1.1 Appendix: SIR Models

All models are wrong; some models are useful.

If we accept the basic wisdom in the quote above, we possibly need to be cautious and question the validity of the results of any model. For models are invariably created because the reality of the system under investigation is too complex, such that a simplified model is often the first step in order to try to predict an outcome. This was especially true in the context of the Covid-19 epidemic, where the mechanisms underpinning the viral infection were initially poorly understood, such that the accuracy of any model has to be eventually judged against real-world statistical data, which was not initially available.

Note: In hindsight we might now recognise that the Imperial College model*, used to advised the UK government in March 2020, was a model that made predictions about the spread of the virus that proved to be wrong, when compared to actual statistics that subsequently emerged.*

However, the phrase '*lies, damned lies, and statistics*' might also be a warning that statistics also have to be questioned as the numbers, in isolation, can be misleading and used promote a weak argument as well as to undermine alternative arguments. Of course, it might be recognised that the '*for and against*' argument can be difficult for the public to make judgement, if access to data is essentially being controlled by one side. As of 22-Oct-2020, the UK is now pursuing a regional 3-tier lockdown system to apparently protect its citizens and limit the spread of the Covid-19 virus. Without going into the details, each tier of the system imposes restrictions based on the assumed infection rate, where the third-tier equates to a total lockdown. So, the question this discussion seeks to understand is whether a *compartmental SIR model* can be shown to support this policy, based on known infection and death statistics, while initially ignoring the obvious social-economic implications.

Note: It will be argued that the use of absolute numbers makes the assessment of risk difficult for many people, who may have little knowledge of a population size or normal all-cause mortality rates. As such, this discussion will attempt to quantify the issues in terms of percentages. In this context, the UK has a population of approximately 67 million and an average annual all-cause mortality rate of 0.94%

So, this discussion will first attempt to quantify the current statistics of infections and deaths in the UK taken from the virusncov.com website, as shown in the table below. The recorded Covid-19 infections and deaths, as both numbers and percentages, are shown in the table left, while the table right details the monthly increase (ΔD), deaths per day (ΔD/day) and deaths per day per million (Dpm/day), which are often used in some references.

Date	Infection	%-I	ΔI	Deaths	%-D	ΔD	Days	ΔD/day	Dpm/day
22-Feb	9	0.000%	0	0	0.000%	0			
22-Mar	4,550	0.007%	4,541	264	0.000%	264	29	9	0
22-Apr	116,911	0.172%	112,361	21,013	0.031%	20,749	31	669	10
22-May	232,140	0.341%	115,229	35,331	0.052%	14,318	30	477	7
22-Jun	276,581	0.407%	44,441	39,892	0.059%	4,561	31	147	2
22-Jul	296,377	0.436%	19,796	41,047	0.060%	1,155	30	39	1
22-Aug	324,601	0.477%	28,224	41,423	0.061%	376	31	12	0
22-Sep	403,551	0.593%	78,950	41,825	0.062%	402	31	13	0
22-Oct	789,229	1.161%	385,678	44,158	0.065%	2,333	30	78	1

The first 4 rows (Feb-May), shaded in pink, might be described as the rising phase of the epidemic that affected most of Western Europe in a fairly similar manner, although geographic and demographic variance will need to be considered by any SIR model. The next 4 entries (Jun-Sep), shown in white, might then be associated with the falling phase. However, the UK figures in Oct-2020 appear to show a worrying trend, which the UK government cites as its justification for the latest tiered lockdown policy.

Note: We might immediately question the idea that only 0.5% of the UK population had been exposed to the virus by Sep-2020, while also asking whether the jump of over 385,678 infections in the month of October might be correlated to an exponential increase in testing for the virus. Finally, it might also be highlighted that the accuracy of the PCR tests has been questioned by many respected sources.

