Usage Guide for Spreadsheet Model:

R. Shuler, 4/21,2020 – updated 6/5/2020

Contents

User Guide Patch for Lite Version Model

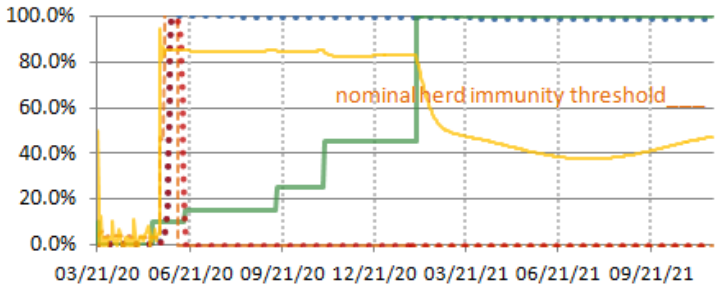
- References to tab **World-likely** can be replaced by any state or country tab.
- There is no longer a **US-likely** tab, use any state or country tab for the US, they are now the same.

Introduction

See ***Partial unlock for COVID-19-like epidemics can save 1-3 million lives worldwide*** for theoretical discussion and examples.

On opening the spreadsheet, you will get a circular reference warning. These are deliberate. Please click cancel:

The circular references are in =IF statements used to "save" predictions for later comparisons on the R0 tracking chart. Most likely you will have no further trouble, but if you get a glitch in a chart like this:



Then some data in columns B, C, D or E has gone missing. All the data is on other tabs in the latest versions, see Data Entry section of this guide at end. If you are missing data, usually near the latter entries, on a model tab, copy the cell above the missing data. This restores the link to the correct data tab.

You will also get notification that you cannot plot ZERO on a LOG plot. Just ignore it. These are from the small 14-year plots, which are best viewed on log scales, but if the infection goes to zero you get this error.

See END for new country & state data selection feature, and DATA ENTRY procedure and links.

- This is a TWO REGION model where one region is a part of the main region. **You can just use one main region if you wish.** Lately I've been using only one region.
- This is a **standard SIR-type epidemiological model which updates current replication rate from the actual data you enter.** It projects next day's rate using the difference in the last two day's calculate rate (new cases / active cases). It adjusts this rate after that by the herd immunity factor which it computes from cumulative cases and population. Because it does not filter data, detailed prediction will vary considerably if there is a significant deviation in today's data. The model is intended for scenario analysis and investigation of case load management techniques. Though if followed over several days and the variations considered together, it will do a reasonable job of prediction. Filters are limited to a 2-day moving average so it is more responsive for case load management (see "Feedback . . ." section below).
- If you have **a geographic region such as a string of villages along a road or coastline**, the virus will spread geometrically or linearly instead of geometrically. This model is not suitable for geometric spreading. It is more for mixed populations, or populations with random travel, or with infection already seeded in most towns and cities.

- The **names of model parameters** (including region names, suggest keeping them short) are along the top row. Values are in second row. E.g.,
- **Region and sub-region names** are near the far right. You will need to enter mortality rates as a percent of vent cases (AJ-AL). Population in AD-AE.

Population main region	Population subregion (e.g. NY St.)
330,000,000	20,000,000

- Then **enter available vents in….**

Vents main region	Vents subregion (if not auto alloc)
100,000	7,500

Initial Known Cases on Vent.
5.0%

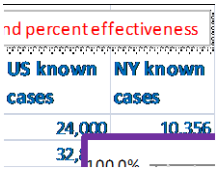
- You can specify **percent of known cases requiring ventilation**.

