

G OPEN ACCESS

Citation: Greenwald SD, Chamoun NG, Manberg PJ, Gray J, Clain D, Maheshwari K, et al. (2022) Covid-19 and excess mortality in medicare beneficiaries. PLoS ONE 17(2): e0262264. https:// doi.org/10.1371/journal.pone.0262264

Editor: Satya Surbhi, University of Tennessee Health Science Center, UNITED STATES

Received: April 7, 2021

Accepted: December 21, 2021

Published: February 2, 2022

Copyright: © 2022 Greenwald et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The individual patient data used to generate the findings of this study are available through restricted access to certain stakeholders as allowed by federal regulations and CMS policy. Requests for access to data to replicate these findings require a research protocol and Data Use Agreement. For more information, contact the Research Data Assistance Center (ResDAC) (http://www.resdac.org. Aggregate data supporting the reported results, tables and figures in this manuscript are available in an Excel workbook on Harvard Dataverse (https://dataverse. RESEARCH ARTICLE

Covid-19 and excess mortality in medicare beneficiaries

Scott D. Greenwald¹[•], Nassib G. Chamoun¹[•], Paul J. Manberg¹[•], Josh Gray¹[•], David Clain¹[•], Kamal Maheshwari^{2,3}[•], Daniel I. Sessler²[•]

 Health Data Analytics Institute, Dedham, MA, United States of America, 2 Department of Outcomes Research, Anesthesiology Institute, Cleveland Clinic, Cleveland, OH, United States of America,
 Department of General Anesthesiology, Anesthesiology Institute, Cleveland Clinic, Cleveland, OH, United States of America

• These authors contributed equally to this work.

* Scott.Greenwald@hda-institute.com

Abstract

We estimated excess mortality in Medicare recipients in the United States with probable and confirmed Covid-19 infections in the general community and amongst residents of long-term care (LTC) facilities. We considered 28,389,098 Medicare and dual-eligible recipients from one year before February 29, 2020 through September 30, 2020, with mortality followed through November 30th, 2020. Probable and confirmed Covid-19 diagnoses, presumably mostly symptomatic, were determined from ICD-10 codes. We developed a Risk Stratification Index (RSI) mortality model which was applied prospectively to establish baseline mortality risk. Excess deaths attributable to Covid-19 were estimated by comparing actual-toexpected deaths based on historical (2017-2019) comparisons and in closely matched concurrent (2020) cohorts with and without Covid-19. Overall, 677,100 (2.4%) beneficiaries had confirmed Covid-19 and 2,917,604 (10.3%) had probable Covid-19. A total of 472,329 confirmed cases were community living and 204,771 were in LTC. Mortality following a probable or confirmed diagnosis in the community increased from an expected incidence of about 4.0% to actual incidence of 7.5%. In long-term care facilities, the corresponding increase was from 20.3% to 24.6%. The absolute increase was therefore similar at 3-4% in the community and in LTC residents. The percentage increase was far greater in the community (89.5%) than among patients in chronic care facilities (21.1%) who had higher baseline risk of mortality. The LTC population without probable or confirmed Covid-19 diagnoses experienced 38,932 excess deaths (34.8%) compared to historical estimates. Limitations in access to Covid-19 testing and disease under-reporting in LTC patients probably were important factors, although social isolation and disruption in usual care presumably also contributed. Remarkably, there were 31,360 (5.4%) fewer deaths than expected in community dwellers without probable or confirmed Covid-19 diagnoses. Disruptions to the healthcare system and avoided medical care were thus apparently offset by other factors, representing overall benefit. The Covid-19 pandemic had marked effects on mortality, but the effects were highly context-dependent.

harvard.edu), https://doi.org/10.7910/DVN/ GFBLK5.

Funding: This study was funded by Health Data Analytics Institute (HDAI) (www.hdainstitute.com). The funder provided support in the form of salaries for authors SDG, NCG, PJM, JG and DC, and HDAI consultant DIS, but did not have any additional role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. The specific roles of these authors are articulated in the 'author contributions' section. The Robert Wood Johnson Foundation provided support for access to the Covid-19 RSI365 risk prediction models by academic researchers for non-commercial purposes via the HDAI API (https://www.hda-institute.com/application-foruse-of-the-hdai-api/).

Competing interests: We have read the journal's policy and the authors of this manuscript have the following competing interests: Greenwald, Chamoun, Manberg, Gray and Clain are employees of and hold equity positions in Health Data Analytics Institute. Dr. Sessler is a paid consultant and holds equity interest in Health Data Analytics Institute. Dr. Maheshwari received no compensation from Health Data Analytics Institute. This does not alter our adherence to PLOS ONE policies on sharing data and materials. There are no patents, products in development or marketed products associated with this research to declare.

Introduction

The Covid-19 pandemic has profoundly influenced US healthcare, especially among Medicare recipients who are mostly at least 65 years old. By March 1, 2021, SARS-CoV-2, the virus responsible for Covid-19, had already infected more than 29 million US-Americans and more than 500,000 deaths associated with infection [1]. However, many people infected with Covid-19 are never tested or have false-negative test results; consequently, the true toll of Covid-19 remains uncertain. Furthermore, while the clinical course is sometimes apparent, Covid-19 also kills people by worsening chronic conditions, with those deaths often being attributed to other causes. Especially early in the pandemic, due to limited testing availability, it was difficult to differentiate deaths caused by Covid-19 from those that may have occurred naturally due to underlying health conditions. It is thus evident that many people who died consequent to Covid-19 infections may not have been diagnosed with the condition or may have died due to underlying causes.

Several teams have estimated "excess" mortality due to Covid-19 in the US population by comparing weekly observed death totals with those that occurred in a prior year or an average from multiple years. For example, Chen et al. estimated that from March 1 through August 22, 2020, 146,557 deaths were recorded in California, with an estimated 19,806 (95% CI, 16,364-23,210) deaths in excess of those predicted by historical trends [2]. Similarly, Faust et al. estimated that from March 1, 2020, to July 31, 2020, a total of 76,088 all-cause deaths occurred among US adults aged 25 to 44 years, which was 11,899 more than the expected 64,189 deaths based on a previous year (incident rate ratio, 1.19 [95% CI, 1.14-1.23]) [3]. Rossen et al. estimated excess mortality from January 26 through October 3 to have decreased 2% for the youngest subjects (aged <25 years) but increased 14.4% for those 45-64 years, 21.1% for those 65-74 years, 21.5% for those 75-84 years, and 14.7% in subjects ≥85 years old [4]. These reports suggest that all-cause mortality in the first six months of the pandemic increased by about 15– 20%. However, historical comparisons do not account for risk at an individual level which may be useful to determine true excess mortality. Furthermore, historical comparisons that incorporate trending over multiple years may provide a more robust baseline estimate of expected deaths over a given time and indicate that Covid had a similar impact in many other high-income countries [5, 6].

On a broad population basis, many risk factors for Covid-19 are now well recognized. For example, the US Centers for Disease Control (CDC) identifies eleven conditions that augment risk for severe forms of Covid-19 [7]. Chronic conditions such as cancer and dementia are reported to be among the most important contributors [8–10]. It is apparent that older members of the population are at special risk of death, although to some extent age may be a surrogate for accumulated comorbidities. However, it is difficult to extrapolate from population risk to individual risk since many people exhibit various combinations of risk factors for Covid-19 mortality, and individual risks attributable to each condition are not necessarily additive. A robust model that considers relevant individual conditions and predicts mortality risk from Covid-19 infections would therefore be valuable.

Numerous groups have proposed individual risk models based on clinical outcomes in various populations studied early in the pandemic (e.g., [11]); however, a consensus model has yet to emerge [12–14]. From a practical perspective, prediction models based on readily available administrative data (e.g., ICD-10 codes) will be most useful since more granular information extracted from clinical health records are neither universally available nor easy to obtain. Our primary goal was therefore to estimate excess risk-specific mortality in people with probable and confirmed Covid-19 infections. An additional consequence of the Covid-19 pandemic has been public health quarantines that have severely disrupted healthcare delivery. The virus may therefore also have caused indirect mortality because patients with chronic diseases and acute exacerbations avoided seeking care due to fear of infection or because health services were overwhelmed or otherwise limiting access [15, 16]. Furthermore, stress related to isolation could increase medical morbidity [17, 18]. In addition, some causes of death such as accidents and homicides may have increased [19, 20], but not among the elderly. The extent to which delayed and disrupted healthcare for non-Covid-related conditions, along with pandemic-related behavioral changes, contribute to mortality remains unclear. Our secondary goal was therefore to estimate whether changes in mortality occurred in people without probable or confirmed Covid-19 infections.

Because Covid-19 is especially lethal in older people, we considered stratification by various age ranges for both our primary and secondary analyses. We also separately considered people residing in the community from those in Long-Term Care facilities, who are expected to have a higher baseline mortality risk and thus may be especially susceptible to Covid-19 infections.

Methods

Data analysis was conducted on the Centers for Medicare and Medicaid Services (CMS) Research Identifiable File (RIF) data using SAS Enterprise Guide (Version 7.15) under a special Data Use Agreement (DUA). Data was handled consistent with this agreement which 1) prohibits identification of any individual in the database, and 2) requires suppression of metrics in downloaded tables for populations smaller than 11 individuals. This project was determined to be exempt from informed consent requirements by the New England Institutional Review Board. Final data analysis of the full cohort was conducted from January 10 to March 11, 2021. This study followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cohort studies [21].

Individual subject data used for our analysis are available to certain stakeholders as allowed by federal regulations and CMS policy. Requests for access to data to replicate these findings require an approved research protocol and DUA with CMS. For more information, contact the Research Data Assistance Center (ResDAC, http://www.resdac.org). Aggregate data supporting the reported results, tables and figures in this manuscript are available in an Excel workbook on Harvard Dataverse (https://dataverse.harvard.edu), https://doi.org/10.7910/ DVN/GFBLK5.

