

1 **Covid-19 and Excess Mortality in Medicare Beneficiaries**

2 Scott Greenwald^{1¶*}, Nassib G Chamoun^{1¶}, Paul J Manberg^{1¶}, Josh Gray^{1¶},
3 David Clain^{1¶}, Kamal Maheshwari^{2,3¶}, Daniel I. Sessler^{2¶}

4

5 ¹ Health Data Analytics Institute, Dedham MA

6 ² Department of **OUTCOMES RESEARCH**, Anesthesiology Institute, Cleveland Clinic,
7 Cleveland, OH

8 ³ Department of General Anesthesiology, Anesthesiology Institute, Cleveland
9 Clinic, Cleveland, OH

10

11 *** Corresponding Author:**

12 Email: Scott.Greenwald@hda-institute.com (SG)

13 ¶ These authors contributed equally to this work.

14

15

16

17 **Abstract**

18 We estimated excess mortality in Medicare recipients with probable and confirmed
19 Covid-19 infections in the general community and amongst residents of long-term care
20 (LTC) facilities. We considered 28,389,098 Medicare and dual-eligible recipients from
21 one year before February 29, 2020 through September 30, 2020, with mortality followed
22 through November 30th, 2020. Probable and confirmed Covid-19 diagnoses,
23 presumably mostly symptomatic, were determined from ICD-10 codes. We developed a
24 Risk Stratification Index (RSI) mortality model which was applied prospectively to
25 establish baseline mortality risk. Excess deaths attributable to Covid-19 were estimated
26 by comparing actual-to-expected deaths based on historical comparisons and in closely
27 matched cohorts with and without Covid-19. 677,100 (2.4%) beneficiaries had
28 confirmed Covid-19 and 2,917,604 (10.3%) had probable Covid-19. 472,329 confirmed
29 cases were community living and 204,771 were in LTC. Mortality following a probable or
30 confirmed diagnosis in the community increased from an expected incidence of about
31 4% to actual incidence of 7.5%. In long-term care facilities, the corresponding increase
32 was from 20.3% to 24.6%. The absolute increase was therefore similar at 3-4% in the
33 community and in LTC residents. The percentage increase was far greater in the
34 community (89%) than among patients in chronic care facilities (21%) who had higher
35 baseline risk. The LTC population without probable or confirmed Covid-19 diagnoses
36 experienced 38,932 excess deaths (35%) compared to historical estimates. Limitations
37 in access to Covid-19 testing and disease under-reporting in LTC patients probably
38 were important factors, although social isolation and disruption in usual care
39 presumably also contributed. Remarkably, there were 31,360 fewer deaths than

40 expected in community dwellers without probable or confirmed Covid-19 diagnoses,
41 representing a 6% reduction. Disruptions to the healthcare system and avoided medical
42 care were thus apparently offset by other factors, representing overall benefit. The
43 Covid-19 pandemic had marked effects on mortality, but the effects were highly context-
44 dependent.

45

46 **Introduction**

47 The Covid-19 pandemic has profoundly influenced US healthcare, especially among
48 Medicare recipients who are mostly at least 65 years old. By March 1, 2021, SARS-
49 CoV-2, the virus responsible for Covid-19, had already infected more than 29 million
50 Americans and more than 500,000 died following infection.(1) However, many people
51 infected with Covid-19 are never tested or have false-negative test results;
52 consequently, the true toll of Covid-19 remains uncertain. Furthermore, while the clinical
53 course is sometimes apparent, Covid-19 also kills people by worsening chronic
54 conditions, with those deaths often being attributed to other causes. Especially early in
55 the pandemic, due to limited testing availability, it was difficult to differentiate deaths
56 caused by Covid-19 from those that may have occurred naturally due to underlying
57 health conditions. It is thus apparent that many people who died consequent to Covid-
58 19 infections may not have been diagnosed with the condition or may have died due to
59 underlying causes.

60 Several teams have estimated “excess” mortality due to Covid-19 by comparing weekly
61 observed death totals with those that occurred in a prior year. For example, Chen et al
62 estimated that from March 1 through August 22, 2020, 146,557 deaths were recorded in
63 California, with an estimated 19,806 (95% CI: 16,364, 23,210) deaths in excess of those
64 predicted by historical trends.(2) Similarly, Faust et al estimated that from March 1,
65 2020, to July 31, 2020, a total of 76,088 all-cause deaths occurred among US adults
66 aged 25 to 44 years, which was 11,899 more than the expected 64,189 deaths based
67 on a previous year (incident rate ratio, 1.19 [95% CI, 1.14-1.23]).(3) Rossen et al
68 estimated excess mortality from January 26 through October 3 to have *decreased* 2%
69 for the youngest subjects (aged <25 years) but increased 14.4% for those 45-64 years,
70 21.1% for those 65-74 years, 21.5% for those 75-84 years, and 14.7% in subjects
71 ≥85 years old.(4) These reports suggest that all-cause mortality in the first six months of
72 the pandemic increased by about 15-20%. . However, historical comparisons do not
73 account for risk at an individual level which may be useful to determine true excess
74 mortality.

75 On a broad population basis, many risk factors for Covid-19 are now well recognized.
76 For example, the CDC identifies eleven conditions that augment risk for severe forms of
77 Covid-19.(5) Chronic conditions such as cancer and dementia are reported to be among
78 the most important contributors.(6-8) It is clear that older members of the population are
79 at special risk, although to some extent age may be a surrogate for accumulated
80 comorbidities. However, it is difficult to extrapolate from population risk to individual risk
81 since many people exhibit various combinations of risk factors for Covid-19 mortality,
82 and individual risks attributable to each condition are not necessarily additive. A robust

83 model that considers relevant individual conditions and predicts mortality risk from
84 Covid-19 infections would therefore be valuable.

85 Numerous groups have proposed individual risk models based on clinical outcomes in
86 various populations studied early in the pandemic, but a consensus model has yet to
87 emerge.(9-11) From a practical perspective, prediction models based on readily
88 available administrative data (e.g., ICD-10 codes) will be most useful since more
89 granular information extracted from clinical health records are neither universally
90 available nor easy to obtain. Our primary goal was therefore to estimate excess risk-
91 specific mortality in people with probable and confirmed Covid-19 infections.

92 An additional consequence of the Covid-19 pandemic has been public health
93 quarantines that have severely disrupted healthcare delivery. The virus may therefore
94 also have caused indirect mortality because patients with chronic diseases and acute
95 exacerbations avoided seekin care due to fear of infection or because health services
96 were overwhelmed or otherwise limiting access. Furthermore, stress related to isolation
97 could increase medical morbidity and provoke suicide and substance abuse. In contrast,
98 some causes of death such as accidents and homicides may have diminished. The
99 extent to which delayed and disrupted healthcare for non-Covid-related conditions,
100 along with pandemic-related behavioral changes, contribute to mortality remains
101 unclear. Our secondary goal was therefore to estimate whether changes in mortality
102 occurred in people without probable or confirmed Covid-19 infections.