Initial Known Cases on Vent.
5.0%

This is really just a scale factor. No need to obsess over it. In reality it will change as testing and other factors change. **If you are managing some other scarce medical resource, just change the name and quantities appropriately.** Make the following two adjustments:
    - CALIBRATE VENT USAGE: Run the model for the actual data you have and adjust either the number of ventilators or the percent of cases on vents so that the ventilator utilization matches what you know it to be. If you don't know this, you can't calibrate the vent utilization and will have to guess.
    - CALIBRATE MORTALITY: Again running the model with the actual data you have, check the estimated deaths in AL. These will not be phased exactly right, but it is close enough to just match the total deaths with the actual total on the last date for which you have data by adjusting the mortality rate (separately for each region). If 80% of those on vents die, you might think the rate will be 80%, but it will be larger because some people die without being put on a vent.
    - **NOTE: The cases on vent will automatically reduce** to about half its initial value in one year, to simulate improving therapeutic technique. Since mortality depends on this, it will reduce also.
- Cell **I2** is used to select region. 0=main region (e.g. US), 1=sub region (e.g. NY), 2=main region without sub region. Various cells are conditional on H1, including some data column headings, region population, ventilators per capita, etc. Do not edit these directly. DO NOT edit gray shaded cells.

Tracking Usage

- Enter cumulative known cases for the main and sub-region in columns **D** and **E**.



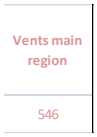
- NEW as of May 14 – ENTER COMPOSITE SUB-REGION DATA ON NYNJCT tab which will sum New York, New Jersey and Connecticut and post to main US tabs.
- If you want to keep a different record of previously predicted data (it's already there for US/NY), you have to paste ***values*** from column **S** into column **U**. Do this before adding data beyond the point from which you are calculating the prediction.

Phased Unlock Amounts and Dates

- **The primary unlock date** and percentage effectiveness is in **D1-D2**.
- **A1** defines when partial daily unlock can start, and the percentage **A2** is a FLOOR not a max.
- The dates in **L3-L9** will enable the daily unlock schedule in column **M**, subject to the overall date in **B1**. This is overridden by the thresholds in column **F** if they are non-zero. The effectiveness of this unlock is in **B1**. I'm not actually using these since the threshold feedback control loop was added. Either one is enabled by the Use Daily Partial Unlock control.
- A *RE-LOCK* date is in **C1**. If not using this, set it far in the future.
- A second unlock date is in cell **E1** and its effectiveness in **E2**. Third and fourth unlock dates are in **F1**, **G1**, with their effectiveness in **F2**, **G2**.

- Scale all unlock percentages with **H2**.
- Unlock everything (override) with **K2**.
- Note that to terminate death counts on a specific date, you should set the unlock percentage back to zero as of that date. You can also do this with the RE-LOCK control.
- The length of calculation of economic impact is 540 days (18 months) embedded in a formula in **P2** where the percentage of locked down economy is displayed. This is calculated as the average of daily unlock percentages.

Other Model Parameters

- **Cell** **provides the option to allocate ventilators by population**, and cell 

is the number of ventilators currently allocated by this algorithm. This is applicable only to a two-region model and most of the models have moved away from that.

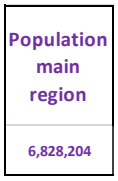
- Enter the percent of **KNOWN cases requiring ventilators** in 

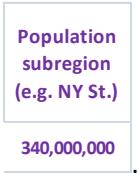
. I already covered this but here is some additional advice. Beware of what people say in interviews. Use actual data on number of ventilators and number of known cases. Note, number of ventilators changes with time. NY had around 3000 initially, and about 5500 in early April. Talking heads were saying 5%, but the NY data indicated it could not be more than 3%. If you don't like that, change it to whatever you want. I have moved away from adjusting this and instead adjusting mortality.

- Cell is an initial estimate of *R0*. It has no bearing on actual data, or even predictions under lockdown (social distancing). But when it comes to unlock, the easing goes the appropriate percentage toward this number. The model reduces it by herd immunity based on cumulative cases.

- is the multiple between known and total cases, or **case ratio**. This is the most important parameter in the model. This was taken from the literature (see my paper for references). 6 gives 16.6%. The published number is more like 14%. Over time with more testing, the number may change. It primarily affects at what level herd immunity reduces *R0*. It is easy to change the number to perform sensitivity analysis.

