Recommendations for Improving the United States Centers for Disease Control (CDC) Data Practices for Pneumonia, Influenza, and COVID-19 (v 1.1)

John F. McGowan[1¶*], Tam Hunt[2¶], and Josh Mitteldorf[3]

[1] Mathematical Software Inc., Sunnyvale, California, United States

[2] Department of Psychological and Biological Sciences, University of California at Santa Barbara, California, United States

[3] Independent Researcher, Philadelphia, Pennsylvania, United States

* Corresponding author:

E-Mail: ceo@mathematical-software.com

[¶] These authors contributed equally to this work

Abstract

During the pandemic, millions of Americans have become acquainted with the CDC because its reports and the data it collects affect their day-to-day lives. But the methodology used and even some of the data collected by CDC remain opaque to the public and even to epidemiologists. In this paper, we highlight areas in which CDC methodology might be improved and where greater transparency could lead to broad collaboration. (1) "Excess" deaths are routinely reported, but not "years of life lost", an easily-computed and more granular datum that is important for public policy. (2) What counts as an "excess death"? The method for computing the number of excess deaths does not include error bars and we show a substantial range of estimates is possible. (3) Pneumonia and influenza death data on different CDC pages is grossly contradictory. (4) The methodology for computing influenza deaths is not described in sufficient detail that an outside analyst might pursue the source of the discrepancy. (5) Guidelines for filling out death certificates have changed during the COVID-19 pandemic, preventing the comparison of 2020-21 death profiles with any previous year. We conclude with a series of explicit recommendations for greater consistency and transparency, and ultimately to make CDC data more useful to the public and epidemiologists and other scientists.

I. Introduction

The United States Centers for Disease Control (CDC) was tasked with a wide array of data tracking and policy recommendations during the course of the COVID-19 pandemic. Many choices were made under extreme time pressure, and CDC personnel did the best they could given the conditions they were tasked with. As a result, a number of CDC practices since the start of the pandemic in early 2020 have not followed common scientific and engineering practice. However, several problems with data presentation and analyses for pneumonia and influenza predate the pandemic.

Common scientific and engineering practices are designed to prevent serious errors and minimize faulty results due to cognitive biases[4,5,6]. Proper use of significant figures and reporting of statistical and systematic errors is generally required for most peer-reviewed journal publications, Ph.D. dissertations, and other scientific and engineering publications. During times of crisis, common scientific and engineering practice should be followed rigorously and uniformly to minimize the chances of serious errors.

For example, CDC analyses and data presentations for pneumonia, influenza, and COVID-19 frequently do not follow common scientific and engineering practice for proper use of significant figures[7,8,9,10,11,12], reporting of statistical and systematic errors[13], clear and consistent definitions of measured quantities, or transparency and reproducibility[14,15,16,17,18,19,20].

This omission of common scientific and engineering practices raises questions about the accuracy of the CDC's data, conclusions, and public health policies in a number of important areas, including the COVID-19 pandemic. These issues may undermine public confidence in the CDC and public health policies if not corrected.

These issues are sometimes shared with other government agencies such as the US Social Security Administration (SSA) and US Census Bureau that work closely with the CDC[21].

As another example, death counts for both individual causes and “all cause” deaths are frequently reported as precise to the last digit without any statistical or systematic errors, despite both known and unknown uncertainties in counting deaths, such as missing persons, unreported deaths due to deceased payee fraud[22], the ~1,000 living Americans incorrectly added to the government Deaths Master File (DMF), each month, for unknown reasons[23,24], considerable uncertainties in assigning the underlying cause of death (UCOD) by coroners and doctors[25,26,27,28,29,30,31,32,33], and other issues.

Similarly, raw counts, adjusted counts, and estimates – often based on incompletely documented computer mathematical models – are often not clearly identified as such. The Deaths Master File, with names and dates of death of deceased persons is exempt from the Freedom of Information Act (FOIA) and unavailable to the general public, independent researchers, and even other government agencies such as the IRS. This confidentiality of data makes independent verification of many CDC numbers, such as the excess deaths numbers tracked during the COVID-19 pandemic, all but impossible.

"Excess" deaths are routinely tracked by CDC, but not "years of life lost" (YLL), an easily-computed and more granular datum that is important for public policy.

This article gives more detail on specific examples of failures to follow common scientific and engineering practice, and related data and policy questions. We conclude with recommended improvements to the CDC's data practices, to improve quality and increase public confidence in the data, analysis, and public health policies where warranted. We review a number of examples in the following sections.

II. Discrepancies in tracking pneumonia and influenza deaths

One of the most striking examples is significant differences in the number of deaths attributed to “pneumonia and influenza” on the CDC FluView website[34] (~188,000 per year), the leading causes of death report[35] (~55,000 per year), and the CDC Excess Deaths website[36] (~55,000 per year). The discrepancy between the FluView website and the leading causes of death report predates the COVID-19 pandemic by several years. It seems likely the weekly pneumonia and influenza death numbers reported on the CDC Excess Deaths website – added during the COVID-19 pandemic – are derived from the same underlying data as the leading causes of deaths reports.

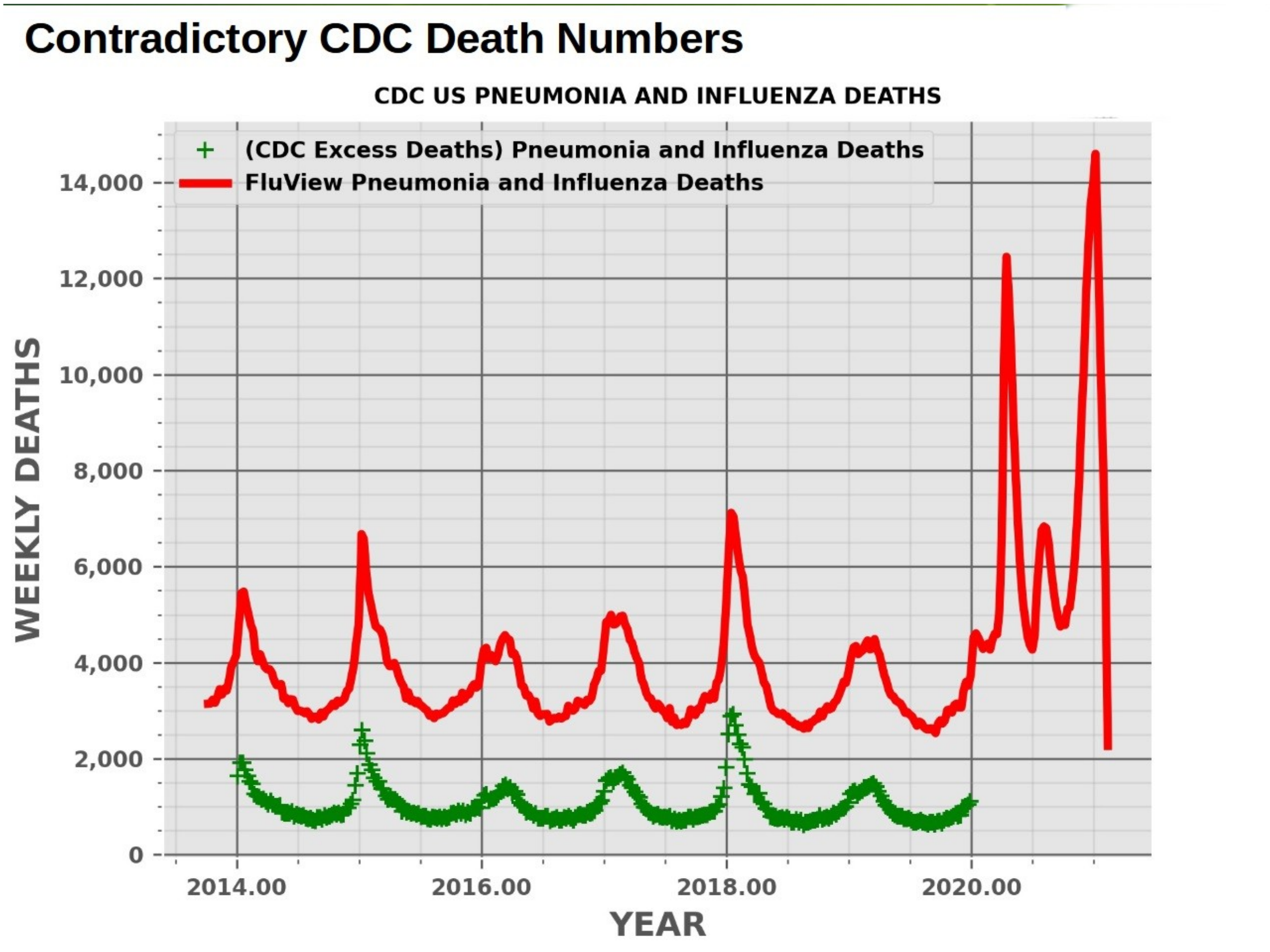


Figure 1: CDC's Contradictory Pneumonia and Influenza Death Numbers, with CDC's excess deaths data showing significantly less than the FluView data. (Our plot of CDC data.)

The CDC FluView website shows that 6-10 percent of all deaths, varying seasonally, are due to pneumonia and influenza (P&I) according to the vertical axis label on the FluView *Pneumonia & Influenza Mortality* plot. The underlying data files from the National Center for Health Statistics (NCHS) list, as mentioned, ~188,000 deaths per year attributed to pneumonia and influenza.

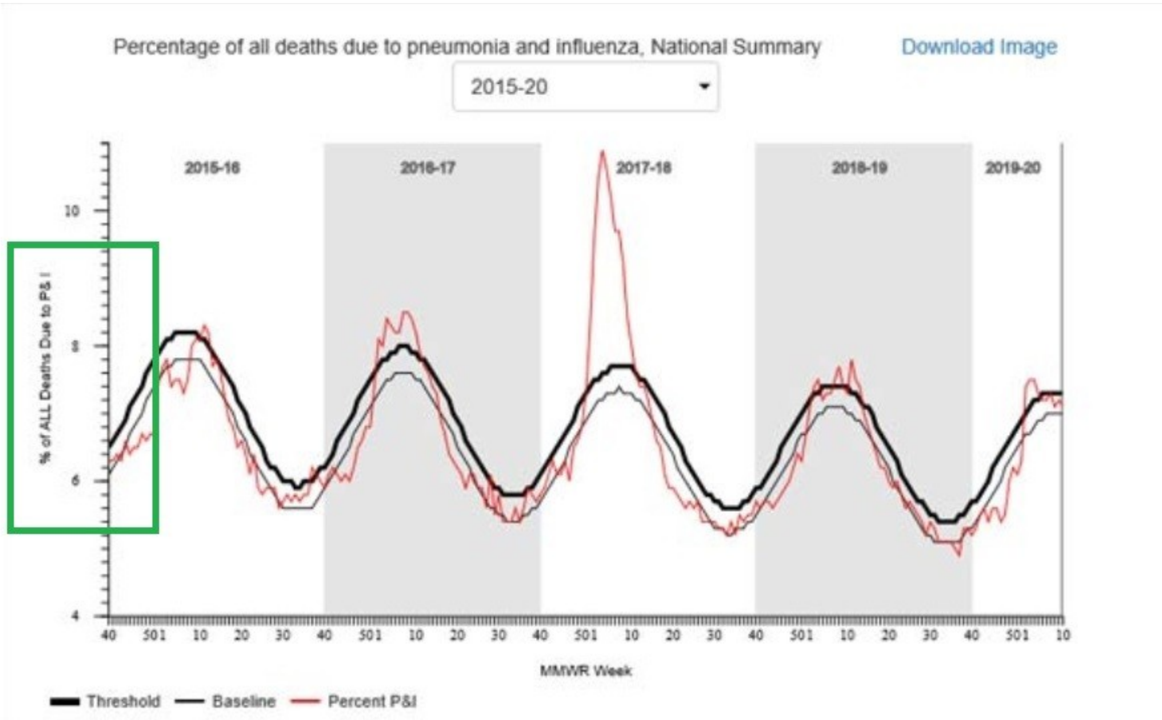


Figure 2: US Centers for Disease Control (CDC) FluView Pneumonia & Influenza Mortality Plot (June 9, 2021)

NOTE: https://www.cdc.gov/flu/weekly/fluviewinteractive.htm and click on *P&I Mortality Tab*

The CDC FluView graphic and underlying data files list no statistical or systematic errors. The counts of deaths in the data files give the numbers to the last significant digit, implying an error of less than one count, one death, based on common scientific and engineering practice.

