The Pandemic is Over

Phil Shea

2022-07-29

Abstract

When does the *science* indicate the pandemic has ended? This paper develops a relatively simple model for the shape of expected weekly deaths, and posits that the pandemic is over when the weekly death curve regains its shape. This occured the last week of March 2022, but was only evident after five or more weeks following the expected death curve. The data is drawn dynamically from the CDC website, a model is fit and demonstrated, and the resolution of the pandemic clearly shown.

Contents

1 Introduction

Tracking when a pandemic is over can be difficult. Recently people have either not been testing when they feel ill, or simply using at-home kits and not reporting the results, so test statistics are useless in this regard. Since the reason the pandemic was taken so seriously was the threat of death, the overall death statistics gathered weekly by the CDC should tell the story. The weekly deaths vary over the year, and generally increase every year (likely due to population growth). This paper shows that the weekly death statistics are predictable,and that total weekly death statistics have reached back to 2019 levels, and therefore the pandemic is over.

1.1 Data Description

This is a comupatable document and the source is available on github at the author's repository. Total death data is retrieved from the CDC website: https://data.cdc.gov/NCHS/Weekly-counts-of-deaths-by-jurisdiction-and-age-gr/y5bj-9g5w.

The data contains deaths by age group, by jurisdiction (including each state, the entire US, and New York city), and also has observations "Predicted (weighted)" (whatever that is) and "Unweighted," (tagged in the `type` column) so many deaths are double counted. Note also that 2020 has 53 weeks (and so hase one week more of death than the other years). Week one of 2020 started on 30 December *2019*, and week 53 covers 28

December 2020 through 3 January *2021* (this follows the rule that weeks are always 7 days, and the week is assigned to the year with four of the seven days).

The code below will filter out just the “US,” “Unweighted” data, and sum the age groups to get total deaths for each week.

```
DS <- Death |> filter( state_abbreviation == 'US', type == 'Unweighted') |>
   group_by( week,  Year=year, date=week_ending_date) |>
   summarise( Total=sum( number_of_deaths), n=n()) # DS has weekly deaths

knitr::kable( DS[1:5, 1:5], caption = "The first five rows of the DS data frame.")
```

2 The Shape of Weekly Deaths

Below is the standard plot that many media outlets have been showing. It does appear that there are a remarkable number of excess deaths starting in 2020. The CDC indicates that the last five weeks or so typically do not have complete data yet (although experience indicates seven are weeks are often incomplete); they should be discounted. That each year’s counts increases is immediately evident.

The code below will perform the filtering and grouping as we did earlier, but will exclude years after 2020 from the data. Now we have to deal with 2020’s week 53 (it had an extra week). For simplicity’s sake, we will simply make the week 53 average the mean of week 1 and week 52 average deaths.