Mortality predictions and outcomes were referenced to an anchor date of Feb 29, 2020, just before the initial wave of documented Covid-19 cases and the week when the first case of potential community spread Covid-19 was reported by CDC. We recognize that scatter undiagnosed cases may have occurred previously.

We used full Medicare fee-for-service and dual eligible (Medicaid and Medicare) files one year before the anchor date through September 30, 2020, with mortality outcomes reported through November 30th, 2020 for the primary analysis of Covid-19 outcomes. This dataset contains longitudinal health records for beneficiaries from across the entire US. We identified beneficiaries with confirmed Covid-19 diagnoses consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. Presumably most subjects with Covid-19 diagnoses were symptomatic, although some may have been tested and received a billing code because of risk or exposure. The confirmed Covid-19 diagnostic code U07.1 requires documentation of positive Covid-19 testing but does not specify which type of test was used.

Demographic characteristics (age, sex, ethnicity, location of care, zip code derived measures) and Medicare coverage information (dates of coverage, enrollment history) were extracted for each subject. Beneficiaries were classified as belonging to the long-term care/ skilled nursing facility (LTC/SNF) cohort if their claims history indicated that they had received services in a LTC/SNF setting at any time in February 2020. The remaining beneficiaries were designated as community dwelling.

We included all Medicaid and Medicare participants alive as of the study anchor date (N = 65,310,173). We excluded beneficiaries who had: 1) ages outside 18–99 years (214,767); 2) non-continuous coverage of Medicare Part A or B (12,445,567) because our predictor assumes a complete medical record in the year prior to the date of prediction; 3) records from insurance programs not in our source database at the time of the analysis [(i.e., Medicare Part C (Medicare Advantage program) coverage (23,633,391)]; 4) missing any variable used in the analysis tabulated in results (i.e., age, sex, race, low income or disability status, median household home of beneficiary's zip code, or RSI) (509,932); and, 5) inconsistent data from source files comprising the medical records [(i.e., inconsistent birth dates, death dates, sex, or race information (117,418)]. The resulting 28,389,098 beneficiaries included 677,100 with confirmed and 2,917,604 with probable diagnosis of Covid-19 (S1 Fig). Mortality was assumed to have occurred on the date-of-death listed in the CMS Common Medicare Environment which is continuously updated from various sources including the Social Security Administration. The causes of death were not available for this study.

Risk stratification

We used an adaptation of the Risk Stratification Index (RSI) to predict nine-month mortality using the prior year's administrative claims. Nine-month mortality was selected based on the availability of curated data at the time of analysis. A model was developed using the full fee-for-service 2018 population for training, with prospective validation of performance on the 2019 dataset as previously described [25]. In brief, RSI calculates a probability of mortality by applying coefficients to sociodemographic factors (i.e., sex, age, Medicaid enrollment (dual eligibility) status) and to indicator variables consisting of: 1) individual diagnostic or procedural codes; and 2) composite classes of clinical conditions or procedures. Specifically, the medical record is represented by individual ICD-10-CM diagnostic and ICD-10-PCS procedure codes, and their respective composite codes as defined by Clinical Classifications Software Refined (CCSR) [26]. More detailed descriptions of the model and prospective testing performance results are provided in S1 File and S2 Fig, respectively. The resulting RSI model was then prospectively applied to all eligible Medicare or dual-eligible beneficiaries as of February 29, 2020 to derive individual RSI scores as of that date—that is, before Covid-19 infections were confirmed in the United States.

For comparative purposes, a second model was similarly developed to predict nine-month mortality from the presence of 27 individual chronic conditions as defined by CMS [27]. Specifically, logistic regression (stepwise selection using p-in of 10^{-3} , p-out of 10^{-2}) was used to select significant predictors from a pool of candidate features (i.e., 27 chronic conditions, age, sex and dual-enrollment status) to create a predictive model using the same training and prospective testing populations described in the previous paragraph. Performance using baseline RSI values or the individual chronic conditions model as predictors of outcomes in 2020 were then tested prospectively (S3 Fig).

Analysis and selection of study cohorts

We conducted a progression of complementary inquiries. To test our primary hypothesis, we first identified the main study cohorts of beneficiaries with diagnoses of probable or confirmed

Covid-19, and then subdivided them based on location of service (community or LTC/SNF) as of February 29, 2020. Within each cohort, we determined 9-month mortality between the anchor date and November 30, 2020.

Our approach was to first define associations between baseline demographic characteristics and health status as characterized by RSI with the risk of mortality following a Covid-19 diagnosis in the overall at-risk population and in pre-defined subpopulations. We initially compared differences in mean baseline RSI scores between survivors and non-survivors, then used univariable and multivariable regression modeling to estimate the relative importance of baseline demographic factors, chronic conditions, and RSI scores as independent predictors of mortality. A similar analysis was conducted to identify risk factors associated with a confirmed diagnosis of Covid-19. We also determined the association between RSI and observed mortality by beneficiary age group and location.

Estimation of expected mortality

Two independent methods were used to estimate expected 2020 mortality in our study population. A historical comparison allowed us to compare year-over-year changes in mortality in Medicare recipients and thus characterize overall effects of Covid-19 and quarantine-induced restrictions in healthcare access on mortality. A case-matched analysis provided an alternate estimate of Covid-19-related excess mortality within the 9 months of 2020 that we considered. A potential advantage of precise case matching (i.e., "digital twins") to form a concurrent control population with similar underlying mortality risk is that it allows attribution of Covid-19 infection as the primary factor responsible for any observed differences in outcome during the study period. This therefore complements the historical comparisons that may require adjustments for variations in population size, demographics, or mortality trends [5].

Historical comparison

As in previous studies [3, 4], the first approach estimated expected 2020 mortality figures from historical records. The daily observed mortality for the Medicare population from 2017–2019 was used to prepare a model with optimal fit to capture seasonality and account for annual trends using a three-year moving average adjustment (S2 Fig). This approach better estimates expected mortality than relying on a single year-over-year comparison because the model better captures year-to-year fluctuations consequent to severity of yearly influenza outbreaks and other factors. We calculated predicted mortality for each individual and designated the sum as the historically expected mortality in each subpopulation. Excess deaths thus equaled the difference between observed 2020 deaths and the historical projection of expected deaths (actual minus expected).

Case matching, digital twins

A second method used case matching or "digital twinning" to estimate excess mortality in exposed subjects compared to concurrent controls who had closely matched health profiles. Beneficiaries receiving a diagnosis of probable or confirmed Covid-19 were pairwise exactly matched 1:1 on Feb 29, 2020 with beneficiaries without a Covid-19 diagnosis based on sex, age (within 1-year), ethnicity, location of services in Feb 2020 (community or LTC/SNF), along with RSI as a propensity matching factor (within 0.1%). Because the eligible Medicare population is large, we successfully matched almost the entire infected population (98.8% overall, 91.22 to 99.77% among subgroups analyzed). Excess deaths were estimated as the difference between the observed number of deaths in probable or confirmed Covid-19 subjects and their matched non-Covid-19 digital twins over the concurrent period. Matching may be more

reliable than the historical comparison for estimating true excess mortality because it accounts for population variation over time and accounts for the impact of substantial disruptions in public health and lifestyle changes caused by the pandemic restrictions in 2020.

Statistical analysis

Patient demographic and clinical characteristics are summarized descriptively. Baseline characteristics were compared using *t* or χ^2 tests, as appropriate. Mortality rates within the study period post-Covid-19 diagnosis are reported as odds ratios with 95% confidence intervals. P values <0.05 defined statistical significance for both the primary and secondary outcomes. No adjustments were made for multiple comparisons. Sample size requirements were not estimated *a priori* because the intention was to include all qualifying 2020 beneficiaries available in the 100% nationwide Medicare files.

Results

Patient characteristics and outcomes

As of Feb 29, 2020, a total of 28,389,098 Medicare or dual eligible beneficiaries met inclusion criteria for this study. Among them, 677,100 (2.4%) beneficiaries had a diagnosis of confirmed Covid-19 and 2,917,604 (10.3%) had a diagnosis of probable Covid-19 during the study period (**S1 Fig**). Among the confirmed cases, 472,329 were in the Community group while 204,771 received care in a long-term care setting.

Tables 1 and 2 compare demographic and clinical profiles for various subgroups. Compared to survivors, patients who died after a Covid-19 diagnosis were older, more often male, not white, received Medicaid, lived in zip-codes associated with lower median income, received services in February 2020 in a long-term care facility, and had higher baseline risk of mortality as defined by RSI. Age and baseline RSI scores were both strongly related to risk of infection and adverse Covid-19 outcomes. Residence in a LTC/SNF location and presence of end-stage renal disease were strong risk factors for acquiring a confirmed diagnosis of Covid-19 (S4 Fig). As shown in S5 Fig, RSI scores were associated with increasing mortality in a consistent rank ordered manner across each age group, thereby suggesting that RSI provides a significant and sensitive measure of co-morbidities and mortality risk that is independent of age.

Mortality risk prediction

Fig 1 presents the relative importance of factors that contributed to mortality in both univariable and multivariable models. Quintiles of RSI, age, LTC/SNF services status, sex, and race were the factors most associated with relative risk of mortality. Status of lung cancer and end-stage renal disease appear to carry meaningful incremental risk after adjustment. Mortality prediction models based primarily on baseline RSI levels performed better than models based on the presence of individual chronic conditions for predicting mortality risk (S3 Fig). Case matching identified a cohort of beneficiaries from the general population who were closely matched with subjects who had a diagnosis of probable or confirmed Covid-19 based on their RSI scores as of Feb 29, 2020 (Table 3). Characteristics of matched and unmatched subjects are tabulated in S1 Table.

