**A daily model for infections such as Covid-19, Version 2.1**

**1 The model**

1.1 In this note I describe a daily model for infections such as Covid-19 which I have constructed in Excel and make available for anyone else to use. It is based on my actuarial experience with income protection insurance (which provides a daily benefit while the insured is sick) and with modelling for AIDS in the 1980s. The model is a simple one, in that it assumes one homogeneous population, with no separation by age or sex or any other factor, but it allows me to get it onto one Excel worksheet.

1.2 The very simple continuous model which produces a logistic curve: eat / (1 + eat) for the total number of infections assume implicitly that an infected person is always infectious, but with many diseases an infected person may be infectious for only a limited period, and generally ceases to be infectious on death. This model allows for these features.

1.3 My model is a state-space model, with daily possible transitions between states. Time, *t*, is measured in days. The period of infection, *d*, is also measured in days. The states are *H*, healthy, *I*, infected and *D*, dead. The whole population starts initially in state *H*, and on the first day, *t* = 1, a new infection appears (from where, I do not specify). By infections from infected and infectious persons, people can transfer to state *I* on becoming infected. They may then die, and go to state *D*. I ignore entirely other deaths in the population.

1.4 My model is probabilistic, but does not allow for random variation. I calculate expected values throughout, and make no assumptions about the variation about those values. This does not matter when the numbers become large, but it may appear incongruous when the numbers are small and one speaks of fractional numbers of deaths or inceptions; every time I mean an expected number.

1.5 Those infected are kept separate according to the number of days infected, starting in day *d* = 1, and going up to *d* = 21, and all those infected for more than 21 days are put in a separate bundle. The number 21 is arbitrary, but it seems enough for Covid-19. At the end of each day an infected person may die. Those infected may be infectious or not infectious on any day; normally one would assume that they were infectious for a consecutive number of days at the beginning of the illness, and this number can be varied. The dead are assumed to be not infectious.

1.6 The number of new infections each day depends on the number of infected persons on that day, *t*, and duration, *d*, times a factor that depends on the day, *t*, (so can be varied with time), times the proportion of non-infected in the total population, times an indicator to show whether they are still infectious or not.

1.7 To explain the detail I now refer to the Excel workbook Cv19DailyModel.xlsx, worksheet Calculations. Below the headings there is one row for each day, *t*. Days, starting at day 0, are numbered in column D, and this is repeated for convenience in columns AG and BM. Move ahead to column G, where the number of healthy persons each day, *H*(*t*), is given. The number in the total population at the start, *H*(0) is copied from cell C2, where the number is given in bold. I use bold for all input variables, so that it is easy to identify them.

1.8 The number infected and still alive on day *t* at duration *d* is denoted *I*(*t*, *d*), and is given in Columns I (for *d* = 1) to AD (for *d* ≥ 22). For *t* = 0 they are all zero. The total *I*(*t*) is given in column AE. On day 1 a number of initial cases, usually 1, is put into *I*(1,1) in cell I12. If one were to insert zero there, there would be no infections at all.

1.9 Apart from the first duration day (*d* = 1) which I deal with later, the number infected at time *t*, duration *d*, moves to time *t*+1, day *d*+1, but may die or survive (but still infected) on the way. Deaths are assumed to take place at the end of each day, and infections at the beginning of each day.

1.10 In row 8, columns I to AC, are the values of *m*(*d*), the probability that a person who is infected at duration *d* dies at the end of that day, for *d* = 1 to *d* = 21. Once infected persons get into the “over 21” box, they are assumed not to die (from this disease). The probability of survival from day *t* to day *t*+1 is therefore 1 – *m*(*d*). Thus we get the expected number of infected at *t*+1, *d*+1, *I*(*t*+1, *d*+1) = *I*(*t*, *d*).(1 – *m*(*d*)). So in cell J13, for *t* = 2, *d* = 2, we get the survivors of *t* = 1, *d* = 1. With *I*(1, 1) = 1 in cell I12 and *m*(1) = 0.04 in cell I8, we get 0.96 survivors (expected). Look along to cell AJ13 and we get the number of new deaths that day, *D*(*t*, *d*) = 0.04 (expected), calculated as *D*(*t*+1, *d*+1) = *I*(*t*, *d*).*m*(*d*). I shall return to the deaths later.

