

THE FAILURE OF THE 2021 FMCO IN MALAYSIA

DAVID BENJAMIN LIM

ABSTRACT. We show for Pulau Pinang, Johor and Negeri Sembilan that during the first 6 weeks of total lockdown, workplace clusters contributed significantly to the state R_0 . A similar statement is still true for the Klang Valley, though in the first 4 weeks rather than 6. This suggests that policy-makers in Malaysia should consider measures targeted at such clusters rather than blanket approaches to lockdowns.

CONTENTS

1. Introduction	1
2. Acknowledgments	2
3. The dangers of overly simplistic thinking	2
4. Summary of results	2
4.1. Conclusion for Johor, Pulau Pinang and Negeri Sembilan	2
4.2. The Klang Valley	3
4.3. Final comments	3
Appendix A. Measuring the total contribution of a particular sector to the total R_0	4
A.1. Theoretical Foundations	4
A.2. Real-world applications of Theorem A.1	4
A.3. An example calculation	5
Appendix B. Factory clusters or sporadic cases: A Bayesian viewpoint	6
References	7

1. INTRODUCTION

The COVID-19 pandemic has forced governments around the world to take measures to contain the spread of the disease. In Malaysia, the government has already implemented several rounds of lockdowns, the most recent of which is the FMCO (full movement control order) which started on 1/6/2021. The FMCO is to this date mostly at full strength albeit for a few states that have been scheduled to move on to what the Malaysian government calls “Phase 3” (Sarawak, Perlis and Labuan). During this phase, things such as indoor dining and gyms are allowed to operate among other things.

For any action taken to control the spread of COVID-19, there is an associated cost. Businesses ordered to close result in lost income and in worst-case scenarios the destruction of careers that have been built over many years, while sedentary lifestyles associated with stay-at-home orders impact mental/physical health. Given these facts, it is absolutely crucial that any lockdown initiated be effective in controlling the spread of COVID-19. Nonetheless, the total lockdown initiated on 1/6/2021 has largely failed. Case numbers dipped initially for the first 2 - 3 weeks, before rising again and hitting a new record-high of 11,618 cases on 14/7/2021.¹ Therefore, it is absolutely important to understand *why* the 2021 FMCO failed. In this article, we argue that in states with a large number of workplace clusters, namely the Klang Valley (Selangor and Kuala Lumpur),² Johor, Pulau Pinang and Negeri Sembilan, that the primary reason for this is due to the continued generation of such clusters during the crucial first 6 weeks of lockdown.

Our method is very simple. For each of these states, we calculate the basic reproduction number R_0 . Then, using data from the official Ministry of Health GitHub, we calculate the contribution of workplace

Date: September 13, 2021.

¹Cases have continued to rise since then, hitting a record high of 21,668 on 12/8/2021.

²Given how interlinked Selangor and Kuala Lumpur are, it is necessary to consider them together as one “state.”

clusters to R_0 . If upon subtracting this contribution from the total R_0 , we find a number significantly less than 1, this gives strong evidence that *if* there were no workplace clusters during this time period, COVID-19 would have *decreased*. See Section 3 and 4 for further discussion and results.

2. ACKNOWLEDGMENTS

We thank Eric Cooper of the Schleier-Smith Lab, Stanford University and Samuel Tenka of the Computer Science and Artificial Intelligence Lab, MIT for many insightful discussions concerning the mathematics behind COVID-19. We also thank Yang Liu (Stanford University), Mark Sellke (Stanford University) and Brian Lawrence (University of Chicago) for many helpful discussions. We are also grateful to have received the support of Latifah Hani Hamzah (Stanford University). We thank Dr. Arvinder Singh for pointing out that the nature of workplace clusters is such that one necessarily over-samples from these and under-samples from the general population.

3. THE DANGERS OF OVERLY SIMPLISTIC THINKING

In a speech in parliament on 27/7/2021, the minister of international trade and industry, Dato’ Seri Azmin Ali said that the manufacturing sector contributed to only 8.8% of all cases between 1/6/2021 and 23/7/2021 [Ubah21, 3:07:00]. After saying this, he says (in Malay): “Saya boleh bagi lagi sampling, Datuk Speaker sama ada harian, mingguan ataupun bulanan, yang membuktikan sumbangan sektor perkilangan kepada kes-kes harian (mic cuts out)” [Ubah21, 3:07:10]. In english, this translates to: “I can give you a sampling, Datuk Speaker be it daily, weekly or monthly, that proves that the contribution of the manufacturing sector to daily cases (mic cuts out).” Although the rest of this sentence cannot be heard, it can be inferred that the minister claims that because the manufacturing sector only contributed to 8.8% of all COVID-19 cases (during the time period mentioned above), that it *cannot* be the cause of the spread of COVID-19 in Malaysia.

