

Unpacking the Drop in COVID-19 Case Fatality Rates: A Study of National and Florida Line-Level Data

Cheng Cheng*, B.S., Helen Zhou*, M.S., M.Eng.,
Jeremy C. Weiss, M.D., Ph.D., Zachary C. Lipton, Ph.D.
Carnegie Mellon University, Pittsburgh, PA, USA

Abstract Since the COVID-19 pandemic began, the United States's case fatality rate (CFR) has plummeted. Using national and Florida data, we unpack the drop in CFR between April and December 2020, accounting for such confounders as expanded testing, age distribution shift, and detection-to-death lags. Guided by the insight that treatment improvements in this period should correspond to decreases in hospitalization fatality rate (HFR), and using a block-bootstrapping procedure to quantify uncertainty, we find that although treatment improvements do not follow the same trajectory in Florida and nationally (with Florida undergoing a comparatively severe second peak), by December, significant improvements are observed both in Florida and nationally (at least 17% and 55% respectively). These estimates paint a more realistic picture of improvements than the drop in aggregate CFR (70.8%–91.1%). We publish a website where users can apply our analyses to selected demographics, regions, and dates of interest.

Introduction

Over the past year, the coronavirus (COVID-19) pandemic has continually evolved, with disease outbreaks expanding and contracting; lockdown measures tightening and loosening; testing capacity (mostly) increasing; and treatments protocols evolving. Decision-makers trying to maintain a grasp of the rapidly unfolding situation face a few key questions (among others): Have new treatment protocols improved outcomes over time? Is COVID-19 fatality decreasing? How does the actual infection rate compare to at previous dates? To what extent are rising case counts artifacts of expanded testing?

The most widely reported statistics for monitoring the pandemic are confirmed cases and deaths. At first glance, the two appear to tell divergent stories about the trajectory of the pandemic over time. Confirmed cases peak for the first time in April (with a 7-day average of nearly 32,000 daily cases), peak again with more than twice as many cases in a second wave in July (nearly 67,000), and yet again in a much larger third wave in December (nearly 230,000) (Figure 1, left panel). However, reported deaths appear to tell a contradictory story concerning the relative severity of the three waves (Figure 1, middle panel). Dividing deaths by cases, we observe that the reported case fatality rate has fallen dramatically after the first wave, from nearly 7.9% at the height of the first wave in mid-April to the 0.7%–2.3% range since July. (Figure 1, right panel). There are several plausible explanations for this decline, each with significant policy implications. Thus far, academic and public discourse has centered around the following hypotheses:

- (H1) The *age distribution* of infected patients has shifted, altering the CFR due to the higher risk among the elderly.^{1–4}
- (H2) *Increases in testing capacity* have driven down the CFR due to a rising number of tests catching milder cases.^{5–7}
- (H3) Apparent shifts in case fatality rate, are artifacts due to the *delay between detection and fatality*.^{1,7,8}
- (H4) *Treatment has improved* as doctors grow more experienced and new therapeutics become available.^{3,9–12}
- (H5) The disease itself is *mutating*, leading to changes in the actual infection fatality rate over time.^{5,13}
- (H6) Social distancing has reduced the *viral load* that individuals are exposed to, resulting in milder infections.^{13–15}

Note that H1–H3 can be misleading if not sufficiently accounted for. Due to Simpson's paradox, since there are large differences in fatality between different age groups, if the age distribution shifts (H1) we could easily observe an *overall decrease* in CFR despite *increasing* CFRs in every age group (or vice versa). Additionally, testing ramp-up (H2) and delays between detection and fatality (H3) can cause the behavior of case fatality rate to diverge substantially from the behavior of the true infection fatality rate. Thus, CFR can be a poor proxy for actual infection fatality rate (IFR).

On the other hand, the last three phenomena—improved treatments, disease mutation, and changing viral load—correspond to actual reductions in mortality and could be grounds for policy changes. This work **demonstrates how**

*Equal contribution.

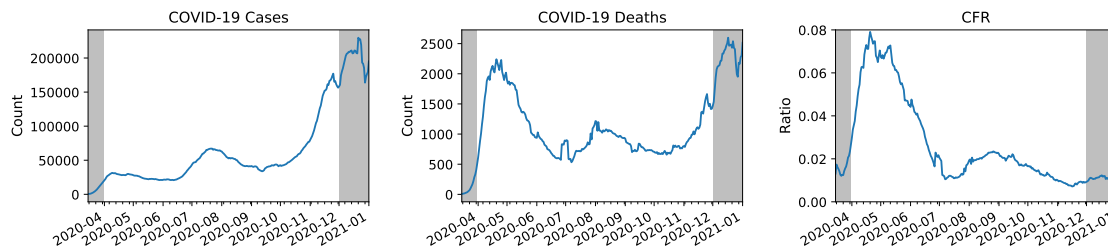


Figure 1: From left to right: confirmed cases, deaths, and case fatality rate, calculated using 7-day trailing averages based on national reporting data available via USAFacts data pulled from the Carnegie Mellon Delphi project’s COVIDcast API¹⁶. Data outside the April 1st to December 1st time range considered in this study is grayed out.

given sufficiently granular data, H1–H3 (“artifacts”) can be accounted for to quantify true improvements in treatment (H4), even though H5 and H6 cannot be decisively separated out from H4 without additional data.

In particular, we argue that complete and accurate *age-stratified, line-level hospitalization data* is pivotal for distinguishing true improvements from artifacts. Hospitalizations should be less influenced by testing capacity than confirmed cases, and less influenced by treatment efficacy than deaths. While there may have been changes in admitting criteria at the very worst moments,^{17,18} for example, when New York hospital demand exceeded capacity in late March, for the most part, criteria for inpatient hospitalization is relatively consistent across time periods. Additionally, in the study time period (April 1st to December 1st), treatment improvements mostly targeted hospitalized COVID-19 patients.