Some have also questioned the accuracy of the 44,158 total UK deaths attributed to Covid-19 as there is evidence that many of these deaths involved elderly patients already at the end of life suffering with multiple comorbidities. However, even if we simply accept the number of UK Covid-19 deaths, as reported, the total number only equates to 0.065% of the UK population. It might also be worth highlighting that most northern latitude countries, like the UK, are now entering the winter months, where an increase in deaths associated with the normal spread of existing viral infections is to be expected. Therefore, we might attempt to contextualise the *all-cause mortality risks* as shown in the following charts.

Note: In the chart left, the red bars reflect a percentage breakdown by age of death from all-cause mortality, which is estimated to be 637,057 lives lost in a single year in the UK. The chart right then shows a breakdown of all-cause mortality in years prior to the virus outbreak. The black curve in the chart left then shows the Covid-19 mortality rates in 2020 by age, such that we might now make some relative judgment of the 44,158 recorded Covid-19 deaths that approximate to 6.9% of all-cause mortalities, where all were heavily biased towards age and existing comorbidities.

From the official statistical data, it is difficult not to conclude that there may have been some over-reaction by the UK government, as in many other Western countries with the notable exception of Sweden, which did not impose a hard lockdown policy. As such, we might compare the statistical outcomes in the UK and Sweden in the following table.

| Country	Population	Infections	%	Deaths	%-Pop
UK	67,772,000	789,229	1.16%	44,158	0.07%
Sweden	10,343,403	107,355	1.04%	5,929	0.06%

While this discussion has questioned the accuracy of both the reported infections and deaths, the number of deaths may be, at least, in the right ballpark, while the number of infections is probably meaningless when considered as a % of the populations.

Can SIR models be used to provide a better estimate of infections?

The reason why a more accurate estimate of infections in any given population is important is that it would provide a measure by which to judge the degree of '*herd-immunity*' that might be protecting the population overall. Therefore, we will start with the chart below showing 3 different SIR infection curves, normalised to a percentage of yet some undefined population. From a modelling perspective, these curves only differ in terms of the initial value of [R0].
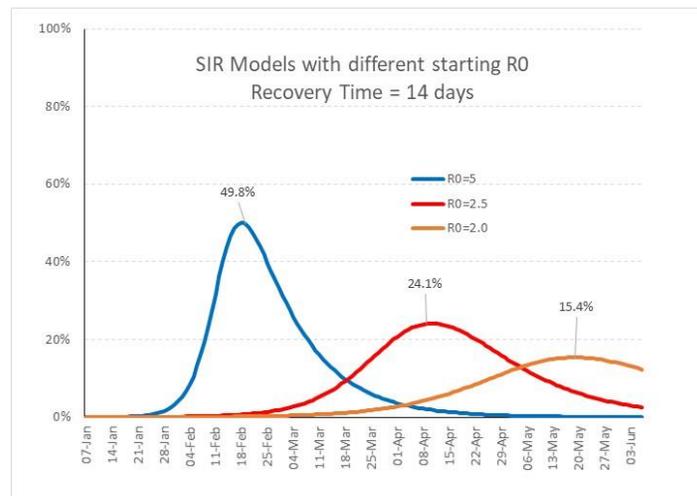
The value of [R0] of a specific virus is generally defined as the number of cases generated by a single case, where all individuals might be susceptible to infection. However, the initial value of [R0] may differ due to geographic and demographic variances within the population under consideration, e.g. a ship like the Diamond Princess, a small rural village, a large urban city or an entire national population. Deferring on such details for the moment, the value of [R0] changes as a function of time as the number of susceptibles [S] falls, as shown in [1].

[1]
$$R0 \Rightarrow R_E = S \frac{\beta}{\gamma}; \ where \ S = susceptibles; \ \beta = infection \ rate; \ \gamma = Recovery \ rate;$$

*So, which of the initial [R0] values in the chart above best fits the statistical data to-date?*

Let us start by reviewing what we think we know. First, there appears to be a reasonably well-established recovery time of about 14 days. Second, the epidemic started in early Jan-2020 and reached a peak by 10-Apr-2020, as supported by the statistics. If so, the red infection curve in the chart above, associated with [R0=2.5], appears to be the only one that comes close to peaking at the right time. However, we might also consider the case of the *Diamond Princess* as a basic sanity check before making any further assumptions.