In contrast, the CDC's leading causes of death report Table C, *Deaths and percentage of total deaths for the 10 leading causes of death: United States, 2016 and 2017* on Page Nine (see Figure 3) attributes *only 2 percent* of annual deaths (about 55,000 in 2017) to "influenza and pneumonia."

The difference between the CDC FluView and leading causes of death report numbers seems to be due to the requirement that pneumonia or influenza be listed as "t*he underlying cause of death*" in the leading causes of death report and only "*a cause of death*" in the FluView data. This is not, however, clear. Many deaths have multiple "causes of death." The assignment of an "underlying cause of death" may be quite arbitrary in some or even many cases. Despite this, none of these official numbers, either in the leading causes of death report or the FluView website, are reported with error bars or error estimates, as is the common scientific and engineering practice when numbers are uncertain. The leading causes of death report for 2017 reports *exactly* 55,672 deaths from "influenza and pneumonia" in 2017 with no errors– as shown in Figure 1.

National Vital Statistics Reports, Vol. 68, No. 6, June 24, 2019 **9**

Table C. Deaths and percentage of total deaths for the 10 leading causes of death: United States, 2016 and 2017

[An asterisk (*) preceding a cause-of-death code indicates that the code is not included in the *International Classification of Diseases, 10th Revision* (ICD–10)]

Cause of death (based on ICD–10)	Rank[1]	2017 Deaths	2017 Percent of total deaths	2016 Deaths	2016 Percent of total deaths
All causes	...	2,813,503	100.0	2,744,248	100.0
Diseases of heart (I00–I09,I11,I13,I20–I51)	1	647,457	23.0	635,260	23.1
Malignant neoplasms (C00–C97)	2	599,108	21.3	598,038	21.8
Accidents (unintentional injuries) (V01–X59,Y85–Y86)	3	169,936	6.0	161,374	5.9
Chronic lower respiratory diseases (J40–J47)	4	160,201	5.7	154,596	5.6
Cerebrovascular diseases (I60–I69)	5	146,383	5.2	142,142	5.2
Alzheimer disease (G30)	6	121,404	4.3	116,103	4.2
Diabetes mellitus (E10–E14)	7	83,564	3.0	80,058	2.9
Influenza and pneumonia (J09–J18)	8	55,672	2.0	51,537	1.9
Nephritis, nephrotic syndrome and nephrosis (N00–N07,N17–N19,N25–N27)	9	50,633	1.8	50,046	1.8
Intentional self-harm (suicide) (*U03,X60–X84,Y87.0)	10	47,173	1.7	44,965	1.6

... Category not applicable.

[1]Based on number of deaths.

SOURCE: NCHS, National Vital Statistics System, Mortality.

Figure 3: CDC's leading causes of deaths report suggests accuracy of death counts to the single digit level, with no error bars or uncertainties reported.

Death certificates frequently have multiple causes of death. One of these is assigned as the underlying cause of death. This may be quite arbitrary in some cases. Indeed, the concept of "underlying cause of death" may not be well defined for some deaths because elderly patients will often develop multiple health problems in parallel that are fatal either in combination or due to one of the comorbidities reaching a level of severity sufficient to induce death. (See the discussion of the uncertain assignment of the underlying cause of death for deaths where pneumonia is present or a cause of death in the CDC's Medical Examiners' and Coroners' Handbook on Death Registration and Fetal Death Reporting (2003 Revision)[37] and Randy Hanzlick's *Cause of Death and the Death Certificate: Important Information for Physicians, Coroners, Medical Examiners, And the Public, Randy Hanzlick Editor (2006), College of American Pathologists* [38] below for examples of this problem.)

In contrast, the FluView site, with a much larger number of deaths, appears to count deaths where pneumonia or influenza is listed as "*a cause of death*," even if it is not the "*underlying cause of death.*" The FluView website and the leading causes of death report use semantically equivalent names for the two grossly different numbers: "influenza and pneumonia" in the leading causes of death report and "pneumonia and influenza" in the FluView website graphics and text. There is no indication in the graphs, tables, or immediately adjacent text that they are different values.

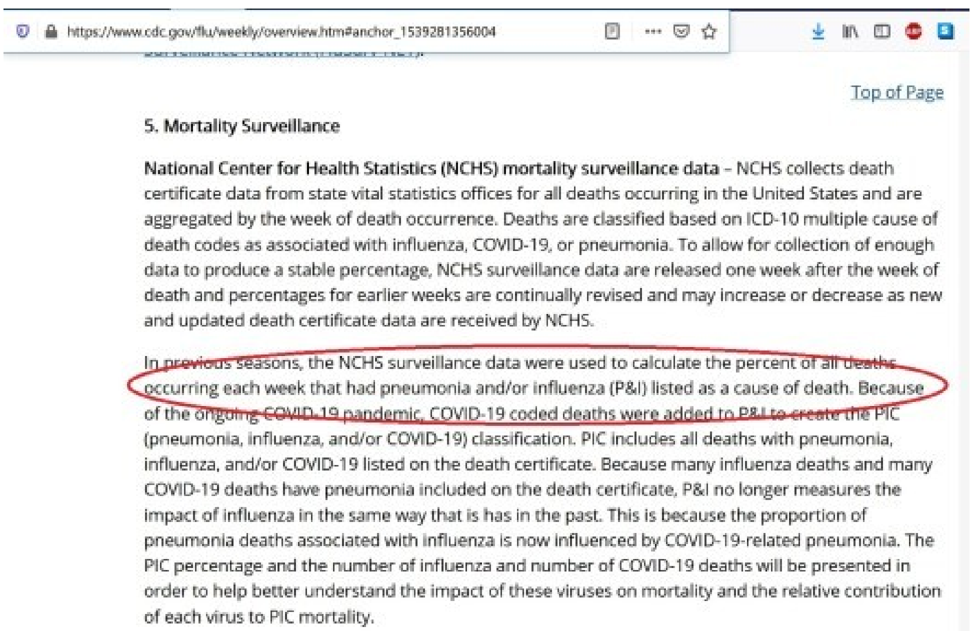
Both of these sources, especially the FluView website, are intended for the public, busy health professionals, policy makers and others, all of whom have limited time or knowledge to decipher the technical notes provided by CDC and whose confidence in these numbers may be significantly diminished if they notice the gross discrepancy in these two sets of numbers that are not clearly distinguished.

In "Peer Review in Scientific Publications: Benefits, Critiques, and A Survival Guide" Kelly et al note in their section on "Common Errors in Scientific Papers"[39]:

Another common fault is the author's failure to define terms or use words with precision, as these practices can mislead readers.

The scientific and medical distinction between the numbers is substantial if the FluView website is listing deaths where "pneumonia and influenza" are only "a cause of death." The FluView numbers likely include large numbers of deaths of persons with chronic obstructive pulmonary disease (COPD), mostly late-stage chronic bronchitis and emphysema, a terminal condition, as well as other often terminal conditions, who are much more likely to die from a respiratory infection than most healthy persons – presumably the "influenza and pneumonia" deaths listed in the leading causes of death report.

Note that the label on the vertical axis of the FluView graph (Figure 2) uses the language "% of All Deaths *Due to* P&I" – where P&I is an abbreviation for "pneumonia and influenza" – not "Deaths *Involving* P&I" or "Deaths *With* P&I." There is no suggestion of any difference between these quite divergent mortality figures.


https://www.cdc.gov/flu/weekly/overview.htm#anchor_1539281356004

Top of Page

5. Mortality Surveillance

National Center for Health Statistics (NCHS) mortality surveillance data – NCHS collects death certificate data from state vital statistics offices for all deaths occurring in the United States and are aggregated by the week of death occurrence. Deaths are classified based on ICD-10 multiple cause of death codes as associated with influenza, COVID-19, or pneumonia. To allow for collection of enough data to produce a stable percentage, NCHS surveillance data are released one week after the week of death and percentages for earlier weeks are continually revised and may increase or decrease as new and updated death certificate data are received by NCHS.

In previous seasons, the NCHS surveillance data were used to calculate the percent of all deaths occurring each week that had pneumonia and/or influenza (P&I) listed as a cause of death. Because of the ongoing COVID-19 pandemic, COVID-19 coded deaths were added to P&I to create the PIC (pneumonia, influenza, and/or COVID-19) classification. PIC includes all deaths with pneumonia, influenza, and/or COVID-19 listed on the death certificate. Because many influenza deaths and many COVID-19 deaths have pneumonia included on the death certificate, P&I no longer measures the impact of influenza in the same way that is has in the past. This is because the proportion of pneumonia deaths associated with influenza is now influenced by COVID-19-related pneumonia. The PIC percentage and the number of influenza and number of COVID-19 deaths will be presented in order to help better understand the impact of these viruses on mortality and the relative contribution of each virus to PIC mortality.

Figure 4: FluView Mortality Surveillance notes with "A Cause of Death" Language Circled in Red (Dec. 18, 2020)

III. The CDC Influenza Virus Deaths Model

The CDC uses an incompletely documented mathematical model that attributes roughly 55,000 deaths from pneumonia and influenza to the influenza virus as the underlying cause of death, a number roughly comparable to the total pneumonia and influenza deaths in the leading causes of death data[40]. The presence of the influenza virus is confirmed by laboratory tests, however, in only a small fraction of pneumonia and influenza deaths, ~6,000 per year in most years.

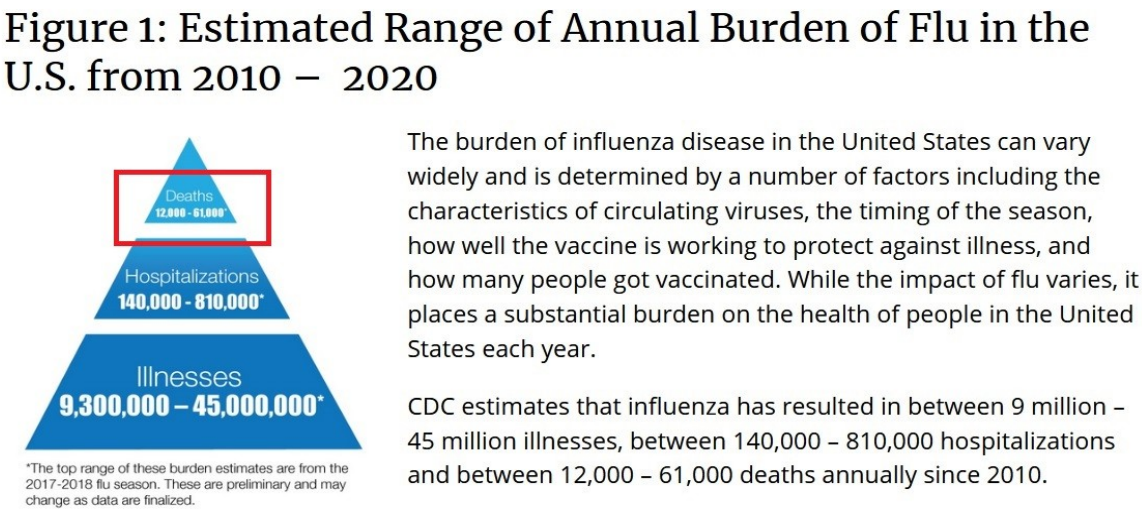


Figure 5: The US CDC Attributes 12,000 to 61,000 Pneumonia Deaths to Influenza on Their website: https://www.cdc.gov/flu/about/burden/

Although the language is often unclear in the CDC documents and websites, the CDC appears to claim that there is substantial under-testing for the influenza virus (see the discussion of the influenza deaths model below) and that an initial influenza infection, which often disappears or becomes undetectable in laboratory tests, leads to the subsequent pneumonia, presumably a bacterial pneumonia, although other viruses would be consistent with some lab tests. Based on this argument, the CDC appears to attribute most pneumonia deaths where, historically, pneumonia was listed as the *"underlying cause of death,"* to the influenza virus — even though laboratory tests frequently fail to confirm influenza or even detect other viruses or bacteria as the cause of death instead of influenza. The "underlying cause of death" issue is discussed in more detail below.

As shown in Figure 5 above, the CDC website Disease Burden of Influenza appears to give a range from 12,000 to 61,000 influenza deaths from this model. The graphic does not indicate if this range is a 95 percent confidence interval — another common scientific and engineering practice — or some other error estimate. The range in the graphic does not appear to match any of the 95 percent confidence levels for estimated deaths attributed to influenza in Table 1 on the CDC Disease Burden of Influenza website.