However, this line of reasoning is wrong: More precisely, just because a particular group of people make up a small percentage of all COVID-19 cases, *does not mean* that they *cannot* be the source of the problem (see Appendix A). The reason for this is that it is entirely possible for there to be a small subset of the population that is very infectious (i.e. have very high R_0), and therefore contributes significantly to the *total* R_0 of Malaysia (Theorem A.1). To be maximally precise, we do not claim that the manufacturing sector *is* the source of the COVID-19 problem in Malaysia. All we are doing is pointing out that the line of reasoning employed by Dato’ Seri Azmin Ali is too simplistic to have any meaningful content.

4. SUMMARY OF RESULTS

This section relies on the theoretical ideas developed in Appendix A. All our results below are for the time period of 1/6/2021 to 14/7/2021.

TABLE 1. R_0 for each state

State	R_0	Contribution from workplace clusters	Residual
Klang Valley	1.121	0.220	0.903
Johor*	0.969	0.473	0.500
Pulau Pinang*	0.958	0.715	0.244
Negeri Sembilan	1.110	0.403	0.708

Table 1 summarizes the R_0 for each state, the contribution to the R_0 from workplace clusters, and the residual, i.e. the contribution from non-workplace clusters. For Johor, Pulau Pinang and Negeri Sembilan, the residual is significantly less than 1, while for the Klang Valley it is less than 1 but by not very much. The asterisks next to Pulau Pinang and Johor simply denote corrections that one must apply when calculating these numbers due to the high proportion of workplace cases to all cases. We refer the reader to A.3 for details.

4.1. Conclusion for Johor, Pulau Pinang and Negeri Sembilan. Given the residual for these states is significantly less than 1, we obtain *very strong* evidence that if there were no more workplace clusters during this time period, COVID-19 would have decreased. Indeed, even the *proportion* of cases from workplaces during this time period was very high:

TABLE 2. Breakdown of cases from Johor, Pulau Pinang and Negeri Sembilan

State	Total cases	Cases from workplace clusters	Proportion from workplace clusters
Johor	17654	8347	$\approx 50\%$
Pulau Pinang	8668	6222	$\approx 70\%$
Negeri Sembilan	29451	8475	$\approx 30\%$

Without even referring to R_0 , the reader can clearly see that for Pulau Pinang, workplaces were the problem - they contributed to 70% of all cases during this time period! Negeri Sembilan on the other hand is where this R_0 stuff hammers home the point. Cases from workplace clusters only constitute 30% of total cases, but the contribution to the R_0 is significant enough that the *residual* is 0.708 - much less than 1!

4.2. The Klang Valley. Exploring the Klang Valley in greater detail, we note that if we restrict to the first 4 (rather than 6) weeks of lockdown in the Klang Valley, we get a much more significant contribution to the R_0 from workplace clusters:

TABLE 3. R_0 for the Klang Valley from 1/6/2021 - 1/7/2021

State	R_0	Contribution from workplace clusters	Residual
Klang Valley	1.017	0.227	0.793

During this time period (1/6/2021 - 1/7/2021) we note that cases from workplace clusters only constitute 11979/64710 $\approx 18.5\%$ of all cases, but their contribution to the R_0 is sufficiently big that the residual is much less than 1. At any rate, regardless of whether we look at the first 4-6 weeks of the 2021 FMCO, one can still make the argument that the contribution from workplace clusters is still significant, though not as strong Johor, Pulau Pinang and Negeri Sembilan. Nonetheless, we would like to point out that due to the higher case numbers in the Klang Valley compared to these other three states ($5\times$ higher population adjusted), it is entirely possible that contact tracing simply became impossible after the first generation, thus undercounting the true case numbers from these clusters.