- **The average duration of a case** is in This affects only the number of active cases, not vents.

- **The spreading days is in** . Leave it alone! This is not really a parameter. It is a scale factor so that true R0 can be used in the model. It converts R0 to a time-based parameter. It is **also used to compute ventilator utilization** but you cannot change it without messing up R0, sorry.
- Region population is a formula (do not edit it) dependent on region selection and the region populations which you enter in 

and 

. On World-likely tab, population is pulled from a database when you enter country name.

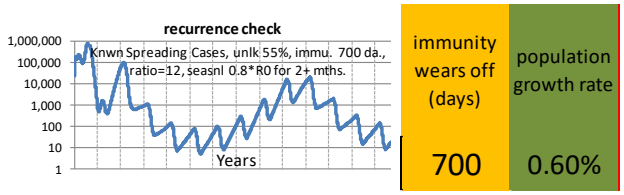
Analytical Features & Profiles

- It can be tedious to get all the unlock percentages and other parameters set conveniently. Even I forget what they do. **Useful templates are on the PROFILES tab**, with descriptions of what they do.
- **The US data in the World-likely tab is LINKED to the US-likely tab**. You only need to enter the world data (total cumulative cases).
- **ALL OTHER TABS ARE LINKED**, their parameters copied from either the US or WORLD root tab (likely). But if something looks wrong, see if the links are working. You can paste again from a profile.
- **The GRAPH on the US (left most** if you change region names) tab displays **data from the second tab**, which is linked and displays the same region. These are the final two columns in the graph. They can be set to choose columns from the second tab in cell **N3**, mysteriously shaded gray with a small number to one decimal place. The first column to the right of the shaded square is ZERO. But look at the second tab to see what columns you are actually getting. For example, **a useful choice is 2.3 which selects unlock percent and vent utilization**. Some results are displayed with the data legend title (deaths and cost on % unlock for example).
- The **ANALYSIS tab** contains charts and graphs. Many do not automatically update. Read the paper (available on the website) for discussion. Two of them are useful for estimating whether a peak is forming as you expect, and gaining a rough idea of case ratio.
    - About row 123 is **plot of actual and projected new cases** from the left most tab. It is linked and displays whatever region is set on that tab, at whatever case ratio and other parameters. Tabs labeled "-exp" are linked copies of the first tab with only partial data, for comparing predicted and actual new cases to see if actual follow the projected peak. At the moment I have some scaled World data in this chart. You can click on the chart (twice, slowly) and then drag the data coverage area to remove these. The data from the first tab auto updates, but not from the World tab.
    - To the right of that is a **less sophisticated World chart for case ratio analysis** with actual to a certain point and then predicted data at different case ratios. On some days this looks useful, and on others not. Column S updates from the World-likely tab. Copy and paste values into the other columns, with the parameters set however you want. Manually update the column headers to reflect what you pasted.

Population, Immunity, Seasonal & Mortality Trend Parameters

Most of these controls are on the far right unless otherwise noted.

- RESIDUAL ENDEMIC CHECK – a plot inserted in a corner of the main plot, first tab only, see below

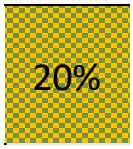


- POPULATION GROWTH RATE – 0.6% for US, enables check for low level infection persisting if the herd immunity target is hit too closely (whether through vaccination or cases) because of population growth (which is not immune). If you have known in-migration from non-immune regions, add that percentage. But if there is any immunity in the migrants, reduce toward zero at

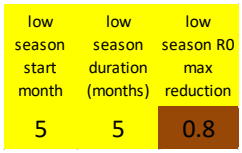
the herd immunity threshold (i.e. migration from a herd immune population does not unbalance your immunity).

- IMMUNITY WEARS OFF (DAYS) – This adds to the usefulness of the endemic check. Susceptibility is increased by the number of new cases that appeared this many days earlier. For two years, set 730 days, etc.

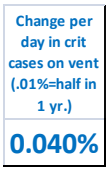
(col. N) (col. AQ etc. . . .)