The website does not provide the source code for the model, nor the data used to produce the model except for the seasons 2010-2011 and 2011-12 provided in the single reference cited. The model was apparently implemented in the proprietary and quite expensive SAS statistics tool based on references to use of the freely available SAS macro BETABIN for fitting a beta-binomial distribution to the data. We did not find any goodness of fit statistics for the beta-binomial model. For example, supplemental figure 1 shows plots of the fitted beta distributions, no error bars on the fitted models, and no goodness of fit statistics or tests:

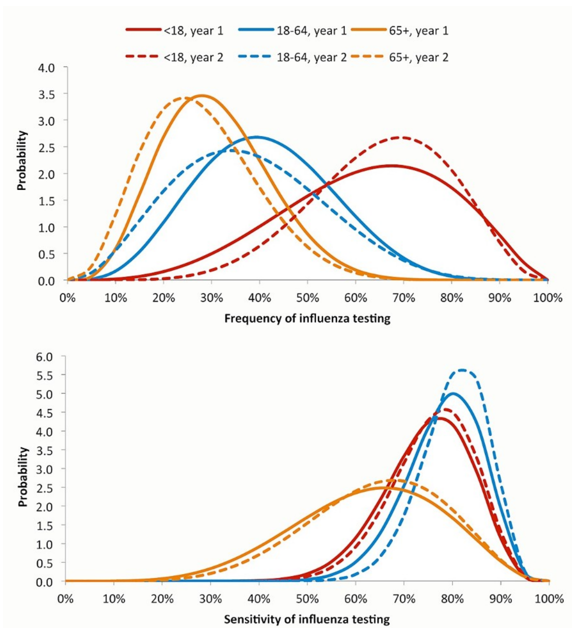


Figure 6: Beta-binomial probability distributions of the summary proportion of patients tested for influenza and sensitivity of influenza testing across six FluSurv-NET sites, by age group and year. *https://doi.org/10.1371/journal.pone.0118369.s001*

The beta distributions shown are fitted to data from *only six* sites in 2010-2011 with a total of 5,458 hospitalized patients and *five* sites in 2011-2012 with a total of 2,502 patients. In both seasons, all but one site in New York are in the western United States. The paper estimated 114,018–633,001 total hospitalizations per season for the 2010-11, 2011-12, and 2012-13 seasons. Thus the sample is a small fraction of the actual hospitalizations and testing for influenza. The beta distribution is a model of the distribution of fractions of patients with respiratory illnesses tested for the influenza virus at different sites.

The sites have widely differing frequency of influenza testing and significant variation in the sensitivity of the influenza testing, ranging from a high of 54% of adults aged 65+ at one site in California (with 1,049 patients) to a low of 18% of adults aged 65+ in New Mexico (with only 102 patients). These fitted beta binomial distributions then appear to be extrapolated nationwide to produce the estimates of influenza deaths which report wide 95% confidence levels. The model is used to adjust the reported deaths with laboratory confirmed influenza by a large multiplier:

$$Multiplier = \frac{1.0}{frequency of influenza testing \times influenza test sensitivity}$$

The paper notes (page 11/13 in the PDF version):

> Our analysis was subject to some limitations. First, we assumed that the probability of a person with influenza being tested for influenza was the same as all persons with a respiratory illness. **If physicians were more likely to recognize influenza patients clinically and select those patients for testing, we may have over-estimated the magnitude of under-detection. (Emphasis Added)**

The CDC's Cold Versus Flu web page (retrieved Sep 27, 2021) presents a graphic that seems to imply that a cold and flu (influenza virus) can be distinguished based on clinical symptoms, absent a diagnostic test[41]:

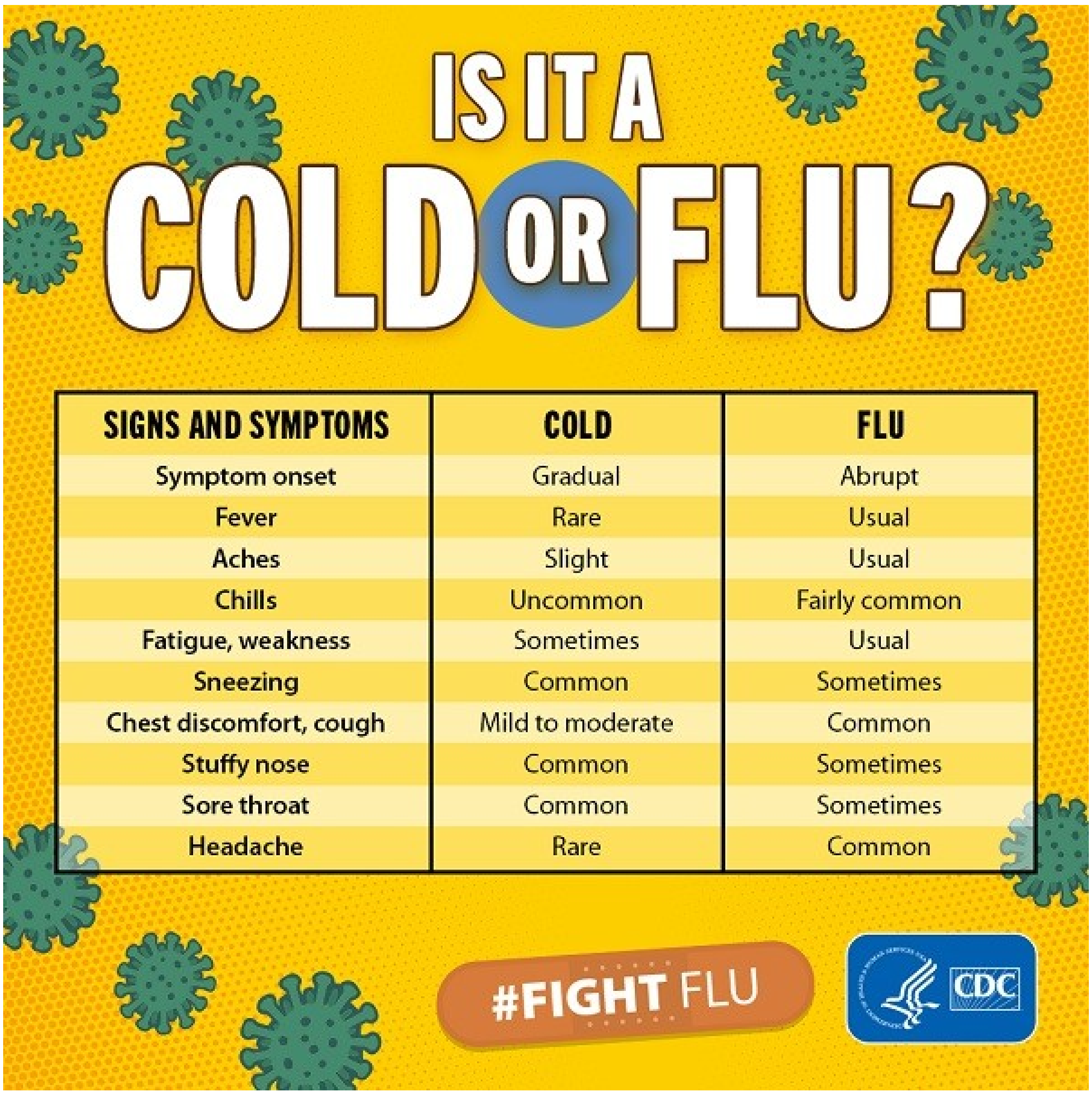


Figure 7: CDC Cold Versus Flu Graphic

We were unable to locate any obvious references for this graphic or any of the statements on the "Cold versus Flu" page which includes contradictory text next to the graphic:

> *Because colds and flu share many symptoms, it can be difficult (or even impossible) to tell the difference between them based on symptoms alone. Special tests can tell if a person is sick with flu.*

The influenza deaths model reference contains a remarkable and counter-intuitive statement with no reference or obvious source (also on page 11/13):

> *Likewise, our estimate of deaths may also be underestimated because we did not adjust for* ***the finding that patients who died in the hospital were less likely to have been tested for influenza than other hospitalized patients. (Emphasis Added)***

One might expect, however, that deaths would be more severe cases of pneumonia and influenza where doctors would order more tests.

There is a substantial history of serious criticism of the CDC's influenza death numbers by medical scientists and others[42,43,44]. One prominent critic is Peter Doshi, currently a professor at the University of Maryland and a senior editor at the British Medical Journal (BMJ). Citing the results of actual laboratory tests of deceased patients, critics of the CDC's flu death numbers such as Doshi have argued that pneumonia deaths are actually due to a range of different viruses, bacteria, other pathogens, and even toxins, rather than predominantly influenza, as implied by the CDC's influenza deaths model. The output of this model appears to be the basis of the baseline "flu" deaths numbers used in most popular and public policy discussions of COVID-19 deaths — although the leading causes of death report number may also be used.

CDC scientists have published rebuttals to some of Doshi's arguments[45]. The unresolved controversy illustrates the difficulties with using models instead of direct measurement, especially models that change consequential results by large factors rather than small few percent improvements in accuracy. We recommend reducing the use of models in this area as much as possible. Ideally, testing all patients with respiratory illnesses for influenza and other respiratory viruses is the preferred solution; improvements in PCR and other molecular technologies may make this feasible now or in the near future. In the short term, comprehensive influenza testing is probably not possible, but a better option is to randomly test symptomatic patients from a representative sample of the entire country for influenza and other respiratory viruses to determine the fraction with influenza and the fraction of those who die with influenza.

IV. CDC Excess Deaths Website Data Presentation and Analysis Issues

Turning to the COVID-19 pandemic data, the CDC Excess Deaths website presents an estimate of the excess deaths due to the COVID-19 pandemic or the pandemic response – associated with COVID-19 in CDC language – based on a mathematical model, the Noufaily or “extended Farrington” model, developed and used for early epidemic detection by the UK Public Health Service[46]. The CDC’s website technical notes indicate the CDC has modified the Noufaily algorithm to “zero out” negative excess deaths in any categories – a statistically invalid procedure for estimating excess deaths that ensures that excess deaths will always be zero or positive even if the actual deaths are lower than expected based on historical deaths data – although this zeroing may be justified as a conservative measure for outbreak detection rather than evaluating the impact of the pandemic and the policy responses to the pandemic.

*Estimates of excess deaths for the US overall were computed as a sum of jurisdiction-specific numbers of excess deaths (**with negative values set to zero**), and not directly estimated using the Farrington surveillance algorithms. (CDC Excess Deaths website, Technical Notes, Retrieved June 7, 2021, emphasis added)*

One purpose of the excess deaths analysis is to verify that reported COVID-19 deaths are an actual increase in the all-cause mortality rate rather than relabeling of deaths due to other causes such as chronic obstructive pulmonary disease (COPD). In the absence of lockdowns, aggressive intubation,[47] and other novel responses to the COVID-19 pandemic, this would be a straightforward inference from a positive excess deaths value larger than the modeling error on the predicted/expected number of deaths from the Noufaily and other models – see the discussion of modeling below. In this context, the problem with the zeroing procedure seems clear. Consider the US has fifty state jurisdictions. For example, if there is no actual increase in the mortality rate between 2019 and 2020, the zeroing procedure can still produce a spurious estimate of increased mortality in 2020. There will be statistical fluctuations in the number of deaths in each state. With no overall increase in all cause mortality, about half the states will see more deaths in 2020 than 2019, ***balanced*** by declines in the number of deaths in the other states. If the negative “excess deaths” in these states with purely statistical declines in the number of deaths are set to zero, however, an overall positive excess death will be incorrectly reported because CDC’s current procedure doesn’t account for negative excess deaths in individual jurisdictions.

Note also that it is theoretically possible for a new virus to lower the all cause mortality rate if it out-competes and crowds out a more dangerous virus or viruses. It could, for example, become the immediate cause of death in COPD patients and yet lower the number of total deaths. In this case, most jurisdictions could show a decrease in deaths (negative excess deaths) but the zeroing procedure would still show positive excess deaths if some jurisdictions showed increases due to chance.

