Every afternoon, Public Health England (PHE) publishes a count of the COVID-19 deaths that occurred on the previous day. The count is an aggregate of individual counts provided to PHE by multiple local authorities.

However, only a few local authorities are able to collect and submit their count to PHE in time for the first release. The remaining counts arrive staggered over subsequent days.

As a result of the delay, the first count published for each date is lower than the number of deaths that actually occurred on that day. It’s only after several additional days that the published count starts to approach the actual number of people who passed away.

For example, figure 1 shows how the reported COVID-19 death count for 22 November 2020 was initially reported to be around 70, whereas the later counts are closer to 360.

Figure :

Shape

Description automatically generated with low confidence

This curve is pretty typical for dates of death in the midst of a pandemic ‘wave’, whereas dates in periods of relative quiescence do not see such a large change in reported counts. The difference between initial and latest reported figures for each unique date of death between October and November (early- to mid-second wave) varies from near 0 to nearly 300.

Text

Description automatically generated with low confidence

**Figure 2: Initial (blue) and latest (as of 30-Nov; green) counts of deaths occurring 28-Aug to 28-Nov**. In the lower panel, the dotted line represents the difference between initial and latest counts; i.e. the magnitude of the initial underestimation. The steep decline of the ‘latest’ counts at the end of November is a result of the reported counts still being within the underreporting phase. The counts in this portion of the curve are expected to increase significantly in later releases.

The extent of the underestimation appears to grow as the rate of death accelerates through the peak of the second wave. In late-August to September, when deaths were lower, the difference between initial and final counts are in the single- and low-double digits. In contrast, at the height of the second wave in the middle of November, the initial figure underestimates the true count by nearly 300. It takes about 3-5 days to close this reporting gap.

A sense of the spread of the death counts for each date of death as reported in subsequent releases, and the time it takes to close the reporting gap, is conveyed in the chart below.

Figure :

Chart, scatter chart

Description automatically generated

Similar to figure 2, this chart also shows the initial and latest counts over time. In addition, the counts as reported on intervening days are also shown. The colour of each dot represents the number of days after the date of death a count has been reported. At the height of the death rate, it commonly takes 4-6 days for the published count to reflect the actual count.

On their public COVID-19 dashboard, PHE addresses this phenomenon by labelling the most recent 5 days of data as “incomplete”. This is represented by a grey-shaded region, shown below (up to 14 January 2021).

Figure : PHE counts of deaths by date of **occurrence**

Chart, histogram

Description automatically generated

The counts continue to change after 5 days, especially when the death rate is high, but on average the published count at 5 days after the date of death accounts for 85% of the true figure. For example: the count of deaths that occurred on 15 November as reported on 20 November was 87% of the latest count as of 30 November (325 vs. 374).

PHE also publishes data on COVID-19 deaths by aggregating the counts that have been reported to it on each day, regardless of the actual date of death. By definition, this count is complete at the time of publication. This way of reporting is more vulnerable to the rising and falling activity through a work week—e.g. fewer local authorities are able to submit their data at the weekend—but can give a better idea of the short term trend.

Figure : PHE counts of deaths by date of **reporting**

Chart, histogram

Description automatically generated

The fact that the count of deaths occurring on a given day can change so radically from first to final reporting reflects the challenge of collecting data in a large, complex system such as national healthcare. This brief analysis investigates how the accuracy of the count changes over time.

The charts above show that when the death rate is low, the discrepancy between initial and actual number of deaths is small but increases dramatically once actual deaths-per-day is more than ~100 (corresponding to initial reports in the single and low-double digits). Consequently, the most recent 2-3 days of counts (by date of occurrence) should be all but disregarded, and the remaining counts in the 5-day period viewed with caution.

I would consider any uncritical use of the most recent 5 days’ counts to be misleading. The expectation is that counts will increase substantially within the first few days of reporting and so cannot be used to reliably detect trends. Instead, counts that are more than 5 days mature should be used for preliminary forecasting. Building a model for such forecasting might be an interesting future project.