Importantly, this analysis yields several important observations: (i) large increases in testing do occur between the waves but do not explain them away; (ii) since age distributions shifted substantially between the first and second waves (and have fluctuated since), age must be accounted for in order to separate out the effects of treatment from age shift; (iii) between the first and second waves age-stratified HFRs improved substantially in the national data (with HFR decreasing by as little as 27% in the 80+ age group and as much as 37% in the 30-39 age group), but were relatively unchanged in Florida (with a slight *increase* in HFR by as little as 2.9% in the 80+ age group and as much as 13% in the 60-69 age group); (iv) by December 1st, both Florida and national data suggest significant decreases in HFR since April 1st—at least 17% in Florida and at least 55% nationally in every age group; and (v) comprehensive age-stratified hospitalization data is of central importance to providing situational awareness during the COVID-19 pandemic.

As far as we are aware, this is the largest national-scale (588,126 hospitalizations, 10.3 million cases) data-driven analysis to quantify and account for all three artifacts (age distribution shift, increased testing, and detection-to-fatality delay) when estimating treatment improvements. To allow users to apply our analyses to time ranges, states, and demographics of interest, we publicly release an interactive web application. Finally, we release a full pre-processing, visualization, and analytical code pipeline for both the CDC and Florida COVID-19 line-level public datasets.

Related Work

Several COVID-19 treatments were developed over the study time range (April 1st to December 1st), and each underwent randomized controlled trials testing for its individual efficacy. Dexamethosone resulted in a lower 28-day mortality among COVID-19 inpatients receiving respiratory support.¹¹ Remdesivir was associated with shortened recovery time among adults hospitalized with lower respiratory tract infection.^{10,19} Clinical trials for hydroxychloroquine^{12,20} and convalescent plasma²¹ found no positive results in prevention of disease progression or mortality. In November, monoclonal antibody treatments bamlanivimab and the combination therapy casirivimab and imdevimab were approved for emergency use authorization.²² These therapies, unlike dexamethosone and remdesivir, are not recommended for inpatients,²³ but were shown to have benefits in outpatients likely to progress to severe COVID-19 (for bamlanivimab),²⁴ and in patients who have not yet mounted their own immune response or have high viral load (for casirivimab and imdevimab).²⁵ Note that monoclonal antibodies were only approved for emergency use within the last month of our study time range, before therapeutic distribution had ramped up.²⁶ More recently in December (outside of our study time range), the first coronavirus vaccines from Pfizer-BioNTech²⁷ and Moderna²⁸ were approved for emergency use, with the Pfizer-BioNTech vaccine clinically proven to achieve 95% efficacy²⁹ and the Moderna vaccine, 94.1% efficacy.³⁰ While these clinical trials have evaluated the effects of specific treatments in their identified target populations, our work studies the broader impacts of treatment improvements over time at a larger national scale.

To get a holistic sense of improvements over time several studies have examined CFRs. In a study of 53 countries, all but ten were found to have lower CFRs in the second wave compared to the first.⁵ However, as delineated in our introduction, confounding factors such as shifting age distribution (H1), testing capacity (H2), and detection-to-death lags (H3) can lead to misleading interpretations of the CFR.¹⁻⁸ For example, when comparing CFR by age group in Italy and China, Onder et. al. suggested variation in testing strategies as a possible explanation for discrepancies.³¹ In study of COVID-19 cases in Germany, an apparent discrepancy between cases and deaths was attributed to shifting age distribution, testing capacity, or true effectiveness of government-issued directives³². While these country-level studies identify the three “artifacts” (H1-H3) as limitations of interpreting the CFR, none explicitly account for them.

To account for changes in testing capacity, we examine hospitalization data. While (as far as we are aware) no nation-wide studies in the U.S. account for all three artifacts, some hospital systems have controlled for them by conducting age-stratified cohort studies. Among 5,121 hospitalized COVID-19 patients in a single New York health system, Horwitz et. al. demonstrated that after adjusting for age, sex, ethnicity, and other clinical factors, HFR between March 1st and June 20th decreased but not as much as observed before adjusting for these factors.³ In another New York hospital system, Mehta et. al. demonstrated that cancer and older age were associated with increased risk of case fatality, finding no significant associations between race and mortality or gender and mortality.³³ In a study conducted among 21,082 COVID-19 patients admitted to 108 English critical care units between March 1st and June 27th, mortality risk in mid-April and May was found markedly lower than earlier in the pandemic even after adjusting for age, sex, ethnicity, comorbidities, and geographic region.⁴ While these studies provide thorough estimates of mortality for their respective regions during their specific time periods, we analyze data over a longer time range and larger scale (588,126 hospitalizations, 10.3 million cases) that purportedly captures all of Florida and most of the United States.

Our data does not contain information about viral mutations and viral loads (H5 and H6). However, a few studies have begun to investigate their impact. The B.1.1.7 and B.1.351 variants of COVID-19 were first reported in the U.S. at the end of December 2020 and January 2021, respectively.³⁴ While it is unclear how these variants will ultimately impact HFRs, recent studies indicate that vaccines may be more effective against the B.1.1.7 variant than the B.1.351 variant.^{35,36} Regarding social distancing precautions reducing viral load, in a study across seven countries, declining CFR was found to be correlated with strict lockdown policies and widespread PCR testing.¹³ At a hospital system in northern Italy, Piubelli et. al. found that among patients diagnosed with COVID-19 in their emergency room, the proportion of patients requiring intensive care decreased over time, also having lower viral load.¹⁵ Our analysis does not attempt to separate out the effects of viral mutations (H5) and changing viral loads (H6), but we note that these are factors that can affect the true infection fatality rate, and therefore can be reflected in our estimates as well.