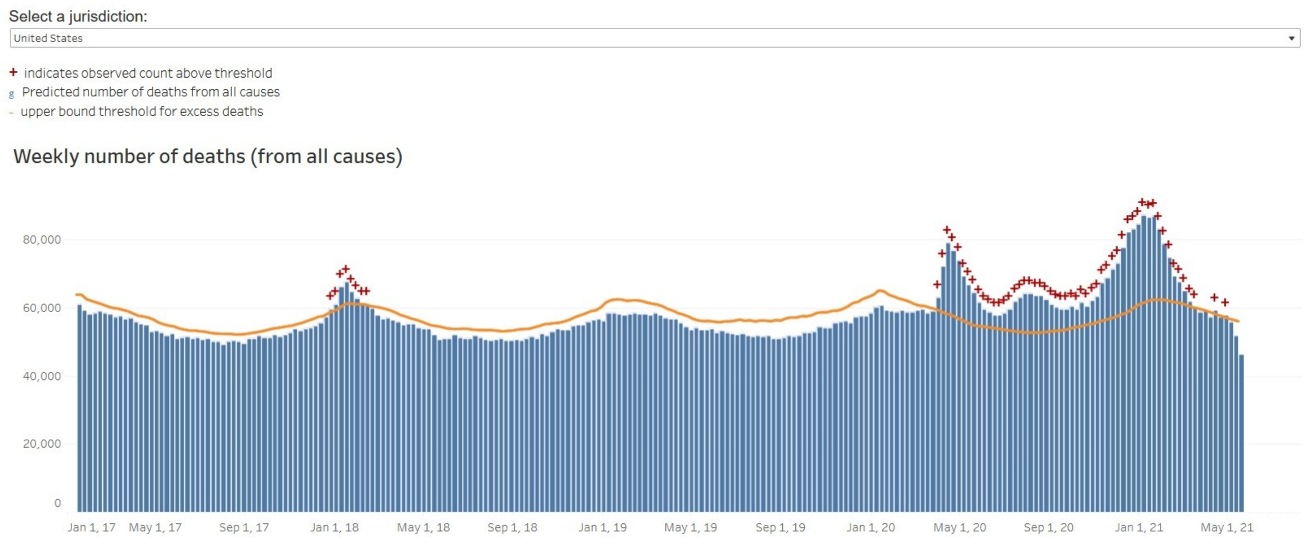


Figure 8: CDC excess deaths website is an interactive tool that allows various displays of data relevant to excess deaths since early 2020 in the U.S.

Note that the graph in Figure 8 can be confusing. The legend in the upper left corner (the blue “g”) seems to indicate that the blue bars are the predicted number of deaths from all causes according to the CDC’s Noufaily, “improved Farrington,” algorithm, but show spikes in the spring, summer, and fall of 2020 suggesting these are the actual weekly deaths during the pandemic. A model based on data before March of 2020 should resemble the beige line, showing a predicted drop in weekly deaths from all causes during the summer of 2020 and no spikes. The legend indicates that the red plus signs are the actual weekly deaths when these exceed the threshold. In common scientific and engineering practice, a plot will show both the model, meaning the predicted deaths, and the data for actual deaths, for the full range of the data – in this case January 2017 through May 2021.

The confusing “Predicted number of deaths from all causes” label refers to a second model used to adjust the weekly death counts for delays in receiving all death certificates based on past experience with the delays[48]. This is distinct from the Noufaily model used to predict expected deaths – the beige line – and compute the excess deaths. This is another example of confusing language on the CDC web site and in some documents where it is unclear what is actually meant. In our recommendations section, we suggest some practices to improve names, labeling, and avoid confusion between different models.

As noted previously, the data on the CDC excess deaths website provides a significantly lower historical (pre-2020) number of deaths attributed to “pneumonia and influenza” (~55,000 per year) than the FluView website (~188,000 per year).

The website does not report the coefficient of determination[49] (usually denoted R^2 or r^2 and pronounced “R squared” in statistics, sometimes denoted R**2 in plain text and statistical programming) or other goodness of fit statistics for their model, nor does it give any estimate or illustration of the systematic modeling error. It is common scientific and engineering practice to report a goodness of fit statistic, frequently the chi-squared statistic X^2 or the coefficient of determination R^2, for any models and rank the models by the goodness of fit statistic for comparison[50,51]. The goodness of fit statistic such as R^2 is itself an estimate, and errors on this measure, usually a 95% confidence interval, should also be reported.

We obtained the algorithm from CDC's GitHub and performed a series of sensitivity analyses under various data assumptions. Figure 10 below shows different possible results under the Noufaily algorithm without the CDC's inappropriate zeroing procedure and with different parameters and using simple alternative models. Our version of the Noufaily model finds about 411,000 excess deaths with the set of parameters that produces the best R^2 value of 0.94. There is an error on the computation of R^2 which is shown as a ninety-five percent confidence level range: 0.91 to 0.96. The largest and smallest number of excess deaths with R^2 in this range are also shown: about 390,000 deaths and 423,000 deaths. This is based on data from the FluView website downloaded on May 17, 2021, through the period ending January 1, 2021.

Note that CDC uses a different set of model parameters with a lower R^2 of about 0.74 (i.e. not as good a fit) to produce their estimate of ~500,000 excess deaths in 2020[52]. The CDC parameters are shown in the white line in Figure 10 below. These results and graph are presented as an illustration of the excess deaths data analysis and presentation that we recommend for the CDC excess deaths website and documents.

Annual deaths in the United States began to rise significantly from 2010 to 2017, at which time the decrease slowed dramatically, and almost stopped prior to the COVID pandemic in 2020. The 2010 to 2017 rise appears to reflect the aging and expected increase in mortality of the 1947-1964 “baby boom” generation. The flattening in 2017-2019 is unexpected and appears to reflect declining death rates, notably for heart and other blood coagulation related conditions, possibly due to reductions in risk factors and improved medical therapies.[53]

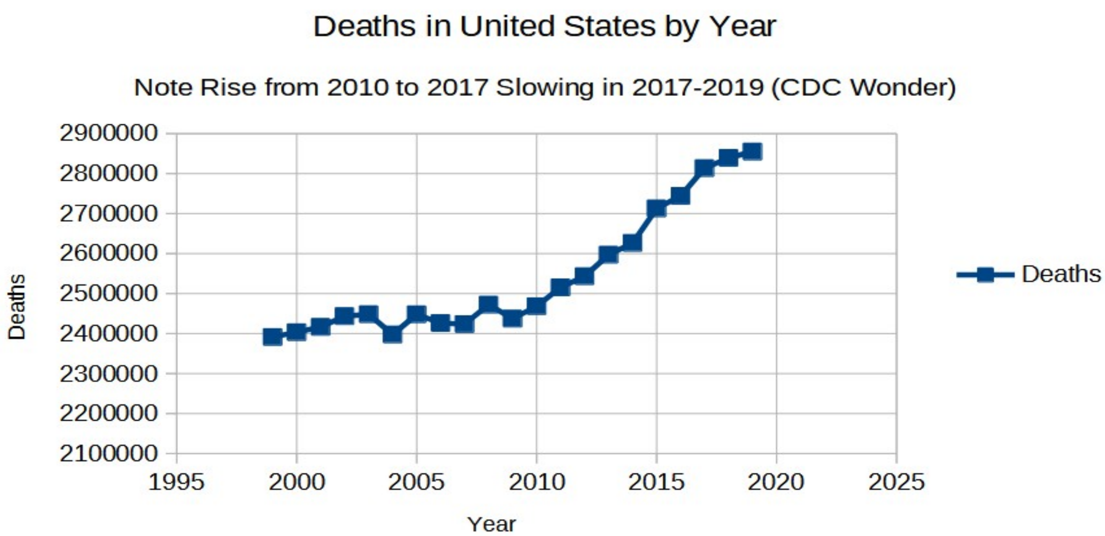


Figure 9: Annual Deaths in USA (CDC Wonder)

The Noufaily model and other simple trend detection models are unable to realistically model this complex evolution of mortality rates. However, the Noufaily model will more accurately match this behavior with the higher R^2 shown below with the longer lookback period of five or more years than the CDC’s default of four years. The shorter lookback period used by the CDC weights the slow, almost minimal growth in the death rate during the anomalous, unexpected 2017-2019 period.

We recommend the use of medically-based models that explicitly incorporate and model demographics and aging as well as trends in specific cause mortality rates such as the reported declining mortality from heart attacks for excess deaths modeling and calculations. See the recommendations sections at the end of this paper.

Figure 10 below shows the results of fitting the Noufaily algorithm in the R surveillance package with different parameters and two simple trend models implemented in Python to the FluView deaths data. The excess deaths, the coefficient of determination R^2 goodness of fit statistics, and the 95% confidence interval for R^2 are given for each model. For the Noufaily models, *b*, *w*, and *t* in the model name refer to key parameters of the model. The most consequential is *b*, the number of previous years used in the prediction as discussed above. The *noufaily_b4_w2_t2.58* white line model is the CDC’s choice of parameters. The FluView weekly death counts data are shown as black plus signs. The date in years is indicated on the horizontal axis.

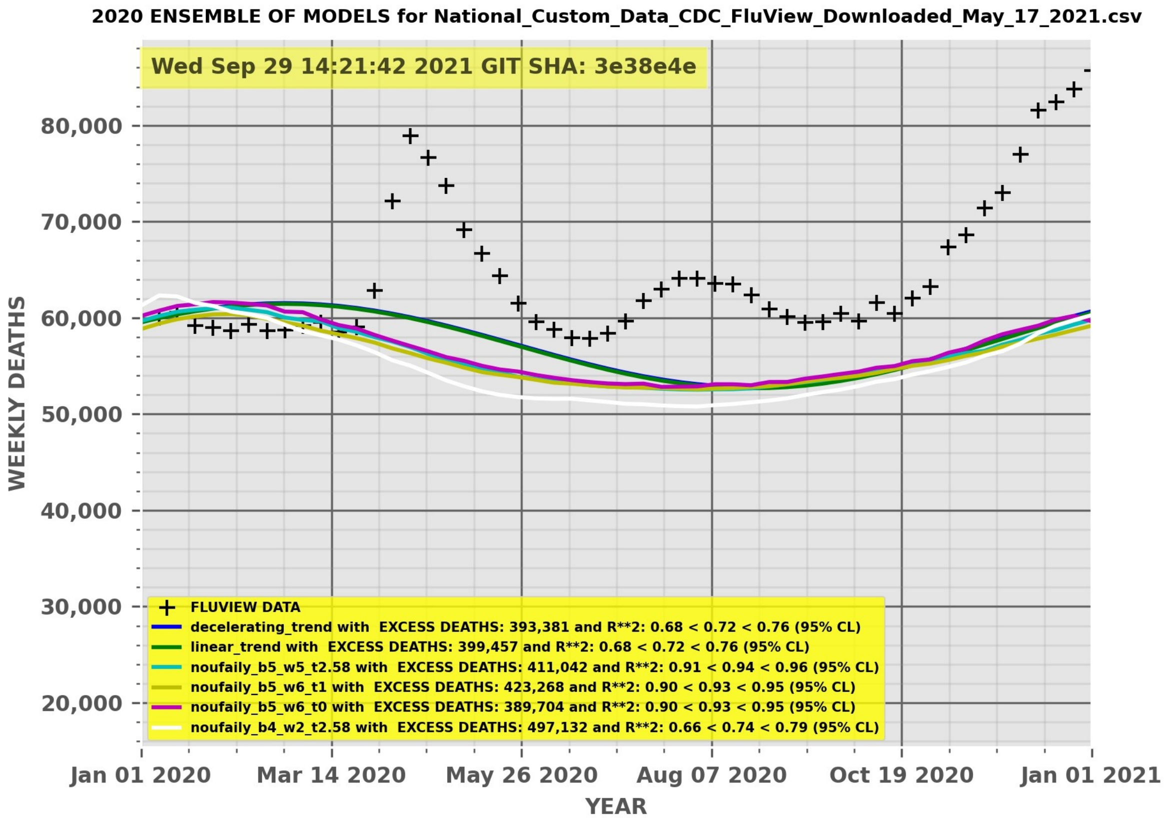


Figure 10: U.S. excess deaths using various statistical models, including Noufaily, with best fit parameters and Alternative Models, Feb. 1 2020-Jan. 1 2021.

V. Lack of Reproducibility of CDC Excess Deaths

It does not appear possible to independently reproduce the CDC excess deaths graph (Figure 8) or the numerical results from raw data such as actual death certificates. The full Deaths Master File (DMF) used by the Social Security Administration (SSA) is not public and not subject to the Freedom of Information Act (FOIA). Even most other government agencies, including the IRS, lack access to this data that includes the names and dates of deaths of all persons reported deceased to the US government[54].

The ostensible reason for this secrecy is that much of the data is reported to the CDC's National Center for Health Statistics (NCHS) by the vital registration offices (VRO's) of individual states and is considered property of the states and not the federal government. The federal government reportedly pays for limited access to this data, instead of general access for the government and general public, as transparency and scientific reproducibility would require.

The CDC provides data files that appear to contain de-identified information on each death on their website. Verifying these files requires the actual names, dates of death, and possibly other identifying information on the deceased persons. A complete verification of all deaths could involve substantial cost and time, but verification of a random sample of the reported deaths provides an affordable alternative. The CDC is not involved in collecting the Deaths Master File – a Social Security Administration project – which means the DMF provides an independent check on CDC tabulations[55].

