A Parsimonious Risk Model to Assist Managers in Deciding When to Close and Reopen Business Premises Based on the Probability of Viral Infection

Stephen Duchesne
Castellan Capital

Kingsley Jones
Jevons Global

The human, business and economic impact of the COVID-19 pandemic has been unprecedented. Here it is argued that basic probability theory, combined with simple scenario planning, can be of value in the management of business risk alongside human resources planning. Specifically, we develop guidance on when to close and later reopen group-work office spaces based on the group size and estimated probability that one or more persons within a given group may have contracted the virus. Examples are included from the live scenario planning exercise conducted by the authors in managing their own businesses.

Keywords: Covid-19, business risk, probability, remote working, risk management, health and safety, OH&S

INTRODUCTION

Government countermeasures to the SARS-Coronavirus 2 (“COVID-19”) pandemic have resulted in global social and business disruption not seen since the World Wars. The economic cost, as a result of the deaths of so many that become infected, and its burden on the health and hospital system, is significant. The economic cost as a result of the countermeasures (travel bans, reduced gathering numbers, social distancing, business closures etc), and the explicit costs of Government financial support, subsidies and tax relief, as well as the impact of what is almost surely to be a deep recession, is enormous, both here in Australia, and globally. The impact on society at large and the fiscal and monetary policies of almost every Government will be felt for decades but, without over exaggerating the situation, for many businesses, small and large, how they react to this crisis will determine their financial survival as it will not come without material explicit and implicit cost.

Businesses critically rely on workers’ productivity in the office environment, as well as the benefits of formal and informal face-to-face interactions. This is certainly true for customer facing businesses, but also for enterprise facing businesses, and any decision to mandatorily require workers to work remotely from home as a countermeasure to becoming infected, while in a social clustering environment, is one that could determine the future viability of the business, and may also carry additional legal liability. Obviously, some businesses have no physical capacity to become a “virtual office”, such as cafes or restaurants, and the impact on those that can, in relation to the explicit costs and potential loss of productivity, will vary from business to business, and will also depend on their relative size, capital resources, liquidly positions, and prevailing Government policies

Therefore, it is a vitally important decision for managers to determine the point they should close the office and transition those that can into remote working employees. Critically, while Governments and health authorities have been generally very good at providing actionable advice in respect of human health, there has been no advice on how managers should go about making this decision for their business. In order to make this decision, managers should come to an understanding of the risk to their employee's health in the office by remaining open, balance that with the business productivity, and make the decision to close when this risk breaches a certain predetermined risk threshold that they are willing to accept.

The purpose of this note is to provide some kind of objective, probability and data-based framework to assist in this important, arguably existential, decision. We indicate how the general level of infection drives the estimation of probability for finding one or more infected persons in a work group. Then we show how the estimation of the general level of infection follows from observed case data, with the proviso that account is made of the likely unseen number of cases already in the community.

RISK-BASED METHODOLOGY WITH SCENARIO PLANNING

Our purpose is to assist managerial decision making in two planning stages: 1) when to close the office and to work remotely; and 2) when to restart or reopen the office. The approach we take employs a Bernoulli trial to estimate the probability that one or more persons in an office cluster of size N may be infected, given an assumed background rate of infection. Since the COVID-19 virus has an estimated fourteen-day incubation period, we must also provide estimates of the likely background infection rate in terms of the currently reported cases. Since this grows over time, it is not sufficient to use the presently reported case load, as it takes around two weeks for the true number of infected people today to become apparent.

Additionally, in circumstances where an epidemic outbreak is approaching the end, we must adjust our model of the estimated probability of infected persons to remove excess counting of those who have since recovered. This other element of the problem is relevant to the *restart decision*.

Finally, in all circumstances, we will need a model of the future development of cases given the phase at which the COVID-19 outbreak is situated. In simple terms, we distinguish between the *managed growth phase,* which is characterised by a growth rate which declines over time, and the *initiatory growth phase*, which is characterised by *near constant exponential growth*. These levels of growth are not constant to the pathogen, but appear to be a function also of social connection and activity. This ability to reduce the growth rate through the deliberate reduction of social connection via travel bans; social distancing; levels of personal hygiene; and self-isolation, or quarantine is the reason why epidemic management is possible at all. Absent a vaccine, or cure, these are the only control measures available, at this time.