. . .

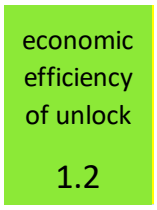


- LOW SEASON START/DURATION/R0 REDUCTION – For the example shown above, months 1-4 and 10-12 have the normal specified BASE R0 (before modification by empirical data and susceptible). Months 6,7,8 (all months greater than the start but less than start + duration) have 0.8*R0. The first and last month of the low season, 5 and 9, have 0.9*R0. Set the amount in N2 as a percentage seasonality (20% in the example). Generalize from that.
- CHANGE PER DAY IN CRITICAL CASES ON VENT – While this is not strictly new, a change in formulas limits it to 6 months. Currently (which I may change without updating this document) the value is set to 0.02% per day which cuts vent requirements in half, and mortality in half by implication, within 6 months. After that I assume no further improvement. Otherwise it drops toward zero exponentially which is not realistic, and strips data from the 14-year residual endemic check chart.



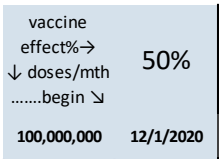
- By the way, endemic deaths are NOT included in the death total since that cuts off at 18 months assuming a vaccine comes to the rescue, and the point is to calculate the cost of waiting. But you can look at the DEATHS column AJ and compute the difference between any two dates.

- You can increase economic unlock more than social distance. This was necessary after May 4, 2020 when it became apparent that masks and other means of restoring economic activity without significantly raising R0 were (at least at that time) working.



VACCINATION MODEL

- Look for these controls …



You can set the data vaccination starts, the number of doses per month, and the effectiveness of the vaccine. On most charts the person-immunity-equivalents will show as a blue dot-dash line (e.g. 10% of the population was vaccinated at 50% effectiveness, it will show as 5% of the population). Vaccines which are not fully effective are annoying to individuals because they aren't prevented from getting ill. But public health officials are actually going for population immunity, and they are quite effective at that. Flu vaccine is 30-70% effective. So that explains why you get the flu anyway. You may get COVID-19 anyway as well. Especially if you are at high risk, is my guess.

COUNTRY DATA SELECTION

Data Select (3=World) see W-data tab	start date displacement
113	0

- Country data from [OurWorldInData](). The spreadsheet now has a tab **W-data** with country case data and population.

Country (spell exactly, copy from W-data)
Lebanon

- Now the preferred method to select a country or a US state is to type its NAME here: It must be the exact name, see W-data tab. You can always copy the name from that tab.

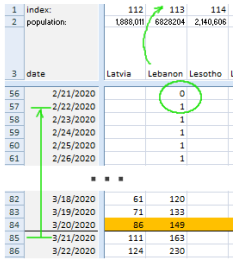
The first time you load country data

0

Once you have set what you think is the initial May 15 unlock amount, use the predicted data save control to reset the R0 tracking chart. Set it to 1, wait a few seconds, and set it back to 0. (This is what causes all the circular formula warnings).

The following is provided for information for the curious. Best to use the name method above. Once you overwrite a formula, you will have to download again and copy the cell to get it back

- Any country can be selected from the **World-likely** tab (not from other tabs). Using the fields shown above. Data select 3 (for column 3 on the data tab) gives whole world data. Start displacement 0 uses 3/21/2020 which we have been using all along, but many countries need to be started earlier to calibrate social distancing (unlock 0% calibration). There is an entry on the data tab called **index** which gives the column number. I will do an example. For Lebanon, we have . . .



Data Select (3=D=World) see W-data tab	start date displacement
113	-28

Note, there is a formula in start date displacement which will copy a value from the data, usually near the start of data. If you override this you will lose the formula. You can always download another copy of the model and copy the formula.

US STATE DATA SELECTION

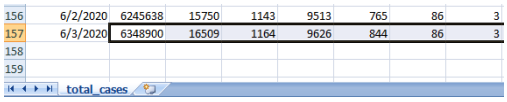
- US STATE DATA SELECTION – State data for the US is now also on the W-data tab. The World-likely tab is set up to access it in the same manner as country data. See below for data updating.