Note: The Diamond Princess was a cruise ship that was held in quarantine between 20-Jan-2020 and 1-Mar-2020. Over this limited time, 567 out of 2666 passengers and 145 of the 1045 crew were infected and 14 died. As such, 21.2% of the passengers and 13.8% of the crew were infected. While there were possibly many reasons for the difference in these infection rates, the age demographics of the passengers and crew might understandably account for some of this difference. Overall, 19.1% of the 3711 people onboard were infected, where the 14 deaths represented 1.96% of those infected, but only 0.37% of the ship's overall population. We might also wish to consider the possibility of herd-immunity within this population approaching 20% after just 41 days against the 101 days between 01-Jan-2020 and 10-Apr-2020.

On the basis of the assumptions outlined, e.g. recovery time, peak infection date and the outbreak onboard the Diamond Princess, we might assume an initial value of [R0=2.5] as a starting estimate and then try to use the next SIR chart to expand on some of the details. So, by way of explanation, the scale on the left is normalised to %, such that we might better visualise the changing number of susceptibles, infected and recoveries within this abstracted population measured against the known timescale of infections peaking 10-Apr-2020, as shown in red. The scale on the right shows how the value of [R_E] changes against the timescale as a function of the number of susceptibles.
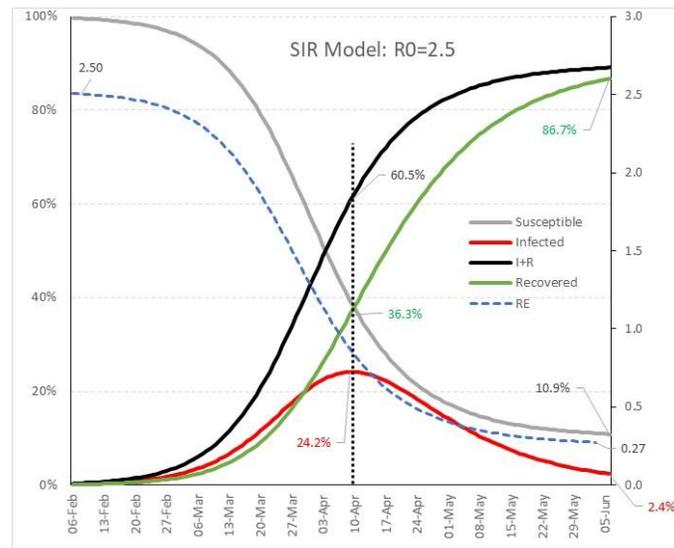
In addition to the infected curve, the % of susceptibles in grey falls as people are infected and recover, as show in green. However, we might wish to consider the curve in black, which represents the sum of those infected and recovered, which equates to 60.5% on 10-Apr-2020. While a highly speculative assumption, there may be an inference that this curve would approximate the level of herd-immunity at any point in time, if we ignore the 14-day lag between infection and recovery.

But, does this model bear any resemblance to the earlier statistics?

In order to address this question, we possibly need to return back to the issue of geographic and demographic variance within the population under consideration, especially when trying to model a national population. First, we need to recognise that people within a large national population can live in very different geographic regions, e.g. rural and urban, such that we might expect a wider range in the initial value of [R0]. Likewise, if the demographics of the population is represented by different ethnic groups with different susceptibilities to the virus, then [R0] might again vary. If so, we might start to recognise that a national population could only be modelled as the sum total of a large number of *compartmental SIR models*. For example, in some urban cities with a higher ethnic susceptibility to the virus, e.g. the Somali population in Stockholm, the value of [R0] might be higher than for a small, isolated rural community with no specific weakness to the virus. Of course, in practice, there may be a multitude of other factors, e.g. age and health, which may affect the accuracy of any SIR model.

So, are SIR models simply too inaccurate to be useful?