Methodologically, prior studies on the COVID-19 fatality have employed logistic regression,³ Cox proportional hazards,⁴ and propensity matching³³ to adjust for age and other comorbidities. While logistic regression and propensity matching can quantify risk of death averaged over their study time period, we are interested in the evolution of fatality risk over time. While the standard Cox proportional hazards does model risk over time, it assumes a simple linear form between the covariates and the log hazard. Without making this assumption, we leverage techniques in time series literature to reduce noise in the raw signal, and compute uncertainty around the estimates at any given time. Assuming smoothness in the true underlying trend, the moving average can obtain a better estimate of the trend than the raw signal.³⁷ To quantify uncertainty in time series, the moving block bootstrap technique was developed in lieu of standard bootstrapping targeting independent and identically distributed observations. In this technique, the time series is chunked into blocks to reduce dependence among them. This reduced dependence is hard to ensure, however, thereby suggesting a strategy between model-based and block resampling. To reduce much of the dependence between original observations, one can “pre-whiten” the data by fitting a model to the data, and computing the residuals. Instead of block resampling the original dependent series, the residuals can be resampled, and added back to the model estimates. This intermediate solution, termed post-blackening, has been shown to work more consistently in practice³⁸.

Materials and Methods

Data Description We center our analysis on (1) state-level COVID-19 Case Line Data made available by the Florida Department of Health (FDOH)³⁹ and (2) national-level COVID-19 Case Surveillance Data made available by the United States Centers for Disease Control and Prevention (CDC).⁴⁰ Both datasets are line-level, including date of detection, demographics (including age and gender), and indicators of eventual outcomes for each case (Table 1). In our defined

Table 1: Demographics and outcomes of the Florida and national cohorts. Data are provided as counts (percentage). For Hospitalized and Died, the recoded “No” category is merged from the original “No”, “Unknown”, and “Missing” categories. Unknown corresponds to checking an “unknown” box in the reporting form, and missing corresponds to leaving the question empty.

	Florida	National
COVID-19 Cases	1,004,818	10,332,725
Age		
0-9	37,153 (3.7%)	388,867 (3.8%)
10-19	91,887 (9.1%)	1,045,408 (10.1%)
20-29	192,790 (19.2%)	2,021,007 (19.6%)
30-39	172,027 (17.1%)	1,699,260 (16.4%)
40-49	153,717 (15.3%)	1,546,212 (15.0%)
50-59	149,530 (14.9%)	1,485,318 (14.4%)
60-69	102,477 (10.2%)	1,043,294 (10.1%)
70-79	61,436 (6.1%)	583,837 (5.7%)
80+	42,696 (4.2%)	453,229 (4.4%)
Unknown	1,105 (0.1%)	66,293 (0.6%)
Gender		
Female	517,640 (51.5%)	5,297,936 (51.3%)
Male	482,227 (48.0%)	4,861,508 (47.0%)
Missing	0 (0.0%)	17,071 (0.2%)
Unknown	4,951 (0.5%)	80,093 (0.8%)

	Florida	National
Outcome		
Hospitalized		
Yes	56,673 (5.6%)	588,126 (5.7%)
No (recoded)	948,145 (94.4%)	9,744,599 (94.3%)
No	538,428 (53.6%)	3,989,850 (38.6%)
Unknown	402,616 (40.1%)	1,593,887 (15.4%)
Missing	7,101 (0.7%)	4,160,862 (40.3%)
Died		
Yes	21,028 (2.1%)	199,677 (1.9%)
No (recoded)	983,790 (97.9%)	10,133,048 (98.1%)
No	0 (0.0%)	4,847,491 (46.9%)
Unknown	0 (0.0%)	1,185,926 (11.5%)
Missing	983,790 (97.9%)	4,099,631 (39.7%)

cohorts, all COVID-19 cases are confirmed with a positive PCR lab result. Each FDOH case is marked with the date it was confirmed, and each CDC case is marked with the date it was reported. Overall, there are 1,004,818 confirmed cases in the FDOH data, and 10,332,725 in the CDC data.

Signal Smoothing To smooth out daily fluctuations in data reporting, for each date, we compute the 7-day lagged averages for COVID-19 cases, hospitalizations, and deaths. From this point onward, any discussion of these quantities or fatality rates based on them, unless otherwise stated, refers to the smoothed signal. Thus, in both FDOH and CDC data, we collect data extending back to March 26th in order to analyze the April 1st to December 1st time range.

Separating Artifacts from True Improvements. We argue that three main phenomena fuel an “artificial” decrease in CFR: increased testing capacity (H1), shifting age distributions (H2), and delays between detection and fatality (H3).

Since testing (H1) is not included in the FDOH or CDC data, we pull in data from The COVID Tracking Project⁴¹ to quantify increased testing capacity in Florida and nationally. We plot 7-day lagged averages of tests administered and the positive test rates. To avoid artifacts from increased testing, we examine changes in *HFRs* rather than *CFRs*.

To establish and account for shifting age distributions (H2), we examine cases, hospitalizations, and deaths stratified by age groups: 0-9, 10-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, and 80+. Naturally, age stratification reduces the amount of data for each estimate, so we omit HFR estimates that are based on fewer than two deaths.

Finally, to account for delays between detection and fatality (H3), we take advantage of the line-level nature of the data in order to perform a cohort-based analysis. For each date, we extract the cohort of individuals confirmed positive on that date, as well as whether those individuals *eventually* died or were hospitalized. By contrast, publicly reported case fatality rates are typically not cohort-based^{1,7,8}—the patients whose deaths are reported in the numerator are not in general the same patients whose confirmed infections show up in the denominator. Because case confirmation tends to precede reported deaths, these signals tend to be misaligned and are subject to fluctuation, even if the actual case fatality rate were fixed (so long as incidence does change). Line-level data enables us to circumvent this problem.