DATA ENTRY AND UPDATING

Currently the tab containing all the latest features (seasonality AND vaccine model AND ability to select any country or state data) is the **World-likely** tab. I suggest you use this tab exclusively for the time being and for only one country or region at a time. Here is how to enter the data for it. No data is any longer entered directly on the model tabs. All data is entered on the **W-data** tab.

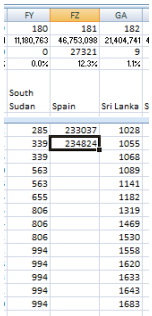
Country Data:

- Go to https://ourworldindata.org/coronavirus-source-data and scroll down and click on the Total confirmed cases file, saving it as **total_cases.csv**. This is usually up to the current day by 5pm CDT.
- Open this file in Excel and select the last line, EXCEPT the date, like this…



- Press **ctrl-C** to copy.
- Open the model file **Coronavirus cases directR0.xlsx** and go to the **W-data** tab.
- Select the empty cell in the D (World) column next to the date you just copied data from

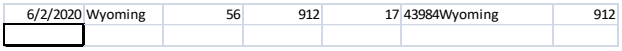
- Press **ctrl-V** to paste the data there.
- Scroll right to find any **blanks** in the data, like these…

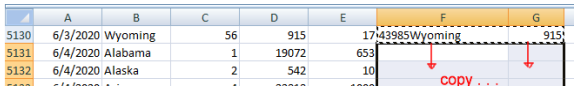


- Use copy and paste to extend the latest data into the blanks.

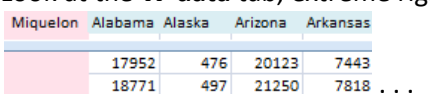
US State Data:

- Go to https://github.com/nytimes/covid-19-data and scroll down and click **us-states.csv**. This is usually updated by the next morning.
- In the headings about mid page on the right click **Raw**. Comma-separated-value data is displayed.
- From your browser menus select **Save Page As** and be sure **text document** is selected.
- Open the file **us-states.csv** in Excel. Scroll down and **select and copy** ALL data for any dates you do not have. This data is 5-columns wide.
- Open the model file **Coronavirus cases directR0.xlsx** and go to the **US-states** tab.
- **Scroll to the end and paste** the data you copied.





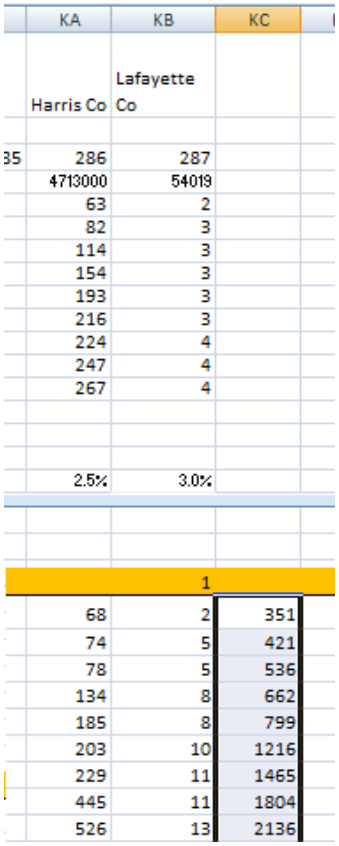
- Copy last entry in cols F-G beside the new data.
- Look at the **W-data** tab, extreme right, and verify state data has been updated.



HOW TO ADD A US COUNTY

Over 200 countries and the 50 US states are preloaded in the spreadsheet on the **W-data** tab. You can add any other region by adding a column for it on this tab, and it will be automatically accessible by typing the region name on the **World-likely** tab in cell **A2**. We will show how to do this by adding Los Angeles County.

