medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A Multi-Factor Risk Model for Severe Covid-19 for Vaccine Prioritization and Monitoring Based on a 15 Million Medicare Cohort

AUTHORS

Bettina Experton, M.D., M.P.H., Hassan A. Tetteh, M.D., Nicole Lurie, M.D., M.S.P.H., Peter Walker, Ph.D., Colin J. Carroll, M.S., Adrien Elena Ph.D., Christopher S. Hein, Ph.D., Blake Schwendiman B.S., Christopher R. Burrow, M.D.

Author Affiliations

From the Project Salus healthcare analytics group at Humetrix (B.E., A.E, C.S.H, B.S., C.R.B), Del Mar, CA, the Department of Defense Joint Artificial Intelligence Center (JAIC) Warfighter Health Mission Team (H.A.T., P.W.), Washington, D.C., the Coalition for Epidemic Preparedness Innovation (CEPI), Oslo, and Harvard Medical School (N.L.), Boston, MA, and the Johns Hopkins University Applied Physics Laboratory (C.J.C.), Baltimore, MD.

ABSTRACT

Background:
Public Health interventions to slow the spread of the Covid-19 pandemic focus on protecting individuals at risk for severe disease. Risk categorization is essential to effective pandemic response. However, existing risk models for severe Covid-19 lack needed integration of both socio-demographic and clinical risk factors, and geographic characteristics.

Methods:
We present an integrated multi-factor risk model for severe Covid-19 using de-identified Medicare claims from which we extracted demographic and clinical data for a cohort of 15 million Medicare beneficiaries with 770,000 Covid-19 cases, and socio-economic data at the county and zip code level from the CDC Social Vulnerability Index. The model and associated digital maps were developed as part of Project Salus of the Department of Defense Joint Artificial Intelligence Center, for use by the National Guard and other military personnel in their support mission to hospitals and local jurisdictions impacted by the pandemic.

Results:
The model affirms ethnicity (Black: OR 1.64; 95% CI 1.61–1.68, American Indian: OR 2.21; 95% CI 2.01-2.42), age over 85 (OR 1.75, 95% CI 1.69-1.81), the socio-economic factor of residing in a zip code in the lowest quartile of income (OR 1.23; 95% CI 1.21-1.26), ESRD (OR 2.35; 95% CI 2.25-2.45) and chronic lung disease (OR 1.95; 95% CI 1.90-2.00) as leading risk factors for Covid-19 hospitalizations, but reveals low risk for COPD (OR 1.15; 95% CI 1.13 -1.17) and minimal or no risk for diabetes (OR 1.03; CI 1.01-1.05), CHF (OR 1.10, 95% CI 1.08-1.12) or hypertension (OR 0.96; 95% CI 0.94-0.98), and demonstrates an association between prior herpes zoster immunization (OR 0.74; 95% CI 0.71-0.77), and to a lesser degree prior influenza and pneumococcal vaccines with less severe Covid-19.

Conclusions:
This multi-factor risk model and derived digital maps can be applied for use by national and local health authorities to augment existing tools for pandemic response, including monitoring of post Covid-19 sequelae, prioritization of Covid-19 vaccine, and vaccine monitoring for both safety and efficacy.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

INTRODUCTION:

Over eight million people in the United States have been infected with SARS-CoV-2, with more than 225,000 deaths of whom 80% are over the age of 65.[1,2] Public health interventions to slow the spread of the pandemic have focused on protecting individuals most at risk for exposure to SARS-CoV-2, or at risk for severe Covid-19. The Centers for Disease Control (CDC) has identified people of color, individuals with certain comorbidities and those over the age of 65 at highest risk.[3]

This risk-based approach to pandemic response is now directing Covid-19 vaccine prioritization to allocate the first available vaccine doses to individuals most at risk. All Covid-19 vaccine prioritization frameworks reviewed by the Advisory Committee for Immunization Practices (ACIP) are risk-driven with a time-phased approach.[4] The National Academies of Science, Engineering and Medicine recommend a four-phase equitable plan determined by risk categories.[5] The "jumpstart phase 1a" allocates the first vaccine doses to front line healthcare professionals and emergency responders, followed by phase 1b for high risk individuals with comorbidities or older individuals living in congregate or overcrowded settings. In addition, the National Academies recommend special efforts to deliver vaccines to residents of high-vulnerability areas by using the CDC's Social Vulnerability Index (SVI)[6] or the Covid-19 Community Vulnerability Index (CCVI).[7]

Multiple-factor risk categorization is needed for vaccine allocation schemes to simultaneously include individual demographics, comorbidities, and socioeconomic characteristics. Integration of these various data should also consider overlap of risk categories, as the CDC highlighted during the ACIP October meeting.[8] This is important when considering that more than half of the 53 million Americans over age 65 suffer from two or more chronic conditions,[9] qualifying them for phase 1b vaccine allocation.

At a population level, both clinical and sociodemographic risk need to be considered in tandem. The current CDC listing of risk factors for severe Covid-19[10] is derived from single hospital-based studies with limited sample sizes,[11,12] or hospital reporting, both of which lack nationwide representation.[13] The Center for Medicare and Medicaid Services (CMS) issues monthly Medicare Covid-19 Data Snapshots that include national demographic characteristics and prevalence of common chronic conditions of hospitalized fee-for-service (FFS) Medicare beneficiaries, but lack more detailed diagnosis, medication, and procedure data as well as socio-economic data needed for identifying risk factors for severe Covid-19 at the individual and population levels.[14] There are no analyses that fully support operationalization of the National Academies recommendation to use both clinical and SVI or CCVI risk factor data. The CCVI provides an integrated risk model of both CDC identified risk factors for severe Covid-19 and SVI data, at the state and county, but not at the local zip code level. This and other models of this type, such as the Pandemic Vulnerability Index,[15] have further limitations. They offer population (not individual) based risk data, include limited clinical data with certain chronic conditions with their geographic distributions, and use dissociated geographic data sources (national data for comorbidities, and census tract level data for SVI socio-economic risk factors).

This study integrates multiple factors in its risk models for severe Covid-19 in the Medicare population, based on de-identified Medicare claim data from which detailed clinical and past medical history data were extracted, the CDC SVI and other public datasets. The risk models were developed in part based on prior analysis of Medicare patient outcomes in the frail elderly population,[16] and the Medicare claim data processing conducted by Humetrix with our CMS approved technology as part of the CMS Blue Button program since 2011.[17] The resulting risk models are both individual and population based and has been used to create an interactive dashboard currently hosted on a secure government intranet, enabling visualization of calculated population risk at national, county and zip code levels.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The risk models and associated mapping were developed as part of a Department of Defense (DoD) Joint Artificial Intelligence Center (JAIC) Covid-19 related data analytics project named Salus (the Roman goddess for safety and health), to provide predictive visualization tools to the National Guard and other military personnel in charge of resourcing Covid-19-related healthcare and other vital sector supply and personnel assets, especially serving local hospitals and health departments affected by the pandemic.[18]

Following on the National Academies recommendation for use of existing systems across all levels of government, we discuss the possible use of the Salus Medicare risk models and its derived dashboard as a tool for supporting equitable allocation, distribution of Covid-19 vaccine and vaccine monitoring, as well as for Covid-19 surveillance in the higher risk Medicare population.

METHODS:

Study design and data sources

The Salus Medicare risk models are based on an observational study of all FFS Medicare beneficiaries who since January 1, 2020, either had a Covid-19 test or diagnosis, or for any medical reason had an emergency department, urgent care, or telehealth visit, or were hospitalized. The CMS Office of Enterprise Data and Analytics provides for this cohort weekly Medicare claim outputs from the CMS Chronic Condition Warehouse with Project Salus specified outpatient and inpatient institutional, Part B professional, skilled nursing facility, and hospice claims extending back to October 1, 2019, and Part D claims starting January 1, 2020. De-identified claims are received in the ECS Federal Secure Unclassified Network (SUNet) for processing and analysis by Salus partner Humetrix to build risk models for severe Covid-19. In addition to the clinical and demographic data generated from Medicare claims, the models used selected socio-economic variables from the CDC SVI.