Taking the above three adjustments into account, our primary quantity of interest for treatment improvements is the age-stratified HFR. For the rest of the paper, we define CFR and HFR at day t as follows:

$$\text{CFR}_t = \frac{\text{cases at day } t \text{ that eventually die}}{\text{cases at day } t}, \quad \text{HFR}_t = \frac{\text{cases at day } t \text{ that eventually get hospitalized and die}}{\text{cases at day } t \text{ that eventually get hospitalized}}$$

Quantifying True Improvements Thus far, news and academic sources have highlighted three main “true improvements”: improvements in treatment (H4), disease mutations (H5), and reduced viral loads due to social distancing (H6). We seek to quantify treatment improvements (H4) by computing the drop in HFR.

Estimation using Block-Bootstrap and Cubic Splines. We use a cubic spline to fit the trend of the 7-day lagged average HFRs, and use a moving block-bootstrapping technique with post-blackening³⁸ to estimate uncertainty around this trend. As described in the related work, 7-day lagged averages provide better estimate of trend by assuming smoothness rather than a specific functional form. Block-bootstrapping with post-blackening enables us to estimate uncertainty around this trend, with a weaker assumption than the standard i.i.d. assumption. We use a 7-day block size, based on the length of the time series⁴² and our observations that reporting follows a weekly cadence. After block-bootstrap resampling the residuals, the residuals are added back to the cubic spline, creating the replicates needed for estimating HFR with uncertainty. A visual walkthrough of this procedure is in the “HFR estimation” section of our web tool.

Visualization Tool. We publish an interactive web tool, available at acmilab.org/unpack_cfr, for dynamically applying our analyses to any demographic or date range of interest. On both FDOH data and CDC data, it displays plots for aggregate and age-stratified cases, hospitalizations, and deaths over time (Figure 3); plots for age distributions of cases, hospitalizations, and deaths over time (Figure 5); and estimates of age-stratified HFR as well as the change between two user-provided dates (Table 2 and 3). The user can select gender, race/ethnicity, and state to form cohorts of interest. For HFR estimates, the user can use a date selector to obtain new estimates for their date range of interest.

Results

Cases, Hospitalizations, and Deaths

In both the FDOH and the CDC data, one can discern three waves of COVID-19 cases. The first wave peaks in mid-April, the second wave peaks in mid-July, and cases leading up to December indicate a third wave (Figure 3).

Increased Testing. Between April 1st and December 1st, testing increases significantly, by approximately 964% in Florida and 1080% nationally (Figure 2). Florida observes a spike in testing near the second peak, whereas national testing rises more smoothly. We note that although peaks in Florida testing (Figure 2, left) coincide with spikes in Florida cases (Figure 3a, left), these spikes in cases cannot be *entirely* attributed to increased testing because there are also rises in positive test rates in April, July, and December (Figure 2, right).

Cases. Across all age strata, as measured by cases, Florida’s second wave is the most severe (Figure 3a, left) out of the three waves. In aggregate, it has approximately 1153% more cases than in the first peak and 46% more cases than in the ongoing third wave (Figure 3a, left). In contrast, nationally the third wave has substantially more cases than the first two peaks—392% more than the first peak and 150% more than the second peak (Figure 3b, left). Also, note that the relative jump in cases between the first two peaks is 96%, much less than the 1153% jump seen in Florida. This could be due to a combination of the spike in Florida’s testing in the second peak, as well as variation in the trajectories of different states (e.g. the populous state of New York was particularly hard-hit in the first wave).

Hospitalizations and Deaths. Overall, hospitalizations and deaths corroborate the story told by positive test rate (Figure 3a, center and right). In Florida, hospitalizations and deaths indicate a more severe second peak than first peak, though the contrast in peak size is not as dramatic as in the plot of cases. By contrast, in the national data, the second peak is *smaller* than the first. Much of the discrepancies of trends seen in cases versus in hospitalizations and deaths are likely attributable to increases in testing (Figure 2). Towards the third wave in December, Florida hospitalizations and deaths are at similar levels to that of the first wave. Nationally, the third wave appears to be worse than the second.

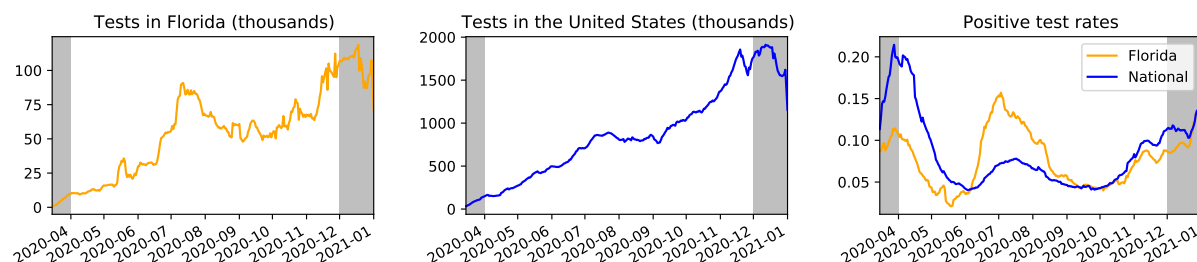


Figure 2: COVID-19 positive test rates (right) and tests (left and middle) for Florida and the United States, calculated using 7-day trailing averages and pulled from the COVID Tracking Project⁴¹. Positive test rate is calculated by dividing new positives by total new tests on each day. Data outside the April 1st to December 1st time range considered in this study is grayed out.