Excess mortality estimates

The distribution of observed and expected mortality by diagnosis, category, and location of care is presented in Fig 2. As expected, subjects with high baseline mortality risk in the LTC/ SNF cohort had actual mortality that exceeded all other groups. Those with confirmed Covid-

					Group Size (N,%)			
Characteristic	All Beneficiaries (28,389,098)	COVID Confirmed (677,100)	Pr < Z	Died (131,460)	Survived (545,640)	COVID Probable (2,917,604)	Pr < Z	Died (214,602)	Survived (2,703,002)
Age Group									
18-55	1,938,606 (6.8)	43,637 (6.4)	<.0001	3,070 (2.3)	40,567 (7.4)	210,430 (7.2)	< .0001	6,155 (2.9)	204,275 (7.6)
56–65	1,912,571 (6.7)	57,655 (8.5)	<.0001	7,630 (5.8)	50,025 (9.2)	224,790 (7.7)	< .0001	12,503 (5.8)	212,287 (7.9)
66–70	6,900,760 (24.3)	118,536 (17.5)	<.0001	12,145 (9.2)	106,391 (19.5)	643,015 (22.0)	<.0001	22,527 (10.5)	620,488 (23.0)
71–75	6,707,830 (23.6)	123,535 (18.2)	<.0001	16,740 (12.7)	106,795 (19.6)	645,540 (22.1)	< .0001	30,035 (14.0)	615,505 (22.8)
76-80	4,688,185 (16.5)	103,925 (15.3)	<.0001	20,067 (15.3)	83,858 (15.4)	476,297 (16.3)	<.0001	34,007 (15.8)	442,290 (16.4)
81-85	3,140,391 (11.1)	88,874 (13.1)	<.0001	22,686 (17.3)	66,188 (12.1)	333,813 (11.4)	<.0001	36,602 (17.1)	297,211 (11.0)
86–90	1,914,030 (6.7)	74,295 (11.0)	<.0001	23,529 (17.9)	50,766 (9.3)	224,030 (7.7)	<.0001	35,949 (16.8)	188,081 (7.0)
91–95	947,362 (3.3)	50,455 (7.5)	<.0001	18,665 (14.2)	31,790 (5.8)	125,819 (4.3)	<.0001	27,433 (12.8)	98,386 (3.6)
96–99	239,363 (0.8)	16,188 (2.4)	<.0001	6,928 (5.3)	9,260 (1.7)	33,870 (1.2)	<.0001	9,391 (4.4)	24,479 (0.9)
Sex									
Female	15,538,151 (54.7)	385,062 (56.9)	<.0001	68,726 (52.3)	316,336 (58.0)	1,655,208 (56.7)	<.0001	110,919 (51.7)	1,544,289 (57.1)
Male	12,850,947 (45.3)	292,038 (43.1)	<.0001	62,734 (47.7)	229,304 (42.0)	1,262,396 (43.3)	<.0001	103,683 (48.3)	1,158,713 (42.9)
Race/Ethnicity									
American Indian/ Alaskan Native	161,358 (0.6)	5,194 (0.8)	<.0001	1,171 (0.9)	4,023 (0.7)	17,312 (0.6)	< .0001	1,380 (0.6)	15,932 (0.6)
Asian/Pacific Islander	766,965 (2.7)	16,894 (2.5)	<.0001	3,679 (2.8)	13,215 (2.4)	57,742 (2.0)	<.0001	4,270 (2.0)	53,472 (2.0)
Black	2,308,293 (8.1)	97,484 (14.4)	<.0001	21,394 (16.3)	76,090 (13.9)	263,726 (9.0)	< .0001	20,962 (9.8)	242,764 (9.0)
Hispanic	1,577,063 (5.6)	65,953 (9.7)	<.0001	12,760 (9.7)	53,193 (9.7)	151,585 (5.2)	<.0001	9,898 (4.6)	141,687 (5.2)
Non-Hispanic White	22,806,998 (80.3)	477,687 (70.5)	<.0001	90,660 (69.0)	387,027 (70.9)	2,352,811 (80.6)	<.0001	175,365 (81.7)	2,177,446 (80.6)
Other	228,402 (0.8)	5,164 (0.8)	0.0001	1,004 (0.8)	4,160 (0.8)	20,356 (0.7)	<.0001	1,272 (0.6)	19,084 (0.7)
Unknown	540,019 (1.9)	8,724 (1.3)	<.0001	792 (0.6)	7,932 (1.5)	54,072 (1.9)	< .0001	1,455 (0.7)	52,617 (1.9)
Low Income/ Disabled Status									
Low Income or Disabled	8,222,944 (29.0)	329,258 (48.6)	< .0001	69,107 (52.6)	260,151 (47.7)	982,097 (33.7)	< .0001	82,852 (38.6)	899,245 (33.3)
Not Low Income/ Disabled	20,166,154 (71.0)	347,842 (51.4)	< .0001	62,353 (47.4)	285,489 (52.3)	1,935,507 (66.3)	< .0001	131,750 (61.4)	1,803,757 (66.7)
Median Household Income (Imputed from Zip code) (Quintile)									
20-40%	5,674,738 (20.0)	121,354 (17.9)	<.0001	23,044 (17.5)	98,310 (18.0)	520,937 (17.9)	<.0001	42,043 (19.6)	478,894 (17.7)
40-60%	5,679,860 (20.0)	122,059 (18.0)	<.0001	23,585 (17.9)	98,474 (18.0)	533,097 (18.3)	<.0001	41,889 (19.5)	491,208 (18.2)
60-80%	5,675,343 (20.0)	122,827 (18.1)	<.0001	23,660 (18.0)	99,167 (18.2)	599,106 (20.5)	< .0001	42,578 (19.8)	556,528 (20.6)
Highest 20%	5,679,506 (20.0)	152,382 (22.5)	<.0001	28,850 (21.9)	123,532 (22.6)	725,509 (24.9)	< .0001	42,077 (19.6)	683,432 (25.3)
Lowest 20%	5,679,651 (20.0)	158,478 (23.4)	<.0001	32,321 (24.6)	126,157 (23.1)	538,955 (18.5)	< .0001	46,015 (21.4)	492,940 (18.2)
Community vs LTC/ SNF									
Community	27,400,582 (96.5)	472,329 (69.8)	<.0001	69,488 (52.9)	402,841 (73.8)	2,676,502 (91.7)	<.0001	166,800 (77.7)	2,509,702 (92.8)
In Long-Term Care Facility or SNF	988,516 (3.5)	204,771 (30.2)	<.0001	61,972 (47.1)	142,799 (26.2)	241,102 (8.3)	< .0001	47,802 (22.3)	193,300 (7.2)
RSI (9mo)									
MEAN (SD)	0.032 (.071)	0.096 (.127)	<.0001	0.181 (.153)	0.075 (.110)	0.052 (.095)	<.0001	0.168 (.155)	0.042 (.082)
Mortality (Through Nov 30)									

Table 1. Comparison of baseline demographic characteristics of all study populations.

(Continued)

					Group Size (N,%	»)			
Characteristic	All Beneficiaries (28,389,098)	COVID Confirmed (677,100)	Pr < Z	Died (131,460)	Survived (545,640)	COVID Probable (2,917,604)	Pr < Z	Died (214,602)	Survived (2,703,002)
Died	1,045,326 (3.7)	131,460 (19.4)	< .0001	131,460 (100.0)	0 (0.0)	214,602 (7.4)	< .0001	214,602 (100.0)	0 (0.0)
Survived	27,343,772 (96.3)	545,640 (80.6)	< .0001	0 (0.0)	545,640 (100.0)	2,703,002 (92.6)	< .0001	0 (0.0)	2,703,002 (100.0)

Table 1. (Continued)

*p-values represent the level of significance comparing attributes between Confirmed and Probable Covid infected populations relative to all Beneficiaries. Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. The baseline risk of 9-month mortality defined by the Risk Stratification Index (RSI) calculated on February 29, 2020 was 3.2% in the entire population and significantly higher among those who died compared to those who survived.

https://doi.org/10.1371/journal.pone.0262264.t001

19 showed similarly increased mortality above expected levels in both the LTC/SNF and community setting. Among community dwelling subjects, mortality also exceeded expected risk in subjects with possible Covid-19.

There was an estimated excess of 130,702 (historical comparison method) or 101,482 (case matching method) deaths attributable to probable or confirmed Covid-19 across the full population in the 9 months of 2020 that we considered. In the matched analysis, half the deaths (50,793) occurred in patients with a confirmed diagnosis of Covid-19 and half (50,689) occurred in those with a probable Covid-19 diagnosis. In contrast, 31,360 fewer subjects without a Covid-19 diagnosis died than expected, representing a 5.4% mortality reduction (Table 4).

Public access

Our model is highly predictive for mortality in Medicare beneficiaries with documented Covid-19 infections. Because baseline RSI scores can help to identify Medicare beneficiaries at highest risk for mortality due to Covid-19, we make the models publicly available in the following formats:

- Access to RSI risk calculators will be provided free of charge for authorized non-commercial uses via the HDAI API website (https://www.hda-institute.com/application-for-use-ofthe-hdai-api/).
- Medicare beneficiaries or their health advocates may access their personalized health history and risk assessment by signing into Health Picture (https://my.healthpicture.com). Health picture is an easy-to-use tool that allows Medicare beneficiaries and their family members a way to access their health histories and understand their Covid-19 risks.
- 3. Coefficients for a public version of a one-year RSI mortality model are provided at–(<u>Risk</u> Stratification Index | Cleveland Clinic).

Discussion

Age, sex, care location, and comorbidities were significant predictors of mortality. The strongest individual predictor following a diagnosis of Covid-19 across all age categories, and in both community and long-term care settings was the integrated measure of patient coTable 2. Comparison of baseline demographic characteristics for no Covid-19, confirmed Covid-19 and probable Covid-19 populations in the community and LTC/SNF subgroups.

