medium.com /@mrabkin/covid-related-deaths-are-much-higher-than-reported-and-soon-well-know-how-much-64bc9fffbe89

COVID-Related Deaths Are Much Higher Than Reported — And Soon We'll Know How Much

Mark Rabkin

16-20 minutes

(**Updated Apr 8, 9am.)**

Latest updates:

- **UK**: Very thorough Apr 7 data release has big jump in total mortality through Mar 27. FT points out that NHS (hospital) data a **78% higher death toll** than the government reports (this is not using total mortality). Direct look at mortality figures suggest a **61% undercounting on a very naive reading**. We'll know much more when the ONS releases the next week's data set on Tue Apr 14 (that week has ~5x more deaths).
- **Spain:** Experiencing huge reporting delays after Mar 24. They're doing daily updates, but with incomplete data — showing huge drop in deaths after mar 24 whereas official counts almost doubled from that date. We'll have to wait to see when the data catches up.
- **France:** Next national data release is Fri Apr 10.
- **US**: NYC death tolls don't count people who died at home, leading to undercounting from 35% to 49% according to FDNY data (from NPR, Gothamist, NYT). City authorities vow to fix this.

For a variety of good reasons, much of the public discourse has shifted to deaths rather than cases, given how fickle case counts are. Cases reflect the testing regime just as much as they reflect the reality of disease spread. As the epidemic grows, even countries that were doing sometimes won't be able to keep up with testing. So now we face a grim reality: at this stage, deaths are a more reliable way to see how a given state or country is affected by the epidemic.

This makes sense. Deaths are much harder to miss or under-count than an infection. And since there's 10–100x fewer deaths than cases, counting deaths is also more tractable.

But emerging evidence from the last few days suggests that our death tolls still have some of the same undercounting problems as confirmed cases. In democracies, the undercounting seems to stem mostly from missing deaths outside hospitals (at home, in nursing homes) or simply from insufficient testing or overly stringent diagnosis requirements.

Until just now, it was hard to tell how big this undercounting was — but initial data is showing that in hard-hit areas in Italy/Spain, **real deaths are 2–4x higher than official tallies**. Meanwhile, broad national data is starting to show 61–78% under-counting in the UK and **~40% under-counting in France** (see section below). Similarly, we're seeing ~35% under-counting in NYC due to not counting deaths at home.

This data right now is really early, since mortality figures take at least a week to collect in most countries. To make the data better, some countries are pushing ahead: France now has online updates with 4 days delay; Spain has a daily bulletin full of information, but alas big data delays. Unfortunately, the United States is a sitting duck with credible data available only after about 3 weeks.

(To be clear, this new data doesn't mean the disease is more fatal to the average person — it mostly just confirms the pandemic has spread further than we know, and caused a lot of cases and deaths that weren't properly attributed.)

This is important — for decision-making by governments, for people to take it seriously, and most importantly so we don't miss the second-order effects of the devastation of this epidemic.

For the clearest view of this, we can look at **total excess mortality**. This just means: compare how many people died in, say, the 2nd week of March 2020 to that same week the last 5–10 years. In big populations deaths typically vary only 10–15% a year, mostly due to changes in climate and diseases like the flu.

The good thing about this method is that it captures every death that got a death certificate. You don't need to argue about why fatality rates per case differ. You don't need to dispel Elon's conspiracy theories on whether cause of death in Italy is overstated by 8x (hint: nope, army trucks for body hauling don't come in for no reason). You just need to know: how many total people died total — and is it unusual?

The bad thing is it doesn't explain the mechanism and thus it doesn't explain the solution. Many of the extra deaths might be from the non-COVID conditions that don't get treatment with a lockdown with an overrun health system. Or it maybe the economic disruption from the lockdowns is leading to lack of food or medicine for many people. We don't know and this data doesn't say.

There's also surprising inverse effects from the lockdowns: in some places, mortalities from other causes are way down. Car crashes and fatalities are down by half in California in March. Flu cases are down 2–3x in Japan from from social distancing. Great air quality due to lack of traffic might save lives.

So again — this method shows us the net total, with all the pluses and minuses (without explaining them).

Roughly this same method produces the large yearly flu deaths estimates — the ones that led to all the terrible "this is just like a flu, it's even smaller than the flu" narratives that crippled response for weeks or months in many countries. The CDC estimates ~24k-63k deaths in the US from flu in 2019, using a model that (as far as I understand) projects out from a small # of actual flu tests, and correlates with excess mortality to calibrate. So those flu deaths were a shitty comparison: after-the-fact fully modeled flu mortality vs. solely directly confirmed COVID deaths, way early in an exponential epidemic.