VI. Including Years of Life Lost analysis alongside excess deaths analysis

Years of Life Lost (YLL) is a granular mortality impact measure[56] that considers age and comorbidities in relation to mortality. Excess deaths analysis, in contrast, does not consider age or comorbidities, just the number of deaths. The average age at death of U.S. COVID-19 victims is ~76[57] and the average comorbidities is ~4,[58] according to CDC data. ~38 percent of all U.S. COVID-19-related deaths occurred in nursing homes,[59] and an even higher proportion occurred in long-term care homes more generally (1.3 million people lived in skilled nursing homes and another 1.7 million in other assisted living and other long-term care[60]).

We note that the CDC Wonder database of deaths in the United States shows an average age of death of ~74 years in 2019, the year before the start of the COVID-19 pandemic, suggesting the YLL from COVID-19 may be quite small (COVID-19 average age of death, as just mentioned, was ~76).[61]

Methodology and assumptions are important for YLL analysis, and will affect outcomes significantly. Briggs et al. 2020 found, for example, a weighted mean of 7.33 YLL for COVID-19 deaths through July of 2020 in the United Kingdom, and 8.42 for the United States. Quast et al. 2021 found an average of 9.2 YLL for U.S. COVID-19 deaths in 2020. Both of these analyses are significantly larger than might be expected from the average age of death of COVID-19 victims. We updated Briggs et al.'s data with CDC's 4.0 average comordibities/additional causes of death (their analysis assumed just 2.0 average comorbidities) and this results in a weighted mean of 5.3 YLL for U.S. COVID-19-related deaths.

A YLL analysis is not as simple as counting deaths and age of death. A YLL analysis is also sensitive to assumptions about pre-existing conditions that generally shorten life expectancy such as obesity, diabetes, chronic obstructive pulmonary disease (COPD), and others common in COVID-19 victims. A proper YLL analysis should show the YLL results for different reasonable assumptions about pre-existing conditions, similar to the ensemble of models shown in Figure 10 for a simple excess deaths analysis.

In order to enable evaluation of the costs and benefits of the pandemic response, the CDC should compare the direct COVID-19 YLL to the YLL due to overdose deaths, homicides, suicides, and other deaths reasonably attributed primarily to the pandemic response (such as "lockdown" policies). For example, we calculate, based on an average age of death of ~43 years for overdose deaths[62], an average 36.8 YLL for overdose deaths (those living to 43 years old have an average of 36.8 additional years to live, based on the Social Security Administration actuarial life table; SSA 2020[63]).

Average age at death is even younger, at ~30 for 2019 homicide deaths.[64] Average YLL for these homicide deaths is significantly higher than overdose deaths, at 49.8. These non-COVID-19 YLL figures are significantly higher than COVID-19 average YLL figures (in the middle or high single digits in the various analyses mentioned) because the age of death is so much younger for these other causes of death.

Figure 10 shows a sharp increase, the highest on record at over 30% annually, from 70,357 overdose deaths in the 12 months preceding November 2019, to over 93,000 overdose deaths in 2020 (and still rising through February 2021, to over 95,000, which is the extent of the data available as of September 2021).[65] Based on these trends, we estimate conservatively 22,000 excess overdose deaths for the full year 2020.

There were ~10,000 excess homicides for 2020 through the third quarter[66] (figure 11), for a preliminary total of ~32,000 excess overdoses and homicides that correlate with the pandemic in 2020.

Figure 10. *Preliminary drug overdose death trends in the U.S., through Nov. 2020 (Source: NVSS 2021). Note that the trend was level for 2018 and 2019 until the end of 2019, pre-pandemic, but started rising steeply at the start of the pandemic and related "lockdowns."*

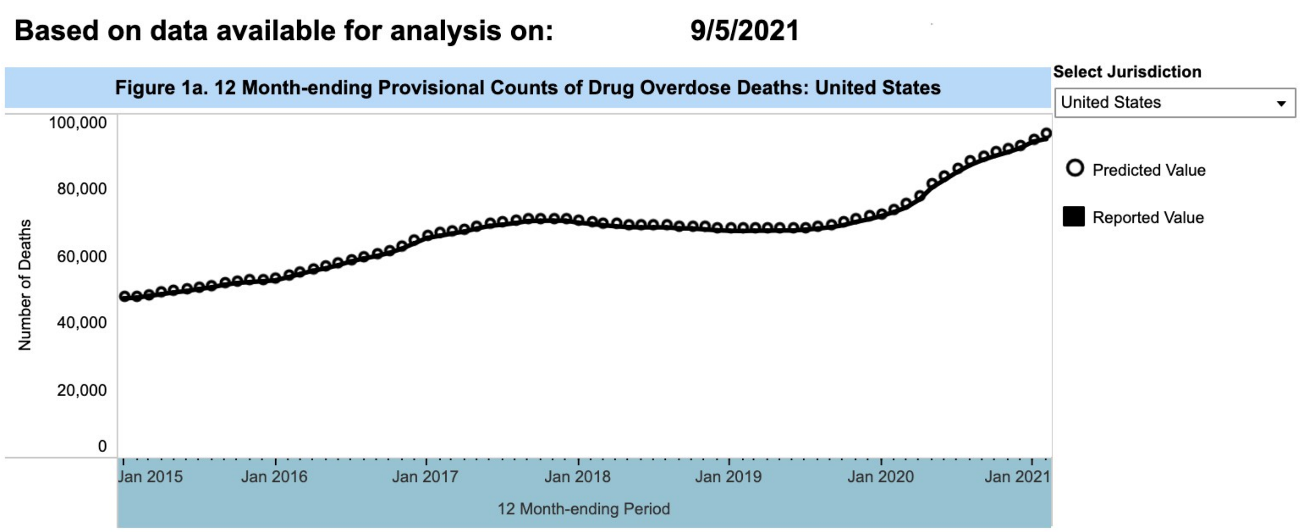


Figure 11: US Overdose Deaths in 2020 compared to previous years (Source: Ahmad et al. 2021)

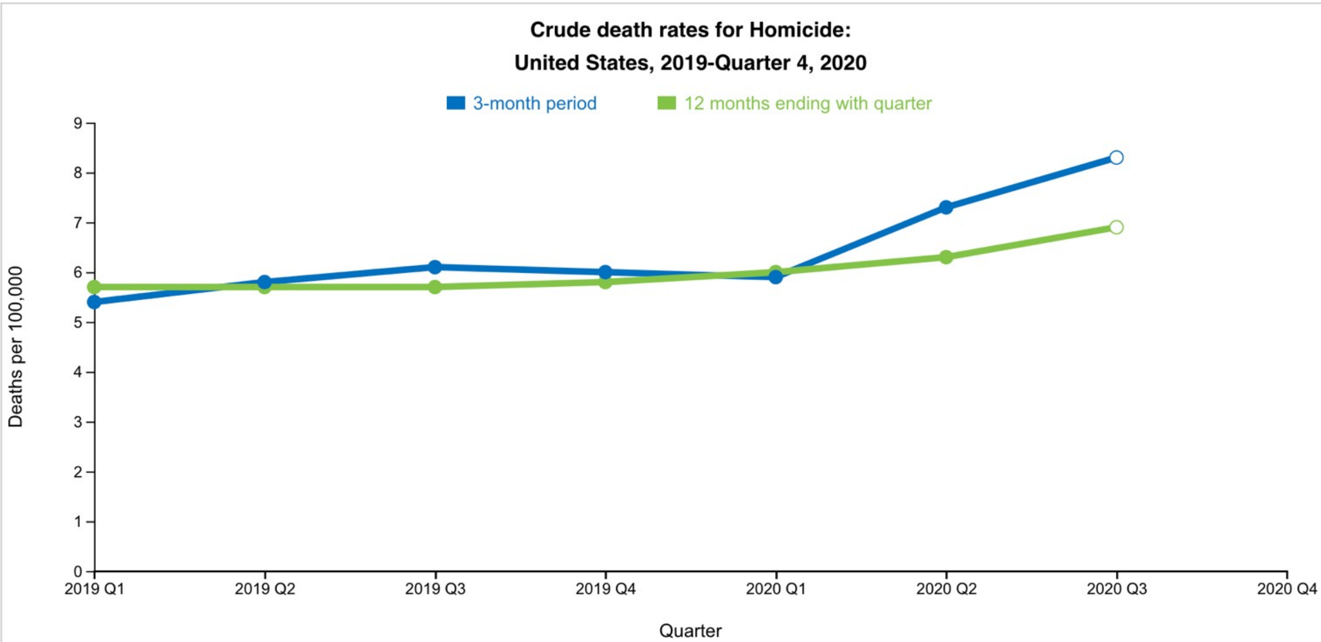


Figure 12: US Homicide Deaths in 2020 Compared to 2019

Using this ~32,000 excess overdose deaths and homicides in 2020 yields ~1.3 million total YLL for just these two categories of non-COVID-19 excess deaths.

Due to the high impact on YLL from pre-existing conditions that shorten life expectancy and from causes of death like overdoses and homicides that affect younger people at a higher rate, it is highly important to include COVID-19 YLL figures alongside, or possibly instead of, excess deaths figures, due primarily to the higher granularity of the YLL measure.

VII. Changing Death Certification Guidelines

During the COVID-19 pandemic the CDC (through its the National Vital Statistics System or NVSS) adopted new death certification guidelines, and related practices, in ways that appear inconsistent with prior practice, and without soliciting public review or comment on these very significant changes (*see, e.g., Florida v. Becerra 2021,* finding that CDC acting in an "arbitrary and capricious" manner in imposing cruise ship restrictions without adequate notice and review[67]). These changes in death certification guidelines, and related coding practices by CDC, make comparing historical (pre-2020) pneumonia and influenza death numbers with COVID-19 pandemic numbers difficult or impossible. It also makes highly important public health policy decisions largely immune from public review and comment.

The Rules for Assigning the Underlying Cause of Death Before COVID-19

Prior to 2020 and COVID-19, most pneumonia deaths did ***not*** list pneumonia or the pneumonia-causing pathogen, if known, as the underlying cause of death. This will be discussed in detail below. The only common *partial* exception was HIV/AIDS where *pneumocystis carinii pneumonia* (a common fungus) was often the *immediate cause of death* and the Human Immunodeficiency Virus (HIV) is almost always listed as t*he underlying cause of death.*

However, HIV is *not the pneumonia-causing pathogen*, which is the *pneumocystis* fungus. Instead, most pneumonia deaths, those included in the FluView numbers but *not* included in the leading causes of death numbers, were attributed to a cause such as a chronic lower respiratory disease, heart disease, cancer, even accidents, and other usually pre-existing conditions as the *underlying cause of death*.

The CDC follows the World Health Organization (WHO)'s definition of the underlying cause of death. WHO defines *the underlying cause of death* as **"the disease or injury which initiated the train of morbid events leading directly to death, or the circumstances of the accident or violence which produced the fatal injury"** in accordance with the rules of the International Classification of Diseases (ICD).[68] In the United States, the *underlying cause of death* is listed at the bottom of the list of causes of death in part I of the death certificate. The *immediate cause of death* is listed first. Part 2 lists other conditions that are considered contributing factors but not implicated in the causal chain leading to death. Pneumonia is often the immediate cause of death in part 1 of the death certificate.

In principle, death certificates and the assignment of causes of death, including the underlying cause of death, is governed or at least guided by the CDC's Medical Examiners' and Coroners' Handbook on Death Registration and Fetal Death Reporting (2003 Revision)[69]. This one-hundred and thirty-eight (138) page manual provides, however, limited guidance on how to assign the underlying cause of death in cases where pneumonia is present. Page 17 of the document contains the only detailed discussion of deaths involving pneumonia, as follows:

> Multiple conditions and sequences of conditions resulting in death are common, particularly among the elderly. When there are two or more possible sequences resulting in death, or if two conditions seem to have added together (e.g., stabbing caused both right intrathoracic hemorrhage and air embolism), choose and report in Part I the sequence or condition thought to have had the greatest impact (7). Other conditions or conditions from the other sequence(s) should be reported in Part II. For example, in the case of a diabetic male with chronic ischemic heart disease who dies from pneumonia, the medical examiner or coroner must choose the sequence of conditions that had the greatest impact and report this sequence in Part I. One possible sequence that the certifier might report would be pneumonia due to diabetes mellitus in Part I with chronic ischemic heart disease reported in Part II. Another possibility would be pneumonia due to the chronic ischemic heart disease entered in Part I with diabetes mellitus reported in Part II. Or the certifier might consider the pneumonia to be due to the ischemic heart disease that was due to the diabetes mellitus and report this entire sequence in Part I. Because these three different possibilities would be coded very differently, it is very important for the certifying medical examiner or coroner to decide which sequence most accurately describes the conditions causing death.