1.11 When we reach column AD for the infected over 21 days, the calculation needs to include those carried forward from the previous day in that box, as well as those ones entering that box from day 21. So *I*(*t*+1, *d*>21) = *I*(*t*, *d*>21) + *I*(*t*, 21).(1 – *m*(21)). The first non-zero cell in that is AD33, which includes zero from AD32 plus the 0.96 carried forward with no further deaths (with the given parameters) from the original infected. Column AE gives the total of all those infected at different durations, *I*(*t*).

1.12 Going further right in the sheet we have the numbers of new deaths at each duration, *D*(*t*, *d*). If the value of *m*(*d*–1) is zero, then there are no new deaths for that day. There are no new deaths for day *d* = 1, so the cells in column AI are blank. Depending on the mortality rates assumed, there may be deaths for each day up to *d* = 21, in columns AJ to BC. There may be deaths at the end of day 21, and they are placed in column BD. Then the number of deaths on that day is summed in column BE and the cumulative number of deaths so far, denoted *DT*(*t*), is in Column BF.

1.13 In column BG is the total number of infected plus the total number of dead so far, *I*(*t*) + *DT*(*t*), which can be compared with the published number of infections so far, just as the total deaths so far can be compared. In column BH is the “Ratio per week”. After day *t* = 7 it is the number of infections so far divided by the number seven days previously, which is a useful indicator of the initial rate of growth in the population. This can be compared with the figures form real populations which are published. In Column BI I show the total deaths so far divided by the total cases so far, as a percentage

1.14 In column BK is a “check total”. This is the total number of Healthy, Infected, and Dead so far by day *t*. and equals *H*(*t*) + *I*(*t*) + *DT*(*t*). This should always equal the starting number, *H*(0), and is a useful check on the arithmetic.

1.15 I now come to the calculation of new infections. In column BN I show the proportion not infected, *p*(*t*), which equals the number of healthy on that day *H*(*t*) divided by the number living on that day, *H*(*t*) + *I*(*t*), so excluding the deaths. In columns BO to CI are shown the numbers of new infections arising from those infected on each day for *d* = 1 to *d* > 21, denoted *N*(*t*, *d*). This depends on the number infected at duration *d*, *I*(*t*, *d*), the infection rate, *r*(*t*), the probability that that a person of duration *d* can be infectious, *j*(*d*), and the proportion of the population available to be infected, *p*(*t*).

1.16 The infection rate, *r*(*t*), can vary with the time, *t*, but in my example it is constant. It is shown in column E, and the first value in cell E12 is copied down. But to change it, one can insert a number in the cell for the appropriate day and it will be copied thereafter. For convenience I shown in column F what I call R7. = *exp*(7.*r*(*t*)), the weekly multiplier in a continuous model, This can be compared with the weekly rate in the actual model in Column BH.

1.17 The values of the probabilities that a person of duration *d* can be infectious, *j*(*d*), are shown in row 7, Columns I to AD. They are shown there for convenience in altering them, but are not used until later, when they are copied into row 7 columns BN to CI. I have only used values of 1 and 0 to show infectious or not infections, but intermediate values could be used to show a person becoming less or more infectious as the duration of infection goes by.

1.18 We then have the formula: *N*(*t*, *d*) = *I*(*t*, *d*).*r*(*t.*.*j*(*d*).*p*(*t*). The values are shown in columns BN to CI and summed in column CJ. This shows *N*(*t*), the number of new infections that day. I assume that they start the infection on the following day, so *I*(*t*+1, 1) = *N*(*t*). the value in column CJ is copied back into column I in the row below, and we start the next day.

1.19 The input variables, or parameters, that one can readily change are:

 (a) The initial susceptible population *H*(0) in cell C2

(b) The initial number of inceptions on day 1, in cell I12.

 (c) The values of *j*(*d*), the infectiousness at duration *d*, in cells I7 to AD7

 (d) The values of *m*(*d*), the probabilities of death at duration *d*, in cells I8 to AC8.

 (e) The values of *r*(*t*), the infection rate at time *t*, in cells E12 and below.

 (f) The value of the calendar date for time *t* = 0, in cell B11.

The values in all other cells are calculated from these inputs.