103 Because Covid-19 is especially lethal in older people, we considered stratification by
104 various age ranges for both our primary and secondary analyses. We also separately

105 considered people residing in the community from those in long term care facilities, who
106 are expected to have a higher baseline mortality risk and thus may be especially
107 susceptible to Covid-19 infections.

108 **Methods**

109 Data analysis was conducted on the Center for Medicare and Medicaid Services (CMS)
110 Research Identifiable File (RIF) data using SAS Enterprise Guide (Version 7.15) under
111 a special Data Use Agreement (DUA). This project was determined to be exempt from
112 informed consent requirements by the New England Institutional Review Board. Final
113 data analysis of the full cohort was conducted from January 10 to March 11, 2021. This
114 study followed the Strengthening the Reporting of Observational Studies in
115 Epidemiology ([STROBE](https://www.strobe-reportingguideline.org/)) reporting guideline for cohort studies.(12)

116 Individual subject data used for our analysis are available to certain stakeholders as
117 allowed by federal regulations and CMS policy. Requests for access to data to replicate
118 these findings require an approved research protocol and DUA with CMS. For more
119 information, contact the Research Data Assistance Center (ResDAC,
120 <http://www.resdac.org>).

121 Mortality predictions and outcomes were referenced to an anchor date of Feb 29, 2020,
122 just before the initial wave of documented Covid-19 cases and the week when the first
123 case of potential community spread Covid-19 was reported by CDC. We recognize that
124 undiagnosed cases may have occurred previously, but it is unlikely that there were
125 many.

126 We used full Medicare fee-for-service and dual eligible (Medicaid and Medicare) files
127 one year before the anchor date through September 30, 2020, with mortality outcomes
128 reported through November 30th, 2020 for the primary analysis of Covid-19 outcomes.
129 We identified beneficiaries with confirmed Covid-19 diagnoses consistent with CMS
130 guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1
131 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September
132 30, 2020.(13) Probable Covid-19 infection cases were identified using ICD-10-CM
133 codes consistent with the CDC guidance (Z20.828) and WHO recommendations
134 (U07.2).(14, 15) Presumably most subjects with Covid-19 diagnoses were symptomatic,
135 although some may have been tested because of risk or exposure.

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

142 We included all Medicaid and Medicare participants alive as of the study anchor date
143 (N=65,310,173). We excluded beneficiaries who had: 1) ages outside 18-99 years
144 (214,767); 2) non-continuous coverage of Medicare Part A or B (12,445,567) or any
145 Medicare Part C coverage (23,633,391) in the year before the Feb 29, 2020 anchor
146 date; 3) missing data for any variable used in the analysis (509,932); and,
147 4) inconsistent data (117,418). The resulting 28,389,098 beneficiaries included 677,100
148 with confirmed and 2,917,604 with probable diagnosis of Covid-19 (**S1 Fig**). Mortality

149 was assumed to have occurred on the date-of-death listed in the CMS Common
150 Medicare Environment which is continuously updated from various sources including
151 the Social Security Administration.

152 ***Risk Stratification***

153 We used an adaptation of the Risk Stratification Index (RSI) to predict nine-month
154 mortality using the prior year's administrative claims. A model was developed using the
155 full fee-for-service 2018 population for training, with prospective validation of
156 performance on the 2019 dataset as previously described.(16) More detailed
157 descriptions of the model and prospective testing performance results are provided in
158 **S1 File** and **S2 Fig**, respectively. The resulting RSI model was then prospectively
159 applied to all eligible Medicare or dual-eligible beneficiaries as of February 29, 2020 to
160 derive individual RSI scores as of that date — that is, before Covid-19 infections were
161 confirmed in the United States.

162 For comparative purposes, a second model was similarly developed to predict nine-
163 month mortality from the presence of 27 individual chronic conditions as defined by
164 CMS.(17) Specifically, logistic regression (stepwise selection using p-in of 10^{-3} , p-out of
165 10^{-2}) was used to select significant predictors from a pool of candidate features (i.e., 27
166 chronic conditions, age, sex and dual-enrollment status) to create a predictive model
167 using the same training and prospective testing populations described in the previous
168 paragraph. Performance using baseline RSI values or the individual chronic conditions
169 model as predictors of outcomes in 2020 were then tested prospectively (**S3 Fig**).

170 **Analysis and Selection of Study Cohorts**

171 We conducted a progression of complementary inquiries. To test our primary
172 hypothesis, we first identified the main study cohorts of beneficiaries with diagnoses of
173 probable or confirmed Covid-19, and then subdivided them based on location of service
174 (community or LTC/SNF) as of February 29, 2020. Within each cohort, we determined
175 9-month mortality between the anchor date and November 30, 2020.

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

185 **Estimation of Expected Mortality**

186 Two independent methods were used to estimate expected 2020 mortality in our study
187 population. A historical comparison allowed us to compare year-over-year changes in
188 mortality in Medicare recipients and thus characterize overall effects of Covid-19 and
189 quarantine-induced restrictions in healthcare access on mortality. A case-matched
190 analysis provided an alternate estimate of Covid-19-related excess mortality within the
191 9 months of 2020 that we considered.

192

193

194 **Historical Comparison**

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

206 **Case Matching, Digital Twins**

207 A second method used case matching or “digital twinning” to estimate excess mortality
208 in exposed subjects compared to concurrent controls who had closely matched health
209 profiles. Beneficiaries receiving a diagnosis of probable or confirmed Covid-19 were
210 pairwise exactly matched 1:1 on Feb 29, 2020 with beneficiaries without a Covid-19
211 diagnosis based on sex, age (within 1-year), ethnicity, location of services in Feb 2020
212 (community or LTC/SNF), along with RSI as a propensity matching factor (within 0.1%).

213 Because the eligible Medicare population is large, we successfully matched almost the
214 entire infected population. Excess deaths were estimated as the difference between the
215 observed number of deaths in probable or confirmed Covid-19 subjects and their
216 matched non-Covid-19 digital twins over the concurrent period. Matching may be more
217 reliable than the historical comparison for estimating true excess mortality because it
218 better accounts for population variation over time and accounts for the impact of
219 substantial disruptions in public health and everyday life caused by the pandemic
220 restrictions in 2020.

221 **Statistical Analysis**

222 We used SAS Enterprise Guide (Version 7.15) to build models and conduct analyzes.
223 We excluded patients who lacked any variable used in our analyses including birth date,
224 sex, race, Medicaid enrollment status, zip code, and baseline RSI. Additionally, we
225 excluded subjects whose records had inconsistent values among source files containing
226 similar variables such as birth date and sex.