1. For a US county, go to https://github.com/nytimes/covid-19-data and click **us-counties.csv** and click **DOWNLOAD**.
2. Open this file and expand the date field so you can read it. Select the column headings **A-F**. Filter this data for the state and county you want by going to **Sort & Filter** on the **Home** menu bar and selecting **Filter**.
3. Buttons with down arrows will appear next to the column headings. First select the STATE. Then select the COUNTY.
4. Select (by clicking and dragging) the CASES data for the dates you want, usually beginning with 3/21 and continuing until the latest date. Use CTRL-C to copy.
5. Go to the model spreadsheet W-data tab, and click in the first cell you want to copy TO, usually the 3/21 date in a new column at the extreme right. Then use CTRL-V to paste. The result will look like this:



	KA	KB	KC
	Harris Co	Lafayette Co	
35	286	287	
	4713000	54019	
	63	2	
	82	3	
	114	3	
	154	3	
	193	3	
	216	3	
	224	4	
	247	4	
	267	4	
	2.5%	3.0%	
		1	
	68	2	351
	74	5	421
	78	5	536
	134	8	662
	185	8	799
	203	10	1216
	229	11	1465
	445	11	1804
	526	13	2136

6. Now you have to fill in the blanks at the top of the column. Here is the list of parameters:

	A	KC
1	date	Los Angeles Co
2		
3	index:	288
4	population:	10400000
5	4/16/20	576
6	4/23/20	895
7	Deaths 4/30/20	1209
8	5/7/20	1512
9	5/14/20	1793
10	5/21/20	2090
11	5/28/20	2338
12	6/4/20	2620
13	6/11/20	2890
14	6/18/20	3110
15	6/25/20	
16	7/2/20	
17	mortality kn cases 5	7.0%
18	mortality reduction	0.061%
19	ventilators	5000
20	seasonality	10%
21	test positivity	9%
22	case ratio	8.7
23	start displacement	0
24	date	Los Angeles Co

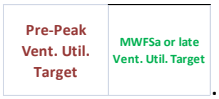
7. The INDEX (ROW 3) should already be there. Type the name of the county in ROW 1. It should automatically appear in a blue square also in ROW 24.
8. Enter the population of the region in ROW 4.
9. You already have the deaths data in us-counties.csv. Copy deaths for the dates shown.
10. Manually estimate the seasonality and enter in ROW 20. Look at the seasonality for other areas and pick one similar. For Southern Hemisphere use a negative percent. For tropical countries that are the same year round enter zero.
11. For start displacement enter zero, unless you used a start date other than 3/21. These numbers aren't automatically used. Look at other regions to see how it works.
12. **CASE RATIO** - Try to determine test positivity for the region. For LA County we found a number of 9%. Assuming testing has been going on for a while, this will be the percent of the population infected. For LA county that means 9% of 10.4 million people = 936,000 people have COVID. But the known active cases was only 104 thousand. So the case ratio is 936/104=9. Enter in ROW 22. Hmm, looks like the number we entered earlier (8.7) needs updating slightly. This is close enough, however.
13. Find the number of ICU beds available in the region (or guess at it) and enter in ROW 19.
14. Go to the World-Likely tab and type the name of the region and make sure it comes up.
15. Look at the mortality plot and take the first number and enter as initial mortality on ROW 17.
16. Enter a guess at the rate of improvement in mortality in ROW 18, something like 0.06%.
17. Now find the current total deaths, and a recent daily deaths amount for the region which is typical. You will CALIBRATE the two mortality numbers you just entered. On the World-likely tab look for this square: 

. The top number is the model calculation of total deaths, and the bottom number a recent daily deaths amount. Adjust the initial mortality to get the top number approximately. Then adjust the mortality improvement number to get close to the bottom one. You will have to adjust these iteratively about three times to get close, as each affects the other.
18. You are now set up. Try different unlock amounts to get the result you want.
19. Enable macros and use CTRL-T to update the R0 chart.

Feedback Based (1-week lag) Daily Unlock Schedule

NOTE: This feature is not actively supported at the moment and may not work.

