**01/21/2022**

# ESTIMATING THE EFFECT OF VACCINATION ON THE CASE-FATALITY RATE FOR COVID-19

Glen H. Murata, MD*

Allison E. Murata, MSIM^

Heather M. Campbell, PharmD, PhD^[&]

Jenny T. Mao, MD, FCCP*[#]

^ New Mexico VA Health Care System

^ VA Cooperative Studies Program - Clinical Research Pharmacy Coordinating Center

[&] University of New Mexico College of Pharmacy

[#] University of New Mexico School of Medicine

**ABSTRACT**

Objective: To evaluate the effectiveness of vaccination on the case fatality rate (CFR) for COVID-19 infection. Unlike infection or mortality rates, CFR is not affected by unmeasured patient behaviors or environmental factors that affect the risk of exposure.

Methods: Cases were identified through the COVID Shared Data Resource (CSDR) of the Department of Veterans Affairs. Patients were included in this study if they had baseline data available for risk stratification. The primary outcome was death within 60 days of the first positive nucleic acid amplification test (NAAT). A patient was considered fully vaccinated if they had received one dose of the Johnson & Johnson product or two doses of any other formulation, at least, 14 days prior to the NAAT. Cases diagnosed in July, August, or September of 2021 were considered to have the delta variant. We used novel methods to control for confounders in multiple domains of the electronic medical record, including ICD10 codes, vital signs, baseline laboratory tests and outpatient medications. These procedures included retrieving all entries more than 14 days prior to the NAAT; deriving summary measures of their contribution to the risk of death; and including these measures as covariates in a logistic regression model evaluating vaccine effectiveness. PDeathDx refers to the risk of death based on 153 root ICD10 codes previously shown to be independent predictors of mortality. PDeathLabs refers to the risk based on 49 parameters from 4 vital signs and 7 baseline laboratory tests. AggRiskDx refers to the aggregate effect of 8 drug classes shown to have protective effects. Other predictors in the model included demographic characteristics and comorbidity scores. Logistic regression was used to derive adjusted odds ratios for the vaccination and delta terms. Separate models were developed for early COVID variants and the delta variant. Split sample validation was used to determine if the estimates for vaccine and delta effects were stable across independent patient samples.

Results: On September 30, 2021, there were 339,772 patients in the COVID CSDR who met the criteria for this study. 9.1% had been fully vaccinated, while 21.5% were presumed to have

the delta variant. The median time from vaccination to diagnosis was 154 days. Overall, 18,120 patients (5.33%) died within 60 days of their diagnosis. Multivariate modeling showed that age, gender, race, ethnicity, veteran status, PDeathDx, PDeathLabs, AggRiskRx and 3 of 4 comorbidity measures were independent risk factors for death within 60 days. The adjusted odds ratio for delta infection was 1.87 ± 0.05, which corresponds to a relative risk of 1.78. The adjusted odds ratio for prior vaccination was 0.280 ± 0.011, corresponding to a relative risk of 0.291. Separate models showed that vaccination had even greater benefits for delta infections than for earlier variants. Split sample procedures showed that the estimates for vaccine and delta effects were stable across independent samples.

Conclusions: Estimates of vaccine effectiveness are valid to the extent that they exclude non-vaccine effects and control for confounding. Infection and mortality rates depend upon the risk of exposure which, in turn, depends upon the extent to which the patient adheres to COVID precautions and environmental factors. Moreover, there are hundreds of confounders that may promote higher vaccination rates in those suspected to have poor outcomes if they contract the virus. Our study used CFR and novel procedures to mitigate these problems. Although delta is substantially more lethal than earlier variants, vaccination reduces the risk of death by over 70%. Moreover, the benefit of such was observed at a median of 5 months after vaccination. Our study using CFR confirms that vaccination is an effective means of preventing COVID death and suggests that CFR would better identify changes in virulence of new variants.

## INTRODUCTION

Recent studies have shown an alarming decrease in the effectiveness of COVID vaccines over time (1-3). The metrics for effectiveness included infection and mortality rates (4, 5). It is imperative that we understand the mechanisms by which vaccines begin to fail. Such efforts can be facilitated by a robust framework for classifying vaccine effects. The most straightforward method is to use a probabilistic approach. Patients who succumb to a contagious disease must first be exposed, then develop an infection as a result of the exposure, and then die as a consequence of the infection. The risk of death in a population observed for a given time is thus the joint probability of these 3 events or P(exposure, infection, death). From the multiplicative rule of probability theory:

$$P(\text{exposure, disease, death}) = P(\text{exposure}) * P(\text{infection} \mid \text{exposure}) * P(\text{death} \mid \text{exposure, disease})$$

Likewise, the risk of infection is the joint probability of the first 2:

$$P(\text{exposure, disease}) = P(\text{exposure}) * P(\text{infection} \mid \text{exposure})$$

It is important to separate these metrics into their underlying risks because the latter represent separate targets for interventions. For example, COVID precautions focus on P(exposure), while anti-virals target P(death | exposure, disease). Vaccination has favorable effects on P(infection | exposure) and P(death | exposure, disease) by promoting an immune response. It will hopefully decrease P(exposure) when herd immunity is achieved. Thus, vaccine effectiveness might vary depending upon which risk is targeted. Moreover, changes in one risk may be offset by changes in another - leading to erroneous conclusions about vaccine effectiveness. For example, beneficial effects of the vaccine on P(infection | exposure) may be diminished by abandoning COVID precautions which increases P(exposure).