OBSERVATIONAL DATA FROM THE COVID-19 OUTBREAK

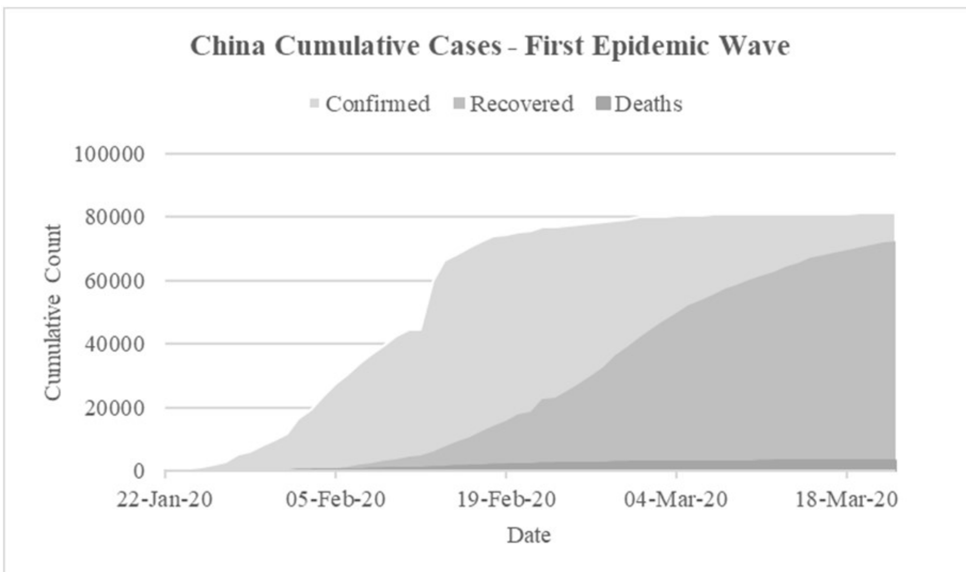
Ahead of laying out the method for computing exposure risk, we need to consider the form in which our input data is provided. It is typical of epidemiological data that we are provided with three time-series of measurements of the number of people affected by an outbreak. In the first instance, there is the *actual total number of cases* that have been observed to date. This is called the *cumulative case curve*, and will typically show a rapidly rising exponential phase, followed by a moderation, and then a fade to a final cumulative case load. The result is an S-Curve of case counts Figure 1 (Source: WHO data Ref. 1 and Johns Hopkins University data, JHU Ref. 2). For those that recover, or die, the count of infectious people goes down. For COVID-19, most people, around 80%, have mild cases, and can be returned to the population when the virus in their body subsides, as they are then no longer infectious.

The proportion who die varies between 1% and 10%, depending on the location, age, comorbidity, and quality of health care. At any time, we can subtract from the cumulative case curve, those who have either *died or recovered*. These are the so-called removals or resolved cases. The result will be the number of *active cases*, which is an estimate of those known to have it who may be infectious.

$$Active\ Cases\ (A) = Cumulative\ Cases(C) - Deaths(D) - Recovered(R) \tag{1}$$

In this sense, what must drive our risk calculation is an *estimate of the true active cases in circulation*, who may well, unknowingly, be carriers of COVID-19 in the workplace. While the above detail is very important for detailed risk assessment, as we shall see, the dominant factor at the start of any outbreak is the tendency for rapid exponential growth to overwhelm such detail. For the practical purpose of prudent risk assessment, we can generally rely on the *overestimate* that is presented by using the *full cumulative case count*, without adjusting for recovery, as a base infection probability.

FIGURE 1
EPIDEMIC PROGRESSION IN CHINA VIEWED AS AT 22-MAR-2020



BERNOULLI TRIAL RISK ESTIMATE OF INFECTION PROBABILITY

The foregoing detail can complicate the final calculations, but only through an adjustment to how we estimate the one key parameter of our risk model: the *base infection probability*. It is possible to write a simple general expression in terms of this parameter. For this reason, we present the general risk formula first, and detail how the above epidemiological growth considerations affect estimation of the key parameter in the subsequent sections. This way we can distinguish between a proper understanding of the *probability theory that drives risk,* from the *epidemiological modelling of key risk parameters*.

The key risk we wish to understand is the probability that a randomly chosen group of people of size N (such as in a commercial office) may contain *at least one infected person*. Naively, we might think that this is the same as the chance that any randomly chosen person in the general population might be infected. However, this is not the case, due to their being *multiple different possible outcomes* from the joint trial.

Consider the process of selecting N people at random from a population with base infection probability p. The probability that we select an infected person at random is p for each trial. Therefore, the probability that we select N uninfected people, requires that we select an uninfected person *for every* *trial*. Following the basic laws of probability, the probability that we select one uninfected person has to be $1 - p$. Therefore, the probability of selecting no infected people N times is:

$$Q = (1 - p)^N \tag{2}$$

where Q is the probability of no infected person in a group of size N. With some rearrangement, we can now find the probability that a group of size N contains at least one infected person. That is:

$$R = 1 - Q = 1 - (1 - p)^N . \tag{3}$$

Here we use the symbol R to denote "risk probability" and Q to denote the probability of an effectively disease-free work group. Note that these are *probabilities*, which are *not certainties*. Through setting R to a given *threshold level of group risk that managers may regard as acceptable*, we can back-solve for N given p, or p given N. Let us now do that and discuss some simple examples.

Now, we consider solutions for different planning scenarios of unknowns. Firstly, let us find p given N:

$$p = 1 - (1 - R)^{1/N}. \tag{4}$$

Secondly, we can solve for N given p:

$$N = \frac{\ln(1-R)}{\ln(1-p)}. \tag{5}$$

The easy way to verify these equations is to substitute the expressions back into Equation (3).