Data processing

The Humetrix Enterprise Platform with its claim and other health data analytics software provides weekly data processing of over 100 million CMS de-identified Medicare claim records in its secure JAIC SUNet enclave, including data upload into a database to generate Covid-19 outcome datasets for: Covid-19 confirmed cases and related hospitalizations, ICU and ventilator use, outpatient care only, and death. Medicare claim-generated clinical, demographic and outcome variables are then merged with CDC SVI variables for the residential zip codes of individuals in the cohort.

Independent variables

The independent variables included in our severe Covid-19 risk models are: beneficiary age, sex, ethnicity, insurance coverage and residential zip code, prior health care utilization (prior hospitalization(s), skilled nursing home admissions, etc.) as a measure for disease severity and frailty, the individual's multiple chronic conditions with in addition to the CMS chronic condition flags, Humetrix compiled diagnostic categories using specific ICD-10 code algorithms, medications grouped by pharmaceutical class, vaccinations before Covid-19 diagnosis, and other variables starting October 1, 2019 (see Supplemental Methods in the Appendix). Socio-economic variables (e.g., income quartile, education, residential density, and other factors) were defined at the individual residential zip code level, after conversion from the census track based data found in the CDC Social Vulnerability Index.

Statistical analysis, variable selection and risk model

To determine significant predictors of Covid-19 related hospitalization, ICU admission and death outcomes, we used logistic regressions (R statistical software, version 3.6 with rms, glmnet and pROC packages). For these regressions, we defined the following groups of Covid-19 cases: those who received

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

outpatient care only (defined as cases that did not require hospitalization or didn't die at least thirty days after diagnosis), hospitalizations attributed to Covid-19, and deaths (defined as cases who died of SARS-CoV-2 infection within 60 days of diagnosis).

The logistic regression models were built on 60% training and 40% validation sets obtained by random selection of Covid-19 confirmed cases. We used a 50:50 ratio of Covid-19 cases who required either hospitalization, or who only received outpatient care for training the hospitalization model, and a 50:50 ratio for Covid-19 survivors versus deceased cases, for training the death model. Determination of correlation coefficients between independent variables, as well as lasso regression, were used to eliminate independent variables which demonstrated significant correlations and collinearities in both models. A stepwise backward variable selection procedure based on the Akaike Information Criterion (AIC) was also used to remove non-significant variables, verifying that independent variables discarded by the lasso regression corresponded with variables identified by the stepwise backward selection procedure. Computation of the 95% confidence intervals for the coefficient estimates were generated by bootstrapping (2000 repetitions) on the training set.

Model performance was measured by the Area Under the Receiver Operating Characteristics (AUROC) curve value. The hospitalization risk model AUROC was computed on a validation set composed of cases that were not used in the training set, with adjustments to give a 60:40 ratio of outpatient to hospitalized cases (which is the observed ratio for hospitalization in the Covid-19 study population). For the death model AUROC we used an 85:15 ratio (which is the observed case fatality ratio in the Covid-19 study population).

Logistic regression coefficients were used to compute individual predicted probabilities of hospitalization in the event of SARS-CoV-2 infection for the entire Salus Medicare cohort of 15 million beneficiaries. The regional percentage of the cohort population over a predicted probability of 0.55, is displayed on a digital risk map at the county and zip code levels. Additional mapping data was also produced to display Covid-19 confirmed cases and their outcomes at the national and local level, with recent, cumulative and times series displays.

RESULTS:

Study population characteristics

Socio-demographic and clinical characteristics of the 15 million study population, as of September 16, 2020, and its subsets of Covid-19 cases (n=770,000) who were either hospitalized with severe disease (n=264,000) or who only required outpatient Covid-19 care (n=299,000), or who died (n=105,000) are summarized in Table 1.

Table 1 displays comparisons between non-SARS-CoV-2 infected FFS Medicare beneficiaries and Covid-19 cases, and among them, hospitalized vs. non-hospitalized and deceased cases. Among Covid-19 cases, the severe disease groups of hospitalized and deceased patients show higher frequencies of diabetes, COPD, ESRD, hypertension, ischemic heart disease, cerebrovascular disease, chronic kidney disease, chronic lung disease, chronic liver disease, and congestive heart failure although the effect sizes for these differences are small (see Supplemental Appendix).

Predictors

We used logistic regression analysis to determine the relative importance of significant predictors of severe Covid-19. The predictor variables for hospitalization are shown on Figure 1 and for death in Figure 2.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Demographic factors are among the strongest predictors of hospitalization and death: ethnicity with American Indian race (hospitalization OR 2.21; 95% CI 2.01-2.42) and Black race (hospitalization OR 1.64; CI 1.61-1.68), and the 85+ age group (death 0R 5.51; 95% CI 5.20-5.83). Among socio-economic factors extracted from SVI data, living in a zip code with the lowest quartile of income is associated with and increased probability of severe Covid-19 (hospitalization OR 1.23; 95% CI 1.21-1.26).

On the clinical side, End Stage Renal Disease (OR 2.35; 95% CI 2.25-2.45), chronic lung disease (OR 1.95; 95% CI 1.90-2.00) which includes idiopathic pulmonary fibrosis and pneumoconiosis, and morbid obesity (OR 1.53; 95% CI 1.49-1.57) are leading risk factors for hospitalization. COPD (hospitalization OR 1.15; 95% 1.13-1.17) and beta-2 agonist bronchodilators (OR 1.18; 95% 1.15-1.22) are modest and independent risk factors for severe Covid-19. On the other hand, diabetes is not a hospitalization predictor (OR 1.03; 95% CI 1.01-1.05) but diabetic Covid-19 patients are at a slightly higher risk of dying from the disease (OR 1.07; 95% CI 1.04-1.10). Hypertension and asthma are not associated with higher odds of Covid-19 hospitalizations or death. ACE inhibitors, angiotensin II blockers, and NSAIDs have modest associations with lower rates of Covid-19-related hospitalization or death.

The regression analyses indicate that prior herpes zoster immunization (with the recombinant GlaxoSmithKline Shingrix® vaccine) is also associated with lower odds of Covid-19 hospitalization or death (hospitalization OR 0.74; 95% CI 0.71- 0.77; death OR 0.71; 95% CI 0.66-0.77). Use of this vaccine was not correlated with any other demographic, socioeconomic or other variable in the model. We observed a smaller association of prior influenza vaccination on hospitalization (OR 0.89; 95% CI 0.88-0.91), but none against death. Pneumococcal vaccination is associated with reduced odds of death (OR 0.87; 95% CI 0.84-0.91), as recently proposed,[19] but not hospitalization.

To test the hypothesis that these vaccinations simply represent preventive behaviors that are also associated with less hospitalization and death, we created a preventive behavior variable based on the use of preventive screenings (colonoscopy and mammography) as well as vaccinations. Indeed, these preventive behaviors were associated with lower odds of developing severe Covid-19 (hospitalization OR 0.73; 95% CI 0.71-0.74, death OR 0.65; 95% CI 0.63-0.67). However, the use of Shingrix was poorly correlated with these other behaviors (preventive screenings and other vaccinations), and vaccination with Shingrix retained a significant reduction in odds ratio, even when preventive behaviors were accounted for in the models. This raises the intriguing hypothesis that this recombinant adjuvanted vaccine may have an unrecognized biological effect which reduces the probability of severe Covid-19.