4.3. Final comments. As pointed out by Dr. Arvinder Singh, there is an inherent bias in the way that testing is done in workplaces. Namely, the ministry of health tends to sample more in workplace clusters as compared to sporadic cases. This affects the calculations above by way of the probability p of a workplace cluster. Examining Appendix A.3 more closely, we see that the denominator in the calculation of p depends on the *recorded number* of total COVID-19 cases, of which there could be many more (test positivity rates $> 10\%$ during this time period suggest this). Nonetheless, even if we multiply the denominator of p by 2 (or even 3), i.e. we believe that the “true” number of COVID-19 cases is double (or triple) what is currently recorded, not much changes for Pulau Pinang, Johor and Negeri Sembilan. The same is true for the Klang Valley if we look at the first 4 weeks of the 2021 FMCO. However, if we expand our time horizon to 6 weeks from the start of the 2021 FMCO, the residual for the Klang Valley is very close to 1, and so inserting a factor of 2 or 3 in the denominator for p brings the residual above 1, weakening the argument for the Klang Valley.

We end this article with a question that the curious reader may wish to think about:

Question 4.1. When we find a bunch of cases from a factory, is it true that these are started by a single (or multiple) person (people) within the factory? Maybe it so happened to be the case that these people got COVID-19 in the wild, but showed up upon mass screening done at the factory?

In Appendix B, we show that when we find 50 positive cases out of 500 tests in a factory, that it is extremely likely that these came from a cluster, rather than the wild.

APPENDIX A. MEASURING THE TOTAL CONTRIBUTION OF A PARTICULAR SECTOR TO THE TOTAL R_0

A.1. Theoretical Foundations.

Theorem A.1. Consider the random tree generated by the following algorithm:

```
# Initialize k nodes
# Initialize number of generations n
# Initialize probabilities p and q with p + q = 1
# Initialize growth rates R_p and R_q

do n times:
  for node in nodes:
    node.color = red w/ probability p, blue w/ probability q
    if node.color is red:
      add R_p children to node
    else:
      add Pois(R_q) children to node
  nodes = nodes.children
```

Then as $n \rightarrow \infty$, the expected number of nodes is infinite if $pR_p + qR_q > 1$, and $k/(1 - (pR_p + qR_q))$ (in particular finite!) if $pR_p + qR_q < 1$.

Proof. We use 0-indexing. Let X_i be the random variable that counts the number of nodes at generation i ; by definition $X_0 = k$. Conditioning on whether a node at generation $i - 1$ is blue or red, we have

$$\mathbf{E}[X_i | X_{i-1}] = X_{i-1}(pR_p + qR_q).$$

Hence, by the law of total expectation,

$$\mathbf{E}[X_i] = \mathbf{E}[\mathbf{E}[X_i | X_{i-1}]] = \mathbf{E}[X_{i-1}(pR_p + qR_q)] = \mathbf{E}[X_{i-1}](pR_p + qR_q).$$

It follows that $\mathbf{E}[X_i] = k(pR_p + qR_q)^i$ and hence

$$\begin{aligned} \text{Expected total nodes after } n\text{-generations} &= \sum_{i=0}^{n-1} \mathbf{E}[X_i] \\ &= \sum_{i=0}^{n-1} k(pR_p + qR_q)^i \\ &= k \left[\frac{(pR_p + qR_q)^n - 1}{(pR_p + qR_q) - 1} \right]. \end{aligned}$$

Letting $n \rightarrow \infty$, this computation shows that if $pR_p + qR_q > 1$, then on average the tree has an infinite number of nodes, and if it is less than 1, it has $k/(1 - (pR_p + qR_q))$ nodes. \square

A.2. Real-world applications of Theorem A.1. Theorem A.1 above shows the following. It is possible to have a proportion p of red nodes, with $p \ll q$, but with large R_p in a way that $pR_p + qR_q > 1$. The random tree we generate in Theorem A.1 above will grow indefinitely even though the total proportion of red nodes is small. Furthermore, one may also arrange so that $R_q < 1$, so that if there are no red nodes present (set $p = 0$ and $q = 1$), then the number of nodes in the tree will be finite. This shows that even though the red nodes make up a very small portion of the tree, it is precisely these nodes that contribute to the tree growing indefinitely.