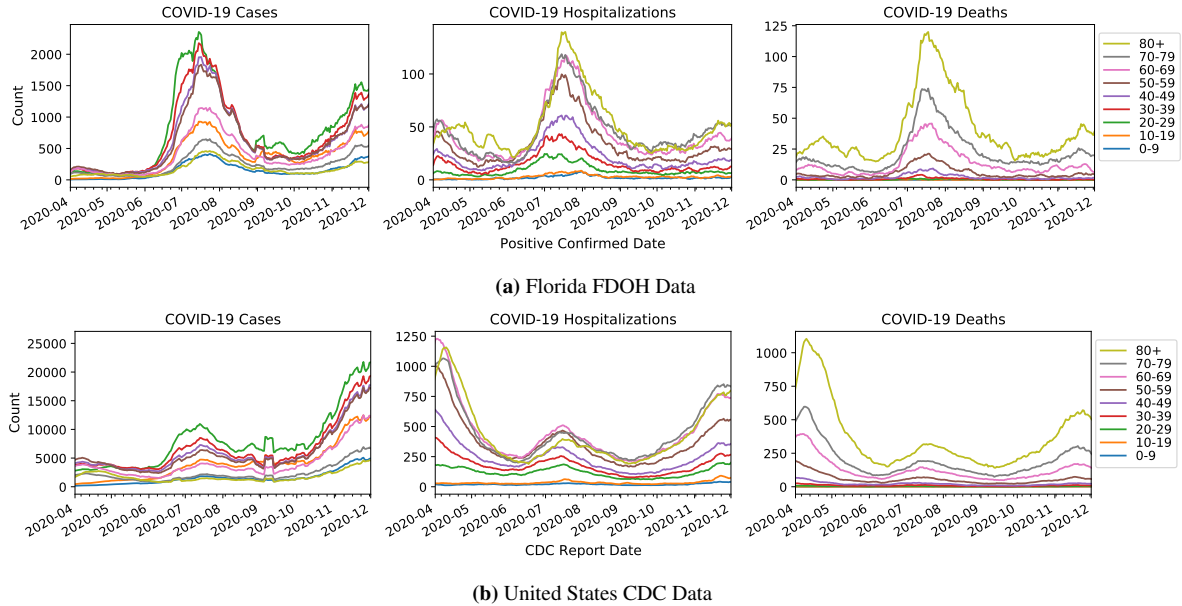


Figure 3: Age-stratified cases, (eventual) deaths, and (eventual) hospitalizations in Florida and in the U.S., by the date of first positive test result (Florida) and date of report to the CDC (U.S.). Note that the x axis is *not* the date of death or date of hospitalization.

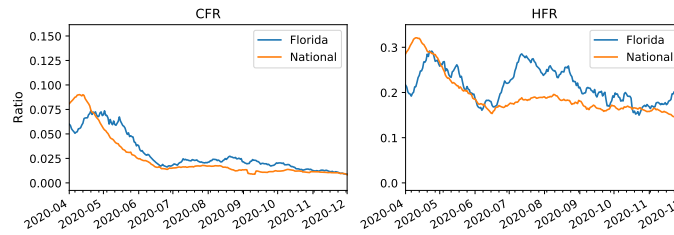


Figure 4: Aggregate case fatality rate (left) and hospitalization fatality rate (right) in Florida and in the U.S., by the date of first positive test result (Florida) or date of report to the CDC (U.S.).

Aggregate Fatality Rates. As expected, aggregate CFRs are lower than aggregate HFRs both in Florida and nationally (Figure 4). Unlike the national CFR which primarily decreases across the time range, the Florida HFR oscillates and reaches similar levels in both April and July. A deeper analysis of age-stratified fatality follows in later sections.

Confounding Due to Demographic Shift

Age. Between the first two peaks, the age distribution of cases shifts substantially, with the median age in Florida falling from 51 to 40, and the median national age group, from 50-59 to 30-39. After the second peak, the age distributions of continue to fluctuate. In September, younger cases increase, possibly related to the start of the school year (Figure 5, left). By December 1st, the Florida median age remains at 40 but the national median age group rises to 40-49. Older individuals comprise a disproportionate share of hospitalization and death counts (Figure 5, middle and right).

Gender. The gender ratios in each age group’s cases, hospitalizations, and deaths appear relatively flat over time. Thus, in this paper we choose not to stratify by gender due to the reasonably small shift in the gender distribution over time, and practically to have more support in each group. However, we do provide this option in our web tool.

Age-stratified HFR

Consistent with prior literature,³³ we find that as the age group increases so does the corresponding HFR (Tables 2 and 3). Measuring treatment improvements by HFR drop (computed as $\frac{HFR_{new} - HFR_{old}}{HFR_{old}}$), we observe larger treatment