			St	udy group,	No. (%)			
		Communit	у			SNF/LTC		
Characteristic	No COVID (24,251,751)	COVID Confirmed (472,329)	COVID Probable (2,676,502)	P- value*	No COVID (542,643)	COVID Confirmed (204,771)	COVID Probable (241,102)	P- value*
Age Group								
18-55	1,663,926 (6.9)	36,472 (7.7)	200,110 (7.5)	<.0001	20,613 (3.8)	7,165 (3.5)	10,320 (4.3)	<.0001
56-65	1,595,016 (6.6)	40,231 (8.5)	205,372 (7.7)		35,110 (6.5)	17,424 (8.5)	19,418 (8.1)	
66-70	6,096,885 (25.1)	98,565 (20.9)	621,310 (23.2)		42,324 (7.8)	19,971 (9.8)	21,705 (9.0)	
71–75	5,881,587 (24.3)	99,089 (21.0)	618,427 (23.1)		57,168 (10.5)	24,446 (11.9)	27,113 (11.2)	
76-80	4,036,407 (16.6)	75,063 (15.9)	443,657 (16.6)		71,556 (13.2)	28,862 (14.1)	32,640 (13.5)	
81-85	2,626,817 (10.8)	55,687 (11.8)	294,796 (11.0)		90,887 (16.7)	33,187 (16.2)	39,017 (16.2)	
86-90	1,510,656 (6.2)	38,824 (8.2)	180,497 (6.7)		105,049 (19.4)	35,471 (17.3)	43,533 (18.1)	
91–95	683,500 (2.8)	22,331 (4.7)	90,337 (3.4)		87,588 (16.1)	28,124 (13.7)	35,482 (14.7)	
96–99	156,957 (0.6)	6,067 (1.3)	21,996 (0.8)		32,348 (6.0)	10,121 (4.9)	11,874 (4.9)	
Sex								
Female	13,146,555 (54.2)	256,524 (54.3)	1,498,825 (56.0)	<.0001	351,326 (64.7)	128,538 (62.8)	156,383 (64.9)	<.0001
Male	11,105,196 (45.8)	215,805 (45.7)	1,177,677 (44.0)		191,317 (35.3)	76,233 (37.2)	84,719 (35.1)	
Race/Ethnicity								
American Indian/Alaskan Native	136,028 (0.6)	4,440 (0.9)	16,237 (0.6)	<.0001	2,824 (0.5)	754 (0.4)	1,075 (0.4)	<.0001
Asian/Pacific Islander	683,510 (2.8)	11,804 (2.5)	53,768 (2.0)		8,819 (1.6)	5,090 (2.5)	3,974 (1.6)	
Black	1,902,790 (7.8)	64,609 (13.7)	236,985 (8.9)		44,293 (8.2)	32,875 (16.1)	26,741 (11.1)	
Hispanic	1,337,729 (5.5)	50,565 (10.7)	140,218 (5.2)		21,796 (4.0)	15,388 (7.5)	11,367 (4.7)	
Non-Hispanic White	19,516,975 (80.5)	329,288 (69.7)	2,157,239 (80.6)		459,525 (84.7)	148,399 (72.5)	195,572 (81.1)	
Other	199,925 (0.8)	3,860 (0.8)	19,064 (0.7)		2,957 (0.5)	1,304 (0.6)	1,292 (0.5)	
Unknown	474,794 (2.0)	7,763 (1.6)	52,991 (2.0)		2,429 (0.4)	961 (0.5)	1,081 (0.4)	
Low Income/Disabled Status								
LowIncome or Disabled	6,653,284 (27.4)	189,901 (40.2)	846,357 (31.6)	<.0001	258,305 (47.6)	139,357 (68.1)	135,740 (56.3)	<.0001
Not LowIncome/Disabled	17,598,467 (72.6)	282,428 (59.8)	1,830,145 (68.4)		284,338 (52.4)	65,414 (31.9)	105,362 (43.7)	
Median Household Income (Imputed from Zip code) (Quintile)								
20-40%	4 917 968 (20.3)	86 115 (18.2)	473 870 (17.7)	<.0001	114 479 (21.1)	35 239 (17.2)	47 067 (19.5)	<.0001
40-60%	4 913 071 (20.3)	84 952 (18.0)	487 217 (18.2)		111 633 (20.6)	37 107 (18.1)	45 880 (19.0)	
60-80%	4 848 516 (20.0)	85 749 (18.2)	553 075 (20.7)		104 894 (19.3)	37 078 (18.1)	46 031 (19.1)	
Highest 20%	4 701 144 (19.4)	105 045 (22.2)	675 456 (25.2)		100 471 (18.5)	47 337 (23.1)	50 053 (20.8)	
Lowest 20%	4 871 052 (20.1)	110 468 (23.4)	486 884 (18.2)		111 166 (20.5)	48 010 (23.4)	52 071 (21.6)	
RSI (9mo)								
MEAN (SD)	0.024 (.055)	0.048 (.086)	0.038 (.077)	<.0001	0.206 (.154)	0.206 (.137)	0.201 (.141)	<.0001
Mortality (Through Nov 30)								
Died	548 556 (2.3)	69 488 (14.7)	166 800 (6.2)	<.0001	150 708 (27.8)	61 972 (30.3)	47 802 (19.8)	<.0001
Survived	23 703 195 (97.7)	402 841 (85.3)	2 509 702 (93.8)		391 935 (72.2)	142 799 (69.7)	193 300 (80.2)	

*p-value: < 0.05 indicates the distribution of patients among subgroups (or the mean values of metrics) differ among the No Covid-19, Confirmed Covid-19 and Probable Covid-19 groups.

Subjects were categorized as "LTC/SNF" if they received services in either a Long-Term Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the "Community." Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. The baseline risk of 9-month mortality defined by the Risk Stratification Index (RSI) calculated on February 29, 2020 was multiple times higher among subjects who received services in LTC/SNF than in the community.

https://doi.org/10.1371/journal.pone.0262264.t002

Variable	Unadjusted Odds I	Ratio (95% CI)	Adjusted Odds F	Ratio (95% CI)
RSI9(mo) Highest 20% vs Lowest 20%	6 17.0 (16.3, 17.8)		8.38 (8.01, 8.78)	
Age: 96-99 vs 66-70	6.55 (6.32, 6.80)		3.25 (3.12, 3.38)	
	5.78 (5.53, 6.05)		4.04 (3.85, 4.23)	Н
Age: 91-95 vs 66-70	5.14 (5.01, 5.28)		2.50 (2.43, 2.58)	н
Age: 86-90 vs 66-70	4.06 (3.96, 4.16)		2.03 (1.98, 2.09)	4
RSI9(mo) 40-60% vs Lowest 20%	3.29 (3.13, 3.46)	н	2.61 (2.49, 2.75)	н
Age: 81-85 vs 66-70	3.00 (2.93, 3.08)		1.72 (1.68, 1.77)	
Services Setting: LTC/SNF vs Commun	il 2.52 (2.48, 2.55)		1.08 (1.06, 1.10)	
Dementia	2.23 (2.20, 2.26)		1.16 (1.13, 1.18)	4
Age: 76-80 vs 66-70	2.10 (2.05, 2.15)		1.50 (1.46, 1.54)	4
Lung Cancer	2.06 (1.97, 2.16)	н	1.48 (1.41, 1.56)	
Acute Myocardiac Infarction	2.05 (1.97, 2.14)		1.22 (1.16, 1.27)	H
Hypertension	1.92 (1.89, 1.96)		(,)	
RSI9(mo) 20-40% vs Lowest 20%	1.89 (1.80, 2.00)	Ĥ	1.66 (1.58, 1.75)	н
Heart Failure	1.83 (1.81, 1.85)	4	1.05 (1.04, 1.07)	1 N N N N N N N N N N N N N N N N N N N
Hip/Pelvic Fracture	1.82 (1.75, 1.89)		(,)	
Anemia	1.81 (1.79, 1.84)		1.06 (1.05, 1.08)	
Chronic Kidney Disease	1.79 (1.77, 1.82)		1.16 (1.14, 1.18)	i i
Atrial Fibrillation	1.79 (1.76, 1.82)		1.06 (1.04, 1.08)	
End-Stage Renal Disease	1.72 (1.67, 1.76)		1.30 (1.26, 1.34)	1 E
Chronic Obstructive Pulmonary Diseas	e 1.66 (1.64, 1.69)		1.10 (1.09, 1.12)	4
Stroke or TIA	1.65 (1.62, 1.69)		1.02 (1.00, 1.04)	4
Ischemic Heart Disease	1.53 (1.51, 1.55)		(,)	
Benign Prostatic Hyperplasia	1.53 (1.50, 1.55)		1.03 (1.01, 1.05)	4
Colorectal Cancer	1.51 (1.44, 1.58)	Ĥ	1.05 (1.01, 1.11)	4
Alzheimers Disease	1.46 (1.44, 1.48)	i i	1.07 (1.06, 1.09)	
Depression	1.44 (1.42, 1.46)	1	0.98 (0.96, 0.99)	
Age: 71-75 vs 66-70	1.37 (1.34, 1.41)	8	1.27 (1.24, 1.30)	
Age: 56-65 vs 66-70	1.34 (1.30, 1.38)	i i	0.85 (0.82, 0.88)	4
Prostate Cancer	1.33 (1.29, 1.37)		(,)	
Diabetes	1.29 (1.27, 1.30)	4	1.06 (1.04, 1.07)	
Sex: Male vs Female	1.26 (1.24, 1.27)		1.44 (1.42, 1.46)	4
Race: N.American Native vs White	1.24 (1.16, 1.33)	н	1.66 (1.55, 1.79)	н
Medicaid: Enrolled vs Not Enrolled	1.22 (1.20, 1.23)		0.94 (0.93, 0.96)	
Race: Black vs White	1.20 (1.18, 1.22)	4	1.31 (1.28, 1.34)	1
Race: Asian vs White	1.19 (1.15, 1.23)		1.31 (1.25, 1.36)	8
Acquired Hyperthyroidism	1.16 (1.15, 1.18)	•	(,)	
Endometrial Cancer	1.14 (1.05, 1.25)	Ĥ	1.16 (1.06, 1.27)	н
Hyperlipidemia	1.10 (1.08, 1.11)	r	0.98 (0.96, 0.99)	l'
Osteoporosis	1.09 (1.06, 1.11)		(,)	
Arthritis	1.04 (1.03, 1.05)		0.92 (0.90, 0.93)	4
Race: Other vs White	1.03 (0.96, 1.10)	Ĥ	1.18 (1.10, 1.27)	н
Race: Hispanic vs White	1.02 (1.00, 1.05)		1.28 (1.25, 1.31)	4
Glaucoma	0.98 (0.96, 1.00)	1	0.95 (0.93, 0.97)	-
Income: 40-60% vs Lowest 20%	0.94 (0.92, 0.96)		0.97 (0.95, 0.99)	1
Income: 60-80% vs Lowest 20%	0.93 (0.91, 0.95)		0.97 (0.95, 0.99)	1
Income: 20-40% vs Lowest 20%	0.92 (0.90, 0.94)		0.97 (0.96, 0.99)	
Income: Highest 20% vs Lowest 20%	0.91 (0.89, 0.93)	1	0.94 (0.93, 0.96)	4
Breast Cancer	0.91 (0.88, 0.94)		(,)	
Obesity	0.91 (0.90, 0.92)	4	1.15 (1.13, 1.17)	4
Cataracts	0.86 (0.85, 0.88)	+	0.91 (0.89, 0.92)	+
Asthma	0.84 (0.82, 0.86)	4	0.95 (0.93, 0.98)	
Age: 18-55 vs 66-70	0.66 (0.64, 0.69)	1	0.58 (0.55, 0.60)	н
Race: Unknown vs White	0.43 (0.40, 0.46)	н	0.94 (0.87, 1.01)	H
			_	
		0.25 1 2 4 8 16		0.25 1 2 4 8 16
	Morta	ality Less Likely Mortality More L	ikely Mo	rtality Less Likely Mortality More Like