Mapping

The hospitalization model logistic regression coefficients were used to calculate the predicted probability for hospitalization in the event of SARS-CoV-2 infection, at the individual level, for the entire 15 million cohort. We mapped areas in which the percentage of this cohort with a predicted probability of hospitalization was greater than 0.55 for every residential zip code in the U.S. Figure 3 shows a portion of the Los Angeles metropolitan area with zip codes displaying a wide range of population percentages over this threshold which were positively correlated with the cumulative percentage of Covid-19 patients requiring hospitalization in these zip codes (Pearson correlation coefficient 0.65 $p < 0.001$). We further conducted the same correlation analysis in a set of zip codes drawn from other metro areas shown in Figure S1 in the Appendix. All correlation coefficients in these regions (Washington D.C., 0.61; Houston, 0.66; Miami, 0.72; Phoenix, 0.72; and New York City, 0.52) had p values < 0.001.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

DISCUSSION:

The severe Covid-19 risk models and the mapping of their outcomes are based, to our knowledge, on the largest Covid-19 dataset assembled to date for this purpose. All members of the study cohort are active users of healthcare services, presenting on average a frailer clinical profile than the general Medicare population,[20] as shown by their clinical characteristics in Table 1. While our observed 770,000 Medicare FFS cases represent a little more than 10% of the total number of Covid-19 cases in the U.S., they contribute close to 50% of all CDC estimated Covid-19-related hospitalizations and a majority of Medicare hospitalizations.[21] The scale of this dataset has allowed us to quantitively and qualitatively validate our hospitalization risk model with actual hospitalizations in multiple geographic areas.

The countrywide distribution, and local representation of the Salus Medicare cohort also enabled Project Salus to offer the first county and zip code level mapping for this population at higher risk for severe Covid-19 and related hospitalizations. Beyond its use for epidemic mapping, the Salus Medicare Covid-19 platform provides a disease surveillance tool for Covid-19, which can be extended to influenza to evaluate the potential compounding effect of these two respiratory diseases on hospitalizations as influenza undergoes its seasonal progress in the coming months.

Our logistic regression analysis, as applied to this very large Medicare population, allowed us to test the current CDC listing of risk factors for severe Covid-19. Our analysis affirms the critical risk factors of ethnicity and older age but reveals the lack of or more modest effect of hypertension, diabetes, COPD, asthma, and cardiovascular diseases.

The new findings of the significant and independent association of less severe Covid-19 with prior use of a recombinant herpes zoster vaccine calls for further investigation. If the effect of herpes zoster vaccination is confirmed, use of this vaccination may be even more beneficial than previously recognized. The independent association of prior use of preventive screenings, a possible marker for a higher rate of Covid-19 preventive behavior (social distancing, hand washing, mask wearing), with less severe Covid-19, is a reminder of the importance of such behaviors.

In preparation for a Covid-19 immunization campaign, the models we have developed integrate both socio-economic factors and individual clinical data, and thus respond to the recommendations of the National Academies for prioritization and allocation of Covid-19 vaccines when they will become available. They could be used by the Medicare program, in collaboration with local health officials to affirmatively invite or encourage highest risk beneficiaries to seek early vaccination and can be used as a tool for vaccine allocation. In Figure 4, we show on a histogram the distribution of the predicted probabilities of hospitalization for SARS-CoV-2 infected patients in the 15 million Salus Medicare population which can identify priority groupings for Covid-19 vaccination, according to the quantities of vaccine doses available. Further, once receipt of vaccination is linked to Medicare claims, the system could be used to support post-licensure pharmacovigilance and effectiveness studies. These are of paramount importance, especially in the early phases of a vaccination campaign.

To conclude, we believe that this quantitative analysis of risk for severe Covid-19 provides important new insights which should be considered before finalization of ACIP Covid-19 vaccine recommendations. State and local governments could consider asking the DoD, through the Mission Assignment process, to provide the Salus Medicare risk mapping to their jurisdictions. The data rich Salus Medicare dataset, which underscores the value of Medicare claim data for epidemiologic surveillance, with its size and nationwide representation, can also augment both the ILINet and COVID-NET disease surveillance systems and complement the existing vaccine monitoring systems for tracking both the safety and efficacy of Covid-19 vaccination in the high risk Medicare population.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

We thank Robert Lyden, Travis Yatsko, Adrien Cirou, and Franz Krachtus of Humetrix for their assistance in project management and in the production of the manuscript figures, tables and video.

Supported by the Johns Hopkins University Applied Physics Laboratory (JHU-APL) under a prime contract with the Department of Defense Joint Artificial Intelligence Center (JAIC) for Project Salus.

Department of Defense Disclaimer:

The views expressed in this article are those of the authors and do not necessarily reflect the official policy or position of the Department of Defense, nor the U.S. Government. Authors are military service members (or an employee of the US Government). This work was prepared as part of their official duties. Title 17 USC§105 provides that copyright protection under this title is not available for any work of the US Government. Title 17 USC§101 defines a US Government work as a work prepared by a military service member or employee of the US Government as part of that person's official duties.

REFERENCES

1. Centers for Disease Control and Prevention. CDC COVID Data Tracker (https://covid.cdc.gov/covid-data-tracker/#cases_casesinlast7days).
2. Centers for Disease Control and Prevention. Older Adults. September 11, 2020 (https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/older-adults.html).
3. Centers for Disease Control and Prevention. People at Increased Risk. September 11, 2020 (https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/index.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fneed-extra-precautions%2Fpeople-at-increased-risk.html).
4. Sara Oliver M.D. Centers for Disease Control and Prevention. ACIP COVID-19 Vaccines Work Group. September 22, 2020 (https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2020-09/COVID-06-Oliver.pdf).
5. The National Academies of Sciences, Engineering, and Medicine. National Academies Release Framework for Equitable Allocation of a COVID-19 Vaccine for Adoption by HHS, State, Tribal, Local, and Territorial Authorities. October 2, 2020 (https://www.nationalacademies.org/news/2020/10/national-academies-release-framework-for-equitable-allocation-of-a-covid-19-vaccine-for-adoption-by-hhs-state-tribal-local-and-territorial-authorities).
6. Centers for Disease Control and Prevention. CDC SVI 2018 Documentation. January 31, 2020 (https://svi.cdc.gov/Documents/Data/2018_SVI_Data/SVI2018Documentation.pdf).
7. Surgo Foundation. The COVID-19 Community Vulnerability Index (CCVI) (https://precisionforcovid.org/ccvi).
8. Kathleen Dooling, M.D., M.P.H. Centers for Disease Control and Prevention. ACIP COVID-19 Vaccines Work Group. September 22, 2020 (https://www.cdc.gov/vaccines/acip/meetings/downloads/slides-2020-09/COVID-07-Dooling.pdf).
9. Centers for Disease Control and Prevention. Percent of U.S. Adults 55 and Over with Chronic Conditions (https://www.cdc.gov/nchs/health_policy/adult_chronic_conditions.htm).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