1.20 In addition, in column B is a series of calendar dates, staring with some input date at *t* = 0. This needs to be chosen after all other input values have been inserted, and it depends on the results. In column BG one sees the total number infected (including dead) so far. One can compare these with the latest published numbers for any given country, and set the initial date in cell B11 so that the numbers may match up at the current date. In addition in column C are given “weeks”, in the sense of the given day divided by 7, so that one can see for example how many weeks ahead the peak might be, or the numbers of new infections drop to zero or nearly zero.

1.21 There is also a Chart of New Infections daily, but one can add whatever charts one likes.

**2 The numbers|**

2.1 I need to put in specimen numbers, and these can be changed as you wish. However, I explain here why I have chosen the ones that are there in this version. You may disagree with my choice, but the point of this model is that you can easily try out whatever you would like to put in. It is interesting to try out quite different numbers, for example with very high death rates, to check that such a disease would, with this sort of model, die out very quickly. In fact the initial value are as described in Section1, but you need to change them as described below to get to the final results.

2.2 I have found a useful and comprehensive data source published by the European Centre for Disease Control at:

<https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide>

This gives daily numbers for a large number of countries in a basic, but useable, format. I have transposed this into a different, more informative, format and the corresponding Excel file is available.

2.3 I have chosen numbers for this model that reflect the position in the United Kingdom, according to this data, as of 21 March 2020. The position day by day changes so rapidly that these will be out of data before I finish writing, but I shall persevere. At that date there had been 3,983 reported cases and 177 reported deaths. The deaths were 4.44% of the number of cases. The number of cases since seven days previously (14 March) had multiplied by 4.63 and the number of deaths by 17.7.

2.4 The first input is (a) the initial susceptible population *H*(0) in cell C2. I have put 5,000,000. Why not put in the whole population of the UK? The first reason is that there seems to be evidence that a large number of people can get the infection, and suffer from it very mildly, or not even notice it. It is speculated that there might be 10 times as many of these as get COVID-19 sufficiently badly for it to be noticed. But anyone acutely ill, needing hospitalisation, is almost sure to be recorded, and deaths will also be recorded. So I treat the very mildly affected as if they were “immune”, and they are left out of the calculations.

2.5 Secondly, there are some areas and institutions that could keep entirely clear of the virus, remote islands, remoter country areas, monasteries, prisons, secure mental hospitals. All these could make themselves immune with sufficiently little outside contact, and careful checking, though if they were to be infected the disease might spread rapidly. So I put in 5,000,000 as *H*(0), less than 10% of the UK population.

2.6 The next input is (b), the initial number of inceptions on day 1, in cell I12. I put this at 1, but any other integer would do. Apart from the first person worldwide, for all countries the first infected person is an import. In fact there seem to have been several independent imports into the UK, but in order to investigate the pattern in one country, putting 1 is good enough.

2.7 The next input is (c) the values of *j*(*d*), the infectiousness at duration *d*, in cells I7 to AD7. I have put these at 1 for each day from 1 to 10. I do not know what the best medical opinion is on this, but if quarantine for 14 days is considered sufficient, then the period of infectiousness should be less than this. It is easy to assume 7 or 14 or any other number of days, or to assume that there are different degrees of infectiousness by putting fractional numbers for *j*(*d*), or that there is a delay before an infected person is infectious, by putting zero for the first or subsequent days. However the values of *j*(*d*) remain the same for all days *t*. We get variation by time by varying *r*(*t*), as discussed below.

2.8 The next input is (d) the values of *m*(*d*), the probabilities of death at duration *d*, in cells I8 to AC8. The deaths so far recorded in the U.K. have been about 4% of recorded cases, and this number applies also in China and world-wide, though in Italy it has been much higher. I have seen it argued that the real rate is much lower, because of the unrecorded mild cases; but I feel that we can only look at recorded cases,, and not speculate on the unknown number of unrecorded ones It is also possible that the death rate will turn out to be a little higher, because people do not all die immediately on getting the disease, so there may be deaths in the pipeline from cases already reported.