Fig 1. Forest plot showing the relative risk and 95% CI of significant predictors of mortality of subjects with confirmed Covid-19. Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. Subjects were categorized as "LTC/SNF" if they received services in either a Long-Term Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the "Community." Predictors were assessed at baseline (February 29, 2020) and include quintiles of Risk Stratification Index (RSI), presence of chronic conditions, location of services (LTC/SNF vs Community), and demographic variables (i.e., age, sex, race, and quintiles of median household income imputed by zip code according to 2015 Census data.) Variables not remaining in the adjusted model are indicated by the presence of empty parenthesis under the adjusted odds ratio. RSI, age, and location of services were the strongest (unadjusted) predictors of mortality. RSI and age remain strong predictors following adjustment; however, risks associated with having chronic conditions were typically reduced when adjusted by the presence of RSI and other factors. Status of Lung cancer and end-stage renal disease appear to carry meaningful incremental risk after adjustment.

https://doi.org/10.1371/journal.pone.0262264.g001

morbidities, RSI. Although many individual chronic conditions were also significant risk factors in our unadjusted univariable analysis, the strength of these associations were substantially diminished after adjustment for the primary covariates of age, sex, race, location of services, and RSI. Our analysis is consistent with previous work showing that older age, non-white race

Characteristic				Study group	, No. (%)			
		Con	nmunity			LTC	/SNF	
	COVID	Confirmed	COVID	Probable	COVID C	Confirmed	COVID	Probable
	Infected (470,622)	Control (470,622)	Infected (2,667,931)	Control (2,667,923)	Infected (186,671)	Control (186,671)	Infected (225,970)	Control (225,970)
Age Group								
18-55	35,622 (7.6)	35,622 (7.6)	195,609 (7.3)	195,609 (7.3)	4,591 (2.5)	4,591 (2.5)	6,669 (3.0)	6,669 (3.0)
56-65	40,029 (8.5)	40,029 (8.5)	204,327 (7.7)	204,327 (7.7)	14,739 (7.9)	14,739 (7.9)	16,995 (7.5)	16,995 (7.5)
66–70	98,466 (20.9)	98,466 (20.9)	620,578 (23.3)	620,576 (23.3)	17,537 (9.4)	17,537 (9.4)	19,728 (8.7)	19,728 (8.7)
71–75	98,994 (21.0)	98,994 (21.0)	617,788 (23.2)	617,784 (23.2)	21,884 (11.7)	21,884 (11.7)	25,150 (11.1)	25,150 (11.1)
76-80	74,957 (15.9)	74,957 (15.9)	443,190 (16.6)	443,188 (16.6)	26,201 (14.0)	26,201 (14.0)	30,893 (13.7)	30,893 (13.7)
81-85	55,589 (11.8)	55,589 (11.8)	294,423 (11.0)	294,423 (11.0)	31,053 (16.6)	31,053 (16.6)	37,647 (16.7)	37,647 (16.7)
86-90	38,729 (8.2)	38,729 (8.2)	180,162 (6.8)	180,162 (6.8)	33,772 (18.1)	33,772 (18.1)	42,494 (18.8)	42,494 (18.8)
91–95	22,229 (4.7)	22,229 (4.7)	90,050 (3.4)	90,050 (3.4)	27,166 (14.6)	27,166 (14.6)	34,779 (15.4)	34,779 (15.4)
96–99	6,007 (1.3)	6,007 (1.3)	21,804 (0.8)	21,804 (0.8)	9,728 (5.2)	9,728 (5.2)	11,615 (5.1)	11,615 (5.1)
Sex								
Female	255,638 (54.3)	255,638 (54.3)	1,494,273 (56.0)	1,494,271 (56.0)	119,501 (64.0)	119,501 (64.0)	148,245 (65.6)	148,245 (65.6)
Male	214,984 (45.7)	214,984 (45.7)	1,173,658 (44.0)	1,173,652 (44.0)	67,170 (36.0)	67,170 (36.0)	77,725 (34.4)	77,725 (34.4)
Race/Ethnicity								
American Indian/ Alaskan Native	4,115 (0.9)	4,115 (0.9)	14,858 (0.6)	14,858 (0.6)	263 (0.1)	263 (0.1)	389 (0.2)	389 (0.2)
Asian/Pacific Islander	11,587 (2.5)	11,587 (2.5)	52,726 (2.0)	52,726 (2.0)	2,868 (1.5)	2,868 (1.5)	2,339 (1.0)	2,339 (1.0)
Black	64,355 (13.7)	64,355 (13.7)	235,296 (8.8)	235,296 (8.8)	25,511 (13.7)	25,511 (13.7)	21,847 (9.7)	21,847 (9.7)
Hispanic	50,242 (10.7)	50,242 (10.7)	138,907 (5.2)	138,907 (5.2)	10,616 (5.7)	10,616 (5.7)	8,160 (3.6)	8,160 (3.6)
Non-Hispanic White	329,199 (69.9)	329,199 (69.9)	2,156,397 (80.8)	2,156,389 (80.8)	146,674 (78.6)	146,674 (78.6)	192,426 (85.2)	192,426 (85.2)
Other	3,627 (0.8)	3,627 (0.8)	18,005 (0.7)	18,005 (0.7)	452 (0.2)	452 (0.2)	468 (0.2)	468 (0.2)
Unknown	7,497 (1.6)	7,497 (1.6)	51,742 (1.9)	51,742 (1.9)	287 (0.2)	287 (0.2)	341 (0.2)	341 (0.2)
Low Income/Disabled Status								
Low Income or Disabled	188,443 (40.0)	188,443 (40.0)	839,214 (31.5)	839,214 (31.5)	123,380 (66.1)	123,380 (66.1)	123,488 (54.6)	123,488 (54.6)
Not Low Income/ Disabled	282,179 (60.0)	282,179 (60.0)	1,828,717 (68.5)	1,828,709 (68.5)	63,291 (33.9)	63,291 (33.9)	102,482 (45.4)	102,482 (45.4)
RSI (9mo)								
MEAN (SD)	0.048 (.085)	0.048 (.085)	0.038 (.076)	0.038 (.076)	0.212 (.136)	0.212 (.136)	0.205 (.140)	0.205 (.140)
Mortality (Through Nov 30)								
Died	69,053 (14.7)	21,958 (4.7)	165,491 (6.2)	97,864 (3.7)	57,312 (30.7)	53,614 (28.7)	45,745 (20.2)	62,683 (27.7)
Survived	401,569 (85,3)	448,664 (95,3)	2,502,440 (93,8)	2,570,059 (96.3)	129.359 (69.3)	133,057 (71.3)	180,225 (79.8)	163,287 (72,3)

Table 3. Comparison of matching characteristics for Covid-19 subjects versus non-Covid-19 controls in the community and LTC/SNF subgroups.

Subjects were categorized as "LTC/SNF" if they received services in either a Long-Term Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the "Community." Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. The baseline risk of 9-month mortality defined by the Risk Stratification Index (RSI) calculated on February 29, 2020. Beneficiaries receiving a diagnosis of probable or confirmed Covid-19 were pairwise exactly matched 1:1 on Feb 29, 2020 with beneficiaries without a Covid-19 diagnosis based on sex, age (within 1-year), ethnicity, location of services in Feb 2020 (community or LTC/SNF), along with RSI as a propensity factor (within 0.1%). The tabulated results demonstrate similarity of baseline characteristics between tightly matched populations. The baseline risk of mortality (RSI) was much higher in patients categorized as LTC/SNF than Community subjects.

https://doi.org/10.1371/journal.pone.0262264.t003

[28], male gender [29] and receiving services in Long-Term Care facilities are important risk factors associated with dying from Covid-19 infections.