Figure 13: CDC Medical Examiner and Coroner's Handbook (2003) on pneumonia

Although the CDC's Medical Examiners' Handbook 2003 gives little specific direction on deaths involving pneumonia, it references several books and articles edited or authored by Randy Hanzlick, M.D., now retired Chief of the Fulton County Medical Examiner's Office and former pathologist with the CDC, including *Cause of Death and the Death Certificate: Important Information for Physicians, Coroners, Medical Examiners, And the Public, Randy Hanzlick Editor (2006), College of American Pathologists* [70](the reference seems to have been updated to the year 2006 since the original release of the handbook in 2003), which discusses the cause of death for pneumonia cases in more detail, notably on pages 89 and 90 (emphasis added):

> **Pneumonia is often a nonspecific process that occurs as the terminal event in someone who dies of a more specific underlying cause of death, such as congestive heart failure resulting from ischemic heart disease. In such cases, the specific underlying cause of death should be included in the cause-of-death statement.**
>
> Pneumonia is often designated as either community acquired or hospital or institution acquired (nosocomial). If the community- or institution-acquired nature of the pneumonia is known, the cause-of-death statement should include an indication of which one applies.
>
> The specific bacterial, viral, or other infectious agent, if known, should be cited in the cause-of-death statement.
>
> Relevant risk factors should also be cited in the cause-of-death statement, as might occur in an alcoholic who develops tuberculous pneumonia. Only in those instances where pneumonia has caused death and there is no known underlying cause or risk factor should the underlying cause of death be stated as "Pneumonia," being sure to specify the infectious agent, if known, or specifying that a specific etiology is unknown, if such is the case.

And on page 113 of *Cause of Death and the Death Certificate* by Randy Hanzlick, dementia, cerebrovascular disease, cardiac disease, and lung disease are all listed as common underlying causes of death in cases of deaths due to pneumonia:

Distractors of Which to Beware

There are a number of situations in which the complications of an underlying cause of death may clinically overshadow the underlying condition itself, resulting in the certifier forgetting to include the underlying cause in the cause-of-death statement.

Some of the more common complications that cause such problems, along with commonly associated underlying disease categories, are listed below. Any time one of these complications exists in an elderly person who died, a conscientious attempt should be made to identify and report as the underlying cause of death the condition that caused the complication. The complications should be reported as an intermediate or immediate cause of death, as appropriate for the case.

Pneumonia	Often a complication of dementia, cerebrovascular disease, cardiac disease, lung disease

Figure 14: Hanzlick on assigning pneumonia as underlying cause of death

Thus, traditionally, pre-pandemic, pneumonia deaths were frequently assigned a non-pneumonia *underlying cause of death*, usually a pre-existing condition and not the pneumonia-causing pathogen such as the influenza virus *or SARS-COV-2*, in common medical practice.

Based on the CDC's technical notes mentioned above, these pneumonia and influenza deaths would be included in the FluView death numbers but not in the leading causes of death report.

VIII. Comparing COVID-19 Death Numbers to the Pneumonia and Influenza Death Numbers and Estimates from Previous Years

As shown above, the CDC tracks at least three (3) different pneumonia and influenza death numbers and estimates: the Leading Causes of Death Report (~55,000 deaths per year, about two percent of annual deaths from all causes), the FluView graph and underlying data from the NCHS (~188,000 deaths per year, six to ten percent of annual deaths from all causes, before 2020), and the influenza death model estimates that range from 12,000 to 61,000 deaths per year, with the best estimate close to the number of pneumonia and influenza deaths in the leading causes of death report.

Are any of these the proper ***baseline*** for comparing COVID-19 deaths to prior years or should some other number or estimate be used?

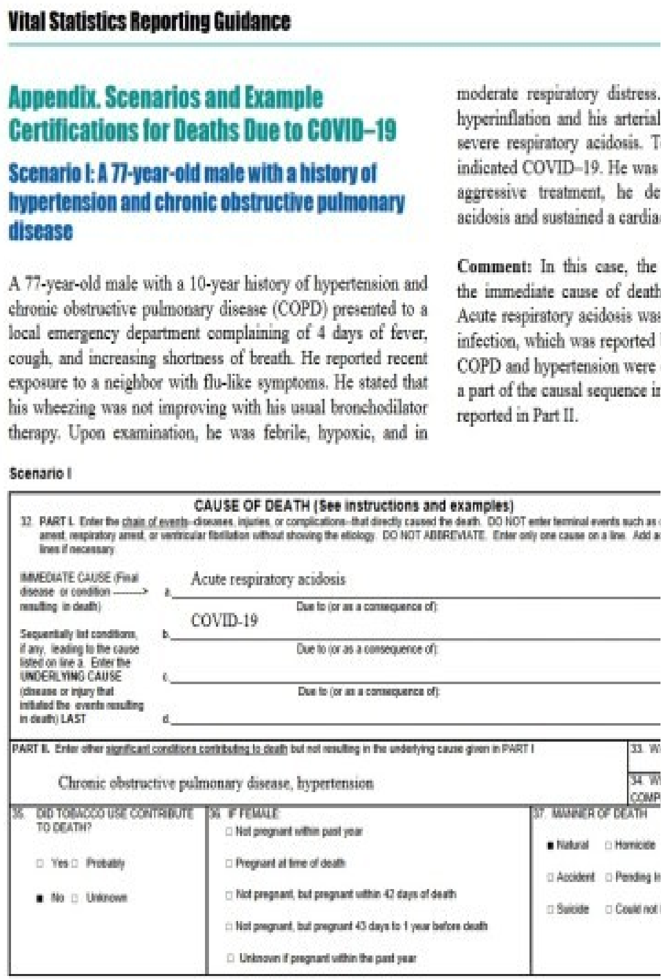
In the absence of the RT-PCR, antigen, and antibody tests for the SARS-COV-2 virus, most COVID-19 deaths would likely have been unexplained pneumonia deaths lacking a laboratory test confirming influenza or other known pathogen. Possibly, some COVID-19 deaths would have been listed as heart attacks or strokes, those COVID-19 deaths attributed to the blood clots and other blood-related anomalies currently blamed on COVID-19[71], or even some other causes.

The rest of this article will focus on the pneumonia deaths that would probably comprise most of the COVID-19 deaths in the absence of Emergency Use Authorization (EUA) laboratory tests for COVID-19, which may be misleading or inaccurate, sometimes to a high degree, depending on how they are employed (*see, e.g.* Skittrall et al. 2021[72], finding, based on a hypothetical application of standard Positive Predictive Value analysis, 25 times more false positives than true positives in testing the United Kingdom population in June 2020, based on measured background prevalence and test sensitivity and specificity).

The US CDC's April 2020 guidelines for reporting COVID-19 deaths (NVSS: Vital Statistics Reporting Guidance, Report 3, April 2020) clearly direct physicians and others not to list chronic obstructive pulmonary disease (COPD) as the underlying cause of death in COVID-19 cases. Instead, it should be included in Part 2 of the death certificate, which is reserved for "non-cause" contributing factors. This guidance differs dramatically from medical practice prior to 2020, as described in Randy Hanzlick's book and implicit in the FluView pneumonia and influenza deaths data above. The April 2020 guidance states, in relevant part:

> In some cases, survival from COVID–19 can be complicated by pre-existing chronic conditions, especially those that result in diminished lung capacity, such as chronic obstructive pulmonary disease (COPD) or asthma. These medical conditions do not cause COVID–19, but can increase the risk of contracting a respiratory infection and death, so these conditions should be reported in Part II and not in Part I.

This guidance also gives a specific example of a COVID-19 death with COPD relegated to Part 2, see Figure 11.


Vital Statistics Reporting Guidance

Appendix. Scenarios and Example Certifications for Deaths Due to COVID–19

Scenario I: A 77-year-old male with a history of hypertension and chronic obstructive pulmonary disease

A 77-year-old male with a 10-year history of hypertension and chronic obstructive pulmonary disease (COPD) presented to a local emergency department complaining of 4 days of fever, cough, and increasing shortness of breath. He reported recent exposure to a neighbor with flu-like symptoms. He stated that his wheezing was not improving with his usual bronchodilator therapy. Upon examination, he was febrile, hypoxic, and in moderate respiratory distress. hyperinflation and his arterial severe respiratory acidosis. T indicated COVID–19. He was aggressive treatment, he de acidosis and sustained a cardia

Comment: In this case, the the immediate cause of death Acute respiratory acidosis was infection, which was reported COPD and hypertension were a part of the causal sequence i reported in Part II.

Scenario I

CAUSE OF DEATH (See instructions and examples)

32. PART I. Enter the chain of events--diseases, injuries, or complications--that directly caused the death. DO NOT enter terminal events such as arrest, respiratory arrest, or ventricular fibrillation without showing the etiology. DO NOT ABBREVIATE. Enter only one cause on a line. Add lines if necessary.

IMMEDIATE CAUSE (Final disease or condition resulting in death) a. Acute respiratory acidosis

Due to (or as a consequence of):

Sequentially list conditions, if any, leading to the cause listed on line a. Enter the UNDERLYING CAUSE (disease or injury that initiated the events resulting in death) LAST b. COVID-19

Due to (or as a consequence of):

c.

Due to (or as a consequence of):

d.

PART II. Enter other significant conditions contributing to death but not resulting in the underlying cause given in PART I

Chronic obstructive pulmonary disease, hypertension

35. DID TOBACCO USE CONTRIBUTE TO DEATH? ☐ Yes ☐ Probably ■ No ☐ Unknown

36. IF FEMALE ☐ Not pregnant within past year ☐ Pregnant at time of death ☐ Not pregnant, but pregnant within 42 days of death ☐ Not pregnant, but pregnant 43 days to 1 year before death ☐ Unknown if pregnant within the past year

37. MANNER OF DEATH ■ Natural ☐ Homicide ☐ Accident ☐ Pending ☐ Suicide ☐ Could not

Figure 15: COVID-19 Death Certificate Guidance Example with COPD as Contributing Factor Only (source: NVSS Vital Statistics Report Guidance April 2020)

Although other causes of death that are often given as the underlying cause of death in pneumonia cases on pre-2020 death certificates are not explicitly identified in the April 2020 guidance document, it seems probable that most physicians would move these pre-existing conditions to Part 2 and not list them as the underlying cause of death for COVID-19, based on the April 2020 CDC guidance document. Note that COPD would fall under the category “lung disease” in the list of “distractors” from Hanzlick’s *Causes of Death and the Death Certificate*, mentioned above (Figure 10).

Thus, COVID-19 deaths since the April 2020 guidance are probably roughly comparable to the FluView ~188,000 pneumonia and influenza deaths per year that occur in a normal flu year. The language “roughly” is used because the April 2020 guidance encourages physicians and others to assign COVID-19 as the underlying cause of death in any death where COVID-19 is detected by tests or even just suspected, raising the possibility that heart attack and stroke deaths might be wrongly classified as COVID-19 deaths, as well as the traditional pneumonia and influenza deaths that would be listed in the FluView data. These would presumably be misclassified (“reassigned”) as the COVID-19 deaths exhibiting the mysterious blood clots and other blood-related problems reported in some COVID-19 cases and deaths. Thus, the *FluView* death numbers may represent a lower bound on COVID-19 deaths rather than an exact baseline.

IX. Recommendations

In light of the previous discussion, we make a number of recommendations to improve CDC’s data practices, including improved observance of common scientific and engineering practice – such as use of significant figures and reporting of statistical and systematic errors. Common scientific and engineering practice is designed to prevent serious errors and should be followed rigorously in a crisis such as the COVID-19 pandemic.

Note that some of these recommendations may require changes in federal or state laws, federal or state regulations, or renegotiation of contracts between the federal government and states. This is probably the case for making the Deaths Master File (DMF), with names and dates of death of persons reported as deceased to the states and federal government, freely available to the public and other government agencies.

- All CDC numbers, where possible, should be clearly identified as estimates, adjusted counts, or raw counts, with statistical errors and systematic errors given, using consistent clear standard language in all documents. The errors should be provided as both ninety-five percent (95%) confidence level intervals and the standard deviation – at least for the statistical errors.