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

234

235 Results

236 Patient characteristics and outcomes

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

242

243 **Table1** and **Table 2** compare demographic and clinical profiles for various subgroups.
244 Compared to survivors, patients who died after a Covid-19 diagnosis were older, more
245 often male, not white, received Medicaid, lived in zip-codes associated with lower
246 median income, received services in February 2020 in a long-term care facility, and had
247 higher baseline risk of mortality as defined by RSI. Age and baseline RSI scores were
248 both strongly related to risk of infection and adverse Covid-19 outcomes. Residence in a
249 LTC/SNF location and presence of end-stage renal disease were strong risk factors for
250 acquiring a confirmed diagnosis of Covid-19 (**S4 Fig**). As shown in **S5 Fig**, RSI scores
251 were associated with increasing mortality in a consistent rank ordered manner across
252 each age group, thereby suggesting that RSI provides a significant and sensitive
253 measure of co-morbidities and mortality risk that is independent of age.

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

Characteristic	Group Size (N,%)								
	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)									
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)

256 *p-values represent the level of significance comparing attributes between Confirmed and Probable Covid infected
 257 populations relative to all Beneficiaries.

258 Confirmed Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29
 259 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September
 260 30, 2020.¹² Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance
 261 (Z20.828) and WHO recommendations (U07.2).^{13,14} The baseline risk of 9-month mortality defined by the Risk
 262 Stratification Index (RSI) calculated on February 29, 2020 was 3.2% in the entire population and significantly higher
 263 among those who died compared to those who survived.

264 **Table 2. Comparison of baseline demographic characteristics for No Covid-19, Confirmed Covid-19 and Probable**
 265 **Covid-19 populations in the Community and LTC/SNF subgroups.**

Characteristic	Study group, No. (%)							
	Community				SNF/LTC			
	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)		2,824 (0.5)	754 (0.4)	1,075 (0.4)	
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)	<.0001	21,796 (4.0)	15,388 (7.5)	11,367 (4.7)	<.0001
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)		258,305 (47.6)	139,357 (68.1)	135,740 (56.3)	
Not LowIncome/Disabled	17,598,467 (72.6)	282,428 (59.8)	1,830,145 (68.4)	<.0001	284,338 (52.4)	65,414 (31.9)	105,362 (43.7)	<.0001
Median Household Income (Imputed from Zip code) (Quintile)								
20-40%	4 917 968 (20.3)	86 115 (18.2)	473 870 (17.7)		114 479 (21.1)	35 239 (17.2)	47 067 (19.5)	
40-60%	4 913 071 (20.3)	84 952 (18.0)	487 217 (18.2)	<.0001	111 633 (20.6)	37 107 (18.1)	45 880 (19.0)	<.0001

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)	
Community vs LTC/SNF								
Community	24 251 751 (100.0)	472 329 (100.0)	2 676 502 (100.0)		0 (0.0)	0 (0.0)	0 (0.0)	
In Long term care Facility or SNF	0 (0.0)	0 (0.0)	0 (0.0)	<.0001	542 643 (100.0)	204 771 (100.0)	241 102 (100.0)	<.0001
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)		150 708 (27.8)	61 972 (30.3)	47 802 (19.8)	
Survived	23 703 195 (97.7)	402 841 (85.3)	2 509 702 (93.8)	<.0001	391 935 (72.2)	142 799 (69.7)	193 300 (80.2)	<.0001

266 *p-value: < 0.05 indicates the distribution of patients among subgroups (or the mean values of metrics) of the No Covid-

267 19, Confirmed Covid-19 and Probable Covid-19 groups.

268 Subjects were categorized as “LTC/SNF” if they received services in either a Long-Term Care (LTC) or Skilled Nursing

269 Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the “Community.” Confirmed

270 Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April

271 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020.¹²

272 Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828)
273 and WHO recommendations (U07.2).^{13,14} The baseline risk of 9-month mortality defined by the Risk Stratification Index
274 (RSI) calculated on February 29, 2020 was multiple times higher among subjects who received services in LTC/SNF than
275 in the community.
276

277 **Mortality Risk Prediction**

278 **Fig 1** presents the relative importance of factors that contributed to mortality in both
279 univariable and multivariable models. Quintiles of RSI, age, LTC/SNF services status,
280 sex, and race were the factors most associated with relative risk of mortality. Status of
281 lung cancer and end-stage renal disease appear to carry meaningful incremental risk
282 after adjustment. Mortality prediction models based primarily on baseline RSI levels
283 performed better than models based on the presence of individual chronic conditions for
284 predicting mortality risk (**S3 Fig**). Case matching identified a cohort of beneficiaries from
285 the general population who were closely matched with subjects who had a diagnosis of
286 probable or confirmed Covid-19 based on their RSI scores as of Feb 29, 2020 (**Table**
287 **3**).

288 **Fig 1. Forest plot showing the relative risk and 95% CI of significant predictors of**
289 **mortality of subjects with confirmed Covid-19.** Confirmed Covid-19 cases were
290 identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29
291 before April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between
292 March 1, 2020 and September 30, 2020.¹² Subjects were categorized as “LTC/SNF” if
293 they received services in either a Long Term Care (LTC) or Skilled Nursing Facility
294 (SNF) in February 2020, otherwise they were categorized as receiving services in the
295 “Community.” Predictors were assessed at baseline (February 29, 2020) and include
296 quintiles of Risk Stratification Index (RSI), presence of chronic conditions, location of
297 services (LTC/SNF vs Community), and demographic variables (i.e., age, sex, race, and
298 quintiles of median household income imputed by zip code according to 2015 Census
299 data.) Variables not remaining in the adjusted model are indicated by the presence of

300 empty parenthesis under the adjusted odds ratio. RSI, age, and location of services
301 were the strongest (unadjusted) predictors of mortality. RSI and age remain strong
302 predictors following adjustment; however, risks associated with having chronic
303 conditions were typically reduced when adjusted by the presence of RSI and other
304 factors. Status of Lung cancer and end-stage renal disease appear to carry meaningful
305 incremental risk after adjustment.
306

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

Characteristic	Study group, No. (%)							
	Community				LTC/SNF			
	COVID Confirmed		COVID Probable		COVID 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)
Community vs LTC/SNF								
Community	470,622 (100.0)	470,622 (100.0)	2,667,931 (100.0)	2,667,923 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)
In Long term care Facility or SNF	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	186,671 (100.0)	186,671 (100.0)	225,970 (100.0)	225,970 (100.0)
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)

309 Subjects were categorized as “LTC/SNF” if they received services in either a Long-Term Care (LTC) or Skilled Nursing
 310 Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the “Community.” Confirmed

311 Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April
312 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020.¹²
313 Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828)
314 and WHO recommendations (U07.2).^{13,14} The baseline risk of 9-month mortality defined by the Risk Stratification Index
315 (RSI) calculated on February 29, 2020. Beneficiaries receiving a diagnosis of probable or confirmed Covid-19 were
316 pairwise exactly matched 1:1 on Feb 29, 2020 with beneficiaries without a Covid-19 diagnosis based on sex, age (within
317 1-year), ethnicity, location of services in Feb 2020 (community or LTC/SNF), along with RSI as a propensity factor (within
318 0.1%). The tabulated results demonstrate similarity of baseline characteristics between tightly matched populations. The
319 baseline risk of mortality (RSI) was much higher in patients categorized as LTC/SNF than Community subjects.
320
321

322 **Excess Mortality Estimates**

323 The distribution of observed and expected mortality by diagnosis, category, and location
324 of care is presented in **Fig 2**. As expected, subjects with high baseline mortality risk in
325 the LTC/SNF cohort had actual mortality that exceeded all other groups. Those with
326 confirmed Covid-19 showed similarly increased mortality above expected levels in both
327 the LTC/SNF and community setting. Among community dwelling subjects, mortality
328 also exceeded expected risk in subjects with possible Covid-19.