For example, if the number of people in the office is 20, and the risk probability is set at the 5% level, then we need to set $N = 20$ and $R = 0.05$ and solve for p. Doing this using the above formula, we find:

$$1 - (1 - p)^{20} = 5\% \tag{6}$$

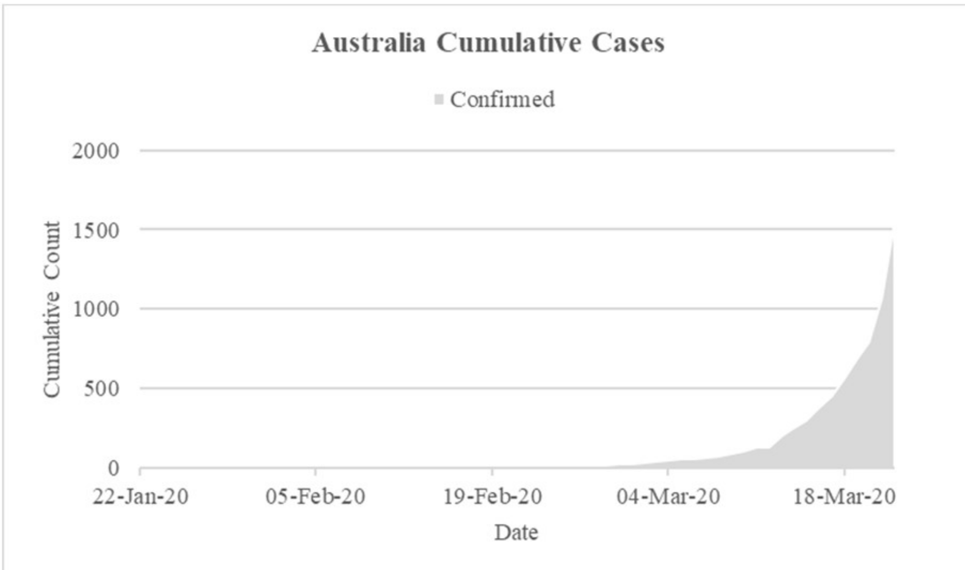
$$p = 0.2561\% \tag{7}$$

To relate this to an epidemic outbreak data, we need to work out how many infectious people need to be generally present in the entire population to produce that *incidence rate of infection*. This can be derived from the *cumulative case number*, denoted by the letter C. At this point, one needs to make assumptions about the uniformity of the infection rate, and size of the reference population P. However, for Australia, we can readily compute $C = p \times P = 0.26\% \times 26{,}000{,}000 = 66{,}680$ people. For caseloads below that number, the risk of a randomly selected group of 20 containing one or more infected people is below the target 5% threshold. If the distribution of infection is non-uniform, then we might adjust the baseline population number in the denominator to equal that of the area of interest, say a single State or City.

The cumulative infections in Australia were 1,549 as at 23rd March 2020 (Source: JHU, Ref. 2) this forecast case number seems relatively large, but as this virus has an incubation or latency period of at least two weeks (during which time the infected person may infect a susceptible), the more realistic estimate of the number of infected people is a fourteen-day forward estimate. This is not a simple matter, especially for COVID-19, as the virus is very contagious, and will naturally grow at a high exponential rate variously between 10% and 40% daily, and *sometimes higher* in the early stages of an outbreak.

When the virus is properly contained, it has been shown that the growth rate can be tamed and brought to zero, provided that there is continual monitoring in place to detect any new cases that might form new clusters of exponential growth. Once we understand this fact about COVID-19, it becomes clear that the risk assessment for people who work in groups is likely an ongoing affair. In spite of that, we can still make some estimates. To do that we look at trend data for the cumulative case curve, see Figure 2 (Source: JHU, Ref. 2). The growth appears exponential, but to assess it properly, we must compute the *rate of growth*.

FIGURE 2
CUMULATIVE GROWTH TREND FOR COVID-19 IN AUSTRALIA AS AT 22-MAR-2020



Importantly, in epidemiology, it is often the case that practitioners look at the so-called *incidence curve*, and focus on prediction of the *peak of the outbreak*. This is natural as their primary advice is to policy makers in health care on the provision of sufficient medical resources for the *likely peak demand*.

Our problem is different, and is the one where there has been no official advice. Those in business are concerned about making risk assessments of likely future infection probability within their work-group in order to make an objective decision on when to close the office. For business, the important risks are at *the start*, and *the end*, of the outbreak. These are where shut-down or remote working decisions need to be made, along with the critical re-start decision. With this in mind, we need to be cautious to not simply copy the standard tools of epidemiologists. They have good tools for their problem, but we need *different* tools for ours. In particular, we need to understand the patterns of growth for the COVID-19 virus when starting from a low base, when the decision risk is most material. This early phase of the outbreak can be difficult to estimate, and so it is helpful to look at real data. The way we shall do this is by plotting the logarithmic growth rate of cumulative cases:

$$G = \log C(t) - \log C(t-1) \ . \tag{8}$$

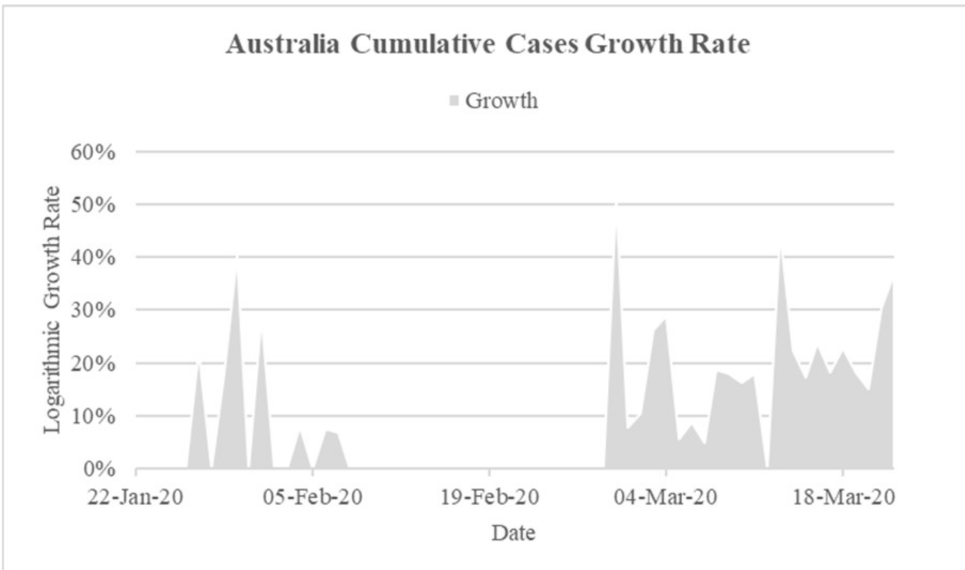
This is a growth rate, measured as the difference between the log of cumulative cases today $C(t)$ and the log of cumulative cases yesterday, $C(t-1)$. Those involved with the financial analysis of stock returns will probably be familiar with such quantities already. The reason we use the logarithm is that the growth rates of COVID-19 are very high from day to day, especially at the start of an outbreak. We therefore need to use the logarithm to get more meaningful geometric averages when estimating growth rates.

One good way to understand the meaning of the growth rate G is to take the exponential:

$$e^{G} = \frac{C(t)}{C(t-1)}. \tag{9}$$

Effectively, the exponential of the growth rate in this form is the ratio of one day to the prior day. For the early part of the Australian outbreak, the epidemic curve looked like Figure 3.

FIGURE 3
LOGARITHMIC GROWTH RATE OF DAILY COVID-19 CASES IN AUSTRALIA



There are a couple of things to notice about this data. Firstly, there was a low-level outbreak back in late January and early February 2020, that soon died out. This was most likely imported cases of the virus which came from China via international arrivals. The important thing to understand about a virus is that it can only reproduce within an active host infection. Research has shown that COVID-19 can survive in the air briefly, within a range of a few meters of someone who coughs or sneezes, and on surfaces, like stainless steel benchtops, handles and public places for hours to days. However, if the viral sources are eliminated, then an outbreak will die out. This is the principle behind quarantine, and isolation of infected persons, or suspected infected persons. Viral agents cannot reproduce without infecting a new host. Absent a vaccine, the goal of infection control is to find and isolate all infectious cases, so as to reduce the pool of active cases, and so the risk of infection. This is a problem with COVID-19 because it is so highly contagious.

Looking again at Figure 3, we see that, likely as not, the first infection wave was isolated. However, there is a second wave that appears later, for which the growth is more vigorous. This started about that time when health officials recognized the presence of so-called *local transmission*. This terminology refers to a self-sustaining transmission between local citizens, without any need for foreign arrivals. In this second phase, the growth rate averaged 20.2% over 23 days from 29th Feb 2020 to 22nd Mar 2020. In that period, the growth was effectively exponential, going from 25 cases to 1549. It is at this stage, during the early stages of an outbreak that the managerial decision is focused on whether or not to close an office. In the following, we treat the two scenarios, the decision to close, and the decision to reopen, *as they appeared* given the data that was available on the dates those sections were written. That was the 22nd March 2020 for the section on closing the office, and the 28th April 2020, for the following section on reopening.

The estimated growth rate fluctuated considerably, since this is the average of all Australian states and territories, which experienced different levels of outbreak at different times. However, travel between states and territories does further spread the virus between population centres. If people were to mix freely, and frequently, among all locations then we would expect the growth rate to be steadier. With the benefit of hindsight, the estimates we now give will be seen to be *overestimates*, but it is helpful to record how the situation appeared *at the time this work was first undertaken*. Later we will update the data to reflect how the subsequent outbreak developed. This point-in-time approach to presenting our discussion may help reinforce the idea that business decision making under uncertainty needs to consider *scenario planning*.