10. Centers for Disease Control and Prevention. People with Certain Medical Conditions. September 11, 2020 (https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html).
11. CDC COVID-19 Response Team. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 — United States, February 12–March 28, 2020. Centers for Disease Control and Prevention 2020;69(13):382–386.
12. Lithander F, Neumann S, Tenison E, et al. COVID-19 in older people: a rapid clinical review. Oxford Academic 2020;49(4):501-515.
13. Garg S, Kim L, Whitaker M, et al. Hospitalization Rates and Characteristics of Patients Hospitalized with Laboratory-Confirmed Coronavirus Disease 2019 — COVID-NET, 14 States, March 1–30, 2020. Centers for Disease Control and Prevention 2020;69(15):458–464.
14. Centers for Medicare & Medicaid Services. Preliminary Medicare COVID-19 Data Snapshot (https://www.cms.gov/research-statistics-data-systems/preliminary-medicare-covid-19-data-snapshot).
15. Marvel S, House J, Wheeler M, et al. The COVID-19 Pandemic Vulnerability Index (PVI) Dashboard: Monitoring county-level vulnerability using visualization, statistical modeling, and machine learning. medRxiv 2020. DOI: 10.1101/2020.08.10.20169649.
16. Experton B, Li Z, Branch L, Ozminkowski R, Mellon-Lacey D. The impact of payor/provider type on health care use and expenditures among the frail elderly. American Journal of Public Health 1997;87(2):210-216.
17. CMS Blue Button 2.0 (https://bluebutton.cms.gov/).
18. AI in Defense DoD's Artificial Intelligence Blog. The JAIC Forges Ahead. May 20, 2020 (https://www.ai.mil/blog_05_20_20-the_jaic_forges_ahead.html).
19. Root-Bernstein R. Possible Cross-Reactivity between SARS-CoV-2 Proteins, CRM197 and Proteins in Pneumococcal Vaccines May Protect Against Symptomatic SARS-CoV-2 Disease and Death. Vaccines 2020;8(4):559-579; DOI:10.3390/vaccines8040559.
20. Centers for Medicare & Medicaid Services. Prevalence of Chronic Conditions among Fee-for-Service Beneficiaries: 2017 (https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Chronic-Conditions/Chartbook_Charts).
21. COVID-NET. COVID-19 Laboratory Confirmed Hospitalizations. September 26, 2020. (https://gis.cdc.gov/grasp/covidnet/COVID19_5.html).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Table 1. **Demographic, Clinical and Socioeconomic Characteristics of the Project Salus Medicare Cohort**

Variable	Non-Covid-19 Cases†	Covid-19 Outpatients‡	Covid-19 Hospitalized¶	Covid-19 Deaths#
Total	14,692,424	298,674	263,568	104,971
Age	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)
Age	72 (67 - 78)✦✦✦	74 (68 - 82)	76 (68 - 84)✦✦✦	82 (73 - 89)✦✦✦
	%	% (no.)	% (no.)	% (no.)
Under 65	17.9%***	14.5% (43,324)	14.8% (38,881)**	7.6% (7,954)***
From 65 to 74	46.6%***	38.8% (115,899)	31.7% (83,500)***	21.4% (22,512)***
From 75 to 84	22.4%***	27.1% (80,955)	30.2% (79,577)***	30.8% (32,360)***
Over 85	13.1%***	19.6% (58,496)	23.4% (61,610)***	40.1% (42,145)***
Sex	%	% (no.)	% (no.)	% (no.)
Male	43.8%***	40.23% (120,231)	48.4% (127,691)***	48.6% (50,982)***
Female	56.2%***	59.70% (178,443)	51.6% (135,877)***	51.4% (53,989)***
Race	%	% (no.)	% (no.)	% (no.)
North Americ. Native	0.7%	0.5% (1,495)	1.0% (2,543)***	0.9% (904)***
Black	10.5%***	13.0% (38,841)	20.1% (52,958)***	17.8% (18,669)***
Hispanic	2.1%***	3.9% (11,656)	5.2% (13,655)***	4.5% (4,757)***
Asian	1.5%***	2.2% (6,441)	2.5% (6,506)***	2.6% (2,709)***
White	81.8%***	76.5% (228,558)	67.9% (178,944)***	71.4% (74,958)***
Other	1.2%***	1.7% (5,107)	1.8% (48,380)***	1.7% (1,836)
Unknown	2.2%***	2.2% (6,576)	1.6% (4,124)***	1.1% (1,138)***
Income (SVI EPI_PCI)	%	% (no.)	% (no.)	% (no.)
Upper quartiles	NA	86.90% (257,978)	82.0% (215,067)***	83.7% (87,420)***
Poorest quartile	NA	13.10% (38,749)	18.0% (47,331)***	16.3% (17,067)***
Poverty (SVI EPL_POV)	%	% (no.)	% (no.)	(no.)
Upper quartiles	88.1%***	86.80% (257,514)	82.6% (216,640)***	83.9% (87,684)***
Poorest quartile	11.9%***	13.20% (39,213)	17.4% (45,758)***	16.1% (16,803)***
Housing	Median (IQR)	Median (IQR)	Median (IQR)	Median (IQR)
Crowded Quarters (SVI EPL_CROWD)	0.43 (0.29 - 0.59)✦✦✦	0.46 (0.31 - 0.67)	0.48 (0.33 - 0.68)✦✦✦	0.47 (0.31 - 0.67)✦
Multi-Unit (SVI EPL_MUNIT)	0.48 (0.32 - 0.63)✦✦✦	0.54 (0.37 - 0.70)	0.52 (0.36 - 0.69)✦✦✦	0.54 (0.37 - 0.70)✦✦✦
	%	% (no.)	% (no.)	% (no.)
Disabled Status	27.7%***	26.4% (78,961)	28.1% (74,029)***	21.9% (23,033)***
Dual Status	24.1%***	39.9% (119,298)	42.8% (112,795)***	49.6% (52,021)***
Prior Hospitalization	%	% (no.)	% (no.)	% (no.)
0	100.0%	86.1% (257,134)	74.5% (196,405)***	73.1% (76,715)***
1 or more	0.0%	13.9% (41,540)	25.5% (67,163)***	26.9% (28,256)***
Clinical Variables	%	% (no.)	% (no.)	% (no.)
ESRD	2.4%***	2.2% (6,610)	6.8% (17,875)***	5.30% (5,545)***
Chronic Kidney Disease	38.7%***	41.5% (123,931)	54.5% (143,598)***	58.90% (61,835)***
Chronic Lung Disease	5.3%***	5.1% (15,361)	12.1% (31,942)***	11.00% (11,588)***
Chronic Liver Disease	2.4%***	2.5% (7,598)	4.2% (11,087)***	3.60% (3,792)***
COPD	28.1%***	29.1% (86,844)	36.8% (97,084)***	40.20% (42,220)***
CHF	26.0%***	32.0% (95,642)	42.5% (111,911)***	49.80% (52,224)***
Stroke/TIA	14.1%***	19.4% (57,923)	23.5% (61,940)***	29.20% (30,666)***
Diabetes	38.6%***	46.7% (139,604)	54.5% (143,639)***	57.20% (60,019)***
Hypertension	77.2%***	80.7% (241,163)	85.5% (225,431)***	89.30% (93,760)***
Acute MI	1.1%	0.9% (2,540)	1.5% (4,070)***	1.70% (1,745)***
Ischemic Heart Disease	45.4%***	51.2% (152,872)	58.9% (155,154)***	65.70% (69,014)***
Asthma	17.5%***	18.3% (54,770)	20.2% (53,208)***	19.20% (20,118)***
Chemotherapy	5.1%***	10.3% (30,774)	12.6% (33,324)***	11.70% (122,66)***
Obesity	11.5%***	12.9% (38,636)	13.4% (35,402)***	9.20% (95,77)***
Morbid Obesity	7.0%***	7.6% (22,837)	11.4% (30,020)***	7.60% (7,996)***

† Frequencies shown for non-COVID-19 cases were determined from a random sample of 500,000 uninfected cases obtained from the 14.7M Salus cohort. Asterisks shown in this column indicate the level of statistical significance of differences in this population versus Covid-19 infected cases.

‡ Covid-19 outpatients are cases which did not require hospitalization and remained alive at least 30 days after diagnosis. Covid-19 cases that do not meet these criteria and are not hospitalized are not shown in this table.

¶ Covid-19 hospitalized cases required inpatient admission for management. Asterisks shown in this column indicating the level of statistical significance of differences in this group versus the Covid-19 outpatient group.