329 **Fig 2. Actual/Expected mortality plot for No Covid-19, Probable Covid-19, and** 330 **Confirmed Covid -19 cohorts in community, LTC/SNF, and combined analysis.**

331 Panels A and B display actual and expected mortality (per 100,000 people) calculated
332 using different methods for Medicare subjects grouped by infection status and location
333 of services. Confirmed Covid-19 cases were identified consistent with CMS guidance
334 using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1 thereafter)
335 as a primary or secondary diagnosis between March 1, 2020 and September 30,
336 2020.¹² Probable Covid-19 infection cases were identified using ICD-10-CM codes
337 consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2).^{13,14}
338 Subjects were categorized as “LTC/SNF” if they received services in either a Long Term
339 Care (LTC) or Skilled Nursing Facility (SNF) in February 2020, otherwise they were
340 categorized as receiving services in the “Community.” Estimated mortality using RSI (A)
341 provides estimates consistent with actual mortality of historical controls (B).

342

343 There was an estimated excess of 130,702 (historical comparison method) or 101,482
344 (case matching method) deaths attributable to probable or confirmed Covid-19 across

345 the full population in the 9 months of 2020 that we considered. In the matched analysis,
346 half the deaths (50,793) occurred in patients with a confirmed diagnosis of Covid-19 and
347 half (50,689) occurred in those with a probable Covid-19 diagnosis. In contrast, 31,360
348 fewer subjects without a Covid-19 diagnosis died than expected, representing a 6%
349 mortality reduction (Table 4).

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

	All Cases				Community				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 2.82%	214,602 7.36%	131,460 19.42%	346,062 9.63%	548,556 2.26%	166,800 6.23%	69,488 14.71%	236,288 7.50%	150,708 27.77%	47,802 19.83%	61,972 30.26%	109,774 24.62%
Expected deaths (%)	691,692 2.79%	150,425 5.16%	64,934 9.59%	215,359 5.99%	579,916 2.39%	101,974 3.81%	22,719 4.81%	124,693 3.96%	111,776 20.60%	48,451 20.10%	42,215 20.62%	90,666 20.33%
Excess deaths (% of actual deaths)	7,572 1%	64,177 30%	66,526 51%	130,703 38%	-31,360 -6%	64,826 39%	46,769 67%	111,595 47%	38,932 26%	-649 -1%	19,757 32%	19,108 17%
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
Actual deaths (%)	NA**	211,236 7.30%	126,365 19.23%	337,601 9.51%	NA**	165,491 6.20%	69,053 14.67%	234,544 7.47%	NA**	45,745 20.24%	57,312 30.70%	103,057 24.97%
Matched twins actual deaths (%)	NA**	160,547 5.55%	75,572 11.50%	236,119 6.65%	NA**	97,864 3.67%	21,958 4.67%	119,822 3.82%	NA**	62,683 27.74%	53,614 28.72%	116,297 28.18%
Excess deaths (% of actual deaths)	NA**	50,689 24%	50,793 40%	101,482 30%	NA**	67,627 41%	47,095 68%	114,722 49%	NA**	-16,938 -37%	3,698 6%	-13,240 -13%

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

354

355 * Excess deaths were calculated as the difference between actual deaths and expected deaths.

356 ** NA: No non-Covid controls were matched to non-Covid patients since the populations were the same.

357 Subjects were categorized as “LTC/SNF” if they received services in either a Long-Term Care (LTC) or Skilled Nursing
358 Facility (SNF) in February 2020, otherwise they were categorized as receiving services in the “Community.” Confirmed
359 Covid-19 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April
360 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September 30, 2020.¹²

361 Probable Covid-19 infection cases were identified using ICD-10-CM codes consistent with the CDC guidance (Z20.828)
362 and WHO recommendations (U07.2).^{13,14} The baseline risk of 9-month mortality defined by the Risk Stratification Index
363 (RSI) calculated on February 29, 2020. Two independent methods were used to estimate expected 2020 mortality as
364 described in the footnote above. The case matching (digital twin) method utilized the baseline risk of 9-month mortality
365 defined by the Risk Stratification Index (RSI). In this method, beneficiaries receiving a diagnosis of probable or confirmed
366 Covid-19 were pairwise exactly matched 1:1 on Feb 29, 2020 with beneficiaries without a Covid-19 diagnosis based on
367 sex, age (within 1-year), ethnicity, location of services in Feb 2020 (community or LTC/SNF), along with RSI as a

368 propensity factor (within 0.1%). The results demonstrate that, within the Medicare population, Covid-19 had a devastating
369 impact by increasing mortality well above what would have been expected based on age and co-morbidities alone.
370 677,100 (2.4%) beneficiaries had confirmed Covid-19 and 2,917,604 (10.3%) had probable Covid-19. 472,329 confirmed
371 cases were community living and 204,771 were in LTC. Mortality following a probable or confirmed diagnosis in the
372 community increased from an expected incidence of about 4% to actual incidence of 7.5%. In long-term care facilities, the
373 corresponding increase was from 20.3% to 24.6%. The absolute increase was therefore similar at 3-4% in the community
374 and in long-term care residents. But the percentage increase was far greater in the community (89%) than among patients
375 in chronic care facilities (21%) who had high baseline risk. The long-term care population without probable or confirmed
376 Covid-19 diagnoses experienced 38,932 excess deaths (35%) compared to historical estimates. Limitations in access to
377 Covid-19 testing and disease under-reporting in long-term care patients probably contributed, although social isolation
378 and disruption in usual care presumably also contributed. Remarkably, there were 31,360 fewer deaths than expected in
379 community dwellers without probable or confirmed Covid-19 diagnoses, representing about a 6% reduction. Disruptions to
380 the healthcare system and avoided medical care were thus apparently offset by other factors, representing overall benefit.
381 The Covid-19 pandemic had marked effects on mortality, but the effects were highly context-dependent.

382 **Public access**

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

387

388 1) Access to RSI risk calculators will be provided free of charge for authorized
389 non-commercial uses via the HDAI API website ([https://www.hda-](https://www.hda-institute.com/api/)
390 [institute.com/api/](https://www.hda-institute.com/api/)).

391 2) Medicare beneficiaries or their health advocates may access their
392 personalized health history and risk assessment by signing into Health
393 Picture (<https://my.healthhpicture.com>). Health picture is an easy-to-use tool
394 that allows Medicare beneficiaries and their family members a way to access
395 their health histories and understand their Covid-19 risks.