THE DECISION TO CLOSE THE OFFICE

Proceeding on the basis of how the data appeared to the authors on 23rd March 2020, we can make some range estimates using an average growth rate $G = 20.2\%$ plus or minus the standard deviation of 12.1%. The result is three scenarios for the likely uplift in cases over a 14-day forward planning horizon, as shown in Table 1. The reason to choose 14 days relates to the standard medical advice for COVID-19 that the incubation period may be as much as two weeks. In the context of our scenario planning, we needed to consider that then observed level of cases (1,549), on the date the planning was undertaken, would not likely reflect the true cases in the community at that time. This is simply due to the delay in case recognition caused by the 14-day incubation period for new infections. There is unavoidable uncertainty, due to the unknown unknowns regarding how many infected people have not yet presented for testing or treatment.

TABLE 1
SCENARIO PLANNING FOR OFFICE CLOSURE DECISION AS AT 23-MAR-2020

Scenario Label	Scenario Assumed Growth Rate (per day)	14-Day growth factor in case number	Forward estimate of the number of cases
Low Range	8.0%	3.08	4,778
Mid-Range	20.2%	16.82	26,058
High Range	32.3%	91.74	142,110

The central estimate of the then case number was about 26,000 cases, on a 14-day forward basis. When those forecasts were made, the limitations on large gatherings in Australia had only happened in the last few days. Therefore, we assumed that the central figure would be a reasonable risk estimate of the true case load at this time. The higher end of the range represented an extreme outbreak, such as happened early on in Italy, while, as it happens, the lower end of the range is what actually transpired in Australia.

Let us now work out the average population-wide incidence rate at this time:

$$p = \frac{21{,}00026058}{26{,}000{,}000} = 0.0100\%. \quad (10)$$

Now let us apply our risk formula for a group of size $N = 20$:

$$R = 1 - (1-p)^N = 1 - (1 - 0.0100\%)^{20} = 1.986\%. \quad (11)$$

This is less than the initial 5% threshold, but over the 1% threshold. If, however, the manager's benchmark is the 5% probability level, they can estimate how long it will take to breach that level by first solving for the required incidence rate:

$$R = 1 - (1-p)^{20} = 5\%, \quad (12)$$

to get

$$p = 0.256\%. \tag{13}$$

This incidence rate gives us an incident number Australia wide of 66,586 (0.256% x 26,000,000), so we need to simply calculate how long it would take the infectious caseloads to grow from 26,000 today to approximately 67.000. Implicitly, the relationship for the time t is given by:

$$67{,}000 \ = 26{,}000\, e^{0.202 \times t} \tag{14}$$

Solving for the unknown, we find $t = 4.69$, or about five days forward in time Accordingly, the manager has estimated that based on these simple assumptions and cumulative growth rate, an office of 20 people may need to be closed in six days.

For completeness, let us apply our risk formula for a group of size N = 500:

$$R = \ 1 - (1 - p)^N = \ 1 - (1 - 0.0100\%)^{500} = 39.4\% \tag{15}$$

It may help to note what that calculation implies. In a group of people, numbering 500, who were randomly selected from the Australian population, and assuming a constant infection incidence of $p = 0.0100\%$, there is a probability of 39.4% that any such group will contain *at least one infected person*. In our view, this is a highly significant risk. It suggests that larger corporations/offices, where there are more people, and possible contacts are at risk for longer, should perhaps close earlier than smaller offices.

THE DECISION TO REOPEN THE OFFICE

An equally important decision that managers will need to make is at what point do they reopen their office premises and require that their staff return to work and attend in person. While broader factors such as a possible level of continuing anxiety the individual worker may experience as a result of a partial renormalisation, or, indeed, a legal directive from the Government for the office to remain closed, are obviously relevant, the purpose of this methodology is to provide an objective, risk-based assessment to assist managers in the absence of any other considerations.

To frame this question properly, on 23rd Mar 2020, the reported cases to date in China numbered 81,498. The population of China is approximately 1.408B people. The average incidence rate is then:

$$p = \frac{81{,}498}{1{,}408{,}000{,}000} = 0.0058\% \tag{16}$$

If we were to simply apply the same reasoning, then the risk for N=500 in China is:

$$R = \ 1 - (1 - p)^N = \ 1 - (1 - 0.0058\%)^{500} = 2.86\% \tag{17}$$

That is still a significant risk and material for Chinese firms considering their re-start decision. However, on 23rd Mar 2020, the number of recoveries in China numbered about 72,814, meaning the *active cases* were only 5,410 and, so:

$$p = \frac{5{,}410}{1{,}408{,}000{,}000} = 0.000038\% \tag{18}$$

For that value of the incidence rate, we find the relevant risk level is only

$$R = \ 1 - (1 - p)^N = \ 1 - (1 - 0.000038\%)^{500} = 0.19\% \tag{19}$$

With the usual caveats on estimation uncertainties, China was at the re-start decision stage for large groups, at that time when nations such as Australia were facing the shut-down decision for small groups.

A significant difference between the approach used to inform the decision to close the office and the decision to reopen the office, concerned the estimate for the number of infectious, or "active", cases. Rather than refer to the cumulative number of confirmed cases, as in the former decision, the decision to reopen should refer to the smaller subset of active cases as they are the ones that are infections because the cumulative cases also include non-infectious "recovered" individuals. The figure for active cases A is defined at Equation (1), it is just the cumulative cases C, minus deaths D, and recovered cases R.