- In the case of adjusted counts, the raw count should be explicitly listed immediately following the adjusted count as well as a brief description of the adjustment and a reference for the adjustment methodology. For example, if the adjusted number of deaths in the United States in 2020 is 3.4 million but the raw count of deaths was 3.3 million with 100,000 deaths added to adjust for unreported deaths of undocumented immigrants, the web pages and reports would say:

    Total deaths (2020): 3.4 million (adjusted, raw count 3.3 million, unreported deaths of undocumented immigrants, adjustment methodology citation: Smith et al, MMWR Volume X, Number Y)

- The distinction between the leading causes of death report "pneumonia and influenza" deaths, ~55,000 per year pre-pandemic, and the FluView website "pneumonia and influenza" deaths, ~188,000 per year pre-pandemic, should be clarified in the labels and legends for the graphics and prominently in the table of leading causes of death or immediately adjacent text. Statistical and systematic errors on these numbers should be provided in graphs and tables.

- In general, where grossly different raw counts, adjusted counts, or estimates are presented in CDC documents and websites with the same name, semantically equivalent or nearly equivalent names such as "pneumonia and influenza" and "influenza and pneumonia," clearly distinct names should be used instead, or the reasons for the gross difference in the values should be prominently listed in the graphs and tables or immediately adjacent text. It should be easy for the public, busy health professionals, policy makers and others to recognize and understand the differences.

- Where mathematical models are fit to data, such as the excess deaths computation, goodness of fit statistics should be reported in results, in or immediately adjacent to any plots, graphs, or tables showing the results. We recommend at least the standard chi-squared and the standard coefficient of determination (R^2), which is often of greater practical utility than the chi-squared statistic, as is common scientific and engineering practice in most fields.

- CDC should provide results for different models for the same data with similar R^2 values – coefficient of determination – to give the audience a quick sense of the systematic modeling errors – since there is no generally accepted methodology for estimating the 95% confidence level for the systematic modeling errors. See Figure 10 above for an example.

- All mathematical models should be free and open source with associated data provided using commonly used free open-source scientific programming languages such as Python or R, made available on the CDC website, GitHub, and other popular sources. The models and data should be provided in a package form such that anyone with access to a standard MS Windows, Mac OS X, or Linux/Unix computer can easily download and run the analysis – similar to the package structure used by the GNU project, for example.

- Specifically, the influenza virus deaths model should be provided to the public as code and data.

- Mathematical models should have distinct short English names where possible. We recommend the use of a unique digital identifier, possibly the DOI (Digital Object Identifier) system for each model and increasing sequential version numbers (e.g. 1.1, 1.2, 2.0, 2.1…) for different versions of the model. The digital identifier should point directly to the free, open source code used. A footnote or link such as (English Model Name, Point of Contact, MODEL ID, Version) should be associated with plots, tables, or other documents generated with the model. For example, (Influenza Deaths Model, Smith, 123423, v 1.12) to enable quick reproducibility of results and avoid confusion between different models. In particular, several different models appear to be used in various aspects of reporting the influenza disease burden, estimating reductions in the burden due to the influenza vaccination program, and other influenza related metrics.

- We recommend minimizing the use of models that produce large changes in the measured value, certainly greater than 100% changes, such as the influenza death model which produces multipliers of 2-12 applied to raw counts of death certificates listing influenza as a cause of death, phasing out such models and switching to direct measurement, or as close to direct measurement as possible.

- With respect to excess deaths tracking, include all major cause of death categories, rather than just the thirteen (13) in the cause-specific excess deaths that CDC tracks, which currently account for about 2/3 of all deaths.

- Include a Years of Lives Lost (YLL) display for COVID-19 deaths[73,74] and non-COVID-19 deaths, as well as excess deaths analysis, due to the higher granularity of YLL analysis when compared to excess deaths analysis. Explain the pros and cons of both analytical tools. Do the same for any future pandemics or health crises.

- Adopt or develop a different algorithm or algorithms for tracking excess deaths which are mostly attributed to non-infectious causes such as heart attacks, cancer, and strokes. The Farrington/Noufaily algorithms were specifically developed as an early warning for often non-lethal infectious disease outbreaks such as salmonella. A medically-based model or models that incorporates population demographics such as the aging "baby boom" and evolving death rates broken down by age, sex, and possibly other factors where known is probably a better practice rather than simple empirical trend models such as the Noufaily algorithm.

- Eliminate the zeroing procedure in calculating excess deaths, in which negative excess deaths in some categories are set to zero, rather than being added to the full excess deaths sum over all categories.

- The anonymized data with causes of death as close to the actual data as possible, e.g. the actual death certificates, should be available on the CDC website in a simple accessible widely used format such as CSV (comma separated values) files. The code used to aggregate the data into summary data such as the FluView website data files should also be public.

- The full Deaths Master File (DMF) including the actual names of the deceased persons and dates of death should be made available to the general public, independent researchers, and others. This is critical to independent verification of many numbers from the CDC, SSA, and US Census.

- COVID-19-related deaths figures should be tracked based on year-specific age of death, rather than 10-year age ranges, as is currently the case.

- CDC frequently changes the structure and layout of the CSV files/spreadsheets on their websites. The CDC should either (1) not do this or (2) provide easy conversion between different file formats with each new format so it is trivial for third parties to quickly adapt to the changes without writing additional code. CDC should provide a program or program in a free and open source language like R to convert between the formats.

- The CDC and other agencies should be required to announce and solicit public comment for changes to case definitions, data collection rules, etc. for key public policy data such as the COVID-19 case definitions, death certification guidelines, and coding rules. Other government agencies have significantly more public participation than CDC, which is appropriate in a modern democracy.

- Any practices and policies imposed in a public emergency, such as case definitions, definitions of measured quantities, data reporting practices, etc. imposed without public comment and review, should have an expiration date (e.g. sixty days) beyond which they must be subject to public review. Public comment, reviews, and cost/benefit analyses should happen during this emergency period.

Enacting these reforms should reduce the risk of serious errors, increase the quality and accuracy of CDC data and analyses, as well as any policies or CDC guidelines based on the data and analysis, and strengthen public confidence in the CDC and public health policies.

Acknowledgements

We thank Dr. Robert Anderson, Dr. Lauren Rossen and other staff of the CDC's National Center for Health Statistics (NCHS) for detailed answers to our questions regarding the CDC's excess deaths analysis and providing a copy of the R statistical programming language code and data used for the excess deaths analysis during this difficult time. All conclusions and opinions in this paper belong to the authors only.

[1] Corresponding Author: John F. McGowan, Ph.D. (E-Mail: ceo@mathematical-software.com)

[2] Tam Hunt, Department of Psychological and Biological Sciences, University of California at Santa Barbara

[3] Josh Mittledorf, Independent Researcher, Philadelphia, Pennsylvania, United States

[4] Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science (New Series), 185, 1124-1131.

[5] Kahneman, Daniel, and Amos Tversky. “Subjective Probability: A Judgment of Representativeness.” Cognitive Psychology 3 (1972): 430–54.

[6] Janis, Irving, *Groupthink: Psychological Studies for Policy Decisions and Fiascoes* (2nd edition), Wadsworth, Cengage Learning, (1982)

[7] *AP Biology Course Exam and Description*, Fall 2020, (discussion of significant figures on pages 167 and 199), The College Board, New York , New York, USA

[8] Silas W. Holman “Discussion of the precision of measurement,” *Technol. Q.* 1, 194-207 (1888).

[9] Silas W. Holman, *Discussion of the Precision of Measurements: With Examples Taken Mainly From Physics And Electrical Engineering* (Ferris Bros. Printers, New York, 1892).

[10] Carter, Ashley, “Evolution of the Significant Figure Rules,” *The Physics Teacher*. 51, 340 (2013); doi: 10.1119/1.4818368

[11] Robert A. Millikan, Duane E. Roller, and Earnest C. Watson, *Mechanics, Molecular Physics, Heat, and Sound, 1st ed.* (M.I.T. Press, Cambridge, 1937). (see Appendix 1, *Significant Figures and Notations by Powers of Two*, page 457)

[12] *Standard Practice for Using Significant Digits in Test Data to Determine Conformance with Specifications*, ASTM Standard E29-08, ASTM International (West Conshohocken, PA, 1993).

[13] ISO (1993). Guide to the Expression of Uncertainty in Measurement. International Organization for Standardization

[14] Taylor, J. R. (1997). *Error analysis: The study of uncertainties in physical measurements (2nd ed.).* Sausalito, CA: University Science Books. (Chapter 2, How to Report and Use Uncertainties, Section 2.2, Significant Figures, page 14, Chapter 4, Statistical Analysis of Random Uncertainties, Section 4.1, Random and Systematic Errors, page 94, Section 4.6, Systematic Errors, page 106)

[15] Bevington, Philip and Robinson, D. Keith, *Data Reduction and Error Analysis for the Physical Sciences (2nd Edition)*, McGraw-Hill Education, New York, New York, USA (1992) (Chapter 1, “Uncertainties in Measurement,” page 1 for statistical and systematic errors, “Significant figures and roundoff” subsection, page 4 for significant figures)

[16] National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Committee on Science, Engineering, Medicine, and Public Policy; Board on Research Data and Information; Division on Engineering and Physical Sciences; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Analytics; Division on Earth and Life Studies; Nuclear and Radiation Studies Board; Division of Behavioral and Social Sciences and Education; Committee on National Statistics; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on Reproducibility and Replicability in Science. *Reproducibility and Replicability in Science.* Washington (DC): National Academies Press (US); 2019 May 7. PMID: 31596559.

[17] Wendy Ginsberg et al., “Government Transparency and Secrecy: An Examination of Meaning and Its Use in the Executive Branch,” Congressional Research Service, November 14, 2012, Online at: https://fas.org/sgp/crs/secrecy/R42817.pdf (accessed June 29, 2021)

[18] Richard W. Oliver, *What is Transparency?* (New York: The McGraw-Hill Companies Inc., 2004

[19] Bradley SH, DeVito NJ, Lloyd KE, Richards GC, Rombey T, Wayant C, Gill PJ. Reducing bias and improving transparency in medical research: a critical overview of the problems, progress and suggested next steps. J R Soc Med. 2020 Nov;113(11):433-443. doi: 10.1177/0141076820956799. PMID: 33167771; PMCID: PMC7673265.

[20] Borja, Angel, “11 Steps to structuring a science paper editors will take seriously”, Elsevier Website, June 24, 2014 https://www.elsevier.com/connect/11-steps-to-structuring-a-science-paper-editors-will-take-seriously

[21] Parish, R. Gibson, *Vital Records and Vital Statistics in the United States: Uses, Users, Systems, and Sources of Revenue, Report prepared for the National Committee on Vital and Heath Statistics Subcommittee on Population Health* (June 10, 2018), Online at: https://www.ncvhs.hhs.gov/wp-content/uploads/2018/02/NCVHS_Vital_Records_Uses_Costs_Feb_23_2018.pdf

[22] Deceased Payee Fraud: Strange but True, Office of Inspector General: Social Security Administration, https://oig.ssa.gov/newsroom/blog/jan2-post

[23] FOLLOW-UP: SURVIVOR BENEFITS PAID IN INSTANCES WHEN THE SOCIAL SECURITY ADMINISTRATION REMOVED THE DEATH ENTRY FROM A PRIMARY WAGE EARNER’S RECORD, AUDIT REPORT, Office of the Inspector General, Social Security Administration, September 2011 A-06-10-20135 https://oig.ssa.gov/sites/default/files/audit/full/pdf/A-06-10-20135.pdf

[24] “What It’s Like to Be Declared Dead by the Government,” Priceonomics.com, June 8, 2015 https://priceonomics.com/what-its-like-to-be-declared-dead-by-the/

[25] Svendsen MT, Bøggild H, Skals RK, Mortensen RN, Kragholm K, Hansen SM, et al Uncertainty in classification of death from fatal myocardial infarction: A nationwide analysis of regional variation in incidence and diagnostic support.

PLoS One. 2020 Jul 27;15(7):e0236322. doi: 10.1371/journal.pone.0236322. PMID: 32716962; PMCID: PMC7384617.

26 Walsh DP, Norton AS, Storm DJ, Van Deelen TR, Heisey DM. Using expert knowledge to incorporate uncertainty in cause-of-death assignments for modeling of cause-specific mortality. Ecol Evol. 2017 Nov 30;8(1):509-520. doi: 10.1002/ece3.3701. PMID: 29321889; PMCID: PMC5756890.

27 Hernández B, Ramírez-Villalobos D, Romero M, Gómez S, Atkinson C, Lozano R. Assessing quality of medical death certification: Concordance between gold standard diagnosis and underlying cause of death in selected Mexican hospitals. Popul Health Metr. 2011 Aug 4;9:38. doi: 10.1186/1478-7954-9-38. PMID: 21816103; PMCID: PMC3160931.

