

Review of: "Underdispersion in the reported Covid-19 case and death numbers may suggest data manipulations"

Oliver Watson¹

¹ Imperial College London

Potential competing interests: The author(s) declared that no potential competing interests exist.

The analysis is a simple approach based on the assumption that weekly COVID-19 mortality is well described by a Poisson distribution. With this assumption, country reported COVID-19 deaths can be assessed for statistical significant departures from this assumption that are suggestive of purposeful data manipulation.

Few general comments (bracketed comments less important):

1. Discussion of what other factors outside of data manipulation may cause departures from a Poisson distribution is needed. For example, testing limitations in a country may violate the suitability of the Poisson distribution. In a number of countries, COVID-19 tests were only available for specific populations (travellers, higher wealth groups etc) and as such mortality reporting may be expected to flatten artefactually.
2. A higher variance will be observed in countries with pronounced weekend effects or other periodic clustering. Would be interested in the outcome of the analysis if the subset of dates not impacted by periodic clustering were used, e.g. if Sat/Sun are significantly lower, assume that Mon-Fri is described by one Poisson distribution and Sat/Sun by another.
3. In weekly periods with increasing transmission and daily deaths, deaths will be increasing exponentially. The variance of an exponentially increasing series will be greater than a Poisson described by the weekly mean, which may mean that the approach taken underestimates the number of periods that are significantly underdispersed.
4. Similarly, the analysis could be more explorative and argue that weekly deaths should be more overdispersed than a Poisson. The general mean underdispersion index around 1 is suggestive that most countries are well explained by a Poisson, but it would be interesting to know how many countries would be better explained by a negative binomial through model comparison, and if there is any pattern here related to country size / undercount ratios etc.
5. National reporting of deaths often reflects the aggregation of multiple subnational reporting systems. As noted in the Russia example, local underdispersion may combine to mask national patterns. Conversely,

subnational reporting may combine to produce significant underdispersion at a national level if subnational patterns are highly asynchronous, e.g. one region experiencing a wave that triggers another wave. The death patterns in a number of countries that appear flat at a weekly level (Brazil 2020 comes to mind and some periods in the Iran epidemic) reflect these dynamics. This is an important limitation to note and I wonder if, for example, Brazil would have appeared underdispersed if reanalysed after removing weekend effects that likely elevate the variance observed.

(6. A vertical line on Figure 3 at $x = 1$ would be great for visualising the variance of the non underdispersed countries.)

(7. Few extra sentences introducing the underdispersion index and why it should be 1 given the Poisson distribution (i.e. mean == variance is a property of the Poisson) would help guide readers more).

Aside from these, I think it's a very interesting paper and important in again bringing to attention that reported numbers are often unable to be relied on for understanding epidemic dynamics. Whether that is due to purposeful manipulation, or other possible factors (testing limits or clustering in testing distribution), this approach would be very useful for incorporating into model forecasting attempts and could be used to caution users about the reliability of possible predictions due to statistically significant underdispersion.