Fig 2. Actual/expected mortality plot for No Covid-19, probable Covid-19, and confirmed Covid -19 cohorts in community, LTC/SNF, and combined analysis. Panels A and B display actual and expected mortality (per 100,000 people) calculated using different methods for Medicare subjects grouped by infection status and location of services. Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. Subjects were categorized as "LTC/SNF" if they received services in either a Long-Term Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the "Community." Estimated mortality using RSI (A) provides estimates consistent with actual mortality of historical controls (B).

https://doi.org/10.1371/journal.pone.0262264.g002

		All C	Cases			Comm	unity			LTC/	SNF	
	No Covid	Probable	Confirmed	Any	No Covid	Probable	Confirmed	Any	No Covid	Probable	Confirmed	Any
Historical Method (Count)	24,794,394	2,917,604	677,100	3,594,704	24,251,751	2,676,502	472,329	3,148,831	542,643	241,102	204,771	445,873
Actual deaths	699,264	214,602	131,460	346,062	548,556	166,800	69,488	236,288	150,708	47,802	61,972	109,774
(%)	2.8%	7.4%	19.4%	9.6%	2.3%	6.2%	14.7%	7.5%	27.8%	19.8%	30.3%	24.6%
Expected deaths	691,692	150,425	64,934	215,359	579,916	101,974	22,719	124,693	111,776	48,451	42,215	90,666
(95% CI)	(690,052– 693,332)	(149,670-151,180)	(64,449-65,419)	(214,459– 216,259)	(578,411– 581,421)	(101, 348 - 102, 600)	(22,425– 23,013)	(124,001– 125,385)	(111,180-112,372)	(48,057-48,845)	(41,849-42,581)	(90,128-91,204)
(%)	2.8%	5.2%	9.6%	6.0%	2.4%	3.8%	4.8%	4.0%	20.6%	20.1%	20.6%	20.3%
(95% CI)	(2.8–2.8)	(5.1-5.2)	(9.5–9.7)	(6.0-6.0)	(2.4 - 2.4)	(3.8-3.8)	(4.7 - 4.9)	(3.9-4.0)	(20.5 - 20.7)	(19.9–20.3)	(20.4 - 20.8)	(20.2 - 20.5)
Excess deaths	7,572	64,177	66,526	130,703	-31,360	64,826	46,769	111,595	38,932	-649	19,757	19,108
(95% CI)	(5,932– 9,212)	(63,422- 64,932)	(66,041– 67,011)	(129,803-131,603)	(-32,865 29,855)	(64,200-65,452)	(46,475-47,063)	(110,903– 112,287)	(38,336– 39,528)	(-1,043 255)	(19, 391 - 20, 123)	(18,570– 19,646)
(% of actual deaths)	1.1%	29.9%	50.6%	37.8%	-5.7%	38.9%	67.3%	47.2%	25.8%	-1.4%	31.9%	17.4%
(95% CI)	(0.8-1.3)	(29.6–30.3)	(50.2 - 51.0)	(37.5 - 38.0)	(-6.05.4)	(38.5-39.2)	(66.9–67.7)	(46.9–47.5)	(25.4-26.2)	(-2.20.5)	(31.3 - 32.5)	(16.9–17.9)
Case Matching (Count)	NA**	2,893,901	657,293	3,551,194	NA**	2,667,931	470,622	3,138,553	NA**	225,970	186,671	412,641
% of cases Matched		99.2%	97.1%	98.8%		99.7%	99.6%	99.7%		93.7%	91.2%	92.5%
Actual deaths	NA^{**}	211,236	126,365	337,601	NA**	165,491	69,053	234,544	NA**	45,745	57,312	103,057
(%)		7.3%	19.2%	9.5%		6.2%	14.7%	7.5%		20.2%	30.7%	25.0%
Matched twins actual deaths		160,547	75,572	236,119		97,864	21,958	119,822		62,683	53,614	116,297
(95% CI)	NA**	(159,768-161,326)	(75,055– 76,089)	(235,180-237,058)	NA**	(97,250– 98,478)	(21,669– 22,247)	(119, 143 - 120, 501)	NA**	(62,257-63,109)	(53,223– 54,005)	(115,719– 116,875)
(%)		5.5%	11.5%	6.6%		3.7%	4.7%	3.8%		27.7%	28.7%	28.2%
(95% CI)		(5.5 - 5.6)	(11.4-11.6)	(6.6–6.7)		(3.6-3.7)	(4.6-4.7)	(3.8 - 3.8)		(27.6–27.9)	(28.5–28.9)	(28.0 - 28.3)
Excess deaths		50,689	50,793	101,482		67,627	47,095	114,722		-16,938	3,698	-13,240
(95% CI)		(49,910-51,468)	(50,276– 51,310)	(100,543-102,421)		(67,013-68,241)	(46,806 - 47,384)	(114,043-115,401)		(-17,364 16,512)	(3,307 - 4,089)	(-13,818 - -12,662)
(% of actual deaths)	NA**	24.0%	40.2%	30.1%	NA**	40.9%	68.2%	48.9%	NA**	-37.0%	6.5%	-12.8%
												(Continued)

Table 4. Number of excess deaths by location, diagnosis, and method of calculation.

Continued)	
Table 4. (

	All C	ases			Comm	nunity			LTC	SNF	
No Covid	Probable	Confirmed	Any	No Covid	Probable	Confirmed	Any	No Covid	Probable	Confirmed	Any
	(23.6–24.4)	(39.8–40.6)	(29.8–30.3)		(40.5-41.2)	(67.8–68.6)	(48.6–49.2)		(-38.0 36.1)	(5.8-7.1)	(-13.4 – -12.3)

Expected and Excess deaths between Covid-19 Confirmed and Probable groups are statistically significantly different. Historical Method: Expected deaths were estimated as the sum of individual xSI-predicted 9-month mortality for the cohort. Case Matching Method: Expected deaths were the actual deaths observed in the case-matched control populations

Excess deaths were calculated as the difference between actual deaths and expected deaths.
 * NA: No non-Covid controls were matched to non-Covid patients since the populations were the same.

deaths (34.8%) compared to historical estimates. Limitations in access to Covid-19 testing and disease under-reporting in long-term care patients probably contributed, although social isolation and services in the "Community." Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a subjects were categorized as "LTC/SNF" if they received services in either a Long-Term Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as receiving Risk Stratification Index (RSI). In this method, beneficiaries receiving a diagnosis of probable or confirmed Covid-19 were pairwise exactly matched 1:1 on Feb 29, 2020 with beneficiaries without a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance 677,100 (2.4%) beneficiaries had confirmed Covid-19 and 2,917,604 (10.3%) had probable Covid-19. 472,329 confirmed cases were community living and 204,771 were in LTC. Mortality following 20.3% to 24.6%. The absolute increase was therefore similar at 3–4% in the community and in long-term care residents. But the percentage increase was far greater in the community (89.5%) demonstrate that, within the Medicare population, Covid-19 had a considerable impact by increasing mortality well above what would have been expected based on age and co-morbidities alone. methods were used to estimate expected 2020 mortality as described in the footnote above. The case matching (digital twin) method utilized the baseline risk of 9-month mortality defined by the [220.828] and WHO recommendations (U07.2) [23,24]. The baseline risk of 9-month mortality defined by the Risk Stratification Index (RSI) calculated on February 29, 2020. Two independent probable or confirmed diagnosis in the community increased from an expected incidence of about 4.0% to actual incidence of 7.5%. In long-term care facilities, the corresponding increase was han among patients in chronic care facilities (21.1%) who had high baseline risk. The long-term care population without probable or confirmed Covid-19 diagnoses experienced 38,932 excess disruption in usual care presumably also contributed. Remarkably, there were 31,360 fewer deaths than expected in community dwellers without probable or confirmed Covid-19 diagnoses. epresenting about a 5.4% reduction. Disruptions to the healthcare system and avoided medical care were thus apparently offset by other factors, representing overall benefit. The Covid-19 Covid-19 diagnosis based on sex, age (within 1-year), ethnicity, location of services in Feb 2020 (community or LTC/SNF), along with RSI as a propensity factor (within 0.1%). The results pandemic had marked effects on mortality, but the effects were highly context-dependent.

1ttps://doi.org/10.1371/journal.pone.0262264.t004

Using RSI as a composite measure of baseline mortality risk permitted precise case-control matching, thereby allowing us to estimate excess deaths attributable Covid-19 by two complementary methods in Medicare recipients with probable or confirmed Covid-19 diagnoses. Using the historical comparison, there was an increase from 215,359 expected to 346,062 actual deaths, representing 130,702 excess deaths and a 61% increase. Using matching, mortality increased from 236,119 expected to 337,601 observed deaths, representing 101,482 excess deaths and a 43% relative increase. Differences between these estimates results largely from excess deaths that occurred in the No Covid subgroup in LTC facilities from which matches were drawn. Nevertheless, both estimates far exceed the 15–20% excess mortality estimated in previous analyses that included younger populations. Our results are therefore consistent with the conclusion that older people with more comorbidities are at much higher risk for developing severe Covid-19—and of dying from it.

Overall, our historical model indicated that mortality following a probable or confirmed diagnosis in the community increased from an expected incidence of about 4.0% to actual incidence of 7.5%. In LTC/SNF's, the corresponding increase was from 20.3% to 24.6%. Therefore, the absolute increase in mortality was similar at 3–4% in the community and in long-term care residents. However, baseline risk (RSI) associated with all individuals in a care setting varied greatly, being only about 2.6% in the community versus 20.5% in long-term care facilities. As a percentage, the relative increase in mortality was thus far greater in the community (89.5%) than among patients in Long-Term Care facilities (21.1%).

Somewhat remarkably, overall mortality *decreased* in Medicare participants without probable or confirmed Covid-19 diagnoses. In fact, among community dwellers, there were 31,360 fewer deaths than expected, representing about a 5.4% reduction. Disruptions to the healthcare system and avoided medical care were thus apparently offset by other factors, representing overall benefit. Obvious health benefits of pandemic isolation include reduced exposure to other airborne illnesses such as influenza, fewer driving accidents and fewer homicides in the over 65-year-old population. However, none seems sufficient to explain the reduction alone. More subtle effects including reduced work or stress-related illness might contribute more, although there is no obvious reason to believe that the pandemic would reduce stress—especially in an over-65-year-old population. Cause of death information would help explain the reasons for decreased mortality in our cohort community dwellers without Covid-19.