In addition, the decision to close also estimated the 14 day forward cumulative cases as the virus has a latency period, while the decision to reopen is based on the *current estimate of known active cases*. The incident rate as a proportion of the population is therefore calculated according to the formula

$$p_a(t) = \frac{A(t)}{P}. \tag{20}$$

The notation $p_a(t)$ indicates that this quantity can be directly monitored as a function of time since it is in turn a function of the observable value of active cases $A(t)$, in a scenario where this is declining. This aspect is most important in highlighting the decision asymmetry between the closure decision and restart. In the closure decision, we need to account for the fact that cases are increasing over the planning horizon. In the restart decision, the calculation is more reactive. It responds to a visible decline in risk. There are still important unknowns, such as the chance of second-wave infections, but the posture is different.

To put the foregoing in context, let us note what the outcome was 14 days later for the planning that was undertaken on 22nd March 2020. On that date the Australian cases to date was 1,549. The outcome on 6th Apr 2020 was 5,797. This is slightly above the low-range forecast of 4,778 but well below the central estimate of 25,058. The social-distancing measures adopted at that time reduced the effective growth rate, but the forward number of 5,797 was still 3.74 times higher than the cumulative case count at the time at which the scenario planning was done. Whereas the central scenario estimate of Table 1 gave a forecast risk level of 1.986% the post-facto value, at that point in time, was more like 0.445%. The scenario planning approach thus provides a conservative framework for informed risk assessment.

Another relevant observation is that this estimate can be for any particular sub-set of the general population, such as at the State or City level. It could therefore accommodate *regional differences* in the geographic location of the specific business, particularly where the sub-set has limited interaction more broadly, such as in a lockdown scenario where very few individuals would be travelling in and out of any particular City or State. Once this estimate has been made, it may be applied on a similar way to the decision to close, and the manager is similarly able to select the required level of probability using

$$R(t) = 1 - \left(1 - p_a(t)\right)^N. \tag{21}$$

The risk level $R(t)$ for a given group size N can be directly monitored by substituting for the value of the incidence probability $p_a(t)$ computed directly from Equation (19). In many ways, the restart decision is considerably less uncertain than the closure decision since the risks involved are falling with time.

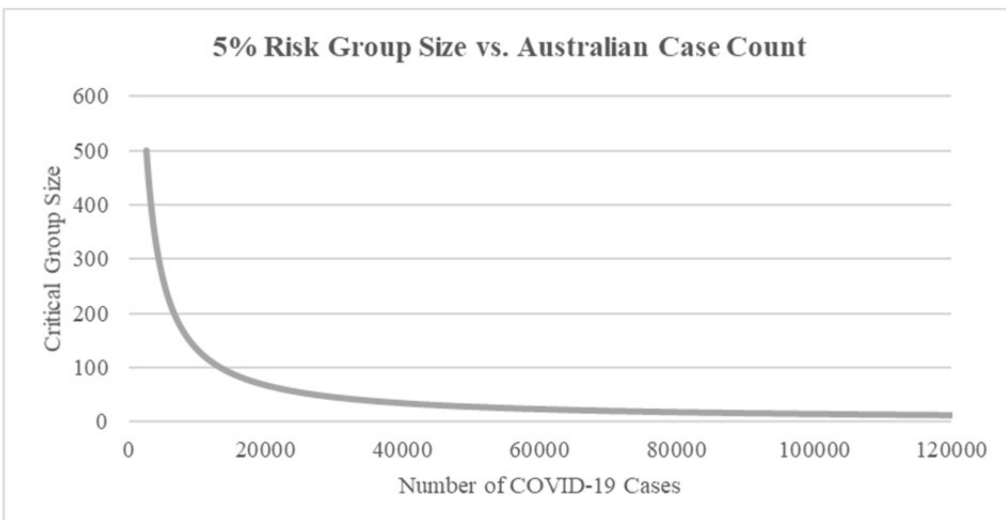
DISCUSSION AND CONCLUSION

By combining approaches from basic probability with early stage exponential growth, a business manager can create an objective, defensible, risk-based framework for making the important decision as to when to instruct staff to work remotely, and when to recall them into the office during a public health emergency such as the one we are currently undergoing. While not only of immediate practical value to the business, it importantly allows a manager to contextualise the risk to the health of the employees. The approach outlined in this note does not attempt to replicate or replace mainstream epidemiological health advice because it does not attempt to model an epidemic. Rather we are concerned about the objective probability of infection given an incident rate, and make a conservative assumption of exponential growth to estimate the forward case number based on data. An advantage of this model is that managers can update this growth rate daily while the office is open and can dynamically adapt to changing conditions.

An interesting implication of this analysis is that larger companies may need to close their premises earlier than smaller companies, and may also remain closed longer than smaller companies for the same level of risk. This property of the risk analysis is illustrated in Figure 4, where we have taken figures for the Australian population, and calculated the critical group size N for different active caseloads C. The graph declines rapidly with the rise in cases, and so tends to affect larger organizations earlier than smaller ones. This may be opposite to the natural intuitive thinking of managers that smaller organizations are inherently riskier than larger organizations. That may be true financially and for many operational issues, but for the group size infection risks posed by an epidemic outbreak, the normal intuition is flipped upside down.