Covid-19 deaths are deaths attributed to Covid-19 based on timing of death in relation to diagnosis. Asterisks shown in this column indicate the level of statistical significance of differences between Covid-19 cases who died of the disease and those who survived.

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ by Chi Square test

✦ $p < 0.05$; ✦✦ $p < 0.01$; ✦✦✦ $p < 0.001$ by Mann-Whitney test

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

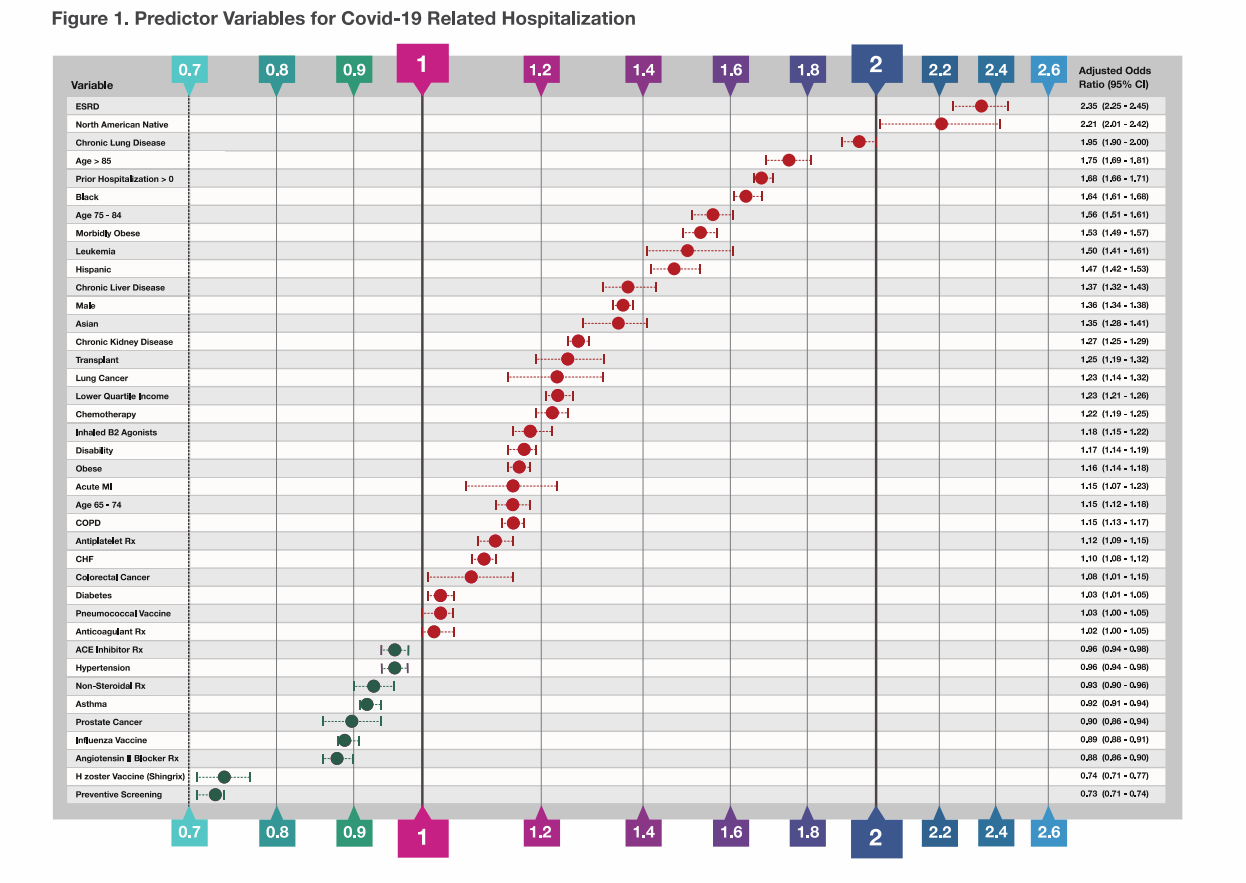


Figure 1: Predictor Variables for Covid-19 Related Hospitalization. The independent variable odds ratios were determined by binary logistic regression analysis of confirmed Covid-19 cases that required hospitalization for the disease and for those that were managed with outpatient care only. In addition to the thirty nine variables shown, HIV (OR 0.90; 95% CI 0.83 – 0.96), a past history of anemia (OR 0.87; 95% CI 0.86 – 0.89); a prescription for Azithromycin overlapping the date of Covid-19 diagnosis (OR 1.18; 95% CI 1.15 – 1.22), Medicare-Medicaid dual coverage and the Medicare defined ethnicity of "Other" were included in the model based on the criteria described in the Methods section. Variables excluded from the model on the basis of these criteria (or for a regression coefficient p value of > 0.05) included prescriptions for Chloroquine or Hydroxychloroquine overlapping the time of Covid-19 diagnosis; the Medicare defined "Unknown" ethnicity, a history of ischemic heart disease, cerebrovascular disease, breast cancer, endometrial cancer; or a history of prescribed immunosuppressive drugs, Histamine type-2 receptor antagonists, inhaled corticosteroids or opioids. The AUROC for the model was 0.65.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