The causes of reduced mortality in community dwelling Medicare participants remains unclear. However, our results suggest that inadequate care for chronic conditions and delayed care of acute events did not produce the feared outcome of higher nine-month mortality in the general population without Covid-19. But due to limited follow-up, we caution that disruptions in healthcare delivery may yet result in adverse longer-term outcomes due to delays in the diagnosis and treatment of new and existing chronic conditions. An additional consideration is that prolonged sequela after severe Covid-19 infections (Long Covid syndrome) appear substantial and is an area requiring urgent further study [30].

There was a distinct disparity between community dwellers and those in long-term care facilities with respect to historical mortality comparisons. In contrast to community Medicare participants, the long-term care population without probable or confirmed Covid-19 diagnoses experienced 38,932 excess deaths (35%) compared to historical estimates. We believe that limitations in access to Covid-19 testing and disease under-reporting in long-term care patients probably were responsible for this finding. It seems likely that many of the excess deaths in this vulnerable population were consequent to undiagnosed Covid-19 infections. But it is also probable that social isolation and disruption in usual care may have contributed as well. The higher-than-expected level of excess deaths observed in this cohort (subjects without a probable or confirmed Covid diagnosis) is reflected in our case matching results, which

indicate a modest relative reduction in deaths in subjects with a Covid related diagnosis. This is most likely due to undiagnosed Covid cases included in the control population, but we cannot rule out the possibility that the focus on care for the Covid patients had an unintended adverse impact on the remaining population.

Limitations

We excluded less than 2.2% of the available population because of missing and inconsistent values. Because data were missing non-systematically, exclusion of these subjects was unlikely to introduce meaningful bias. We relied on administrative diagnostic claims for Covid-19 to assign exposure. Surely these are inexact, especially during our study period early in the pandemic. Furthermore, a new diagnostic code for confirmed Covid-19 (U07.1) was introduced on April 1, 2020, and we assume that there was some uncertainty regarding its proper application. However, Kadri et al. recently reported that this Covid-19 specific code showed high sensitivity and specificity compared with the PCR test results [31]. We elected to address this uncertainty in coding by analyzing subjects with a confirmed Covid-19 code separately from those with a probable or unconfirmed Covid-19 code. A second limitation is that we did not account for temporal changes in risk of exposure to Covid-19 in either setting, nor for improvements in treatment of infected individuals over time [32, 33].

We assigned individuals to either community dwelling or long-term care subgroups based on coding in February 2020. Some participants undoubtedly changed their care settings during the analysis period. Skilled nursing facilities, for example, include patients who remain semipermanently along with others who stay for short periods such during rehabilitation from major orthopedic procedures before resuming community life. But among patients who died, 79% of those who were in a long-term care facility on the anchor date of Feb 29, 2020 had Long-Term Care charges within two months of death.

Our analysis was based on 28,389,098 adults enrolled in the US fee-for-service and Medicare/Medicaid program (43% of total Medicare eligible population in the US). The results are therefore broadly applicable to Medicare eligible adults. One potential selection bias to consider is that excluded subjects, especially those without continuous coverage and those enrolled in Medicare Advantage programs, may have had a higher baseline risk profile than those included in our sample. A consequence would be underestimating the full impact of Covid-19 across the entire US Medicare population. Although our sample included a fair number of dual eligible subjects below age 65, our results should only be cautiously generalized to younger and healthier populations.

Summary

Mortality following a probable or confirmed Covid-19 diagnosis in the community increased from an expected incidence of about 4.0% to actual incidence of 7.5%. In long-term care facilities, the corresponding increase was from 20.3% to 24.6%. The absolute increase was therefore similar at 3–4% in the community and in long-term care residents. But the percentage increase was far greater in the community (89.5%) than among patients in chronic care facilities (21.1%) who had high baseline risk.

The long-term care population without probable or confirmed Covid-19 diagnoses experienced 38,932 excess deaths (34.8%) compared to historical estimates. Limitations in access to Covid-19 testing and disease under-reporting in long-term care patients probably contributed, although social isolation and disruption in usual care presumably contributed. Remarkably, there were 31,360 fewer deaths than expected in community dwellers without probable or confirmed Covid-19 diagnoses, representing about a 5.4% reduction. Disruptions to the healthcare system and avoided medical care were thus apparently offset by other factors, representing overall benefit.

The Covid-19 pandemic had marked effects on mortality, but the effects were highly context-dependent. Among community dwelling Medicare participants with suspected or confirmed Covid-19 diagnoses, mortality nearly doubled, but from a relatively low baseline. Patients in long-term care facilities had a similar absolute increase in mortality, but because their baseline mortality was 20.5%, the relative increase was smaller. In contrast, community dwelling Medicare participants without COVID had about 5.4% lower-than-expected mortality.

Supporting information

S1 Fig. Consort style waterfall flowchart detailing population selection methodology. Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. Subjects were excluded for missing data if values for any baseline characteristic used in the study were missing (i.e., age, sex, ethnicity, location of care, zip code derived measures, dates of coverage, or baseline risk of 9 month mortality assessed with the Risk Stratification Index (RSI).) Additionally, we excluded subjects whose records had inconsistent values among source files containing similar variables such as birth date and sex.



S2 Fig. Performance characteristics of RSI model. Panel A: ROC curve, Panel B Calibration plot, Panel C Sensitivity and Positive Predicted Value vs probability of mortality. (A) Area Under the Receiver Operating Curve (AuROC) for the development Learn Set (80% of 2018 Set) was 0.88 (95% Confidence Interval of [0.88-0.88]). AuROC for the prospective Test Set (20% of 2018 Set) was 0.88 (95% Confidence Interval of [0.88–0.88]. Similar performance in the Test Set compared to the Learn Set supports a lack of overfitting in the development of the predictor. (B) The calibration plot displays the mean actual vs predicted 1 year mortality for populations clustered in increments of 1% probability of mortality. Dark green, light green, and red dots are populations of the lowest 95%, 95%-99%, and top 1% risk of mortality. The diagonal line identifies the domain of ideal performance where actual and expected mortality rates are equal for a population. The performance of this index is very close to ideal performance for approximately 99% of the population. Tabulated metrics: The sample size in this test set (N) was 11,871,985 with an incidence of 1yr mortality (Event_Test) of 4.4%. The Slope and Intercept (INT) fit of the data are 0.94 and 0.01, respectively. The area under the Receiver Operating Curve was 0.88. The Mean Average Error (MEA) from cluster coordinates (i.e., (expected, actual) couplets) to the identity line was calculated for the database divided into populations grouped from the riskiest to least risky subjects using cluster sizes ranging from 1 (i.e., each individual as a cluster) to 1000 neighboring subjects (e.g., MAE to MAE 1000). The 95% Confidence Interval (CI) for the fits of these populations to the identify line is tabulated (i.e., AE_CI to AE_CI_1000). Rsq_unit is a goodness of fit measure of individual results to the ideal line. (C) Positive Predictive Accuracy (blue dots) and Sensitivity (purple dots) versus the fraction of population, sorted by the risk of 1 year mortality. The vertical red line indicates where the number of patients above the risk threshold equals the incidence of mortality in the population. There are several metrics tabulated in the figure: The area under the Receiver Operating Characteristic (ROC) is 0.88. The incidence of mortality (IR) in the population of

11,871,985 was 4.4%. Vertical bars help identify the PPA and sensitivity performance for detectors operating to identify the riskiest 5%, 10%, and 20% patients. The PPV, sensitivity and relative risk (RR) are tabulated for these detector operating points. (TIF)

S3 Fig. Comparison of RSI (Panel A) and Chronic Condition based models (Panel B) on 2020 population. ROC curves for all subjects, No Covid-19, Probable Covid-19 and Confirmed Covid-19 populations. Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. (A,B) ROCs display the sensitivity vs. 1 –specificity in detecting patients who died within 9 months after prediction from February 29,2020 (baseline). The areas under each ROC, with their corresponding 95% confidence intervals, are tabulated in the lower right of each figure. Predictions using RSI yielded better performance (A) than those using a model based on age, sex and chronic conditions (B).

(TIF)

S4 Fig. Forest plot showing the relative risk and 95% CI of significant predictors of confirmed Covid-19 infection. Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. Subjects were categorized as "LTC/SNF" if they received services in either a Long-Term Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the "Community." Predictors were assessed at baseline (February 29, 2020) and include quintiles of Risk Stratification Index (RSI), presence of chronic conditions, location of services (LTC/SNF vs Community), and demographic variables (i.e., age, sex, race, and quintiles of median household income imputed by zip code according to 2015 Census data.) Variables not remaining in the adjusted model are indicated by the presence of empty parenthesis under the adjusted odds ratio. Location of services, age, status of end-stage renal disease (ESRD) and RSI were the strongest (unadjusted) predictors of infection. Location of services and ESRD remained strong predictors following adjustment; however, risks associated with having chronic conditions were typically reduced when adjusted by the presence of other factors.

(TIF)

S5 Fig. Observed mortality rates by age and RSI quintiles. Rates of mortality within 9 months following baseline (February 29, 2020) in Medicare subpopulations categorized by age, location of services, infection status, and quintiles of the baseline risk of mortality assessed using the Risk Stratification Index (RSI). Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. Subjects were categorized as "LTC/SNF" if they received services in either a Long-Term Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the "Community." As expected, subjects in quintiles with higher baseline

risk of mortality had higher rates of observed mortality. For subjects without a Covid diagnosis, mortality rates were lower in the community setting compared to those in the LTC/SNF; however, for subjects with confirmed or probable Covid infection, mortality rates were typically higher in the community setting than in the LTC/SNF. (TIF)

S1 Table. Comparison of characteristics for Covid-19 subjects matched versus unmatched with controls in community and LTC/SNF subgroups. Subjects were categorized as "LTC/ SNF" if they received services in either a Long-Term Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the "Community." Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020 [22]. Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2) [23, 24]. The baseline risk of 9-month mortality defined by the Risk Stratification Index (RSI) calculated on February 29, 2020. Beneficiaries receiving a diagnosis of probable or confirmed Covid-19 were pairwise exactly matched 1:1 on Feb 29, 2020 with beneficiaries without a Covid-19 diagnosis based on sex, age (within 1-year), ethnicity, location of services in Feb 2020 (community or LTC/SNF), along with RSI as a propensity factor (within 0.1%). The tabulated results demonstrate the expected results where the percentage of unmatched cases is generally highest in the subpopulations within a category that are either the smallest (e.g., American Indians/Alaskan Native) or that generally do not use LTC/SNF services (e.g., the youngest population). (PDF)

S1 File. RSI development method for 9-month predictions. (PDF)

Acknowledgments

We thank John Parks, Zhenyu Hong and Douceur Tengu (Data Analysts, Health Data Analytics Institute) for their assistance in developing models and preparing figures and tables.