2.9 I initially use a total mortality rate of 0.04, which neds to be spread over the values of *m*(*d*) in some way. I do not know anything about the incidence of death by duration of illness. At one extreme one could put *m*(1) = 0.04 and all other values zero, or one could put it at the end of the period of infectiousness, or spread it over the day, putting say 0.004 for *m*(1) to *m*(10) inclusive. I decide after further investigation of the actual progress in the UK, as described below. But to start with I put *m*(1) = 0.04 and all other values of *m*(*d*) to zero.

2.10 The next, and very important input is (e) the values of *r*(*t*), the infection rate at time *t*, in cells E12 and below. To set this I need also to consider input (f) the value of the calendar date for time *t* = 0, in cell B11. The first recorded cases in the UK were two on 31 January, so I put *t* = 0 on 30 January, so 21 March is day 51 and I try to match results with that.

2.11 I initially try a value for *r*(*t*), say 0.2, for all values of *t*. I then look at the number of cases so far, the infected plus dead, in column BG. The actual number of cases in the UK reported as at 21 March was 3,983, with 177 deaths. My results for 21 March are in row 62. I get the total number of cases according to the model in cell BG62, and with *r*(*t*) = 0.2 it is it is 1,945, so it is much too low. By trial and error I get 3,812 cases with *r*(*t*) = 0 .215. You can see this by changing cell E12 to 0.215.

2.12 The total number of deaths so far, according to the model, is shown in column BF. The number of deaths as a proportion of the total number of cases is shown in column BI. With *m*(1) = 0.04, the percentage in column BI is always less than this to start with, because of the time lag even if deaths occur at the beginning of duration day 2., though eventually the percentage of deaths catches up. To match the 177 observed deaths by 21 March I need to increase the value of *m*(1). Changing the value of *m*(1) changes all the other numbers. After several trials and errors I use values *r*(*t*) = 0.218 and *m(*1) = 0.05 and get 4,020 cases and 171 deaths by 21 March, and you can replicate this by putting E12 to 0.218 and I8 to 0.05.

2.13 With these values the number of new cases reaches a maximum of 196,887 on 4 May, and the maximum number of deaths is a day later, at 9,844 on 5 May. I would not be surprised if, with these numbers, the recording system were to collapse. Also the funeral systems could not cope, since the normal number of deaths per day in the UK is around 2,000, with some seasonal variation. However, by 15 July the number of new cases expected by the model drops to less than 1.00, and the total number of deaths is neatly 250,000. Every one of the initial *H*(0) has been infected.

2.14 As at the time of writing the various governments in the UK (including those of the devolved nations) have taken various steps to try to control the infection. I cannot estimate the effect of any of the measures. But they have to affect the epidemic either by reducing *r*(*t*) or by changing *H*(*t*). One can reduce the value of *r*(*t*) by putting a chosen value in column E opposite the relevant day. If, for example, you wished to set *r*(*t*) = 0.1 for 24 March onwards, put 0.1 in cell E65. It is copied on for each cell thereafter. With this, one can see that the peak numbers of infections and deaths are about halved, and occur about 3 or 4 July. But the disease continues to November, and there are just as many deaths in total.

2.15 Changing *H*(*t*) can be done in Excel. Assume that the isolation measures proposed put healthy people (and perhaps especially older ones, though I am not allowing for age) into the “immune” category. Imagine that this works with half the healthy population as at 24 March. The formula in cell G65 is “= G64 – I65”, that is the previous days healthy minus the day’s new infections. If we write “=(G64 – I65)/2” then the value of *H*(*t*) is halved. If we include both adjustments, to *r*(*t*) and *H*(*t*), we get the peak a bit sooner, about 23 June, with the number of cases and deaths again halved to about 43,000 and 2,000 per day. The epidemic lasts till about 30 October, but the total number of death is almost halved to a little over 125,000. This is still a lot of deaths. Note that the check total, in column BK is changed by this process.

2.16 I must emphasize that I am not making specific forecasts. I am providing a tool that allows others to try out the results of hypothetical changes in what seem to me to be a plausible set of initial assumptions. If I made it much more realistic by stratifying the population by age, the model would need a full programme, which would be much harder for people to play with. Those who are happy changing the Excel may modify it as they wish. Others may just try altering certain values. My only request is that you recognise the authorship, and do not try copywriting it yourself to prevent others using it.

Copyright A D Wilkie

24 March 2020