 95% we know that the risk-based model is safe.*

Next, we may wonder under what conditions would the test likely succeed? we can argue that if $p^* \geq 4\tilde{p}$ we are likely to see at least 8 positive cases. The reasoning is based on the following thread. The average positive cases we expect to see in a random population of size n is np^*, and the standard deviation is $\sqrt{np^*(1-p^*)}$. It follows that with high probability, the number of positive cases will be at least $np^* - 2\sqrt{np^*(1-p^*)}$ (we took two standard deviations). It is easy to verify that if $np^* \geq 16$ then $np^* - 2\sqrt{np^*(1-p^*)} \geq 8$. Hence, if $p^* \geq 4\tilde{p}$, we are likely to see at least 8 positive cases, and our validation of the risk-based model will succeed.

It is worth noting that in the current formulation, we do not assume anything about p^* and only assume the knowledge of b and k. One can also make some assumptions on what is p^*, for example based on statistics from other countries. In this case, as argued above, the experiment is likely to succeed (in the sense of at least 8 positive cases) if $p^* \geq 4\tilde{p} = 4k/b$, we can switch order and say that it is likely that the experiment will succeed if $b \geq 4k/p^*$. For example, suppose that $k = 1$ and let us assume by prior knowledge that p^* is at least 1%[4]. Then, we can set $b = 400$ and the experiment is likely to succeed (and we gained that we only needed $n = 1600$ tests). Naturally, we can also apply binary search in order to search for the right number of tests.

2.1 Estimating p^* from pooled tests

A pulled test is obtained by taking a sample from T persons, mixing all of these samples, and searching for traces of the virus. If the pulled test is positive, it means that at least 1 of the persons is positive. The following lemma generalizes Lemma 1 to pooled tests.

Lemma 2 *Fix $\delta \in (0,1)$ and denote $\tilde{p} = \frac{k}{b}$. Let T be a pool size and denote*

$$\tilde{\phi} = 1 - (1-\tilde{p})^T \ .$$

Suppose we sample $n = \frac{\frac{4}{3}\log(1/\delta)}{\tilde{\phi}}$ pools, uniformly at random from the low-risk population, test all of them for the virus, and find out that there are at least $\frac{8}{3}\log(1/\delta)$ positive pools. Then, with probability of at least $1-\delta$ we have that $p^* \geq \tilde{p}$.

Proof Let ϕ^* be the probability that for a random pool of size T we will have $S_T \geq 1$, where S_T is the number of positive cases in the pool. Note that by definition of the pool test, it ends up positive if and only if $S_T \geq 1$. Using the same proof of Lemma 1, we have that for the n specified in the lemma, if we get at least $\frac{8}{3}\log(1/\delta)$ positive pools then with probability of at least $1-\delta$ we have that $\phi^* \geq \tilde{\phi}$. Next, observe that

$$\phi^* = \mathbb{P}[S_T \geq 1] = 1 - \mathbb{P}[S_T = 0] = 1 - (1-p^*)^T$$

Hence, if $\phi^* \geq \tilde{\phi}$ we have that

$$1 - (1-p^*)^T = \phi^* \geq \tilde{\phi} = 1 - (1-\tilde{p})^T \ .$$

Rearranging terms, we obtain that $p^* \geq \tilde{p}$. ■

The gain from the pool is clearly observed if $\tilde{p}\,T \ll 1$. In this case, using the approximation $1 - x \approx e^{-x}$ we have that $\tilde{\phi} \approx \tilde{p}\,T$. Hence, the number of pooled tests we need is roughly T times less than the number of regular tests we need, in order to show the same conclusions on p^*.

[3] Note that current tests are PCR tests and not serology tests, which only strengthen our results. Furthermore, PCR tests are known to have a negligible level of false positives.

[4] say, through the experiment done in Iceland.

3 Discussion and Implications

The purpose of this document is to provide decision makers a formal and tight bounds to investigate whether the health system can cope with the number of Severe cases that would reach ICU. Embedded in the reasoning is the idea of *selective* quarantine (based on age groups and existing pre-conditions, but could be any other criteria) where the "high-risk" group (the one we suspect will have a high rate of Severe cases) is quarantined and the other is allowed to spread the virus under certain distancing protocols. The underlying premise is that a full population wide quarantine is not a solution in itself — it is merely a step to buy time followed by a more managed (non brute-force) approach. The managed phase underlying our thinking is to create herd immunity of the low-risk group in a controlled manner while keeping the economy going. It is all about keeping the health system in check and not overwhelming its capacity to handle Severe cases. The question we ask in this document is wether we can estimate in advance, through sampling, that the number of Severe cases arising from the low-risk group would not overwhelm the system?

We conclude that this selective quarantine approach will work if p^*, the percentage of positive cases among the low-risk (non-quarantined) population, is not very small, relative to $\tilde{p} = k/b$ which is the ratio between the known number of Severe cases from existing data and the capacity of the system (number of respiratory systems for example). Using concentration bounds we show that if we randomly sample $n \approx 1/\tilde{p}$ people from the low-risk population and find at least few positive cases, then the selective quarantine approach will succeed. Lemma 1 provides the precise bound with a $1 - \delta$ confidence. In Lemma 2 we extend the result to include "pooled" sampling as well. We also conclude that if we have other means of estimating p^* we can use the result that $p^* \geq 4k/b$ in order to estimate the capacity b. For example, random sampling of a population of 2000 individuals in Iceland indicated that $p^* = 0.01$ from which we deduce that $b = 400k$. Finally, the cut-off age of 67 is an arbitrary number we chose for the sake of concreteness but in reality one can pick other optimal cut-off age between low-risk and high-risk groups.

Acknowledgements

The authors wish to thank the faculty members of the computer science department of the Hebrew University for their feedback and comments on earlier drafts of this manuscript.

References

[1] S. N. Bernstein. On a modification of chebyshev's inequality and of the error formula of laplace. *Annals Science Institute Sav. Ukraine*, Sect. Math. 1, 1924.