This excess mortality method hasn't been used much yet for COVID because it only works once you have a cause of death that's big enough to spike up the total numbers at least 20% or so. Unfortunately, the COVID pandemic in many places is now so big that we're finally able to see total mortality numbers — and they're concerning.

We will know a lot more in the next few weeks as these numbers get more accurate.

Over the last few days, we're now seeing spikes in total mortality in most countries that are heavily affected. That puts us a lot closer to understanding the true impact of COVID on fatalities.

I've been really curious about this and have been looking around at what most countries are sharing. As mentioned above, it's a pretty mixed bag for what countries share, but it's good that we have what we have. The main pain point in this data is the delays. Worldwide, the number of deaths has grown 2x in the last 5 days and 4x in 10 days. So numbers that are 5–10 days slow are extremely out of date.

However, even in hindsight this data is starting to reveal flaws in how deaths are counted that I think will lead to big improvements in the tallies. Good on those countries that are making this transparently available online in as close to real time as they can do it (yay France and Spain!).

Here's what I've found for each country:

Update Apr 8: NYC has now had some revelations, showing **undercounting of 35–49% due to deaths at home being missed**.

This was first investigated and broken by the Gothamist, that revealed deaths at home (not in hospital) weren't counted. NPR estimates that on Monday Apr 7 there were 727 deaths in NYC, with 280 deaths at home versus a normal rate of 20–25 a day in the city, an excess of 255+. This is an **undercounting of ~35%.**

Mark Levine, chair of NYC Council Health Committee confirmed 3days ago:

Today, the Gothamist states the city has reversed policy and will begin to count "probable deaths" from COVID that occurred at home. They share this:

> Earlier in the day, Mayor Bill de Blasio acknowledged that the vast majority of deaths taking place at home were likely also due to the virus, meaning the death toll could be as much as 70 percent higher than currently reported figures.
>
> "We do want to know the truth about every death at home, but it's safe to assume that the vast majority are coronavirus related," he said. "That makes it even more sober, the sense of how many people we are losing."

FDNY says over the last two weeks (Mar20 to Apr5) 2192 NYC residents died in their home versus 453 same period last year (+1739). The cumulative NYC deaths official report stood at 3544 on Apr 8, a **49% undercounting**.

> Data from FDNY via Gothamist

Original from Apr 4: Unfortunately, and unsurprisingly — I have not yet seen any reasonably up-to-date mortality data for the US. If you see anything at all — for different states, for the country overall, for certain counties — please let me know!

We need this data to know the true total effects of both the lockdown and the virus.

The best statistics I have found are from the CDC at the national level. Alas, this data is very delayed. Even worse, it's just incomplete (it's not clear when the data is "done" and safe to use). The description says that there's at least a 1–2 week delay, and may be "several weeks" until all the data is collected from all over the country, and indeed everything less than 4 weeks ago has data missing. This will make it useful for after-the-fact analysis but almost a month delay is pretty useless for tracking how we're doing.

I've tried looking around at the state/county level, but so far have had no luck. New York State health site, for example, is not super encouraging. I guess we'll know in 2022.

Update Apr 8: Spain is continuing daily updates on total mortality, but disappointingly the data is relatively useless. The reporting from the last week or so from several big provinces, most notably the largest one of Madrid, is totally broken (and the update acknowledges data reporting delays). So right now for Spain, we don't have reliable recent mortality numbers.

Their latest update from the ISCiii on Apr 7 still contains huge drops from Mar 30 onward. In a big understatement, it says (translated):

> At this moment, we see a delay in notification of deaths from various regions, notably La Rioja and the Comunidad de Madrid. The results from the most recent days will be updated shortly.

Here's a graph of the nation overall and Madrid. They both show steep declines from Mar 24, which is obviously incorrect as the official counts have almost doubled from that date.

The official trend of deaths in Spain is here — obviously a more reasonable shape:

We'll have to wait until Spain fixes their reporting to see the real counts here — disappointing turn after a lot of good initial transparency.

In **Spain**, in in several hard-hit provinces, the official coronavirus death toll is 50–70% lower than the spike seen in total mortality, according to a deep analysis in El País. The Economist also took this up today. The summary is that excess deaths are being catalogued as generic pneumonia and not labeled coronavirus due to shortage of testing. And, for the graphs below, sources in the health ministry explain that the drop back down at the end of the graphs are "all a problem of data quality. The system, also the registers, is overflowed. The important thing about the study is that it reveals for the first time the real increases in mortality in areas and dates when the data quality is good."

The good news for transparency is that Spain's government is all over this kind of analysis and data. It's great. Their ministry of health, ISCiii, is publishing daily updates with detailed graphs with the realtime data they have. The bad news is that reporting for Madrid and some big provinces seem stuck and totally broken — realtime reporting doesn't help when the data is totally incomplete.

However, they do have a thorough daily update, that shows total mortality vs. previous years as well as some basic gender/age brackets.