Author Contributions

Data curation: Scott D. Greenwald, Nassib G. Chamoun.
Formal analysis: Scott D. Greenwald, Nassib G. Chamoun.
Funding acquisition: Nassib G. Chamoun, Josh Gray.
Investigation: Scott D. Greenwald, Nassib G. Chamoun.
Methodology: Nassib G. Chamoun, Paul J. Manberg, Daniel I. Sessler.
Project administration: Scott D. Greenwald, David Clain.
Resources: Nassib G. Chamoun.
Software: Scott D. Greenwald.
Supervision: Scott D. Greenwald, Nassib G. Chamoun, Paul J. Manberg.
Validation: Paul J. Markang, Lock Gray, David Clain, Kamel Makachurgi

Validation: Paul J. Manberg, Josh Gray, David Clain, Kamal Maheshwari, Daniel I. Sessler.

Visualization: Scott D. Greenwald, David Clain.

Writing - original draft: Scott D. Greenwald, Paul J. Manberg, Daniel I. Sessler.

Writing – review & editing: Scott D. Greenwald, Nassib G. Chamoun, Paul J. Manberg, Josh Gray, David Clain, Kamal Maheshwari, Daniel I. Sessler.

References

- 1. CDC. CDC COVID Data Tracker: Centers for Disease Control and Prevention; 2021 [Available from: https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days.
- Chen YH, Glymour MM, Catalano R, Fernandez A, Nguyen T, Kushel M, et al. Excess Mortality in California During the Coronavirus Disease 2019 Pandemic, March to August 2020. JAMA Intern Med. 2020.
- Faust JS, Krumholz HM, Du C, Mayes KD, Lin Z, Gilman C, et al. All-Cause Excess Mortality and COVID-19–Related Mortality Among US Adults Aged 25–44 Years, March-July 2020. JAMA. 2020.
- Lauren M. Rossen AMB, Farida B. Ahmad, Paul Sutton, Robert N. Anderson. Excess Deaths Associated with COVID-19, by Age and Race and Ethnicity—United States, January 26–October 3, 2020 2021 [Available from: https://www.cdc.gov/mmwr/volumes/69/wr/mm6942e2.htm.
- Stang A, Standl F, Kowall B, Brune B, Bottcher J, Brinkmann M, et al. Excess mortality due to COVID-19 in Germany. J Infect. 2020; 81(5):797–801. https://doi.org/10.1016/j.jinf.2020.09.012 PMID: 32956730
- Islam N, Shkolnikov VM, Acosta RJ, Klimkin I, Kawachi I, Irizarry RA, et al. Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. BMJ. 2021; 373:n1137. https://doi.org/10.1136/bmj.n1137 PMID: 34011491
- CDC. Medical Conditions and Risk of for Severe COVID-19 Illness: Centers for Disease Control and Prevention; 2021 [Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/ people-with-medical-conditions.html.
- Wang Q, Davis PB, Gurney ME, Xu R. COVID-19 and dementia: Analyses of risk, disparity, and outcomes from electronic health records in the US. Alzheimers Dement. 2021.
- Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: A federated electronic medical record analysis. PLoS Med. 2020; 17(9):e1003321. https://doi.org/10.1371/journal.pmed.1003321 PMID: 32911500
- Wang Q, Berger NA, Xu R. Analyses of Risk, Racial Disparity, and Outcomes Among US Patients With Cancer and COVID-19 Infection. JAMA Oncol. 2021; 7(2):220–7. <u>https://doi.org/10.1001/jamaoncol.</u> 2020.6178 PMID: 33300956
- Kowall B, Nonnemacher M, Brune B, Brinkmann M, Dudda M, Bottcher J, et al. A model to identify individuals with a high probability of a SARS-CoV-2 infection. J Infect. 2021; 82(3):e32–e4. https://doi.org/ 10.1016/j.jinf.2020.11.028 PMID: 33245942
- Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ. 2020; 369: m1328. https://doi.org/10.1136/bmj.m1328 PMID: 32265220
- Experton B, Tetteh HA, Lurie N, Walker P, Carroll CJ, Elena A, et al. A Predictive Model for Severe Covid-19 in the Medicare Population: A Tool for Prioritizing Scarce Vaccine Supply. medRxiv. 2020:2020.10.28.20219816.
- Dun C, Walsh CM, Bae S, Adalja A, Toner E, Lash TA, et al. A Machine Learning Study of 534,023 Medicare Beneficiaries with COVID-19: Implications for Personalized Risk Prediction. medRxiv. 2020:2020.10.27.20220970.
- Beaney T, Clarke JM, Jain V, Golestaneh AK, Lyons G, Salman D, et al. Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide? J R Soc Med. 2020; 113(9):329–34. https://doi. org/10.1177/0141076820956802 PMID: 32910871
- Stokes AC, Lundberg DJ, Elo IT, Hempstead K, Bor J, Preston SH. COVID-19 and excess mortality in the United States: A county-level analysis. PLoS Med. 2021; 18(5):e1003571. <u>https://doi.org/10.1371/journal.pmed.1003571</u> PMID: 34014945
- VanderWeele TJ. Challenges Estimating Total Lives Lost in COVID-19 Decisions: Consideration of Mortality Related to Unemployment, Social Isolation, and Depression. JAMA. 2020; 324(5):445–6. https://doi.org/10.1001/jama.2020.12187 PMID: 32639547
- Czeisler ME, Lane RI, Petrosky E, Wiley JF, Christensen A, Njai R, et al. Mental Health, Substance Use, and Suicidal Ideation During the COVID-19 Pandemic—United States, June 24–30, 2020. MMWR Morb Mortal Wkly Rep. 2020; 69(32):1049–57. https://doi.org/10.15585/mmwr.mm6932a1 PMID: 32790653

- US Department of Transportation NHSA. Traffic Safety Facts: Early Estimates of Motor Vehicle Traffic Fatalities and Fatality Rates by Sub-Categories in 2020: NHTSA's National Center for Statistics and Analysis; 2021 [updated June 2021 (Revised). Available from: https://crashstats.nhtsa.dot.gov/Api/ Public/ViewPublication/813118.
- Rosenfeld RL, E. Pandemic Social Unrest and Crime in US Cities: National Commission on COVID-19 and Criminal Justice; [updated Dec 2020. Available from: https://build.neoninspire.com/counciloncj/wpcontent/uploads/sites/96/2021/07/Crime-in-US-Cities-October-Update.pdf.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet. 2007; 370(9596):1453–7. https://doi.org/10.1016/S0140-6736(07) 61602-X PMID: 18064739
- CMS. Preliminary Medicare COVID-19 Data Snapshot | CMS: Centers for Medicare & Medicaid Services; 2021 [Available from: https://www.cms.gov/research-statistics-data-systems/preliminary-medicare-covid-19-data-snapshot.
- 23. CDC. ICD-10-CM Official Coding and Reporting Guidelines April 1, 2020 through September 30, 2020. Centers for Disease Control and Prevention; 2020.
- WHO. COVID-19 coding in ICD-10: World Health Organization; 2020 [cited World Health Organization. Available from: https://www.who.int/classifications/icd/COVID-19-coding-icd10.pdf.
- Chamoun GF, Li L, Chamoun NG, Saini V, Sessler DI. Validation and Calibration of the Risk Stratification Index. Anesthesiology. 2017; 126(4):623–30. https://doi.org/10.1097/ALN.00000000001560 PMID: 28187023
- 26. AHRQ. Clinical Classification Software Refined (CCSR): Agency for Healthcare Research and Quality; 2021 [updated March 5, 2021. Available from: https://www.hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp.
- 27. CMS. Condition Categories—Chronic Conditions Data Warehouse: Centers for Medicare & Medicaid Services; 2021 [Available from: https://www2.ccwdata.org/web/guest/condition-categories.
- Gross CP, Essien UR, Pasha S, Gross JR, Wang SY, Nunez-Smith M. Racial and Ethnic Disparities in Population-Level Covid-19 Mortality. J Gen Intern Med. 2020; 35(10):3097–9. https://doi.org/10.1007/ s11606-020-06081-w PMID: 32754782
- 29. CDC. CDC COVID Data Tracker, COVID -19 Weekly Cases and Deaths per 100,000 Population by Age, Race/Ethnicity, and Sex: Centers for Disease Control and Prevention; 2021 [Available from: https://covid.cdc.gov/covid-data-tracker/#demographicsovertime.
- Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA, Cuapio A, et al. More than 50 Long-term effects of Covid-19: a systematic review and meta-analysis. medRxiv preprint. 2021.
- Kadri SS, Gundrum J, Warner S, Cao Z, Babiker A, Klompas M, et al. Uptake and Accuracy of the Diagnosis Code for COVID-19 Among US Hospitalizations. JAMA. 2020; 324(24):2553–4. <u>https://doi.org/</u> 10.1001/jama.2020.20323 PMID: 33351033
- Nguyen NT, Chinn J, Nahmias J, Yuen S, Kirby KA, Hohmann S, et al. Outcomes and Mortality Among Adults Hospitalized With COVID-19 at US Medical Centers. JAMA Netw Open. 2021; 4(3):e210417. https://doi.org/10.1001/jamanetworkopen.2021.0417 PMID: 33666657
- Tenforde MW, Fisher KA, Patel MM. Identifying COVID-19 Risk Through Observational Studies to Inform Control Measures. JAMA. 2021. https://doi.org/10.1001/jama.2021.1995 PMID: 33616617