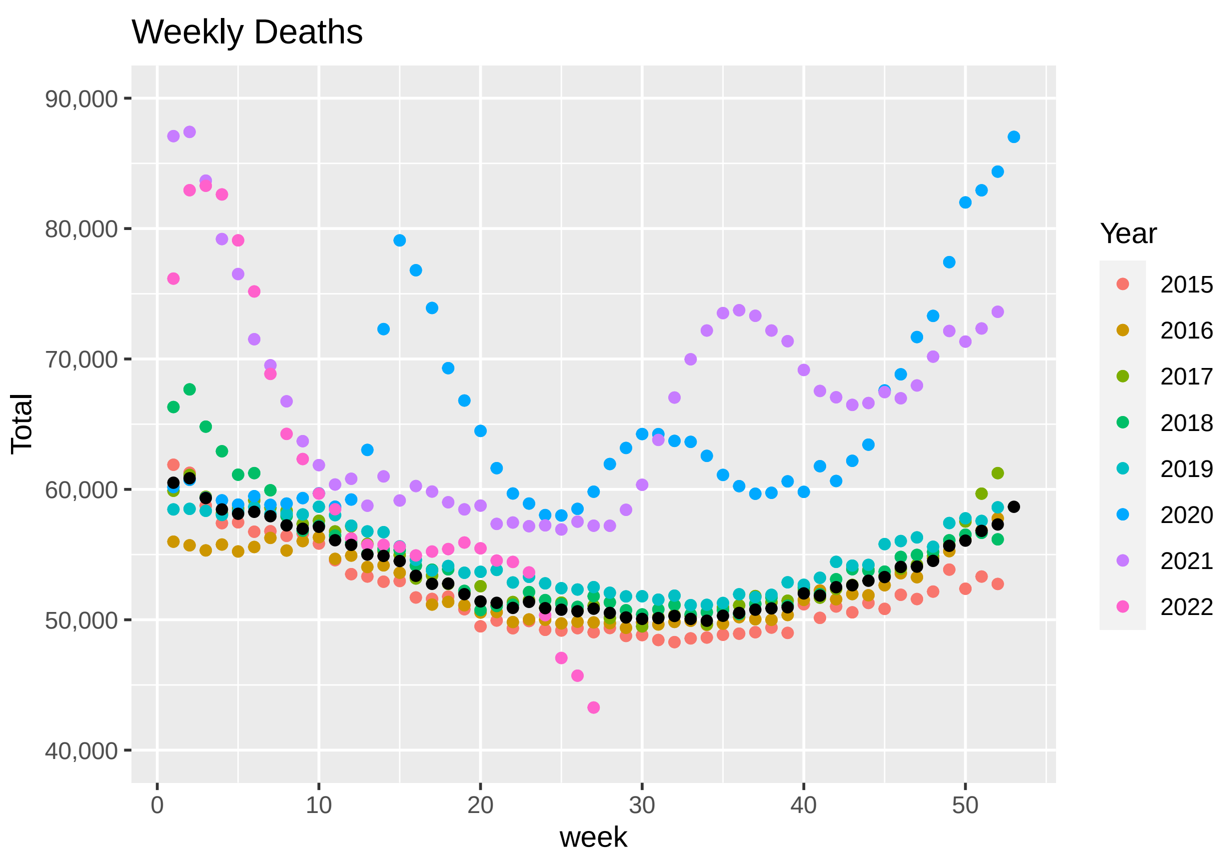

Figure 1: The entire weekly death data plotted by week number. Note that 2020 had 53 weeks, and that the last 7 weeks or so of data is incomplete. Black dots are the avereage for years 2015-2019.

We cut the lower part as any sample below 40,000 is incomplete. The average points have a distinct curvature that can be fit by polynomial.

2.1 Fitting the Shape

In order to make reliable estimates of weekly excess deaths, we need a reliable estimate of the weekly death rate, which obviously varies by week and by year. We'll look at a polynomial approximation the equations:

$$D_e(w, y) = \sum_{n=1}^{4} a_n(w - w_{min})^n + by + D_{ave}$$

where D_e are the expected weekly deaths in week w and year y, w_{min} is the week number of the minimum deaths (and is 34 in this data), b is the yearly increase in deaths, and D_{ave} is the average deaths over the modeled period. 2017 will be taken as year zero. A 4^{th} degree polynomial is used due to some prior work which indicates it is statistically "best." This model has six parameters: the four a_n, b, and D_{ave}.

Now we need to apply this method over the whole data set. The code below creates new `aYear` and `aw` columns, performs the linear model, and displays a summary of the fit. This is a lot of archane technical detail, but there is some interesting numbers there.

Briefly, this tells us that the coefficients are all highly significant, with the 4^{th} degree polynomial being just above 11% significance[1]. The polynomial coefficients are uninteresting[2], but the `aYear` value of `761` indicates that each year the weekly death increases by that amount (yearly: 52 * 760 or about 40,000). The R-squared values tell us that the model explains about 89% of the variation, which is remarkably good.

The following figure extends the fit through current dates.

H1N1 is evident in the beginning of 2018, and drove a January peak of expected deaths. Nevertheless, the curve matches the modeled years pretty well. Note the small highlighted section in March and April of 2022. Note that one should never extend a polynomial fit beyond the range of the original data, as the higher order terms will take over (and just like a Taylor series, are only valid over a small range), and we have not done so here. The polynomial only gives shape over the 52 weeks of the year - the extension is performed only by the *by* term, adding about 760 weekly deaths every year.