396 3) Coefficients for a public version of a one-year RSI mortality model are
397 provided at – ([Risk Stratification Index | Cleveland Clinic](#)).

398

399

400 **Discussion**

401 Age, sex, care location, and comorbidities were significant predictors of mortality. The
402 strongest individual predictor following a diagnosis of Covid-19 across all age
403 categories, and in both community and long-term care settings was the integrated

404 measure of patient co-morbidities, RSI. Although many individual chronic conditions
405 were also significant risk factors in our unadjusted univariable analysis, the strength of
406 these associations were substantially diminished after adjustment for the primary
407 covariates of age, sex, race, location of services, and RSI. Our analysis is consistent
408 with previous work showing that older, non-white men, and patients receiving services
409 in long term care facilities are at special risk of dying from Covid-19 infections.

410 Using RSI as a composite measure of baseline mortality risk permitted precise case-
411 control matching, thereby allowing us to estimate excess deaths attributable Covid-19
412 by two complementary methods in Medicare recipients with probable or confirmed
413 Covid-19 diagnoses. Using the historical comparison, there was an increase from
414 215,359 expected to 346,062 actual deaths, representing 130,702 excess deaths and a
415 61% increase. Using matching, mortality increased from 236,119 expected to 337,601
416 observed deaths, representing 101,482 excess deaths and a 43% increase. Both
417 estimates far exceed the 15-20% excess mortality estimate reported in previous
418 analyses that included younger populations. Our results are therefore consistent with
419 the belief that older people are at much higher risk for developing severe Covid-19
420 illness — and of dying from it.

421 Overall, our historical model indicated that mortality following a probable or confirmed
422 diagnosis in the community increased from an expected incidence of about 4% to actual
423 incidence of 7.5%. In LTC/SNF's, the corresponding increase was from 20.3% to
424 24.6%. Therefore, the absolute increase in mortality was similar at 3-4% in the
425 community and in long-term care residents. However, baseline risk (RSI) associated
426 with all individuals in a care setting varied greatly, being only about 2.6% in the

427 community versus 20.5% in long-term care facilities. As a percentage, the relative
428 increase in mortality was thus far greater in the community (89%) than among patients
429 in long term care facilities (21.0%).

430 Somewhat remarkably, overall mortality *decreased* in Medicare participants without
431 probable or confirmed Covid-19 diagnoses. In fact, among community dwellers, there
432 were 31,360 fewer deaths than expected, representing about a 6% reduction.
433 Disruptions to the healthcare system and avoided medical care were thus apparently
434 offset by other factors, representing overall benefit. Obvious health benefits of pandemic
435 isolation include reduced exposure to other airborne illnesses such as influenza, fewer
436 driving accidents and fewer homicides. However, none seems sufficient to explain the
437 reduction. More subtle effects including reduced work or stress-related illness might
438 contribute more, although there is no obvious reason to believe that the pandemic
439 would reduce stress — especially in an over-65-year-old population.

440 The causes of reduced mortality in community dwelling Medicare participants remains
441 unclear. However, our results suggest that inadequate care for chronic conditions and
442 delayed care of acute events did not produce the feared outcome of higher short-term
443 mortality in the general population without Covid-19. But due to limited follow-up, we
444 caution that disruptions in healthcare delivery may yet result in adverse longer-term
445 outcomes due to delays in the diagnosis and treatment of new and existing chronic
446 conditions. An additional consideration is that prolonged sequela after severe Covid-19
447 infections (Long Covid syndrome) appear substantial and is an area requiring urgent
448 further study.(18)

449 There was a distinct disparity between community dwellers and those in long-term care
450 facilities with respect to historical mortality comparisons. In contrast to community
451 Medicare participants, the long-term care population without probable or confirmed
452 Covid-19 diagnoses experienced 38,932 excess deaths (35%) compared to historical
453 estimates. We believe that limitations in access to Covid-19 testing and disease under-
454 reporting in long-term care patients probably were responsible for this finding. It seems
455 likely that many of the excess deaths in this vulnerable population were consequent to
456 undiagnosed Covid-19 infections. But it is also probable that social isolation and
457 disruption in usual care may have contributed as well. The higher-than-expected level of
458 excess deaths observed in this cohort (subjects without a probable or confirmed Covid
459 diagnosis) is reflected in our case matching results, which indicate a modest relative
460 reduction in deaths in subjects with a Covid related diagnosis. This is most likely due to
461 undiagnosed Covid cases included in the control population, but we cannot rule out the
462 possibility that the focus on care for the Covid patients had an unintended adverse
463 impact on the remaining population.

464

465 **Limitations**

466 We excluded less than 2.2% of the available population because of missing and
467 inconsistent values. Because data were missing non-systematically, exclusion of these
468 subjects was unlikely to introduce meaningful bias. We relied on administrative
469 diagnostic claims for Covid-19 to assign exposure. Surely these are inexact, especially
470 during our study period early in the pandemic. Furthermore, a new diagnostic code for

471 confirmed Covid-19 (U07.1) was introduced on April 1, 2020, and we assume that there
472 was some uncertainty regarding its proper application. However, Kadri et al recently
473 reported that this Covid-19 specific code showed high sensitivity and specificity
474 compared with the PCR test results.(19) A second limitation is that we did not account
475 for temporal changes in risk of exposure to Covid-19 in either setting, nor for
476 improvements in treatment of infected individuals over time.(20, 21)

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

485 Our analysis was based on 28,389,098 adults enrolled in the US fee-for-service and
486 Medicare/Medicaid program. The results are therefore broadly applicable to Medicare
487 eligible adults. Although our sample included a fair number of dual eligible subjects
488 below age 65, our results should only be cautiously generalized to younger and
489 healthier populations.

490 **Summary**

491 Mortality following a probable or confirmed Covid-19 diagnosis in the community
492 increased from an expected incidence of about 4% to actual incidence of 7.5%. In long-

493 term care facilities, the corresponding increase was from 20.3% to 24.6%. The absolute
494 increase was therefore similar at 3-4% in the community and in long-term care
495 residents. But the percentage increase was far greater in the community (89%) than
496 among patients in chronic care facilities (21%) who had high baseline risk.

497 The long-term care population without probable or confirmed Covid-19 diagnoses
498 experienced 38,932 excess deaths (35%) compared to historical estimates. Limitations
499 in access to Covid-19 testing and disease under-reporting in long-term care patients
500 probably contributed, although social isolation and disruption in usual care presumably
501 contributed. Remarkably, there were 31,360 fewer deaths than expected in community
502 dwellers without probable or confirmed Covid-19 diagnoses, representing about a 6%
503 reduction. Disruptions to the healthcare system and avoided medical care were thus
504 apparently offset by other factors, representing overall benefit.

505 The Covid-19 pandemic had marked effects on mortality, but the effects were highly
506 context-dependent. Among community dwelling Medicare participants with suspected or
507 confirmed Covid-19 diagnoses, mortality nearly doubled, but from a relatively low
508 baseline. Patients in long-term care facilities had a similar absolute increase in mortality,
509 but because their baseline mortality was 20.5%, the relative increase was smaller. In
510 contrast, community dwelling Medicare participants without COVID had about 6% lower-
511 than-expected mortality.