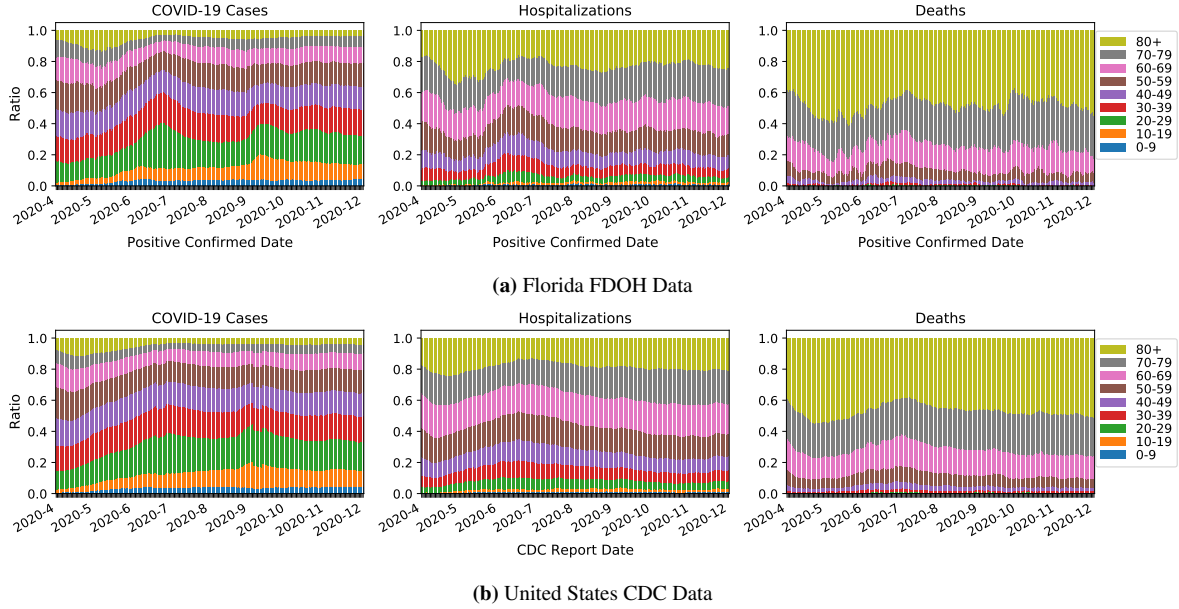


Figure 5: Age distributions among Florida and national cases, (eventual) hospitalizations, and (eventual) deaths, by the date of first positive test result (Florida) and date of report to the CDC (U.S.), respectively.

Table 2: Estimates of HFR and drop in HFR on peak dates. Median and 95% confidence intervals are computed using block bootstrapping. Results with inadequate support are omitted.

Age group	Florida			National		
	2020-04-15	2020-07-15	04-15 to 07-15	2020-04-15	2020-07-15	04-15 to 07-15
aggregate	0.23 (0.21, 0.26)	0.23 (0.21, 0.24)	-0.023 (-0.14, 0.13)	0.29 (0.28, 0.3)	0.17 (0.17, 0.18)	-0.41 (-0.44, -0.37)
20-29	-	-	-	0.021 (0.018, 0.025)	0.014 (0.012, 0.016)	-0.34 (-0.48, -0.16)
30-39	-	-	-	0.044 (0.041, 0.047)	0.027 (0.025, 0.03)	-0.37 (-0.44, -0.3)
40-49	-	-	-	0.079 (0.074, 0.084)	0.056 (0.053, 0.059)	-0.29 (-0.35, -0.22)
50-59	0.092 (0.078, 0.11)	0.1 (0.093, 0.11)	0.12 (-0.085, 0.38)	0.15 (0.14, 0.15)	0.1 (0.096, 0.1)	-0.31 (-0.35, -0.27)
60-69	0.18 (0.15, 0.21)	0.21 (0.19, 0.22)	0.13 (-0.045, 0.38)	0.26 (0.25, 0.27)	0.18 (0.18, 0.19)	-0.3 (-0.33, -0.26)
70-79	0.31 (0.28, 0.34)	0.33 (0.31, 0.34)	0.034 (-0.078, 0.18)	0.4 (0.39, 0.42)	0.27 (0.26, 0.27)	-0.34 (-0.37, -0.31)
80+	0.46 (0.43, 0.49)	0.48 (0.46, 0.49)	0.029 (-0.055, 0.12)	0.57 (0.55, 0.59)	0.41 (0.4, 0.42)	-0.27 (-0.3, -0.24)

improvements between April and December to be correlated with younger age (Table 3). Note, however, that the younger groups have small HFRs to begin with, so the opposite trend might appear when considering *absolute* rather than *relative* improvements. Additionally, the confidence intervals for HFR drops are wider for younger age groups.

Between the first two peaks (Table 2), the national age-stratified HFR estimates from block bootstrapping decrease by as little as 27% in the 80+ age group, and as much as 37% in the 30-39 age group. On the other hand, in Florida the age-stratified HFR actually *increases* in each age group by as little as 2.9% in the 80+ age group, and as much as 13% in the 60-69 age group. Note that the HFR changes between peak dates in Florida are an example of Simpson’s paradox, where in each age group the HFR increase, but the aggregate HFR actually decreases by 2.3%.

Compared to peak-to-peak changes, across the entire time range (Table 3) we observe a more dramatic decrease in HFR. In Florida, the HFR drops by as little as 17% in the 80+ age range, and as much as 42% in the 60-69 age range. Nationally, the HFR drops by as little as 55% in the 80+ age groups, and as much as 73% in the 20-29 age group.

While this paper presents estimates at the two peaks and the endpoints of the study time range, we can easily read off similar estimates with uncertainty for all dates between April 1st and December 1st. This type of interactive functionality is available in our web tool. In our web tool, when stratifying by gender in addition to age, the conclusions surrounding drops in HFR are similar to those when just stratifying by age.

Table 3: Estimates of HFR and drop in HFR between April 1st and December 1st. Median and 95% confidence intervals are computed using block bootstrapping. Results with inadequate support are omitted.

Age group	Florida			National		
	2020-04-01	2020-12-01	04-01 to 12-01	2020-04-01	2020-12-01	04-01 to 12-01
aggregate	0.23 (0.2, 0.27)	0.16 (0.12, 0.19)	-0.33 (-0.52, -0.094)	0.34 (0.32, 0.35)	0.13 (0.11, 0.15)	-0.61 (-0.67, -0.56)
20-29	-	-	-	0.025 (0.019, 0.03)	0.0066 (0.0014, 0.012)	-0.73 (-0.95, -0.43)
30-39	-	-	-	0.049 (0.045, 0.054)	0.019 (0.014, 0.025)	-0.61 (-0.72, -0.48)
40-49	-	-	-	0.087 (0.08, 0.095)	0.031 (0.024, 0.038)	-0.65 (-0.74, -0.54)
50-59	-	-	-	0.16 (0.15, 0.17)	0.06 (0.05, 0.07)	-0.63 (-0.69, -0.55)
60-69	0.18 (0.14, 0.22)	0.1 (0.06, 0.14)	-0.42 (-0.69, -0.096)	0.28 (0.27, 0.3)	0.11 (0.097, 0.13)	-0.61 (-0.66, -0.55)
70-79	0.31 (0.26, 0.35)	0.19 (0.14, 0.23)	-0.38 (-0.57, -0.17)	0.45 (0.43, 0.47)	0.18 (0.16, 0.2)	-0.6 (-0.65, -0.54)
80+	0.44 (0.39, 0.49)	0.37 (0.32, 0.41)	-0.17 (-0.32, -0.0038)	0.62 (0.59, 0.64)	0.28 (0.25, 0.31)	-0.55 (-0.59, -0.49)