While an underlying assumption is that all employees are homogenously circulating in the community and that individuals are equally likely to be infectives or susceptible, so that the population-wide incident rate is equally applicable to every company, an individual employee's social networks and connectedness would change that for different companies. Nevertheless, the idea that larger companies that have more employees have a higher chance of one of them being infected, and therefore potentially infecting the others, makes intuitive sense, and it may also be true that larger companies have deeper resources in which to facilitate a remote working environment. This is where individual judgment must be exercised.

FIGURE 4
CRITICAL GROUP SIZE FOR AUSTRALIAN POPULATION AT 5% RISK LEVEL



Smaller companies are likely to have less resources to withstand a material drop in productivity (or, indeed, a complete shutdown) from a remotely working employees, so the decision as to when to close and restart the office is relatively more important. Additionally, it also allows managers to estimate this risk on a more granular basis by using specific local data that is more relevant to the company's geographic location. In conclusion, the methodology we have presented provides managers with the ability to frame the question of office closure and restart in the context of employee health risk and business viability. This potentially fills a gap for managerial discussions in an environment where there are no official guidelines for how to frame such an important employee health and economic decision.

REFERENCES

2019 Novel Coronavirus COVID-19. (2019). Data Repository by Johns Hopkins CSSE, Johns Hopkins University. Retrieved April 29, 2020 from https://github.com/CSSEGISandData/COVID-19

World Health Organization Situation Report on COVID-19 for 23rd March 2020. (2020, March 23). Retrieved April 29, 2020, from https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

APPENDIX

STEP-BY-STEP GUIDE TO RISK CALCULATIONS

Estimating the Probability of at Least One Employee Being Infected

This estimate involves combining two calculations: the incident rate, or probability of individual infection p_f, and the group risk $\mathrm{R}(N)$ which is a function of this and the group size N.

Estimate of the Incident Rate p_f in d Days Time

Adopting the simple exponential model for the early stage of an outbreak, we estimate the proportion of the total population likely to be infected. The resulting *forward incident rate* depends on the incubation period d, the growth rate of infection G, the number of cases today C, and the population P. The formula is

$$p_f(d) = \frac{Ce^{Gd}}{P} \tag{A1}$$

where

d is the number of days forward to estimate (suggested value is $d = 14$)
C is the population wide current number of cumulative cases today
P is the population
G is the daily exponential growth rate of new cases.

This can be understood as the ratio of the likely number infected in d days time, which is given by the numerator Ce^{Gd}, divided by the total population denominator P. The probability is higher if the growth rate G is higher, or the number of days d is further into the future. Other things being equal, it rises the larger the number of cases C is compared with the total population P. For the choice $d = 0$ we have

$$p_f(0) = \frac{C}{P} \tag{A2}$$

Calculation of the Risk Probability $R(N)$ Given $p_f(d)$

The proportion of the total population that are likely to be infected, p_f is equal to the probability of selecting an infected person at random from the total population P. Using the theory of Bernoulli Trials, we can then estimate the probability, R, of there being at least one person in the office of N employees who is infected. The calculation follows by first estimating the probability of a random trial where *none* of N people, who are randomly selected from a large population P is infected. For a large population, so we can ignore the size of N relative to P, and the formula becomes:

$$\mathrm{Q}(N) = \left(1 - p_f(d)\right)^N \tag{A3}$$

That probability Q is the likelihood that nobody in the group of N is infected. The laws of probability then imply that the $R = 1 - Q$ is the probability of the alternative outcome, which requires that *at least one person* in the group is infected. The required formula for R as a function of group size N is then:

$$\mathrm{R}(N) = 1 - \left(1 - p_f(d)\right)^N \tag{A4}$$

where

R is the estimated probability that at least one person is infected
d is the number of days forward to estimate (suggested value is $d = 14$)
N is the size of the group
$p_f(d)$ is the forward estimate of the population-wide incident rate.

Obviously, this risk is larger in a larger group. Through the dependence on $p_f(d)$ the risk is also dependent on the forecast horizon d and is lower as this is made smaller for the growth phase of the epidemic. The other sensitivities relate to the population size P, the current base of active cases A and the growth rate of the epidemic G. Experience has shown that social-distancing can reduce the growth rate G below what it might otherwise have been. Therefore, these risk estimates are likely to be *conservative* if such measures are likely to be higher going forward. Conversely, they may need to be modified if social-distancing is being relaxed. Monitoring of the level and rate of change of *active* cases, seems most relevant for the waning phase of an epidemic outbreak. As active cases decline, $p_f(d)$ goes down, and so too does the risk $R(N)$.

Worked Example
Assumptions:

$N = 15$
$P = 26{,}000{,}000$
$C = 1{,}549$
$G = 20.2\%$

Then $p_f(14) = 0.0100\%$ and $R(15) = 1.49\%$.