512

513

514 **Acknowledgements**

515 Author Contributions: Greenwald and Chamoun had full access to all the data in the
516 study and take responsibility for the integrity of the data and the accuracy of the data
517 analysis.

518 Data Curation: Chamoun, Greenwald

519 Formal Analysis: Greenwald, Chamoun

520 Funding Acquisition: Chamoun, Gray

521 Investigation: Chamoun, Greenwald

522 Methodology: Manberg, Sessler, Chamoun

523 Project Administration: Greenwald, Clain

524 Resources: Chamoun

525 Software: Chamoun, Greenwald

526 Supervision: Manberg, Greenwald, Chamoun

527 Validation: Sessler, Maheshwari, Manberg, Gray, Clain

528 Visualization: Clain, Greenwald

529 Writing – Original Draft Preparation: Manberg, Greenwald, Sessler

530 Writing – Review and Editing: Sessler, Maheshwari, Gray, Clain, Manberg, Chamoun,
531 Greenwald

532 Conflict of Interest Disclosures: Greenwald, Chamoun, Manberg, Clain and Gray are
533 employees of and hold equity interest in Health Data Analytics Institute. Dr. Sessler is a
534 paid consultant of and holds equity interest in Health Data Analytics Institute. Dr.
535 Maheshwari received no compensation from Health Data Analytics Institute.

536 Funding/Support: The study was funded by Health Data Analytics Institute. The Robert
537 Wood Johnson Foundation provides support for providing access to the Covid-19 RSI₃₆₅
538 risk prediction models by academic researchers for non-commercial purposes via the
539 HDAI API (<https://www.hda-institute.com/api/>).

540 Role of Funder/Sponsor: Other than the authors and Data Analyst Douceur Tengu, no
541 additional members of Health Data Analytics had a role in the design and conduct of the
542 study; collection, management, analysis, and interpretation of the data; preparation,
543 review, or approval of the manuscript; or decision to submit the manuscript for
544 publication.

545
546 We thank John Parks, Zhenyu Hong and Douceur Tengu (Data Analysts, Health Data
547 Analytics Institute) for their assistance in developing models and preparing figures and
548 tables.

549 **References**

- 550 1. CDC. CDC COVID Data Tracker: Centers for Disease Control and Prevention;
551 2021 [Available from: [https://covid.cdc.gov/covid-data-](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days)
552 [tracker/#cases_casesper100klast7days](https://covid.cdc.gov/covid-data-tracker/#cases_casesper100klast7days).
- 553 2. Chen YH, Glymour MM, Catalano R, Fernandez A, Nguyen T, Kushel M, et al.
554 Excess Mortality in California During the Coronavirus Disease 2019 Pandemic, March to
555 August 2020. *JAMA Intern Med.* 2020.
- 556 3. Faust JS, Krumholz HM, Du C, Mayes KD, Lin Z, Gilman C, et al. All-Cause
557 Excess Mortality and COVID-19–Related Mortality Among US Adults Aged 25-44 Years,
558 March-July 2020. *JAMA.* 2020.
- 559 4. Lauren M. Rossen AMB, Farida B. Ahmad, Paul Sutton, Robert N. Anderson.
560 Excess Deaths Associated with COVID-19, by Age and Race and Ethnicity — United
561 States, January 26–October 3, 2020 2021 [Available from:
562 <https://www.cdc.gov/mmwr/volumes/69/wr/mm6942e2.htm>.
- 563 5. CDC. Medical Conditions and Risk of for Severe COVID-19 Illness: Centers for
564 Disease Control and Prevention; 2021 [Available from:
565 [https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html)
566 [medical-conditions.html](https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html).
- 567 6. Wang Q, Davis PB, Gurney ME, Xu R. COVID-19 and dementia: Analyses of
568 risk, disparity, and outcomes from electronic health records in the US. *Alzheimers*
569 *Dement.* 2021.

- 570 7. Harrison SL, Fazio-Eynullayeva E, Lane DA, Underhill P, Lip GYH. Comorbidities
571 associated with mortality in 31,461 adults with COVID-19 in the United States: A
572 federated electronic medical record analysis. PLoS Med. 2020;17(9):e1003321.
- 573 8. Wang Q, Berger NA, Xu R. Analyses of Risk, Racial Disparity, and Outcomes
574 Among US Patients With Cancer and COVID-19 Infection. JAMA Oncol. 2021;7(2):220-
575 7.
- 576 9. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al.
577 Prediction models for diagnosis and prognosis of covid-19 infection: systematic review
578 and critical appraisal. BMJ. 2020;369:m1328.
- 579 10. Experton B, Tetteh HA, Lurie N, Walker P, Carroll CJ, Elena A, et al. A Predictive
580 Model for Severe Covid-19 in the Medicare Population: A Tool for Prioritizing Scarce
581 Vaccine Supply. medRxiv. 2020:2020.10.28.20219816.
- 582 11. Dun C, Walsh CM, Bae S, Adalja A, Toner E, Lash TA, et al. A Machine Learning
583 Study of 534,023 Medicare Beneficiaries with COVID-19: Implications for Personalized
584 Risk Prediction. medRxiv. 2020:2020.10.27.20220970.
- 585 12. von Elm E, Altman DG, Egger M, Pocock SJ, Gotsche PC, Vandenbroucke JP,
586 et al. The Strengthening the Reporting of Observational Studies in Epidemiology
587 (STROBE) statement: guidelines for reporting observational studies. Lancet.
588 2007;370(9596):1453-7.
- 589 13. CMS. Preliminary Medicare COVID-19 Data Snapshot | CMS: Centers for
590 Medicare & Medicaid Services; 2021 [Available from: [https://www.cms.gov/research-](https://www.cms.gov/research-statistics-data-systems/preliminary-medicare-covid-19-data-snapshot)
591 [statistics-data-systems/preliminary-medicare-covid-19-data-snapshot](https://www.cms.gov/research-statistics-data-systems/preliminary-medicare-covid-19-data-snapshot)].

