
**CBMM Memo No. 104** **April 1, 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*

April 1, 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 an intensive care unit (ICU) or specifically respiratory systems, that the health system can handle, say $b = 600$ for a country of the size of Israel[1]. Let $m_q$ be the number of low-risk people that will develop severe symptoms and will require an ICU assuming we will adopt the risk-based selective model. Then, the model is "safe" if $b > m_q$. Our goal is to derive an upper bound on $m_q$, so that we'll be able to ensure that $b > m_q$. Let $m$ be the size of the low-risk group and let $\nu$ be the probability that a person that comes from the low-risk group will develop severe symptoms, specifically require a respiratory system, assuming the person is currently sick. Then,

$$m_q \;=\; m \cdot \nu \,.$$

So, to ensure that $b > m_q$ we will require that $\nu < b/m$.

Before we continue, we note that fully understanding the dynamic of the spread of the virus and the dynamic of the development of the disease (for the sake of knowing when we will need ICUs and for how long) is very challenging and will probably require much more research and time. What we propose here is a *worst-case* analysis. The idea is to adopt a pessimistic view and show that even under this pessimistic view, the health system is not likely to collapse. We already made a pessimistic assumption since we did not take into account the facts that not all of the low-risk population will get sick, and even those that will get sick will not get sick at the same time and will not need an ICU at the same time.

Continuing our derivation, let $p^*$ be the *current*, unknown, percentage of positive cases among the low-risk population and let $k$ be the number of severe cases among the low-risk population from today until a week from now. Assuming that people that are positive cases will either develop severe symptoms within a week or will never develop severe symptoms, then we can estimate $\nu$ as follows:

$$\nu \approx \frac{k}{p^* \, m}.$$

Note that again we are taking a worst-case view. Maybe some of the severe cases in a week's time will be due to people that are not infected today but will get infected tomorrow. This is even likely when the pandemic grows at an exponential rate. Next we turn the approximation into inequalities by upper bounding $k$ (using concentration bounds) and lower-bounding $p^*$ (using sampling of the low-risk population).

**Lemma 1** *Fix some $\delta \in (0,1)$ and let*

$$\tilde{k} \;=\; k + \sqrt{2\log(1/\delta)\, k} + 4\log(1/\delta) \,.$$

*Then, with probability of at least $1-\delta$ we have that*

$$\nu \leq \frac{\tilde{k}}{p^* \, m} \,.$$

The proof of the lemma follows directly from measure concentration bounds (see for example Lemma B.10 in [2]). Tighter bounds can be obtained. A tighter analysis for a particular case study, where $k \geq 15$ and $p^* m > 10,000$, is provided in Appendix A.

Based on this lemma, we can upper bound $m_q$ (with probability of at least $1-\delta$) by

$$m_q \leq \frac{\tilde{k}}{p^* m} \cdot m = \frac{\tilde{k}}{p^*}$$

since even if *all* of the young population will get the virus, the number of severe cases will be the right hand side of the above equation. Of course, in reality we expect $m_q$ to be much smaller than the above, both because not all of the

[1] Israel has around 6 critical beds per 100,000 inhabitants.

young population will get the virus and because not all the severe people will be sick at the same time. Nevertheless, as mentioned previously we adopt a worst-case analysis.

The health system will be able to treat all of these severely sick people if $b > m_q$, from each we obtain the requirement:

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

Since $\tilde{k}$ and $b$ are known, it is left to derive a method for ensuring that $p^*$ is large enough.

This is a classical hypothesis testing problem: denote $\tilde{p} = \frac{\tilde{k}}{b}$, the null hypothesis is that $p^* \leq \tilde{p}$, and the alternative hypothesis is that $p^* > \tilde{p}$. By sampling $n$ persons, uniformly at random from the low-risk population, we would like to find a cutoff value such that if the number of positive cases in our sample is above the cutoff, then we can reject the null hypothesis with a confidence of, say, $95\%$.

While one can find cutoff values for different values of $n$ by using tables of the Binomial distribution, the following theorem specifies formulas for the sample size and for the cutoff value. Another variant, which optimizes the lower bound on $p^*$ rather than the size of the sample, is given in Appendix A.

**Theorem 1** *For $\tilde{p} \in (0, 0.1)$, setting the sample size to be $n = 4.438/\tilde{p}$ and the cutoff to be $10$, we can reject the null hypothesis with a confidence of at least $95\%$. In other words, if we test $n = 4.438\, b/\tilde{k}$ random people for the virus [2] and find out that there are more than $10$ positive cases, then with probability of at least $95\%$ we know that the risk-based model is safe.*

The proof of the theorem follows directly from the following lemma.