Said differently, if a red node represents a workplace cluster and blue node a sporadic case, then it is entirely possible for the number of workplace clusters (red nodes) to be very small, but that it is the main driver behind the spread of COVID-19. In practice, Theorem A.1 can be used to as follows. Fix some time period. Then (for example) red nodes may be workplace clusters and blue sporadic cases. The probability p (and hence $q = 1 - p$) can be empirically calculated from the file `epidemic/clusters.csv` in the (public!) GitHub repository of the Malaysian Ministry of Health [MoH21]. The same is true of R_p , and then $R_q = (R_0 - pR_p)/q$ is calculated using R_p and the total R_0 during this time period. If we find that $R_q < 1$, this gives strong evidence to the fact that if there were *no* workplace clusters during

our fixed time period, that COVID-19 would have “died off,” i.e. workplaces were behind the spread of COVID-19.

As a final note, for the purpose of quick and fast empirical estimates, R_0 during this time period may be calculated as follows. Let t_{start} and t_{end} be the start and end of the time period respectively, and $f(t)$ the number of daily new cases reported at time t (The value $f(t)$ can be obtained empirically from the file `epidemic/cases_malaysia.csv` of *loc. cit.*, or for a breakdown by state, `epidemic/cases_state.csv`). Then

$$(1) \quad R_0 \approx \exp\left(\frac{\log f(t_{\text{end}}) - \log f(t_{\text{start}})}{\text{length (in days) of time period}} \times \tau\right),$$

where τ is the serial interval of COVID-19 (approximately 5.2 days).

A.3. An example calculation. Let us think of a workplace cluster as corresponding to a single red node, and every other COVID-19 case as a single blue node. We now show how the ideas in this appendix can be used in practice.

Recall first that p is the probability of a red node (i.e. workplace cluster) and R_p the number of children of a red node (i.e. the average size of a workplace cluster). We have

$$(2) \quad p = \text{probability of a workplace cluster}$$

$$(3) \quad = \frac{\text{\#people that started a workplace cluster}}{\text{\#people who could have started a workplace cluster}}$$

$$(4) \quad \approx \frac{\text{\#workplace clusters}}{\text{Total cases} - \text{Total workplace cases} + \text{\#workplace clusters}}.$$

When the number of cases from workplaces is very high compared to the total number of cases (e.g. Johor and Pulau Pinang), we need to take the denominator in (4) to be the total number of cases. If we do not do so, p will be so big that $R_0 - pR_p$ will be negative. The quantity R_p , i.e. the average size of a workplace cluster is

$$R_p = \frac{\text{Total workplace cases}}{\text{\#workplace clusters}}.$$

For the sake of concreteness, let us now calculate p and R_p for the whole of Malaysia during the time period 1/6/2021 - 23/7/2021. We choose this time period because all the data we need has already been conveniently laid out in the form of a facebook post by Dato’ Seri Azmin Ali [AA21].³ For the results in Section 4, we simply wrote a simple python script to process data from the files `epidemic/cases_state.csv` and `epidemic/clusters.csv` of the Ministry of Health’s GitHub [MoH21]. The calculation is *mutadis mutandis* the same as what follows. As per Dato’ Seri Azmin Ali’s facebook post, we have

$$\begin{aligned} \text{Total cases} &= 408,134 \\ \text{Total workplace cases} &= 60,587 \\ \text{\#workplace clusters} &= 743 \end{aligned}$$

from which it follows that $p = 0.0021$, $R_p = 81.5437$. Hence the *total contribution* to the R_0 of Malaysia from workplaces is

$$pR_p = 0.1712.$$

On the other hand, by the method of least squares, we find that the R_0 for Malaysia is

$$R_0 = 1.0764.$$

The reader may obtain a similar value without the method of least squares using the estimate (1). The residual R_q (i.e. R_0 from non-workplace clusters) is thus

$$R_q = \frac{R_0 - pR_p}{1 - p} = 0.907.$$

³The data from this facebook page is from the Ministry of Health’s GitHub [MoH21].

APPENDIX B. FACTORY CLUSTERS OR SPORADIC CASES: A BAYESIAN VIEWPOINT

We will answer Question 4.1 from a Bayesian viewpoint. The reader that wishes to skip details may go straight to Punchline B.1. Let us consider two hypotheses, H_0 and H_1 , defined as follows:

H_0 : The 50 cases from this factory are all sporadic.