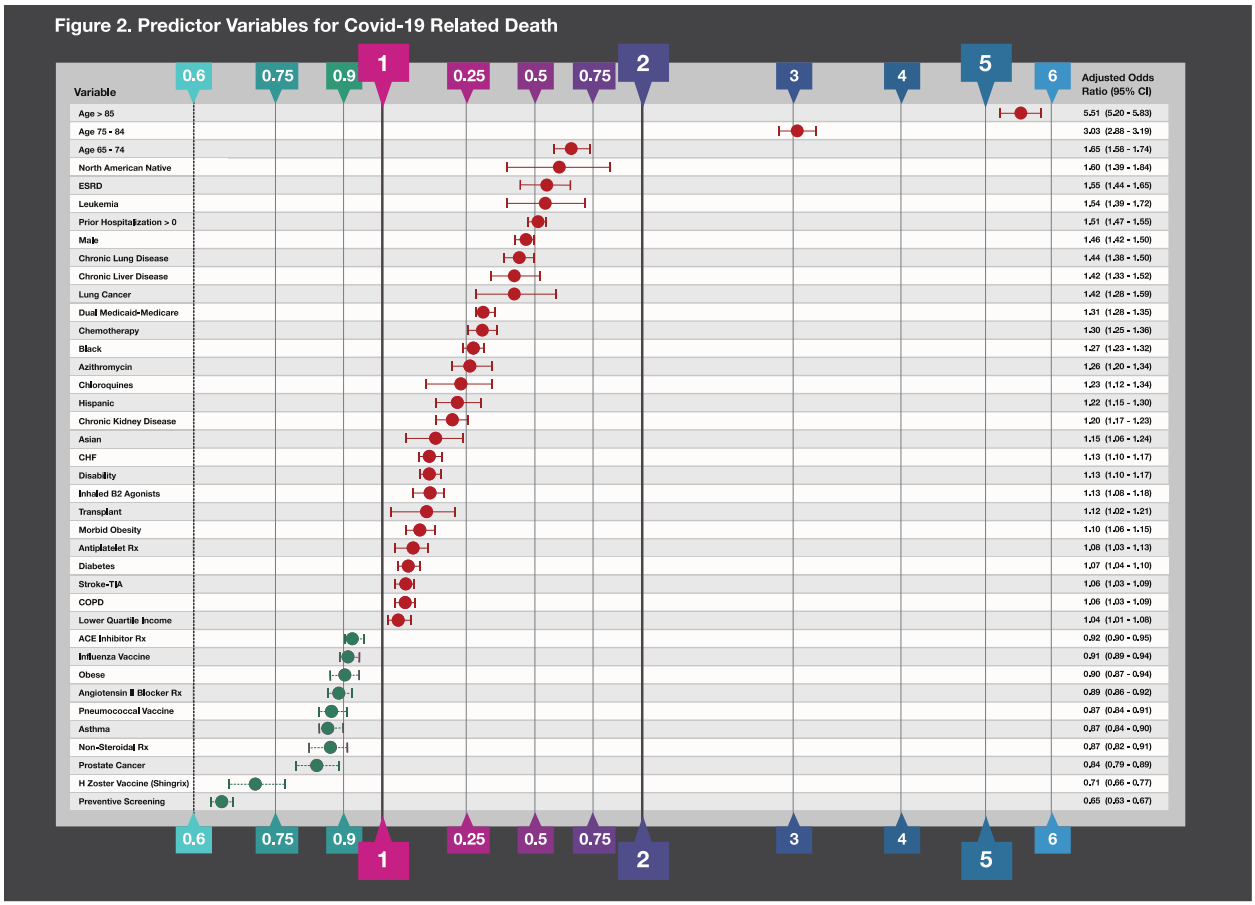
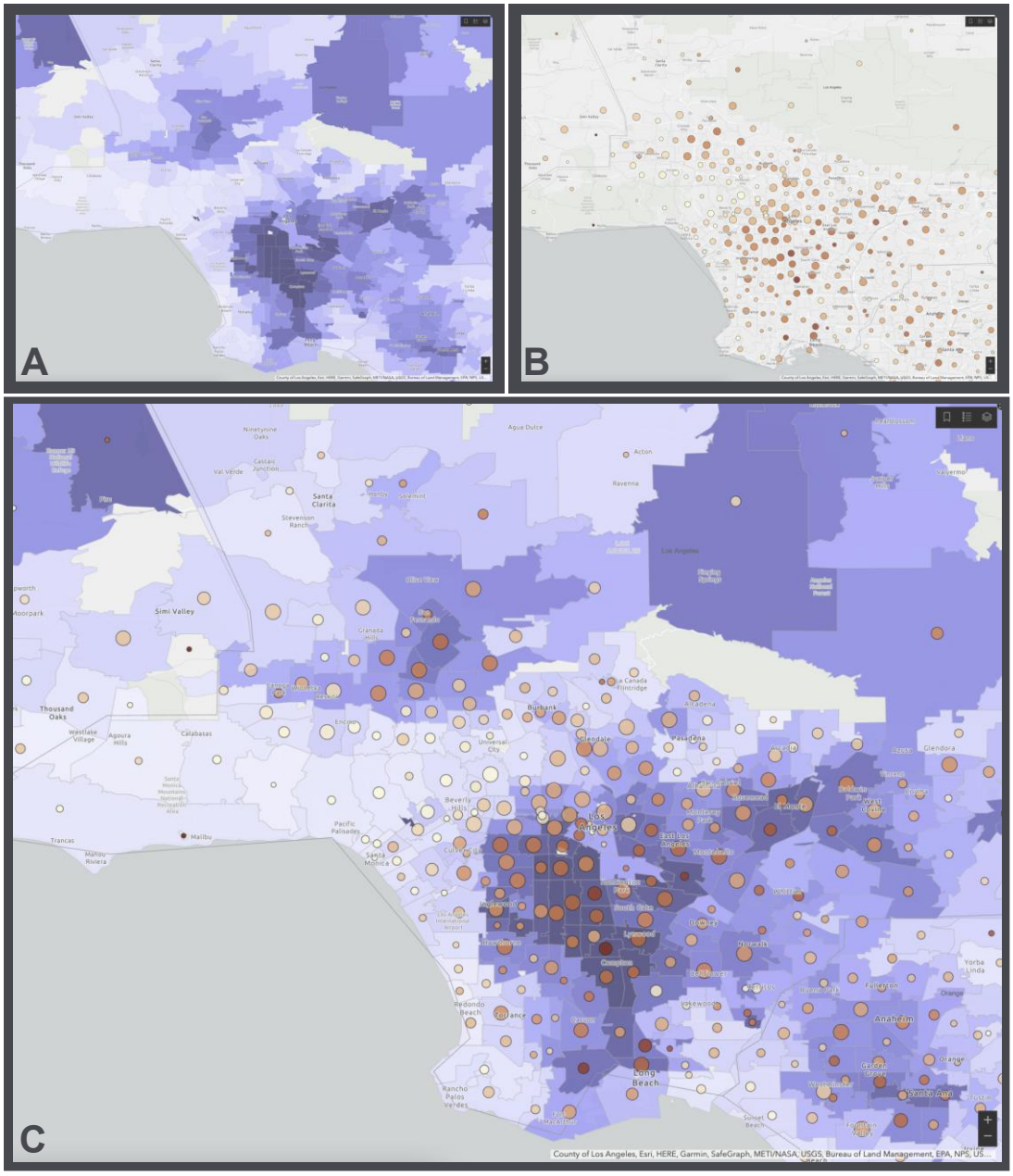


Figure 2: Predictor Variables for Covid-19 Related Death. The independent variable odds ratios were determined by binary logistic regression analysis of confirmed Covid-19 cases that survived and those that died within 60 days of Covid-19 diagnosis. In addition to the thirty nine variables shown, a prescription history which included immunosuppressive drugs (OR 0.93; 95% CI 0.89 – 0.98), and the Medicare defined "Other" ethnicity were included in the model based on the criteria described in the Methods section. Variables excluded from the model on the basis of these criteria (or for a regression coefficient p value of > 0.05) included the Medicare defined ethnicity of "Unknown"; a history of prescribed histamine type-2 receptor antagonists, anticoagulants, inhaled corticosteroids or opioids; a past history of endometrial cancer, colorectal cancer, breast cancer, hypertension, ischemic heart disease, acute myocardial infarction in 2019 or HIV. The AUROC for the model was 0.71.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.


A
B
C
Santa Clarita
Simi Valley
Thousand Oaks
Calabasas
Malibu
Burbank
Glendale
Pasadena
Los Angeles
Beverly Hills
Santa Monica
Long Beach
Anaheim
Santa Ana
Orange
Glendora
Redondo Beach
Torrance
Lakewood
Norwalk
Downey
Inglewood
Compton
Yorba Linda
Fullerton
Rancho Palos Verdes
Agua Dulce
Acton
Ravenna
Ninetynine Oaks
Castaic Junction
Val Verde
Stevenson Ranch
Granada Hills
Encino
Universal City
Pacific Palisades
Trancas
Malibu Riviera
Westlake Village
Agoura Hills
Moorpark
Sunset Beach
Los Alamitos
Garden Grove
Westminster
Cerritos
Hawaiian Gardens
Bellflower
Paramount
Pico Rivera
Vernon
Maywood
Bell
Cudahy
Huntington Park
South Gate
Lynwood
Florence
Hawthorne
Gardena
Alhambra
Monterey Park
Rosemead
El Monte
Baldwin Park
West Covina
Azusa
Duarte
Arcadia
La Canada Flintridge
Altadena
Sierra Madre
Monrovia
Mira Loma
Pedley
Ontario
Rancho Cucamonga
Fontana
Rialto
Bloomington
Colton
Grand Terrace
Highland
Mentone
Redlands
Yucaipa
Calimesa
Beaumont
Cherry Valley
Banning
Cabazon
Desert Hot Springs
Palm Springs
Cathedral City
Rancho Mirage
Palm Desert
Indio
Coachella
Thermal
Mecca
Oasis
Thermal
Vaverma
Pine
Lake Arrowhead
Crestline
Rancho Cucamonga
Ontario
Chino
Chino Hills
Corona
Norco
Eastvale
Jurupa Valley
Mira Loma
Pedley
Riverside
Moreno Valley
Perris
Menifee
Wildomar
Lake Elsinore
Canyon Lake
Murrieta
Temecula
Hemet
San Jacinto
Winchester
Homeland
Sun City
Quail Valley
Menifee
Nuevo
Romoland
Homeland
Green Acres
Perris
Mead Valley
Glen Ivy
Corona
Coronita
Home Garden