Again, let us first consider this question in the context of the quote used at the outset of this discussion: *All models are wrong; some models are useful.* If so, we need to question what could be wrong with the previous SIR model based on what we now know about the Covid-19 virus. The original assumption of a recovery time of 14 days does not seem an unreasonable estimate, nor the peak infection date in the UK of 10-Apr-2020, as both appear to be supported by the statistics outlined. However, there were other assumptions implicit in the previous SIR model that need to be considered. For example:

Will 100% of a population be susceptible to the virus at the start of the outbreak?

Starting with the next chart left, we might realise that people under 50, who represent 53% of the UK population, appear to have relatively little to fear from the virus. Likewise, while speculative, the video entitled '*SARS2 reactive T-cells in unexposed, normal healthy donors*' provides further potential evidence that a percentage of most populations may already have some form of T-cell immunity to a broad range of corona viruses, some of which have been in circulation for the last 20 years. Also, many

of the videos listed at the end of the discussion '*Pandemic Addendum*' provide further causal evidence that the %-susceptibility at the start of the outbreak will probably not have been 100%, as assumed in the first SIR model. Therefore, in the next SIR chart right, the percentage of susceptibles has been reduced to 70%, simply to see whether it might be a better fit to known UK national statistics.
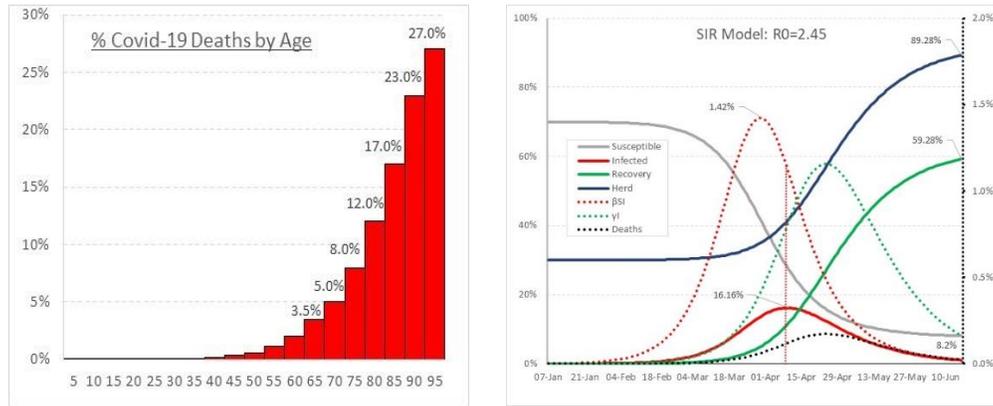
So, what has changed in the SIR model right?

The recovery time assumption of 14 days and peak date of 10-Apr-2020 are retained as they appear to be supported by statistical evidence. The SIR chart also retains the original [R0], although slightly reduced to match with the known peak date. However, the speculative assumption of a reduced susceptibility of 70% changes the primary SIR curves, as shown as solid lines, plus some additional dotted curves are introduced, which require the SIR equation to be modified, as per [2].

[2]
$$S_{t+1} = S_t - \beta S_t I_t; \quad I_{t+1} = I_t + \beta S_t I_t - \gamma I_{t-14}; \quad R_{t+1} = R_t + \gamma I_{t-14}$$

In [2], the incremental number of infections is given by [βSI], which is then offset by the incremental number of recoveries [γI$_{t-14}$], but where the recoveries have to be synchronised to the earlier infections in order to allow for the 14-day lag time assumed by the model. In the chart above right, these values are plotted as the dotted red and green curves respectively against the reduced scale on the right. Now we see the peak daily increment in infections occurs before the 10-Apr-2020, while the peak daily increment in recoveries occurs afterwards. Finally, the black dotted curve, also scaled to the right, reflects deaths that are normally included in the recovery figure, which averages to 0.061% of this population, similar to earlier UK statistics.

So, what if anything does this revised SIR model tell us?

