

Reliability of R Estimates – Progress Report

Visualization

The Winton centre has been contacted and [REDACTED] is happy to help, he has recommended wording for each level of reliability and we will indicate these with ticks on the graphs themselves. This update will be linked to plans to look at a banding system for the estimates themselves to indicate any changes through time.

Calculation of Reliability

Background

This is clearly a research project in and of itself. There are many complex measures and studies that could be undertaken. However, given the pressing time constraints the following strawman assumes that we require a system that can be implemented quickly and rigorously using already available data and without substantial effort from SPI-M members.

We are therefore assuming that the assessment of R is on its applicability to a region as a standalone estimate for further inference, not on a particular models ability to predict an R value, this would be another possible metric but would require detailed examination of every prediction which cannot be conducted given the above constraints.

Rating

The proposed strawman is as follows, here each number in the table below represents a number of ticks from zero to three, and at zero it is recommended that the values are deemed unreliable.

In summary there are two disparate reasons why R might not be a good measure in a particular region, because case numbers are too low or because the cases in a region are too heterogeneous. Therefore this metric combines a measure of these two cases in a standard two way combination. Clearly the best way to measure these would be the incidence from a rigorous testing regime and the spatial heterogeneity from a fully spatially resolved model. However, these are not currently available so we propose using deaths as a proxy for incidence (although many other metrics could be chosen) and we measure heterogeneity as positive tests above homogeneity in a sub-region

For these measures counts are defined as numbers in any region and metric. The suggested measure for overall counts would be numbers of deaths per day in a given region. There are many ways of measuring heterogeneity and it is suggested that the simple approach proposed as an interim solution is improved with a more rigorous metric in due course. Initially it is proposed that the number of positive tests per region normalised by the number of tests and population is used as a proxy.

The metric for counts is based loosely off the normal approximation to a Poisson as this defines a count number for which we can make continuous distributional assumptions reliably and deaths in a region are clearly counts of this form.

The measure of variance accounts for regions where we might have a large outbreak in a single sub-region. Here we would expect a large number in one sub-region and low numbers elsewhere. There are many, many ways of doing this and it is really a research question in and of itself. However, the direction is for a very simple metric as follows.

Work with each region and metric in turn. Find the median (middle value, m) and the maximum (M).

Then if $M < 6m$ the heterogeneity is low, if $M < 9m$ the heterogeneity is medium, if $M < 12m$ it is high and everything else is very high.

This metric is therefore simply defined as the number of times outside of the overall range that the highest data point lies.

Allocation of a Reliability

The allocation of a reliability is then achieved as follows:

Death \ Variability	Low	Medium	High	Very High
100+	3	3	2	0
<100	3	2	1	0
<10	2	1	1	0
<5	0	0	0	0

I would further suggest any region with less than 3 data points is given a score of 1 automatically and any with less than 2 a score of 0.

Suggested Wording for each level

The wording is taken from the DI yardstick as recommended by the Winton centre and in line with the research done by the IPCC.

0 – It is highly unlikely that these estimates can be relied upon due to the low levels of infection and clustered nature of the outbreak in this region

1 – It is unlikely that these estimates can be relied upon due to the low levels of infection and clustered nature of the outbreak in this region

2 – It is likely that these estimates are reliable and are a good measure of the current situation

3 – It is highly likely that these estimates are reliable and are a good measure of the current situation

Conclusion

A basic method for assessing R reliability with a real metric is entirely possible. There are many detailed ways to do this, however, there is a justifiable measure that can be implemented simply and efficiently in the timescales of relevance.