Discussion

We unpack the drop in CFR to quantify improvements reasonably attributable to advances in treatment, accounting for shifting age distributions (H1) by age-stratifying, increased testing capacity (H2) by focusing on the hospitalized, and the detection-to-fatality delay (H3) by conducting a cohort-based analysis. We find that increased testing does not explain away the three waves due to corresponding peaks in hospitalizations, deaths, and positive test rates. We visualize the shifting age distributions, and quantify the decrease in age-stratified HFRs between the first two peaks and across the entire study time range. Combining all these analyses, we arrive at the following narrative:

At the beginning of April, testing was relatively sparse (Figure 2). Cases, hospitalizations, and deaths were rising, and reached peak levels circa April 15th (Figure 3). Roughly one in every ten tests came back positive in Florida, and one in every five tests, nationally. In Florida, the aggregate HFR was approximately 23%, with age-stratified HFRs ranging between 9.2% for the 50-59 age group to 46% for the 80+ age group (Table 2). Nationally, the aggregate HFR was approximately 29%, with the age-stratified HFRs ranging between 2.1% for the 20-29 age group and 57% for the 80+ age group (Table 2). In each age group, the national HFR was higher than the Florida HFR, possibly due to overwhelmed hospital systems in states hit hard during the first wave. In fact, our web tool indicates that 34.5% of national CDC cases between April 1st and April 15th were recorded in New York alone.

Over the next three months, the proportion of younger individuals with COVID-19 grew steadily (Figure 5). Testing continued to rise nationally, and spiked in Florida as it approached a heavier second peak around July 15, with positive test rates also at an all-time high (Figure 2). Florida experienced record hospitalizations and deaths, and the age-stratified HFRs were at least as high as in the first wave (Table 2). While Bill Gates had publicly attributed “a factor-of-two improvement in hospital outcomes” to dexamethosone and remdesivir,⁹ this did not yet appear to be true in Florida. (Alternatively, treatment improvements might have been counterbalanced by strain on the hospital system.) On the other hand, cases in New York had diminished (shown in our web tool) and were starting to surge in other states, forming a smaller second peak nationally (as measured by hospitalizations and deaths). Between the first two peaks, the national HFR had dropped by 41% in aggregate, with age-stratified HFRs dropping as much as 37% in the 30-39 age group and as little as 27% in the 80+ age group. The different stories told here by Florida and the national aggregate data underscore the importance of state-level rather than national analysis.

Finally, come December 1st, a third wave is underway. Approximately 31% of all national cases since April were confirmed in the last month alone (Figure 3b). In terms of hospitalizations, deaths, and positive test rate, this third wave has already surpassed the second wave. For Florida, the third wave is already at least as severe as the first wave. Fortunately, age-stratified HFRs in both Florida and the national aggregate data appear to have dropped significantly since the start of the pandemic, likely indicating treatment improvements (though possibly confounded by H5 and H6). Since April 1st, the age-stratified HFR in Florida has decreased by as much as 42% in the 60-69 age group and as little as 17% in the 80+ age group. Nationally, the age-stratified HFR has decreased by as much as 73% in the 20-29 age group and as little as 55% in the 80+ group. Regarding the CFR, on July 27, former President Donald Trump stated in a press briefing that “Due to the medical advances we’ve already achieved and our increased knowledge in how to treat the virus, the mortality rate for patients over the age of 18 is 85 percent lower than it was in April.”⁴³ Note, however, that none of our estimates of treatment improvements are as large as the 85% touted by Trump. In summary, CFR can be misleading if age distribution shift, increased testing, and detection-to-death delay are unaccounted for.

Limitations

We aim to quantify treatment improvements (H4) in Florida and the U.S. by estimating changes in the age-stratified HFR, but H4 could also be influenced by disease mutation (H5) and changing viral loads (H6). To distinguish their effects in future work, we need additional data. Furthermore, while we listed the six hypotheses we found in literature review, possible alternative explanations may arise in the future as pandemic evolves.

We assume that treatment improvements will be reflected in the HFR because over our study's time range, major treatment improvements (e.g., dexamethasone and remdesivir) targeted hospitalized patients. While the first U.S. vaccination was administered outside of our study time range, vaccines take effects before hospitalization, and so their treatment improvements may not be reflected in the HFR for future studies. While our method could still quantify post-hospitalization treatment improvements, we note that vaccination roll-out criteria (e.g. occupation, age) and other characteristics (e.g. socioeconomic background) could influence who gets hospitalized in the first place.

Other limitations arise from data quality issues. In both the FDOH and CDC data, hospitalization and death are highly missing (Table 1), possibly introducing bias in the HFR estimates if the data are not missing at random. States have different patterns when reporting data to the CDC (seen by filtering for any state in our web tool). The reported CDC cases appear to be incomplete. For instance, the CDC cases reported from Florida only account for 67% of the cases provided by the FDOH, they are reported sporadically even after smoothing, and no death is reported since October. Cross-referencing with the COVIDcast API, we find that in the CDC data reported from Texas, only 4.9% of the cases and 0.01% of the deaths are accounted for¹⁶, and only 14 hospitalizations were recorded. Thus, in our national analysis, we are making the assumption that in aggregate the signal will outweigh the noise. Despite the data limitations, the CDC data appears to be the best available source of line-level cases needed for cohort-based analysis across the United States. We note that in the Florida FDOH data, on the other hand, we use the positive test confirmed date which is not missing at all in this data, making the Florida HFR estimates more reliable than those from the national data.