- 592 14. CDC. ICD-10-CM Official Coding and Reporting Guidelines April 1, 2020 through
593 September 30, 2020. Centers for Disease Control and Prevention; 2020.
- 594 15. WHO. COVID-19 coding in ICD-10: World Health Organization; 2020 [cited World
595 Health Organization. Available from: [https://www.who.int/classifications/icd/COVID-19-](https://www.who.int/classifications/icd/COVID-19-coding-icd10.pdf)
596 [coding-icd10.pdf](https://www.who.int/classifications/icd/COVID-19-coding-icd10.pdf).
- 597 16. Chamoun GF, Li L, Chamoun NG, Saini V, Sessler DI. Validation and Calibration
598 of the Risk Stratification Index. *Anesthesiology*. 2017;126(4):623-30.
- 599 17. CMS. Condition Categories - Chronic Conditions Data Warehouse: Centers for
600 Medicare & Medicaid Services; 2021 [Available from:
601 <https://www2.ccwdata.org/web/guest/condition-categories>.
- 602 18. Lopez-Leon S, Wegman-Ostrosky T, Perelman C, Sepulveda R, Rebolledo PA,
603 Cuapio A, et al. More than 50 Long-term effects of Covid-19: a systematic review and
604 meta-analysis. *medRxiv preprint*. 2021.
- 605 19. Kadri SS, Gundrum J, Warner S, Cao Z, Babiker A, Klompas M, et al. Uptake
606 and Accuracy of the Diagnosis Code for COVID-19 Among US Hospitalizations. *JAMA*.
607 2020;324(24):2553-4.
- 608 20. Nguyen NT, Chinn J, Nahmias J, Yuen S, Kirby KA, Hohmann S, et al. Outcomes
609 and Mortality Among Adults Hospitalized With COVID-19 at US Medical Centers. *JAMA*
610 *Netw Open*. 2021;4(3):e210417.
- 611 21. Tenforde MW, Fisher KA, Patel MM. Identifying COVID-19 Risk Through
612 Observational Studies to Inform Control Measures. *JAMA*. 2021.

613

614 **Supporting Information**

615

616 **S1 Fig. Consort style waterfall flowchart detailing population selection**

617 **methodology.** Confirmed Covid-19 cases were identified consistent with CMS
618 guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020 and U07.1
619 thereafter) as a primary or secondary diagnosis between March 1, 2020 and September
620 30, 2020.¹² Probable Covid-19 infection cases were identified using ICD-10-CM codes
621 consistent with the CDC guidance (Z20.828) and WHO recommendations (U07.2).^{13,14}
622 Subjects were excluded for missing data if values for any baseline characteristic used in
623 the study were missing (i.e., age, sex, ethnicity, location of care, zip code derived
624 measures, dates of coverage, or baseline risk of 9 month mortality assessed with the
625 Risk Stratification Index (RSI).) Additionally, we excluded subjects whose records had
626 inconsistent values among source files containing similar variables such as birth date
627 and sex.

628

629 **S2 Fig. Performance characteristics of RSI model. Panel A: ROC curve, Panel B**
630 **Calibration plot, Panel C Sensitivity and Positive Predicted Value vs probability of**
631 **mortality.** (A) Area Under the Receiver Operating Curve (AuROC) for the development
632 Learn Set (80% of 2018 Set) was 0.88 (95% Confidence Interval of [0.88-0.88]). AuROC
633 for the prospective Test Set (20% of 2018 Set) was 0.88 (95% Confidence Interval of
634 [0.88-0.88]). Similar performance in the Test Set compared to the Learn Set supports a
635 lack of overfitting in the development of the predictor.

636 (B) The calibration plot displays the mean actual vs predicted 1 year mortality for
637 populations clustered in increments of 1% probability of mortality. Dark green, light
638 green, and red dots are populations of the lowest 95%, 95%-99%, and top 1% risk of
639 mortality. The diagonal line identifies the domain of ideal performance where actual and
640 expected mortality rates are equal for a population. The performance of this index is
641 very close to ideal performance for approximately 99% of the population. Tabulated
642 metrics: The sample size in this test set (N) was 11,923,144 with an incidence of 1yr
643 mortality (Event_Test) of 4.5%. The Slope and Intercept (INT) fit of the data are 0.94
644 and 0.01, respectively. The area under the Receiver Operating Curve was 0.88. The
645 Mean Average Error (MEA) from cluster coordinates (i.e., (expected, actual) couplets) to
646 the identity line was calculated for the database divided into populations grouped from
647 the riskiest to least risky subjects using cluster sizes ranging from 1 (i.e., each individual
648 as a cluster) to 1000 neighboring subjects (e.g., MAE to MAE_1000). The 95%
649 Confidence Interval (CI) for the fits of these populations to the identify line is tabulated
650 (i.e., AE_CI to AE_CI_1000). Rsq_unit is a goodness of fit measure of individual results
651 to the ideal line.

652 (C) Positive Predictive Accuracy (blue dots) and Sensitivity (purple dots) versus the
653 fraction of population, sorted by the risk of 1 year mortality. The vertical red line
654 indicates where the number of patients above the risk threshold equals the incidence of
655 mortality in the population. There are several metrics tabulated in the figure: The area
656 under the Receiver Operating Characteristic (ROC) is 0.88. The incidence of mortality
657 (IR) in the population of 11,891,922 was 4.4%. Vertical bars help identify the PPA and
658 sensitivity performance for detectors operating to identify the riskiest 5%, 10%, and 20%

659 patients. The PPV, sensitivity and relative risk (RR) are tabulated for these detector
660 operating points.

661

662

663 **S3 Fig. Comparison of RSI (Panel A) and Chronic Condition based models (Panel**
664 **B) on 2020 population. ROC curves for all subjects, No Covid-19, Probable Covid-**
665 **19 and Confirmed Covid-19 populations.** Confirmed Covid-19 cases were identified
666 consistent with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before
667 April 1, 2020 and U07.1 thereafter) as a primary or secondary diagnosis between March
668 1, 2020 and September 30, 2020.¹² Probable Covid-19 infection cases were identified
669 using ICD-10-CM codes consistent with the CDC guidance (Z20.828) and WHO
670 recommendations (U07.2).^{13,14} (A,B) ROCs display the sensitivity vs. 1 – specificity in
671 detecting patients who died within 9 months after prediction from February 29,2020
672 (baseline). The areas under each ROC, with their corresponding 95% confidence
673 intervals, are tabulated in the lower right of each figure. Predictions using RSI yielded
674 better performance (A) than those using a model based on age, sex and chronic
675 conditions (B).

676

677

678 **S4 Fig. Forest plot showing the relative risk and 95% CI of significant predictors**
679 **of confirmed Covid-19 infection.** Confirmed Covid-19 cases were identified consistent
680 with CMS guidance using ICD-10-CM codes for Covid-19 (B97.29 before April 1, 2020
681 and U07.1 thereafter) as a primary or secondary diagnosis between March 1, 2020 and

682 September 30, 2020.¹² Subjects were categorized as “LTC/SNF” if they received
683 services in either a Long Term Care (LTC) or Skilled Nursing Facility (SNF) in February
684 2020, otherwise they were categorized as receiving services in the “Community.”
685 Predictors were assessed at baseline (February 29, 2020) and include quintiles of Risk
686 Stratification Index (RSI), presence of chronic conditions, location of services (LTC/SNF
687 vs Community), and demographic variables (i.e., age, sex, race, and quintiles of median
688 household income imputed by zip code according to 2015 Census data.) Variables not
689 remaining in the adjusted model are indicated by the presence of empty parenthesis
690 under the adjusted odds ratio. Location of services, age, status of end-stage renal
691 disease (ESRD) and RSI were the strongest (unadjusted) predictors of infection.
692 Location of services and ESRD remained strong predictors following adjustment;
693 however, risks associated with having chronic conditions were typically reduced when
694 adjusted by the presence of other factors.