There is a 1.49% probability that at least one person, is infected.
Given a group size of 15, a population of 26,000,000, a current caseload of 1,549 and a daily growth rate of 20.2%, the fourteen day forward incident rate is estimated at 0.0100% for a threshold risk of 1.49%.

Decision to Close the Office
This estimate involves two calculations: firstly, the number of future cases at which a certain infection probability is reached; and secondly, the time to reach that level.

Estimate the Number of Cases $C_f(d)$ in d Days Time
The formula is

$$C_f(d) = Ce^{Gd} \tag{A5}$$

where,

C is the population wide current number of cumulative cases
G is the exponential growth rate.

This formula is appropriate in the early phase of an epidemic when the decision to close an office is most relevant. Once infection control measures start to take effect, the rate of change of new cases starts to slow down. The effective growth rate G will depend on time, and so the appropriate rule of thumb is to use an estimate of the *geometric average* growth rate, as discussed in the main text.

Estimate the Horizon Time T for Risk Level R

In this guide, we have assumed a threshold probability, $R = 5\%$, of at least one person being infected. This risk parameter is a decision input, and may be set lower or higher. The important variable for forward planning is the *horizon time* T, at which the number of cases might rise to meet a threshold risk level R.

Using the risk formula (A4) we can solve for the value of $p_f(d)$ that corresponds to the threshold R:

$$p_f(d) = 1 - (1 - R)^{1/N} \,. \tag{A6}$$

This value multiplied by the total population P equals the target case number at which R would be met:

$$C_f(d) = P \times \left(1 - (1 - R)^{1/N}\right), \tag{A7}$$

where

P is the population
R is the target risk threshold
N is the number of employees in the office

If we now combine (A5) with (A7) we can solve for the current number of cases C_R for risk level R:

$$C_R = Pe^{-Gd} \times \left(1 - (1 - R)^{1/N}\right), \tag{A8}$$

Now we can estimate the time T at which the number C_R is reached when the starting number of cases is just C. Applying the growth rate of new cases, G, to a starting case number C, we find that

$$C_R = Ce^{GT}, \tag{A9}$$

which is

$$T = \frac{1}{G} \log \frac{C_R}{C}, \tag{A10}$$

The time to closure T is lower for higher growth rates G, and goes as the logarithm of the ratio between the threshold case level C_R and the current case level C.

Worked Example

Assumptions:

$N = 15$
$R = 5\%$
$P = 26{,}000{,}000$
$C = 21{,}000$

Then $C_R = 88{,}757$ and $T = 7.43$ days.

The current case number of 21,000 is expected to grow to 88,757 in around 7.43 days. Given a group size of 15, and a population of 26,000,000, the probability that at least one person in the group is infected then exceeds the nominated 5% threshold.

Decision to Reopen the Office

The decision to reopen an office mirrors that over when to close with one important difference. In the latter stages of an epidemic, when the peak infection wave has passed, the new infection risk is driven by the remaining *active cases*. The definition of active cases was given already at Eqn. (1):

$$A = C - D - R, \tag{A11}$$

where,

A is the active number of cases as an estimate of infectious people
C is the cumulative number of cases reported
D is the cumulative number of deaths reported
R is the cumulative number of recoveries reported

There is a delay between when a person is first infected and when they later recover, or die. This is up to two weeks, and so the number A does not start to decline until a little time after the peak in new cases of reported infection. However, once the growth in the number of new cases peaks the level or removals via R and D starts to reduce the number of active cases A at a near exponential rate. This is generally slower than the rate at which the infections first grew but reduces the new infection risk over time. Very late in the outbreak the number A will be much smaller than C. Therefore, we should use the number A in preference to C when estimating the risk of reopening an office.

Estimation of Risk R Given Current Active Cases A

In this guide, we have assumed a threshold probability $= 5\%$, of at least one person being infected. From the foregoing discussion of Bernoulli trials, the risk of any one person being *infectious* at the tail end of an outbreak is driven by the ratio of the number of currently *active cases* to the total population.

$$p_a = \frac{A}{P}, \tag{A12}$$

where

A is the active number of cases as an estimate of infectious people
P is the total population

Now we can estimate the *current* probability of there being at least one person being infected. Using the single person probability p_a the N-person probability of at least one infected is just Eqn. (A4):

$$\mathrm{R}(N) = 1 - (1 - p_a)^N \tag{A13}$$

Since this number goes down in a waning epidemic it is possible to monitor A and graph the decline in the above risk probability R. Once this falls below the desired risk threshold, say 5% in this example, then the decision to reopen can be taken.

Notice that since A declines in the waning phase we can use *current* values. There is not the same problem of forward estimation that we encounter with the early stage of the outbreak, when A is approximately equal to C and growing exponentially. The two decisions are asymmetric. However, the reopen decision is no less important, since it would be easy to *overestimate* risk if we do not account for the difference between the cumulative number of active cases C, which never goes down, and the current number of *active cases* A, which naturally declines once the epidemic has peaked.

Worked Example

Assumptions:

$N = 15$
$P = 26{,}000{,}000$
$A = 5{,}000$

Then $p_a = 0.0192\%$ and $R = 0.29\%$, which is less than the target 1% threshold risk level.