- Cell **Use Daily Partial Unlock** selects whether feedback is used to enable day-based partial unlock. You must enter the degree of unlock in the phased unlock date controlled cells described in the next section. This feature is no longer heavily used. Write if you need help making it work.
- The numbers in column **Target Vent. Util.** are ventilator utilization percentages. The model simulates a management group locking and unlocking specific days to try and manage the case load to a constant level. Roughly if the utilization is lower easing kicks in, if higher locking resumes.

  These numbers are now calculated from the target you specify in 

. Just make them all the same.
- To avoid over control by a bang-bang controller, each day of the week adds an additional 2% margin to the threshold so that all days do not kick in at the same time. Think of this as part of the gain of the feedback loop. This simulates the actions of a decision maker for simulation. It is not intended to actually manage caseload, see last bullet in this section. A human does that.
- The other aspects of loop gain are the compliance of the population which you have to measure or estimate, and the spreading speed *R0* of the virus. Both of these are calibrated by actual data in this model, it doesn't try to guess them. Well, it does try to guess *R0* one day in advance. Compliance and virus replication are both components of *R0* in column **K**.

Initial & Unlocked R0	Total / Known case ratio
2.5	12.0

- Starting too far before the caseload peaks will lack downward control authority.
- There is a week lag built in. To change this you have to change all the formulas in column **M**. They have to be individually changed for the 2nd week. Then this week can be copied and pasted a week at a time. Copying just one cell (rather than a week) will ruin this column.
- In actually managing caseload, not just running simulations, you would *use new cases data rather than ventilator data to avoid extra lag*. You should be able to forecast ventilator usage from new cases data.

To vary the nominal ventilator threshold to trigger unlock for a day, originally values in column **F** were edited directly, and some time was spent making alternating day patterns. This version of the spreadsheet has simple '=' formulas in F which reference the thresholds mentioned above. Only the first is used at the moment, I believe. Originally it was intended to be used (with some risk) before the peak, but auto peak detection didn't work too well. There can be zig zags and many peaks.

EXCEL ERROR MESSAGES

There are two types of errors. Those you ignore (two of them), and data errors. If there is a model error, you probably won't get a message. If you get a message that is not planned, there is an error in the data.

Planned Error Messages

- ### Zero on Log Plot Error
  There are many log scale plots, most for the 14 year long range view. If cases goes to zero, you get an error message. There is no way to turn them off. Ignore it.

- ### Circular Formula Error
  The R0 plots require remembering predicted data. This is done with a two-state cell which retains its old value in one mode, and remembers a new value in the other. This is supported in Excel, but turning off the message is not. It is too convenient and saves too much time to consider doing it another way. There really isn't another way.

Unplanned (data) errors

These take two forms:

- ### Sharp unrealistic peaks in plots
  These are due to an abrupt change in data in a disallowed direction, such as total cases being less than the previous value (physically impossible in the real world). However, you may get data like this. I found a data decline of this type in Spain case data. There may be others in the database in your model file which are in data I haven't used yet. Find the data that is backing up and replace it, fudge it, or delete it. Your choice. Nothing I can do about it.

  Sometimes the predicted data memory formulas (see above) will cause a zero to be inserted in the data. This used to be common but I haven't seen it lately. Copy the cell above (the data cells on model tabs only contain formulas, not actual data) to restore the correct data.

- ### Messages such as NUM# and VALUE# and DIV/0# in one or many cells
  These are also due to data problems. For example, you have updated world data but not US data for today's date, because the US data isn't available yet. The model sees the world data and thinks "Oh, I have data on this date" and proceeds to subtract US data from world data, or some other thing like that. All hell breaks loose. Delete any data beyond where you have data from everyone and it will go away.

  You will get strange results if you start before 3/21/2020 or if a country doesn't have any data until after that date. Supply the missing data, even if it is zero. Or move the start date.

  The data sources will sometimes place a blank on a day where the data hasn't changed. Excel reads this as ZERO. Doesn't really work, does it? Copy the last data to fill any blank cells.