28 Lindahl BI, Allander E. Problems in the classification of cause of death diagnoses affecting the reliability of mortality statistics for rheumatoid arthritis. J Chronic Dis. 1985;38(5):409-18. doi: 10.1016/0021-9681(85)90136-5. PMID: 3998055.

29 Landes SD, Turk MA, Bisesti E. Uncertainty and the reporting of intellectual disability on death certificates: a cross-sectional study of US mortality data from 2005 to 2017. BMJ Open. 2021 Jan 31;11(1):e045360. doi: 10.1136/bmjopen-2020-045360. PMID: 33518529; PMCID: PMC7853001.

30 Chamberlain J, Coleman D, Ellman R, Moss S. Verification of the cause of death in the trial of early detection of breast cancer. UK Trial of Early Detection of Breast Cancer Group. Trial Co-ordinating Centre. Br J Cancer. 1991 Dec;64(6):1151-6. doi: 10.1038/bjc.1991.480. PMID: 1764379; PMCID: PMC1977852.

31 Tallis GM. Misclassification, correlation, and cause of death studies. Hum Biol. 2002 Feb;74(1):75-81. doi: 10.1353/hub.2002.0013. PMID: 11931580.

32 Mieno MN, Tanaka N, Arai T, Kawahara T, Kuchiba A, Ishikawa et al Accuracy of Death Certificates and Assessment of Factors for Misclassification of Underlying Cause of Death. J Epidemiol. 2016;26(4):191-8. doi: 10.2188/jea.JE20150010. Epub 2015 Dec 5. PMID: 26639750; PMCID: PMC4808686.

33 Ebrahimi N. The effects of misclassification of the actual cause of death in competing risks analysis. Stat Med. 1996 Jul 30;15(14):1557-66. doi: 10.1002/(SICI)1097-0258(19960730)15:14<1557::AID-SIM286>3.0.CO;2-Q. PMID: 8855481.

34 https://www.cdc.gov/flu/weekly/fluviewinteractive.htm

35 “Deaths: Leading Causes for 2017,” by Melanie Heron, Ph.D., Division of Vital Statistics, National Vital Statistics Reports Volume 66, Number 6, June 24, 2019 (https://www.cdc.gov/nchs/data/nvsr/nvsr68/nvsr68_06-508.pdf)

36 https://www.cdc.gov/nchs/nvss/vsrr/COVID19/excess_deaths.htm

37 Hanzlick, Randy, *Medical Examiners' and Coroners' Handbook on Death Registration and Fetal Death Reporting (2003 Revision)*, Centers for Disease Control, National Center for Health Statistics, Hyattsville, MD, April 2003, DHHS Publication No. (PHS) 2003-1110.   Online at: https://www.cdc.gov/nchs/data/misc/hb_me.pdf

38 Hanzlick, Randy, *Cause of Death and the Death Certificate: Important Information for Physicians, Coroners, Medical Examiners, and the Public*, College of American Pathologists, 325 Waukegan Road Northfield, Illinois 60093, (2006), Online at: https://www.health.state.mn.us/people/vitalrecords/physician-me/docs/capcodbook.pdf

39 Kelly J, Sadeghieh T, Adeli K. Peer Review in Scientific Publications: Benefits, Critiques, & A Survival Guide. EJIFCC. 2014;25(3):227-243. Published 2014 Oct 24.

40 Reed C, Chaves SS, Daily Kirley P, Emerson R, Aragon D, Hancock EB, et al Estimating influenza disease burden from population-based surveillance data in the United States. PLoS One. 2015 Mar 4;10(3):e0118369. doi: 10.1371/journal.pone.0118369. PMID: 25738736; PMCID: PMC4349859.

41 https://www.cdc.gov/flu/symptoms/coldflu.htm

42 Doshi, Peter, “Influenza: marketing vaccine by marketing disease,” BMJ 2013;346:f3037, Online at: https://www.bmj.com/content/346/bmj.f3037

43 Doshi, Peter, “Influenza: a study of contemporary medical politics,” Thesis (Ph. D. in History, Anthropology, and Science, Technology and Society (HASTS)) – Massachusetts Institute of Technology, Program in Science, Technology and Society, 2011. Online at: https://dspace.mit.edu/handle/1721.1/69811

44 Doshi P. Trends in recorded influenza mortality: United States, 1900-2004. Am J Public Health. 2008 May;98(5):939-45. doi: 10.2105/AJPH.2007.119933. Epub 2008 Apr 1. Erratum in: Am J Public Health. 2009 Aug;99(8):1353-4. PMID: 18381993; PMCID: PMC2374803.

45 William W. Thompson, Matthew R. Moore, Eric Weintraub, Po-Yung Cheng, Xiaoping Jin, Carolyn B. Bridges, et al, 2009: Estimating Influenza-Associated Deaths in the United States American Journal of Public Health **99**, S225_S230, https://doi.org/10.2105/AJPH.2008.151944

46 Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An Improved Algorithm for Outbreak Detection in Multiple Surveillance Systems. Statistics in Medicine 2012;32(7):1206-1222.

47 Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson KW; Et al Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. 2020 May 26;323(20):2052-2059. doi: 10.1001/jama.2020.6775. Erratum in: JAMA. 2020 May 26;323(20):2098.

PMID: 32320003; PMCID: PMC7177629.

[48] Personal communication, Dr. Robert Anderson, National Center for Health Statistics (NCHS), CDC, September, 2021
[49] Wright, Sewall (January 1921). "Correlation and causation". Journal of Agricultural Research. 20: 557–585.
[50] Bevington, Philip and Robinson, D. Keith, Data Reduction and Error Analysis for the Physical Sciences (2nd Edition), McGraw-Hill Education, New York, New York, USA (1992) (Chapter 11, “Testing the Fit” page 194 for evaluating the goodness of fit, Section 11.1 “X^2 Test for Goodness of Fit,”)
[51] Whitlock, Michael and Schluter, Dolph, The Analysis of Biological Data, 3rd Edition, W.H. Freeman and Company, imprint of Macmillan Higher Education, (January 6, 2020), (Chapter 8, Fitting probability models to frequency data, Section 8.1 X^2 Goodness-of-fit test: the proportional model)
[52] Rossen LM, Branum AM, Ahmad FB, Sutton PD, Anderson RN. Notes from the Field: Update on Excess Deaths Associated with the COVID-19 Pandemic — United States, January 26, 2020–February 27, 2021. MMWR Morb Mortal Wkly Rep 2021;70:570–571. DOI: http://dx.doi.org/10.15585/mmwr.mm7015a4
[53] Ford ES, Ajani UA, Croft JB, Critchley JA, Labarthe DR, Kottke TE, et alExplaining the decrease in U.S. deaths from coronary disease, 1980-2000. N Engl J Med. 2007 Jun 7;356(23):2388-98. doi: 10.1056/NEJMsa053935. PMID: 17554120.

[54] “IRS sent stimulus checks to more than 1 million dead people, government watchdog agency says,” by Michael Collins, USA Today, June 25, 2020 (https://www.usatoday.com/story/news/politics/2020/06/25/coronavirus-irs-sent-stimulus-checks-1-million-dead-people/3256793001/)
[55] Personal communication, Dr. Robert Anderson, NCHS, CDC, September 2021
[56] *See, e.g.* Briggs AH, Goldstein DA, Kirwin E, Meacock R, Pandya A, Vanness DJ, et al. Estimating (quality adjusted) life-year losses associated with deaths: With application to COVID-19. Health Economics. 2020;30(3):699–707; Gardner JW, Sanborn JS. Years of potential life lost (YPLL)--what does it measure? Epidemiology. 1990 Jul;1(4):322-9. doi: 10.1097/00001648-199007000-00012. PMID: 2083312; Quast, T., et al. 2021. Years of life lost associated with COVID-19 deaths in the USA during the first year of the pandemic.
[57] CDC “Provisional COVID-19 deaths by sex and age.” Online at: https://data.cdc.gov/NCHS/Provisional-COVID-19-Deaths-by-Sex-and-Age/9bhg-hcku. Data through end of 2020 was used.
[58] CDC *Weekly Updates by Select Demographic and Geographic Characteristics* Website Online at: https://www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm.
[59] Ioannidis, J., Axfors, C., & Contopoulos-Ioannidis, D. G. (2021). Second versus first wave of COVID-19 deaths: Shifts in age distribution and in nursing home fatalities. Environmental research, 195, 110856. https://doi.org/10.1016/j.envres.2021.110856 (See Table 3: Proportion of COVID-19 deaths in nursing home residents, page 5, USA line) Online at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7875012/
[60] Dooling, K. 2020. CDC presentation, p. 8. Online at: https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2020-12/COVID-02-Dooling.pdf.
[61] CDC Wonder, Multiple Causes of Death ZIP File (2019), https://www.cdc.gov/nchs/data_access/VitalStatsOnline.htm#Mortality_Multiple, https://ftp.cdc.gov/pub/Health_Statistics/NCHS/Datasets/DVS/mortality/mort2019us.zip
[62] American Addiction Centers 2019. Online at: https://www.projectknow.com/discover/cutting-it-short/. We assume, for this preliminary analysis, that the average age of overdose death is a good proxy for homicides and suicides also, partly because the excess deaths for overdoses are considerably higher than the other two categories, and also because it is likely that homicide and suicide deaths may be younger on average than overdose deaths.
[63] Social Security Administration actuarial life table, 2020. Online at: https://www.ssa.gov/oact/STATS/table4c6.html.
[64] Statista.com, for 2019 homicides, online at https://www.statista.com/statistics/251878/murder-victims-in-the-us-by-age/.
[65] NVSS 2021, Provisional drug overdose death counts. Online at: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm.
[66] Ahmad FB, Cisewski JA. Quarterly provisional estimates for selected indicators of mortality, 2018-Quarter 4, 2020. National Center for Health Statistics. National Vital Statistics System, Vital Statistics Rapid Release Program. 2021. Online at: https://www.cdc.gov/nchs/nvss/vsrr/mortality-dashboard.htm#.
[67] Florida v. Becerra, CASE NO. 8:21-cv-839-SDM-AAS, June 18, 2021. Online at: https://www.leagle.com/decision/infdco20210621722
[68] CDC’s website states its adherence to ICD guidelines for coding: “Causes of death are coded according to the *International Classification of Diseases, 10th Revision* (ICD–10). Online at https://www.cdc.gov/nchs/COVID19/coding-and-reporting.htm. Accessed March 9 2021.
[69] Hanzlick, Randy, *Medical Examiners' and Coroners' Handbook on Death Registration and Fetal Death Reporting (2003 Revision)*, Centers for Disease Control, National Center for Health Statistics, Hyattsville, MD, April 2003, DHHS

Publication No. (PHS) 2003-1110.   Online at: https://www.cdc.gov/nchs/data/misc/hb_me.pdf

70 Hanzlick, Randy, *Medical Examiners' and Coroners' Handbook on Death Registration and Fetal Death Reporting (2003 Revision)*, Centers for Disease Control, National Center for Health Statistics, Hyattsville, MD, April 2003, DHHS Publication No. (PHS) 2003-1110.   Online at: https://www.cdc.gov/nchs/data/misc/hb_me.pdf
e/vitalrecords/physician-me/docs/capcodbook.pdf

71 Asakura H, Ogawa H. COVID-19-associated coagulopathy and disseminated intravascular coagulation. Int J Hematol. 2021 Jan;113(1):45-57. doi: 10.1007/s12185-020-03029-y. Epub 2020 Nov 7. PMID: 33161508; PMCID: PMC7648664.

72 Skittrall JP, Fortune MD, Jalal H, Zhang H, Enoch DA, Brown NM, et al. Diagnostic tool or screening programme? Asymptomatic testing for SARS-CoV-2 needs clear goals and protocols. The Lancet Regional Health - Europe. 2021;1:100002.

73 Quast T, Andel R, Gregory S, Storch EA. Years of life lost associated with COVID-19 deaths in the USA during the first year of the pandemic. Journal of Public Health (Oxford, England). 2021 Apr. DOI: 10.1093/pubmed/fdab123. PMID: 33839789; PMCID: PMC8083296.

74 Andrew H. Briggs, Daniel A. Goldstein, Erin Kirwin, Rachel Meacock, Ankur Pandya, David J. Vanness, et al, Estimating (quality-adjusted) life-year losses associated with deaths: With application to COVID-19, Health Economics Letter, 24 December 2020, https://doi.org/10.1002/hec.4208