**Lemma 2** *Fix $\delta \in (0, 1)$ and $\tilde{p} \in (0, 1/2)$. Suppose we sample $n = \frac{\frac{4}{3}\log(1/\delta)}{\tilde{p}(1-\tilde{p})}$ i.i.d. Bernoulli variables with parameter $p^*$ and get at least $\left(\frac{4}{3(1-\tilde{p})} + 2\right)\log(1/\delta)$ positives. 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 $t > 0$,

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

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

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

Setting

$$t = 2\log(1/\delta) \quad , \quad n = \frac{\frac{4}{3}\log(1/\delta)}{\tilde{p}(1-\tilde{p})}$$

we have that

$$n\tilde{p}(1-\tilde{p}) = \frac{4}{3}\log(1/\delta) .$$

Therefore,

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

[2] 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.

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

$$\mathbb{P}[S_n > n\tilde{p} + 2\log(1/\delta)] \leq \mathbb{P}[S_n > np^* + 2\log(1/\delta)] \ .$$

Combining all of the above yields

$$\mathbb{P}[S_n > n\tilde{p} + 2\log(1/\delta)] \leq \delta \ .$$

This means that the probability that $S_n$ will be at least

$$n\tilde{p} + 2\log(1/\delta) = \left(\frac{4}{3(1-\tilde{p})} + 2\right) \log(1/\delta)$$

is at most $\delta$, which concludes our proof. ■

Next, we may wonder under what conditions would the hypothesis testing likely succeed? we can argue that if $p^* \geq 4.438\tilde{p}$ we are likely to see at least 10 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 $np^* \geq 19$ then $np^* - 2\sqrt{np^*(1-p^*)} \geq 10$. Hence, if $p^* \geq 4.438\tilde{p}$, we are likely to see at least 10 positive cases, and our validation of the risk-based model will succeed.

Alternatively, if $p^* < 4.438\tilde{p}$ one can obtain a tighter lower bound on $p^*$ by increasing the sample size $n$. Appendix A.2 shows that with $n \geq 100/\tilde{p}$ and with $S_n$ being the number of positive cases in the sample, then with probability of at least 95% we have

$$p^* \geq \max\{0.83\,\frac{S_n}{n}\ ,\ \tilde{p}\}.$$

Taken together, the tightest bound we have so far for the case of $k \geq 15$, $n > 100/\tilde{p}$, $p^*m > 10,000$ gives us a bound on the capacity $b$ of:

$$b \geq 1.92\frac{k}{S_n/n}.$$

## 2.1 Estimating $p^*$ from pooled tests

A pooled test is obtained by taking a sample from $T$ persons, mixing all of these samples, and searching for traces of the virus. If the pooled test is positive, it means that at least 1 of the persons is positive. The following lemma generalizes Lemma 2 to pooled tests.

**Lemma 3** *Fix $\delta \in (0,1)$ and denote $\tilde{p} = \frac{\tilde{k}}{b}$. Let $T$ be a pool size and denote*

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

*Suppose we sample $n = \frac{\frac{4}{3}\log(1/\delta)}{\tilde{\phi}(1-\tilde{\phi})}$ pools, uniformly at random from the low-risk population, test all of them for the virus, and find out that there are at least $\left(\frac{4}{3(1-\tilde{\phi})} + 2\right) \log(1/\delta)$ positive pools. Then, with probability of at least $1-\delta$ we have that $p^* \geq \tilde{p}$.*

**Proof** Let $\phi^*$ 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 2, we have that for the $n$ specified in the lemma, if we get at least $\left(\frac{4}{3(1-\tilde{\phi})} + 2\right) \log(1/\delta)$ 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 Discussion and Implications

In the event a risk-based quarantine approach would be contemplated by decision makers, 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 whether we can estimate in advance, through sampling, that the number of severe cases arising from the low-risk group would not overwhelm the system?

Taking Israel as a case study, as of March 30, $k = 15$ (out of 74 severe cases in ICU). As of today, there is no scientific estimation of $p^*$ only indications. We believe that $p^* = 0.02$ as of a week ago. Substituting the bound $\tilde{k} = 1.6k$ from Appendix A we obtain:

$$b \geq 1.6\frac{k}{p^*} = 1200.$$

This translates to 12 critical beds per 100,000 inhabitants. Going from 6 beds to 12 is something that Israel can reasonably handle. Talks are to go up to 50 critical beds during this crisis — therefore, the analysis indicates that releasing the low-risk group to achieve herd immunity is not unreasonable. Of course, one cannot do anything operational until an i.i.d. sample of the low-risk population is conducted. From the bounds of Apprendix A.2 we conclude that $n \geq 100\tilde{k}/b$ translates to $n = 5000$ (assuming capacity of 1200 critical beds) which is a reasonable sample size given Israel's current PCR testing capabilities.

