Review of: "A Multi-factor Model of COVID-19 Epidemic in California"

Josef Smolle

Potential competing interests: No potential competing interests to declare.

It is a worthwile undertanking that the authors do not rely on diffusion/differenital equation modeling, but apply straight-forward statistics on original data.

In paragraph 1 of the Abstract "duration of the epidemic" is named as one of the two dependent variables, whereas paragraph 2 refers to "duration of infection". This may be misleading - the use of a common term is recommended.

It is unclear, what "duration of the epidemic" or "duration of infection" means. Does it refer to the duration of the disease in an individual (usually a couple of days)? Or does it refer to the duration of the whole epidemic? In the latter case, the authors should state how the beginning and the end of the epidemic was defined. Was it defined with the occurence of the first and the last case in the region?

As far as income is concerned, the authors should state whether the income of the individual counties is given as mean or median income of the population.

Throughout the study, cumulative cases should not be given in absolute numbers, but should be related to the size of the population, e.g. cases per 1.000.000 people. Relying on absolute numbers, the high correlation with population size is self-evident, and correlation with other single factors cannot be reasonably interpreted. Therefore, the whole statistics, particularly the univariably statistics, should be repeated with relative case numbers as dependent variable.

In the legend of Figure 2, the authors should explain which type of curve was fitted to the different scatter plots.

The authors should provide a table with one line for each variable, denoting at least median and range. Furthermore, the contribution of each independent variable to the prediction of the two dependent variables should be given in an additional table.

In the "What is a multi-factor model" section the authors define "infection rate" as the "probability that contact will transfer the disease". In the "Combining factors" section the authors state that "The infection rate is shown to be uncorrelated and appears as a random variable". Do they use the same definition of infection rate in both instances? Or does the latter statement simply refer to the number of cases divided by population size?

Finally, the authors do not address the fact that the reported number of cases strongly depends on the number and availability of diagnostic tests performed. Therefore, the number of Covid-19 deaths is usually more reliable for comparisons than case numbers, because the likelyhood of severely ill patients to be tested is usually high and does not

differ so much between different regions.