3 End of the COVID 19 Pandemic

When the death curve starts to match the prediction again, the pandemic is truly over. Simply eyeballing the previous graphs, it looks like the death rate has settled back to the 2019 level. The post-pandemic expected death rate for 2022 should be lower than that predicted by the model, as many people who would have died in 2022 (had there been no pandemic) had already died. The following demonstrates this.

This does not mean that people are no longer dying from COVID-19. They are. Just like people have always died from the flu, from colds, and other ailments The point is that they are dying **just like they always have**. Like all biological systems, this may change, and there is, perhaps, evidence of a slightly increasing death rate in weeks 18 and 19 (the last two highlighted weeks). (Before you panic, look at the increase from H1N1 in 2018; this increase is only about 1,500 where the H1N1 increase was over 10,000.) While the CDC claims that after five weeks the statistics should be reliable, experience has shown that seven weeks should be discounted, so the seventh latest point being on or below the 2019 prediction should be ignored.

[1] What this means is that there is only about an 11.2% chance that that 4^{th} degree coefficient would have happened randomly due to the data. The lower the significance the better.

[2] Principally due to the way `R` uses something called *orthogonal polynomials*

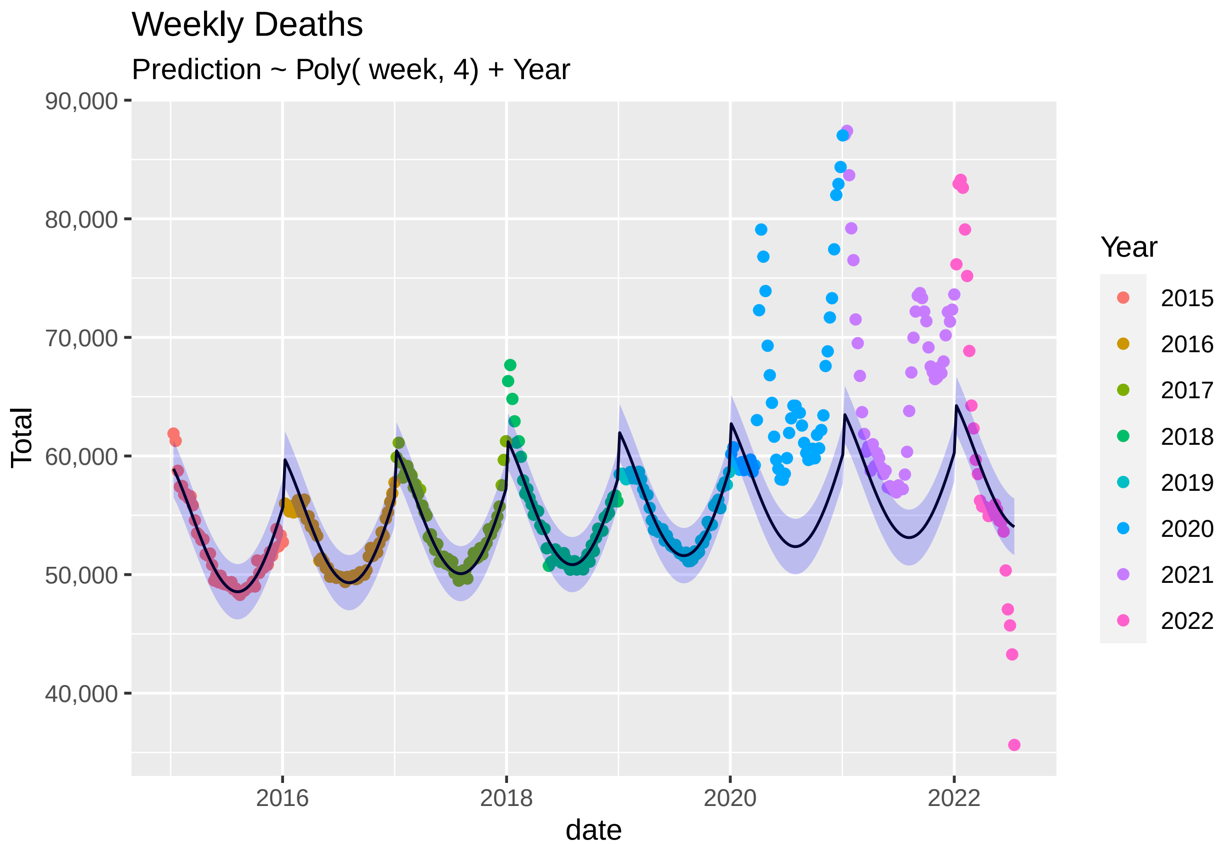

Figure 2: Weekly death with expected death curve extended throough 2022. The shaded ribbon shows the 95% cionfidence interval for hte predictions.