Some additional points worth mentioning. The random variable $k$ could be a very useful indicator to decide on what constitutes the high-risk group. What age cut-off and what preexisting conditions to include. At any given time, we would want the definition of the high-risk group to create a small value of $k$ (as the capacity $b$ monotonically increases with $k$). For example, in our case study we decided on a cut-off age of 67 and looked among the 74 severe cases for those without *any* pre-existing conditions. In order to avoid over-fitting this kind of study should be done with data preferably coming from other countries.

Another point worth mentioning is that the risk-based quarantine model is not only beneficial form the point of view of economical sustainability. Among other selective quarantine ideas (like based on geography or contact-tracing isolating the infected and those around them) the risk-based approach has the better chances of reducing the overall mortality rate. The reason is that the highest mortality is with the high-risk group which in this model is isolated. When the high-risk group is released from isolation they would be facing a largely immune population thus naturally facing a very slow spread of infection with a good chance to whither the storm until a cure or vaccine is available. In all other selective quarantine models the high and low risk are equally susceptible to be infected so that even if the health system is not overwhelmed still the mortality of the high-risk group is likely to be higher than the risk-based model.

Yet another point worth mentioning is that we focused on what is "safe" for the health system in the sense of how to estimate the number of severe cases that would be low enough not to overwhelm the ICUs. We ignored the fact that some severe cases could end up in the mortality statistics even when given proper care. In fact there are two probabilities to estimate (i) the probability of being in the "severe" category among the low-risk group, and (ii) the probability of mortality given proper care. We have bounded the former and ignored the latter. The reason for doing

so is that the latter is beyond the scope of this paper because it is essentially a moral tradeoff between "safety" and "usefulness" that is employed in every aspect of society. For example, society does not put a lockdown on passenger car use in order to significantly reduce car accidents even though such a lockdown will save lives. Likewise, governments do not allocate infinite budgets for the health system even though there is a correlation between increased investments and saving lives.

As a final remark, going out of quarantine is a choice not an obligation. This is no different than people that are afraid of flights and decide not to go on an airplane. Families can decide to stay quarantined either as an extra safety measure or if some members of the family are from the high risk group while the others are from the low risk group.


# Acknowledgements

The authors wish to thank the faculty members of the computer science department of the Hebrew University, as well as to Prof. Peter Bartlett, Prof. Nir Friedman, Prof. Katrina Ligett, Prof. Nati Srebro, Prof. Herve Bercovier, and Dr. Renana Eitan for comments and feedback on earlier drafts of this paper.


# A A Tighter Analysis

## A.1 A tighter bound for $\tilde{k}$

In this section we derive a tighter value for $\tilde{k}$ assuming that we fix $\delta$ to be $0.05$ and assuming that $n := p^* m > 10000$.

**Lemma 4** *Suppose $k \geq 15$, and setting $\tilde{k} = 1.6k$, then with probability of at least $1-\delta$ we have that $\nu \leq \tilde{k}/(p^* m)$.*

To prove the lemma, we rely on the analysis given in [3], in particular, the following theorem due to [4].

**Theorem 2** *Let $S_n \sim \text{Binomial}(n,p)$, where $p \in (0,1)$ and $n > 0$, then for $k \in \{1,\ldots,n-2\}$ we have*

$$\Phi\left(\text{sign}(k-np)\sqrt{2nD(p,k/n)}\right) \leq P[S_n \leq k] \leq \Phi\left(\text{sign}(k+1-np)\sqrt{2nD(p,(k+1)/n)}\right)$$

*where $\Phi$ is the cumulative distribution function of a standard normal variable and $D(p,c) = c\log(c/p) + (1-c)\log((1-c)/(1-p))$ is the KL divergence.*