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

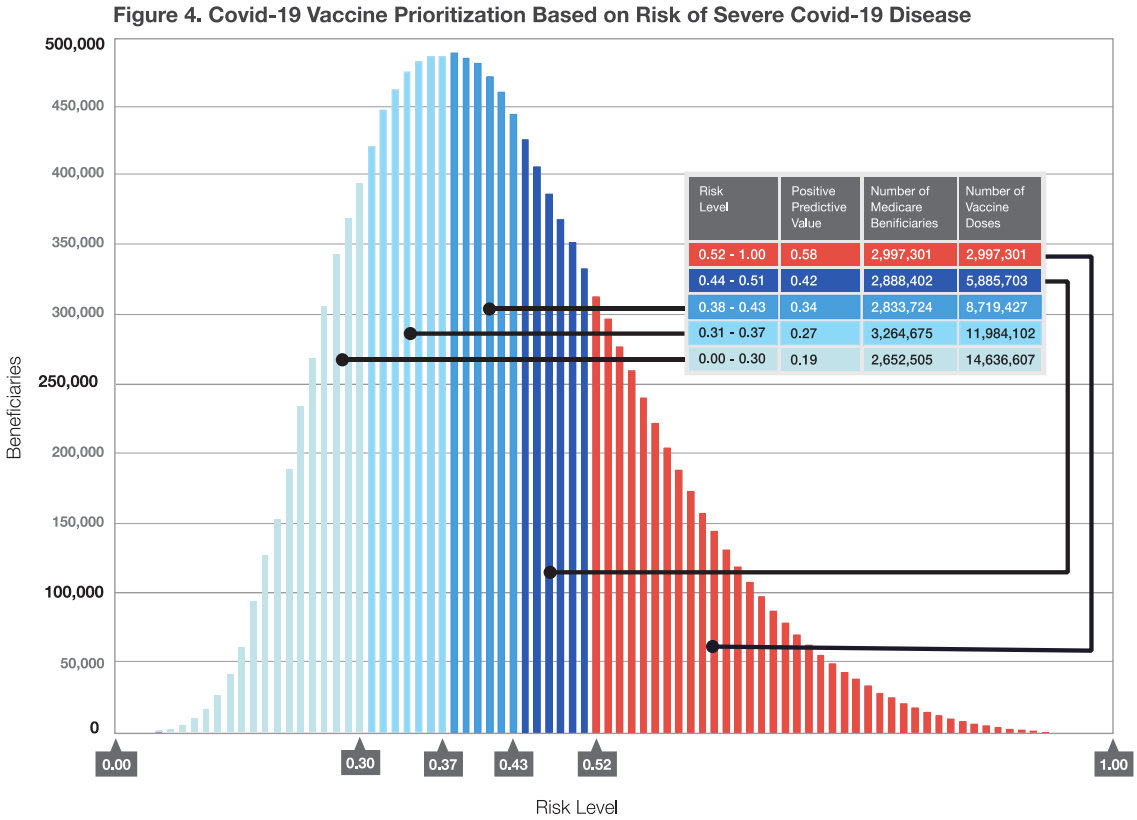


Figure 4 – Covid-19 Vaccine Prioritization Based on Risk of Severe Covid-19 Disease. The logistic regression model coefficients for the independent variables shown in figure 1 were used to calculate the predicted probabilities of hospitalization in the Salus cohort. The distribution of predicted probabilities was split into 5 groups shown in the table of approximately 3M beneficiaries each to enable stratification of the cohort by risk of severe disease in order to prioritize individuals for Covid-19 vaccination.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Supplemental Appendix

Supplemental Methods:

Medicare claim data processing; The Humetrix SaaS platform installed in the SUNet secure classified government network provides automated pre-processing of weekly updates downloaded to the Humetrix SUNet enclave of Medicare Part A inpatient and outpatient, Hospice and SNF, Part B Carrier claims and Part D (PDE) claims data to generate output files containing derived variables used to train logistic regression models as described below. The platform's medical terminology service includes a complete collection of AMA CPT-4, FDA NDC, NNPES NPI, ICD-10-CM, CMS Level II HCPCS, and NLM RxNorm codes and provides automated identification grouping of ICD-10-CM codes to identify chronic condition categories and NDC drug code to RxNorm ingredient code mappings to identify pharmaceutical classes of active pharmaceutical ingredients from Medicare claim data.

Demographic and Coverage Variables: weekly updates of CMS "Master Beneficiary Summary 2020 File" (MBSF_2020) data were processed, extracting variables: ORIG_REASON_FOR_ENTITLEMENT (Disability), ZIP_CD (residential zip code), YOB, SEX_CODE, RACE_CODE, and DUAL_STUS_CD 01-12 (Dual Medicare-Medicaid insurance).

Social Vulnerability Index (SVI) variables associated with zip codes: these were derived from CDC data, which are categorized by census tract. The variables analyzed using binary logistic regression to determine significant predictor variables for hospitalization or death due to Covid-19 included the following SVI variables: EPL_PCI, EPL_POV, EPL_NOHSDP, RPL_THEME1, EPL_CROWD, EPL_GROUPQ, EPL_MUNIT, EPL_THEME4, RPL_THEMES.

Chronic Condition Summary Variables: Chronic conditions in the MBSF_2019 chronic condition segment file was extracted, including variables:

1. acute myocardial infarction July-December 2019 (AMI)
2. chronic kidney disease (CHRONIC_KIDNEY_EVER), COPD (COPD_EVER)
3. congestive heart failure (CHF_EVER)
4. diabetes mellitus (DIABETES_EVER)
5. ischemic heart disease (ISCHEMICHEART_EVER)
6. stroke/transient ischemic attack (STROKE_TIA_EVER)
7. recent breast cancer July-December 2019 (CANCER_BREAST)
8. recent colorectal cancer July-December 2019 (CANCER_COLORECTAL)
9. recent prostate cancer July-December 2019 (CANCER_PROSTATE)
10. recent lung cancer July-December 2019 (CANCER_LUNG)
11. recent endometrial cancer July-December 2019 (CANCER_ENDOMETRIAL)
12. anemia July-December 2019 (ANEMIA)
13. asthma (ASTHMA_EVER)
14. hypertension (HYPERT_EVER)

End Stage Renal Disease (ESRD) was identified by finding a "Y" code in the ESRD_INDICATOR field of the MBSF_2020 file.

Humetrix SaaS platform identification of additional chronic conditions: The following chronic condition variables were identified by analyzing ICD10 codes in these CMS claim files: (Beneficiary Part A

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

institutional Inpatient and Outpatient claims, Part B Carrier claims, SNF and Hospice claims) from October 1, 2019 through September 30, 2020:

1. leukemia (includes acute and chronic myeloid and lymphocytic leukemias as well as less common leukemias),
2. chronic lung disease (includes pulmonary fibrosis, interstitial lung disease, pneumoconioses, pulmonary sarcoidosis, pulmonary hypertension),
3. chronic liver disease (includes alcoholic cirrhosis, primary biliary cirrhosis, chronic viral hepatitis due to hepatitis B and C, alcoholic fatty liver, primary sclerosing cholangitis, Wilson's disease)
4. HIV/AIDS
5. organ transplant (includes following transplants: lung, bone, heart, liver, pancreas, intestine, kidney, bone marrow),
6. morbid obesity (BMI over 40),
7. obesity (BMI 30 – 40).