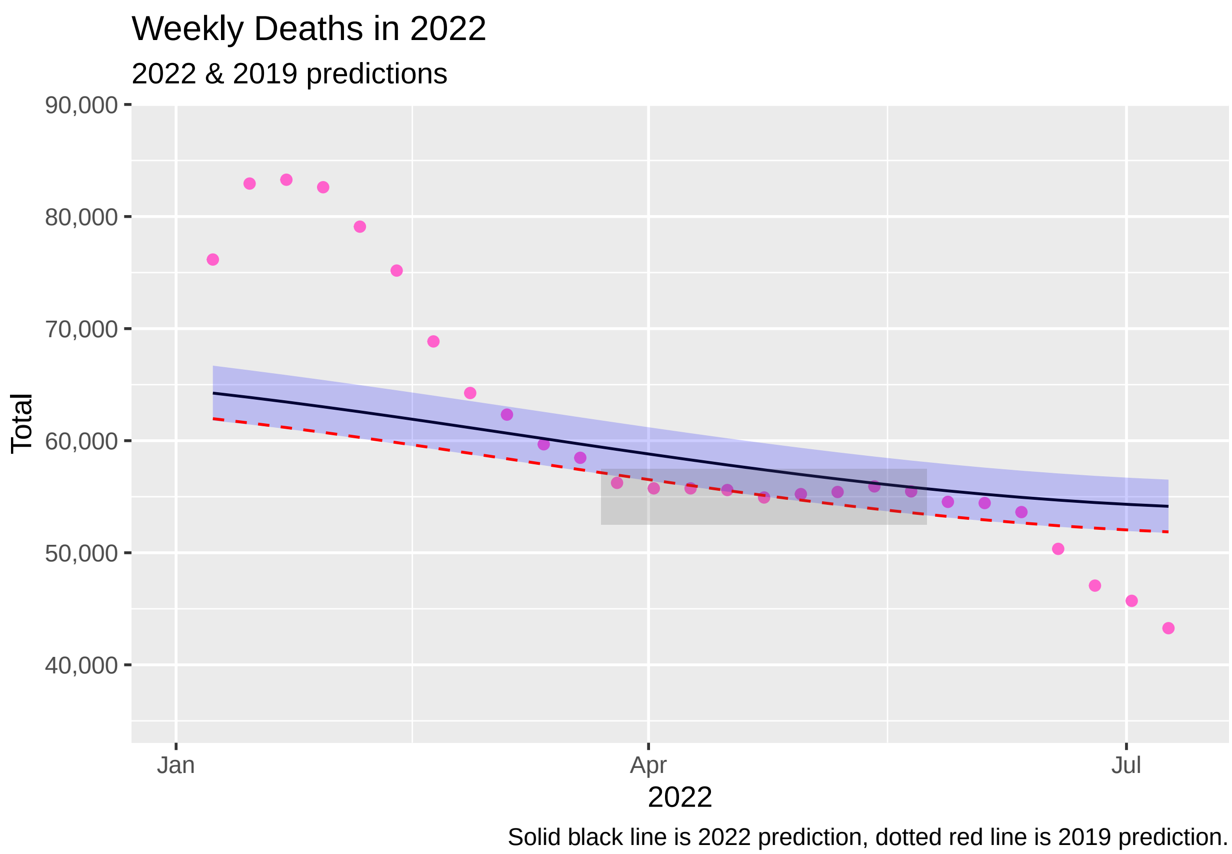

Figure 3: Zoom into 2022, with the red dotted line the predicted death line less three years of increases (essentially, 2019 levels). The shaded area is the significant points surrounding the 2019 level.