695

696

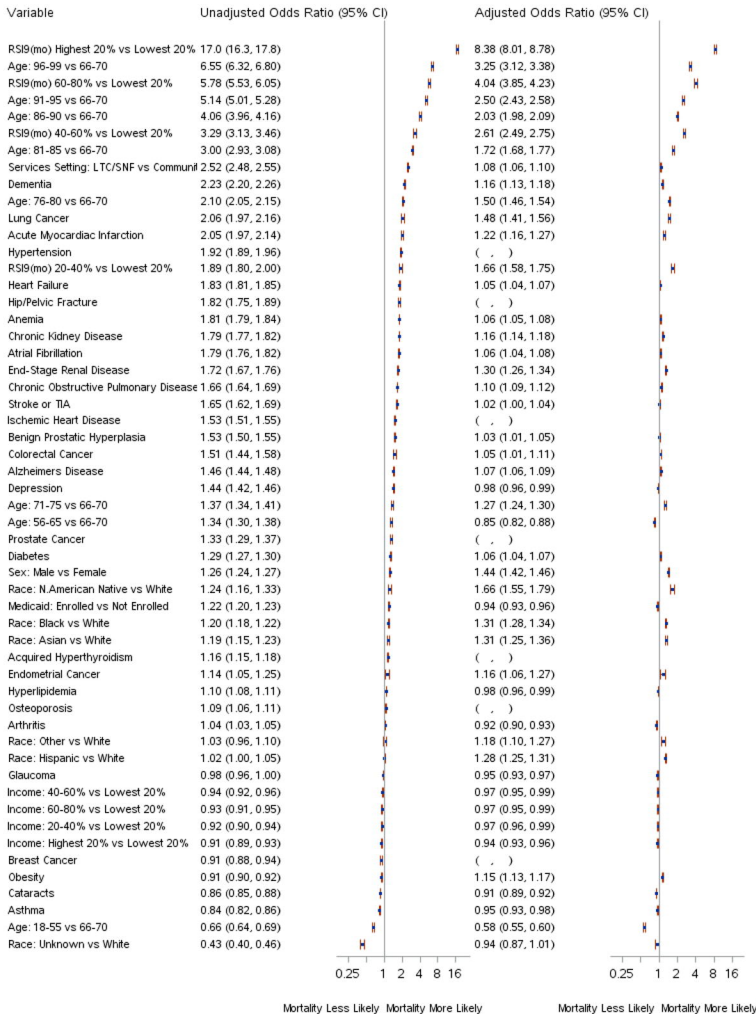
697 **S5 Fig. Observed mortality rates by age and RSI quintiles.** Rates of mortality within
698 9 months following baseline (February 29, 2020) in Medicare subpopulations
699 categorized by age, location of services, infection status, and quintiles of the baseline
700 risk of mortality assessed using the Risk Stratification Index (RSI). Confirmed Covid-19
701 cases were identified consistent with CMS guidance using ICD-10-CM codes for Covid-
702 19 (B97.29 before April 1, 2020 and U07.1 thereafter) as a primary or secondary
703 diagnosis between March 1, 2020 and September 30, 2020.¹² Probable Covid-19
704 infection cases were identified using ICD-10-CM codes consistent with the CDC

705 guidance (Z20.828) and WHO recommendations (U07.2).^{13,14} Subjects were
706 categorized as “LTC/SNF” if they received services in either a Long-Term Care (LTC) or
707 Skilled Nursing Facility (SNF) in February 2020, otherwise they were categorized as
708 receiving services in the “Community.” As expected, subjects in quintiles with higher
709 baseline risk of mortality had higher rates of observed mortality. For subjects without a
710 Covid diagnosis, mortality rates were lower in the community setting compared to those
711 in the LTC/SNF; however, for subjects with confirmed or probable Covid infection,
712 mortality rates were typically higher in the community setting than in the LTC/SNF.

713

714 **S1 File. RSI Development Method for 9-month predictions.**

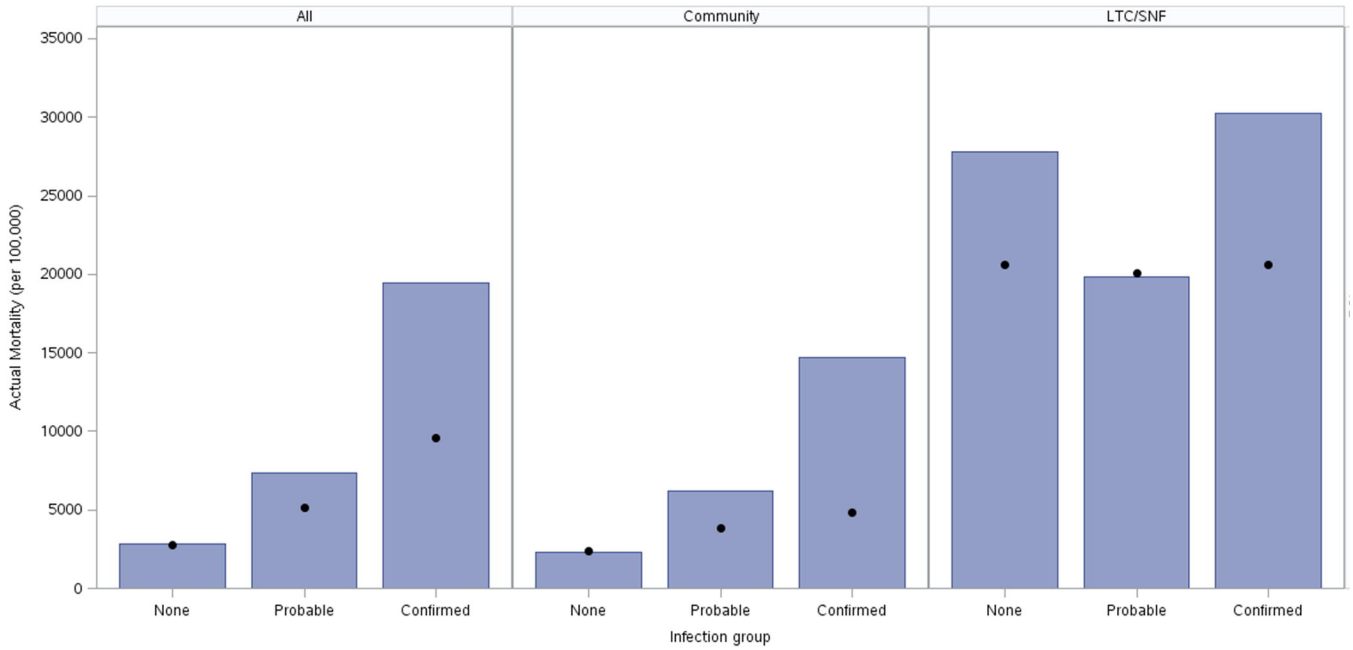
715



Mortality Less Likely Mortality More Likely

Mortality Less Likely Mortality More Likely

(A)



(B)