Medication variables PDE and Claims Data Files were analyzed by the Humetrix SaaS system to derive the factors listed below. Note coding of non-chemotherapy drugs as "True" signifies at least one applicable prescription fill was identified in 2020. These variables were all identified by mapping NDC drug product codes to RxNorm ingredient codes for these pharmaceutical classes of drugs:

1. chemotherapy: a value of "True" signifies that a beneficiary at any time in 2020 either had an ICD10 code for chemotherapy in part A Institutional or Part B Carrier claims, had a CPT-4 code indicating administration of parenteral chemotherapy in a Part B claim, or had a pharmacy (PDE) claim with an NDC code which mapped to an RxNorm ingredient code for an active pharmaceutical ingredient belonging to multiple classes of chemotherapeutic agents.
2. anticoagulant drugs
3. antiplatelet drugs
4. inhaled Beta-2 agonists
5. inhaled corticosteroid drugs
6. opioid drugs
7. histamine type-2 receptor blockers
8. angiotensin converting enzyme inhibitors (ACE inhibitors)
9. angiotensin II receptor blockers
10. non-steroidal anti-inflammatory (NSAID) drugs
11. immunosuppressive drugs of diverse pharmaceutical classes
12. Azithromycin and Chloroquine drugs (includes both Chloroquine and Hydroxychloroquine). A value of "True" was assigned if a sufficient quantity of these drugs was filled to extend up to, or 10 days beyond, the Covid-19 diagnosis date.

Immunization Variables: Influenza and pneumococcal vaccinations were identified using their respective CPT-4 codes found in Part B claims from October 1, 2019 until September 30, 2020. Shingrix® herpes zoster vaccinations were identified by NDC codes found in Part D (PDE) claims from January 1, 2020 through September 30, 2020.

COVID-19 outcome variables:

1. Covid-19 hospitalizations were identified either by Part B Carrier claims with place of service code and CPT codes indicating inpatient care with a date of service no more than 14 days after

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

or 10 days before the Covid-19 diagnosis date, or by finding Part A Inpatient claims where the data of admission no more than 14 days after or 10 days before the Covid-19 diagnosis date.

2. Covid-19 cases managed by outpatient care only were defined as cases who were not hospitalized in the 30 days after the Covid-19 diagnosis date and who did not die within 60 days of the diagnosis date.
3. Deaths due to Covid-19 were defined as either deaths which occurred during a Covid-19 hospitalization using the inpatient PTNT_DSCHRG_STUS_CD code, or a death which occurred within 60 days of the Covid-19 diagnosis as reported in the DEATH_DT field of the MBSF_2020 file.

Supplemental Results:

Effect size calculation (Table 1): Due to the very large sizes of the compared populations, the p values for the statistical tests presented in Table 1 are almost all very significant. We used Cramér's V to calculate effect sizes values for the comparisons between groups for binary and categorical variables (age, sex, race, income, poverty, dual Medicare-Medicaid dual status, disabled status, prior hospitalization, and clinical variables) all of which had a computed Cramér's V values under 0.17 indicating a small effect size. We used Vargha and Delanay's A (VDA) to calculate effect sizes for comparisons between groups for the quantitative variables (age in years, SVI EPL_CROWD and SVI EPL_MUNIT) and found Vargha and Delanay's A values of 0.42 and 0.66 also indicative of small effect size.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20219816; this version posted November 3, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

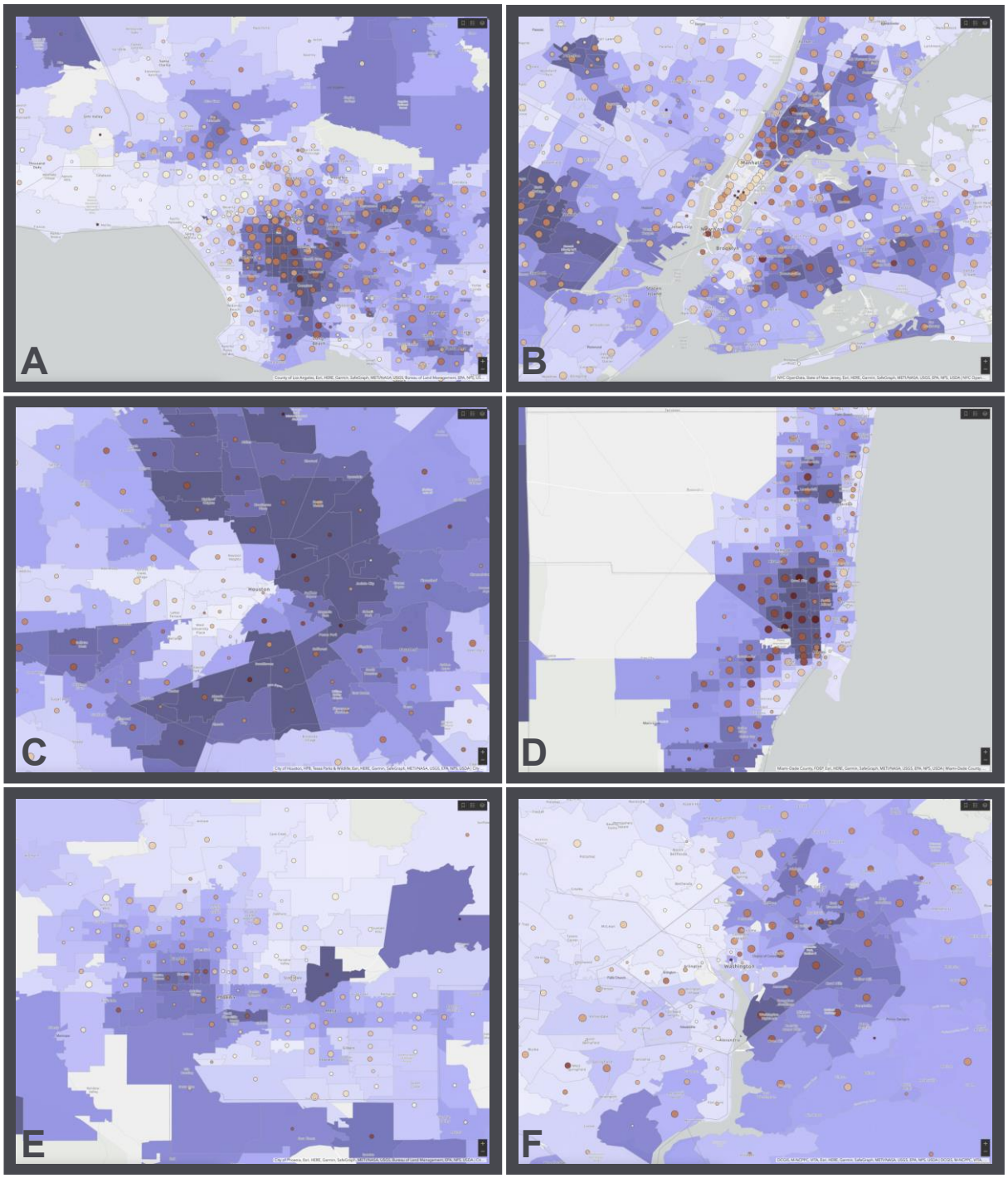


Figure S1: Metropolitan region risk maps for severe Covid-19 (A) Los Angeles County, (B) New York City, (C) Houston, (D) Miami Dade, (E) Phoenix Arizona, (F) Washington DC. In each metropolitan region, zip codes with the higher predicted probabilities of hospitalization with Covid-19 based on the logistic regression hospitalization model are shown in darker shades of lavender. Circles denote observed cumulative hospitalizations due to Covid-19 extracted from claims data with the size of the circles scaled to the number of hospitalizations. The percentage of cases requiring hospitalization is displayed by the color of the circles on a beige to dark orange-red color scale with the darker circles indicating higher zip codes with higher percentages of cases requiring hospitalization for Covid-19.