References

- [1] Derek Thompson. Covid-19 cases are rising, so why are deaths flatlining?, 2020.
- [2] M. Whet. Why Changing COVID-19 Demographics in the US Make Death Trends Harder to Understand, 2020.
- [3] Leora Horwitz and et al. Trends in covid-19 risk-adjusted mortality rates in a single health system. *medRxiv*, 2020.
- [4] John Dennis and et al. Improving covid-19 critical care mortality over time in england: A national cohort study, march to june 2020. *medRxiv*, 2020.
- [5] Guihong Fan and et al. Decreased case fatality rate of covid-19 in the second wave: A study in 53 countries or regions. *Transboundary and emerging diseases*, 2020.
- [6] Alexis Madrigal and Whet Moser. How many americans are about to die?, 2020.
- [7] Piotr Spychalski and et al. Estimating case fatality rates of covid-19. *The Lancet. Infectious Diseases*, 2020.
- [8] Alexis Madrigal. A second coronavirus death surge is coming, 2020.
- [9] Steven Levy. Bill Gates on Covid: Most US Tests Are 'Completely Garbage', 2020.
- [10] John H Beigel and et al. Remdesivir for the treatment of covid-19—preliminary report. *NEJM*, 2020.
- [11] RECOVERY Collaborative Group. Dexamethasone in hospitalized patients with COVID-19. *NEJM*, 2020.
- [12] Wesley H Self and et al. Effect of hydroxychloroquine on clinical status at 14 days in hospitalized patients with covid-19: A randomized clinical trial. *JAMA*, 2020.
- [13] Maria Pachetti and et al. Impact of lockdown on covid-19 case fatality rate and viral mutations spread in 7 countries in europe and north america. *Journal of Translational Medicine*, 18(1):1–7, 2020.
- [14] S El Zein. Declining trend in the initial sars-cov-2 viral load over time: Observations from detroit, michigan.
- [15] Chiara Piubelli and et al. Overall decrease in sars-cov-2 viral load and reduction in clinical burden: the experience of a hospital in northern italy. *Clinical Microbiology and Infection*, 2020.

- [16] CMU Delphi Project, 2020.
- [17] Jason Phua and et al. Intensive care management of coronavirus disease 2019 (COVID-19): challenges and recommendations. *The Lancet Respiratory Medicine*, 8(5):506–517, 2020.
- [18] Pieter Cohen, Joann G Elmore, Jessamyn Blau, and Allyson Bloom. *Coronavirus disease 2019 (COVID-19): Outpatient evaluation and management in adults*. UpToDate, 2020.
- [19] Lone Wulff Madsen et al. Remdesivir for the treatment of covid-19-final report. *NEJM*, 2020.
- [20] Peter Horby and et al. Effect of hydroxychloroquine in hospitalized patients with covid-19: Preliminary results from a multi-centre, randomized, controlled trial. *MedRxiv*, 2020.
- [21] Anup Agarwal and et al. Convalescent plasma in the management of moderate covid-19 in adults in india: open label phase ii multicentre randomised controlled trial (placid trial). *British Medical Journal*, 371, 2020.
- [22] U.S. FDA. COVID-19 Update: FDA Authorizes Monoclonal Antibodies for Treatment of COVID-19. Nov 2020.
- [23] Owen Dyer. Covid-19: Eli Lilly pauses antibody trial for safety reasons, 2020.
- [24] Peter Chen and et al. Sars-cov-2 neutralizing antibody ly-cov555 in outpatients with covid-19. *NEJM*, 2020.
- [25] Regeneron. Regeneron’s casirivimab and imdevimab antibody cocktail for covid-19 is first combination therapy to receive fda emergency use authorization. Nov 2020.
- [26] HHS Launches Web-Based Locator for COVID-19 Outpatient Treatment Sites for Monoclonal Antibodies. 2021.
- [27] The U.S. Food and Drug Administration (FDA). Pfizer-BioNTech COVID-19 Vaccine, 2021.
- [28] The U.S. Food and Drug Administration (FDA). Moderna COVID-19 Vaccine, 2021.
- [29] Fernando P Polack and et al. Safety and efficacy of the bnt162b2 mrna covid-19 vaccine. *NEJM*, 383(27), 2020.
- [30] Lindsey R Baden and et al. Efficacy and safety of the mrna-1273 sars-cov-2 vaccine. *NEJM*, 2020.
- [31] Graziano Onder and et al. Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy. *JAMA*, 323(18):1775–1776, 05 2020.
- [32] Ned Stafford. Covid-19: Why germany’s case fatality rate seems so low. *BMJ*, 369, 2020.
- [33] Vikas Mehta and et al. CFR of cancer patients with COVID-19 in a NY hospital system. *Cancer discovery*, 10(7), 2020.
- [34] Science brief: Emerging sars-cov-2 variants. *cdc.gov*, Jan 2021.
- [35] Yang Liu and et al. Neutralizing activity of bnt162b2-elicited serum. *NEJM*.
- [36] Kai Wu and et al. Serum neutralizing activity elicited by mrna-1273 vaccine-preliminary report. *NEJM*.
- [37] Gidon Eshel. *Spatio-temporal data analysis*. Princeton University Press, 2012.
- [38] A. C. Davison and D. V. Hinkley. *Bootstrap Methods and their Application*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 1997.
- [39] Florida Department of Health. Florida case line data, 2020.
- [40] CDC, COVID-19 Response. Covid-19 case surveillance data access, summary, and limitations, 2020.
- [41] Robinson Meyer and Alexis Madrigal, 2020.
- [42] Cosma Shalizi. Advanced data analysis from an elementary point of view, 2013.
- [43] Donald Trump. Remarks by president trump in press briefing on covid-19, Jul 2020.