First, we possibly need to highlight that this SIR model was not predictive, but rather a retrospective attempt to fit with known statistical data. Second, the original assumption that the SIR model can be driven by just [R0] has been questioned, if it does not account for the degree of natural immunity that might exist within most populations. Third, accuracy of the SIR model will become increasingly questionable as the population size is increased, as any assumptions about the initial parameters driving the model cannot be representative of both geographic and demographic variance. Even if we were to restrict the model to a small population, variance in age and health, would still lead to very different Covid-19 mortality outcomes

> *Note: The population of the UK's largest cities range from London with 8 million, Birmingham with 1.1 million and 100 having populations over 100,000 plus 1000 others having a population over 10,000. However, each of these populations may result in different outcomes to the virus due to geographic and demographic variance, which may be impossible to predict.*

As shown in [1], the expression for [R_E] depends on the changing number of susceptibles, which in-turn depends on the population size and the immunity assumed. If so, the idea that even multiple SIR models, let alone a single model, could be used to predict the collective outcome for all these different cities would appear questionable.

> *Note: The reader may wish to review a video:* Modelling by the Imperial College *by* John Ioannidis*, where he raises some very questionable issues, especially as the Imperial model was used to advise the UK government.*

While this discussion does not pretend to be authoritative, it is difficult not to conclude that predicating public policy on such models appears foolhardy, especially in light of the obvious social, economic and political costs of adopting a policy like lockdown. However, the statistical analysis of the data for both the UK and Sweden, along with numerous other studies, also appears to show no discernible difference in the trajectory of the virus based on government policy.

But what about the second wave?

Based on the evidence provided throughout the various *Covid-19 discussions*, there appears to be little supportive evidence in the actual statistics that the lockdown policy was effective in preventing the spread of the virus in the '*first wave'*. However, as has been pointed out, it appears that statistics related to the number of infections are a gross under-estimation of the real numbers. While there was little substantiative data concerning the level of immunity in any population in early Jan-2020, it might reasonably be assumed that it had to be greater than zero. However, even if this were the case, this cannot be the situation today after 10 months of the virus being in circulation through most parts of the population as a whole. Therefore, the starting assumptions for any form of '*second wave'* might reasonably be based on an existing 30% immunity, where in reality it might be much higher. Of course, if there is a much higher level of immunity within the population, then it would seem that any '*second wave'* may be lost within the normal statistical increases in the winter months for influenza-like infections.

If so, how might this affect the efficacy of the vaccination option?

While speculative, the accuracy of some of the official statistics has been challenged, especially deaths. Again, retrospective analysis has suggested that many of the deaths attributed to Covid-19 may have only reflected the presence of the virus at death, not the cause. Some estimates suggest that possibly 80% of deaths were more attributable to end of life comorbidities, where the presence of the virus had little effect on the outcome. If so, the 44,158 deaths attributed to Covid-19 in the UK might be reflected in a more meaningful figure of 8,831. However, we might assume that this figure would be further reduced in the future, if herd-immunity has already increased above 50%, such that we might reduce the projected deaths attributable to a '*second wave'* in 2021 to 4,415. Based on an earlier discussion entitled '*The Efficacy of Vaccines'* it was estimated that the Covid-19 vaccine may only be 36% effective, if given to the most at-risk age groups, i.e. over 60, which on the basis of the revised 4,415 deaths projected, might save 1,589 lives. Of course, by the same argument that the majority of the people most at risk are already at the end of life with multiple comorbidities, we might question the efficacy of the vaccine, if the extension and quality of life is not really being addressed. However, from a purely statistical perspective, we might understand that 1,589 lives amount to 0.25% of the 637,057 lives lost to all-cause mortality in the UK every year.

But surely every life matters?

While we might all perceive the moral or political reasoning behind this question, we need to question the wisdom of such rhetoric as the basis of public policy. For we might reverse the inference in this question to argue that the quality of life of younger generations also matters, especially in terms of the loss of employment opportunity and, in a wider context, it also matters in terms of the state of the nation as a whole, i.e. social and economic. Therefore, if the government wishes to pursue its lockdown policy it should be required to openly debate its arguments within a much wider spectrum of '*experts*' not simply those that support its assumptions