H_1 : The 50 cases from this factory are from a cluster started from a single seed.

Let D denote the data of having observed 50 positive cases out of the 500 positive tests conducted at this factory. We want to study the ratio

$$(5) \quad \frac{P(H_1|D)}{P(H_0|D)} = \frac{P(D|H_1)}{P(D|H_0)} \times \frac{P(H_1)}{P(H_0)}.$$

Now the sampling theorist might object that we have to choose priors for $P(H_1)$ and $P(H_0)$, but there is no need to do so: We will be *extremely generous* to H_0 and say that

$$(6) \quad P(H_0) = 10 \times P(H_1),$$

i.e. it is 10 *times* more likely that these 50 cases are sporadic rather than from a cluster. Great, so it remains to calculate the Bayes factor

$$(7) \quad \frac{P(D|H_1)}{P(D|H_0)}.$$

Let p_{sporadic} denote the probability that a randomly selected person from the population will get COVID-19, i.e. the number of active cases divided by the total population. The denominator of the Bayes factor (7) is then

$$P(D|H_0) = \binom{500}{50} p_{\text{sporadic}}^{50} (1 - p_{\text{sporadic}})^{450}.$$

For the sake of concreteness, let us again be very generous to H_0 and take $p_{\text{sporadic}} = 0.05$. To convince the reader that this is very generous, consider Pulau Pinang on 19/8/2021 [PLC21]. There were 8613 active cases with a population of about 1.8×10^6 people. So $p_{\text{sporadic}} = 0.05$ means that we believe that there are 90,000 active cases in Penang right now, or almost *ten times* the number of recorded active cases. Assuming this, we now have

$$(8) \quad P(D|H_0) = \binom{500}{50} (0.05)^{50} (0.95)^{450} \approx 2 \times 10^{-6}.$$

On the other hand, the numerator of (7), $P(D|H_1)$ is the probability that we observe these 50 positive out of 500 tests, *given that they all came from a single seed*. Now suppose that a single seed starts a cluster in a factory. We do not know how many cases this cluster will have, but one *reasonable* thing to say is that if we conduct a fixed number of tests (such a 500 in our case), it is equally likely that any of the numbers 2, ..., 500 occurs as a possibility for the number of cases that we detect. In other words,

$$(9) \quad P(D|H_1) = \frac{1}{500}.$$

Plugging (6), (8) and (9) into (5), we get

$$\frac{P(H_1|D)}{P(H_0|D)} = \frac{1}{10} \times \frac{\frac{1}{500}}{2 \times 10^{-6}} = 100.$$

Punchline B.1. Even after being *extremely generous* to the null hypothesis H_0 , the probability that the 50/500 positive tests from this factory came from a cluster, is still 100 times greater than the probability that they are sporadic. In other words, it is *far more likely* that this was a factory cluster than a random sampling of COVID-19 cases.

We remark that our model above is necessarily over simplified in several ways. For example, we assumed that different people are totally independent in the non-cluster situation. The value $P(D|H_1) = 1/500$ is a reasonable "uninformed" choice in the absence of a more precise understanding of typical cluster sizes for COVID 19. However since we obtained a large likelihood ratio of 1000, we expect the punchline to be robust to small changes in the model.

REFERENCES

- [AA21] Dato' Seri Azmin Ali, *Kes COVID-19 dalam Sektor Pembuatan di Malaysia*, <https://www.facebook.com/MohamedAzminAli/photos/4322810617783251>, 2021, [Online; accessed 16-August-2021].
- [PLC21] Penang Lawan COVID-19, *19 Ogos 2021 12pm - Jumlah Kes COVID-19 di Pulau Pinang*, <https://www.facebook.com/penanglawancovid19/photos/397041928606278>, 2021, [Online; accessed 20-August-2021].
- [MoH21] Ministry of Health Malaysia, *Open data on COVID-19 in Malaysia*, <https://github.com/MoH-Malaysia/covid19-public/>, 2021, [Online; accessed 16-August-2021].
- [Ubah21] Ubah TV, *Sidang Khas Dewan Rakyat, 26 Julai 2021 (Sesi Petang)*, <https://www.youtube.com/watch?v=R1FMhYkgBOY>, 2021, [Online; accessed 16-August-2021].