To prove the lemma, let the null hypothesis be that $\nu > 1.6k/n$. The extreme case is when $\nu = 1.6k/n$, so lets assume this is the case. We have that

$$P[S_n \leq k] \leq \Phi\left(-\sqrt{2nD(1.6k/n,(k+1)/n)}\right)$$

Observe:

$$\begin{aligned}
nD(1.6k/n,(k+1)/n) &= n\left(\frac{k+1}{n}\log(\frac{k+1}{1.6k}) + (1-\frac{k+1}{n})\log\left(\frac{n-(k+1)}{n-1.6k}\right)\right) \\
&= (k+1)\log(\frac{k+1}{1.6k}) + (1-\frac{k+1}{n})\,n\,\log\left(\frac{n-(k+1)}{n-1.6k}\right) \\
&= (k+1)\log(\frac{k+1}{1.6k}) + (1-\frac{k+1}{n})\,n\,\log\left(1+\frac{n-(k+1)-(n-1.6k)}{n-1.6k}\right) \\
&= (k+1)\log(\frac{k+1}{1.6k}) + (1-\frac{k+1}{n})\,n\,\log\left(1+\frac{0.6k-1}{n-1.6k}\right) := h(k,n)
\end{aligned}$$

Let's analyze the function $h(k,n)$. First, fix $k$ and let's show that $h(k,n)$ monotonically decreases when $n$ increases.

$$\begin{aligned}\frac{dh(k,n)}{dn} &= \log\left(1+\frac{0.6k-1}{n-1.6k}\right) + (1-\frac{k+1}{n})\, n\, \frac{n-1.6k}{n-(k+1)}\, \frac{1-0.6k}{(n-1.6k)^2} \\ &= \log\left(1+\frac{0.6k-1}{n-1.6k}\right) + \frac{1-0.6k}{n-1.6k} \\ &\leq \frac{0.6k-1}{n-1.6k} + \frac{1-0.6k}{n-1.6k} = 0\end{aligned}$$

where we used $\log(1+x) \leq x$ for $x > 0$.

Therefore,

$$h(k,n) \geq h(k,\infty) = (k+1)\log(\frac{k+1}{1.6k}) + 0.6k - 1$$

Finally, the right-hand side of the above monotonically increases with $k$, and for $k = 15$ we get that $\Phi\left(-\sqrt{2h(k,\infty)}\right) \leq 0.05$, from which the claim follows.

## A.2 A tighter estimation for $p^*$

In this section we derive a tighter lower bound on $p^*$, based on the value of $\tilde{p}$ and $S_n$, assuming that we fix $\delta$ to be $0.05$. Here, $n$ is the size of the i.i.d. sample from the low risk population.

**Lemma 5** *Fis $\delta = 0.05$ and $\tilde{p} \in (0,1)$, let $n \geq 100/\tilde{p}$, and let $S_n \sim \text{Binomial}(n, p^*)$ for some unknown $p^* \in (0,1)$. Then, with probability of at least $1-\delta$ we have that*

$$p^* \geq \max\left\{0.83\,\frac{S_n}{n}\ ,\ \tilde{p}\right\}\ .$$

**Proof** Using Lemma B.10 in [2], we have, w.p. $\geq 1-\delta$

$$p^* + \sqrt{p^*}\cdot\sqrt{\frac{2\log(1/\delta)}{3n}} + \frac{2\log(1/\delta)}{n} - \frac{S_n}{n} \geq 0\ .$$

For $\delta = 0.05$ we have $\log(1/\delta) = 2.9957 < 3$. Then,

$$p^* + \sqrt{p^*}\cdot\sqrt{\frac{2}{n}} + \frac{6}{n} - \frac{S_n}{n} \geq 0\ .$$

Either $p^* \geq \frac{100}{n}$, or, if not, then

$$p^* \geq \frac{S_n}{n} - \frac{\sqrt{200}+6}{n}$$

The extreme case is when $n = 100/\tilde{p}$. Then, either $p^* \geq \tilde{p}$ or $p^* \geq \frac{S_n}{n} - \frac{\sqrt{200}+6}{100}\tilde{p} \geq \frac{S_n}{n} - 0.2p^*$. In the latter case $p^* \geq 0.83\,\frac{S_n}{n}$. We can conclude that w.p. 95% we have $p^* \geq \max\{0.83\,\frac{S_n}{n}\ ,\ \tilde{p}\}$. ■

# 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.

[2] S. Shalev-Shwartz and S. Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

[3] M. Short. Improved inequalities for the poisson and binomial distribution and upper tail quantile functions. *ISRN Probability and Statistics*, 2013, 2013.

[4] A. M. Zubkov and A. A. Serov. A complete proof of universal inequalities for the distribution function of the binomial law. *Theory of Probability & Its Applications*, 57(3):539–544, 2013.