Since the epidemic has progressed furthest in Italy out of western countries, some of the most striking data are from there.

The breakthrough analysis was from Eco Di Bergamo, a local paper in the hardest-hit region. In their study, Bergamo province had 5400 March deaths compared to 900 last year, giving us 5x more excess deaths from COVID than all other causes combined. But, the official death tally for the region stands at

2060, **2.18x too low** than these 4500 excess deaths. The paper explains that most of these uncounted deaths occurred at home, or in assisted living facilities and were recorded as deaths, but not COVID.

In a somewhat grim piece of "positive" news from that analysis, the researchers also estimate that almost 50% of the province might've already been infected at this point. The methodology seems like a rough cut — but, if true, it suggests Bergamo is close to herd immunity and the death toll doesn't have more that much more room to increase from here.

The Economist expanded on that today:

The Italian Ministero della Salute is now starting to issue some transparent data. Here is their full detailed bulletin for March 2020. It has a comparison of the last 4 flu seasons for the 65+ age group. This year started way below for some reason (low flu season?) but is now spiking through to a record.

In France, there is similar data emerging now that total deaths are producing noticeable spikes in overall mortality.

A few days ago from Le Monde we have a graph of increases in the provinces through March 16th. The problem is that Mar 16 data is pretty useless — in those two weeks, deaths have gone from 148 total to 5400. So we will need analysis more recent data to really see how things are doing.

But more recent data is coming. The French national statistics bureau "Insee" now has data available through Mar 27th. I did a little basic summations in Excel, and the data show 2786 excess deaths through Mar 27, compared to the average of the last two years. This compares to the official Mar 27 death toll of 1995, leaving around 800 deaths unaccounted for through that date: **+40% more,** similar to the Spanish data.

Same delay issue here — there were more than 3000 official deaths in the last 5 days alone, so this is not an up-to-date picture. We will know a lot more in the next few days — I fear that this gap between official numbers and excess mortality might increase above the current 40%.

Update Apr 8: FT has analyzed the UK Office of National Statistics release from Tue Apr 7. They find that "Figures show 1,649 people had died in English hospitals by 27 March, **78% higher than previously stated**". Total mortality figures suggest a **61% undercount** on a naive reading.

However, this is a debatable headline number. It's not total mortality — it's just seizing on a disagreement between how the National Health Service (NHS) coded their deaths and the criteria the government had for attributing a death to COVID:

However, the raw data is available on the ONS site. Here's the key data table updated to Mar 27.

Worldometers shows 582 official death toll in Week13 in the UK (177 → 759 total). Meanwhile, on total mortality, 2020w12 was +72 vs same week in 2019; and 2020w13 was +1011, a +939 growth in trend. If accurate, this shows a **potential 61% undercounting** but it's a very simplistic analysis.

I was worried how much this could just be normal variance, so I pulled the last 5 years of data and looked at variance. Highlighted in yellow are all cells with a sizable # of deaths. Overall, standard deviation is 2.6% and we have a 10% move, 3.9 z-score. So while it's still possible that this is noise, it seems extremely unlikely.

There's also some added credibility that it's not just variance given that in the last 5 years, deaths are on a decreasing trend week 10 to week 13 (11498 → 11205 → 10573 → 10130). And, until week 11, 2020 was actually a normal, slightly good year (139k deaths vs 144k avg) and following the usual trends.

Next week will put this question to rest, since official death toll for week 14 was 2846 (almost 5x higher than week 13). If these numbers were accurate, we'll see roughly that increase in total mortality. But, if they're undercounted we will see much more — my current guess is roughly 40% over like in other countries, so I fear we may see something like ~4000 excess deaths next week, and I dearly hope I'm wrong.

Week 14 data is released by the ONS next Tuesday, Apr 14.

The UK has a fantastic data set, with great data quality, and demographic breakdowns which will be super illuminating (gender, age). But it's on a significantly longer delay than other European countries above. They publish detailed updates weekly 10 days after the end of every week:

For now, however, their latest update is as of Mar 20, when the UK had 177 total deaths. And, unsurprisngly, nothing really shows up in these weekly figures yet. We will know in two weeks what the true effects are up to today: there were unfortunately 582 additional deaths 3/20 to 3/27, which should show up in the numbers; and further ~2300 deaths or so this week which will definitely be a huge spike.

The EU has an good overall dashboard on total mortality called “EuroMoMo”.

But, it seems it’s on quite a delay (a couple weeks?). In their latest update, we’re starting to see small spikes in Italy, France, UK, Switzerland, Spain.

Very interestingly, deaths are significantly down in several other countries, presumably due to some social distancing preventing other deaths and a relatively small COVID epidemic for now. This will be important to watch over the next few weeks.