One problem with certain models for vaccine effectiveness is that they do not account for the risk of exposure. Doing so requires adjustments for patient behaviors and community level effects. The former include adherence to COVID precautions such as masking, social distancing, handwashing, avoidance of large crowds, testing of contacts, and working from

home. The latter include the prevalence of the virus, its infectivity, the extent to which the community embraces COVID precautions, and government mandates. As a result, increases in infection and mortality rates may be related to diminished vaccine effectiveness, changes in exposures, or both. In this study, we focused on the third risk or P(death | exposure, disease). This term is analogous to the case fatality rate (CFR). We have chosen this outcome to assess vaccine effectiveness because, unlike infection and mortality rates, it is not affected by unmeasured patient behaviors and environmental factors.

Measuring effectiveness in observational studies requires a robust approach because treatment is not randomly allocated across a population. Confounding is introduced when the treatment (vaccination) affects the outcome, the condition for which treatment is indicated (a pre-existing condition) affects the outcome, and there is an association between the severity of the condition and the likelihood of treatment. The latter can be positively or negatively correlated, in which case treatment effects are under- or overestimated, respectively. In most cases, the benefits of treatment are offset by their preferential use in patients with a poorer prognosis. As such, the benefit can only be revealed when the bias is removed by multivariate analysis which separates the independent effects of treatment and associated comorbidities. The biggest challenge is that there are hundreds of conditions that may serve as confounders. Co-morbidity scores may not be suitable for this purpose because they do not represent all of the conditions that may be high risk. Thus, a patient with an included condition of moderate risk may be given a higher score than another with a rare condition that is fatal. Critical findings on vital signs and laboratory tests may also serve as confounders. The most robust solution is to do a systematic survey of all high-risk conditions from several domains in the medical record and adjust the effect of vaccination by some aggregate measure of their effect. We have developed such procedures for individual ICD10 codes, vital signs, commonly used laboratory tests, and outpatient medications. As a result, our estimates of vaccine effect on the case fatality rate may be one of the least biased approaches of those reported to date.

## METHODS

Cases were identified through VA's COVID Shared Data Resource (CSDR). Membership in this registry requires at least one positive nucleic acid amplification test. Subjects were included in this study if their index infections occurred before October 2021. Delta variants of the virus were considered the infecting agent for those presenting in July, August, or September of 2021. Although sporadic cases of delta were reported in late May, delta was the predominant variant over the time we selected. The primary outcome was death within 60 days of the diagnosis. The outcome was retrieved from the CSDR, which assigns a 1 to those who died and 0 otherwise. The cohort was followed through November 2021 so that each subject reached a definitive endpoint.

VA maintains two databases containing information on COVID vaccination. CSDR has a robust and highly vetted registry of patients who have been vaccinated within and outside of the agency. The Immunization domain of the Corporate Data Warehouse (CDW) contains similar information but is less structured and contains duplicates. The CDW data were scrubbed and re-organized to match the CSDR format. Cases identified in CDW, but not in CSDR, were added to the latter to create a pooled vaccine registry. Patients were considered vaccinated if they had received 1 dose of the Johnson & Johnson product or 2 doses of any other formulation at least 14 days prior to the diagnosis of COVID.

PDeathDx refers to the predicted probability of death based upon 153 ICD10 root diagnoses (Murata GH, doi pending). Pre-existing conditions were identified by reviewing all diagnoses entered into the electronic medical record during outpatient visits, as updates to the patient problem list, or at the time of hospital discharge. "Pre-existing" refers to entries made up to 14 days prior to the COVID diagnosis. ICD9 codes were converted to ICD10 using a crosswalk provided by the Centers for Medicare/ Medicaid Services. A "root diagnosis" was defined as all characters preceding the decimal point for ICD10 codes or the ICD9 equivalent. Each patient was deemed to have (or not have) each root condition. A proprietary computer

program was used to identify all patients with a given condition who died or survived, as well as all patients without the condition who died or survived. The software used these cell frequencies to derive the relative risk (RR) of death associated with the condition along with the confidence interval (CI). CIs were adjusted for multiple comparisons by the Bonferroni method. A root diagnosis was considered to have a significant effect on the outcome if the lower limit for the CI was ≥ 1.5 or the upper limit for the CI was ≤ 0.80. The procedure was thus used to identify conditions that were either high-risk or protective. Stepwise logistic regression identified those diagnoses that were independent predictors of death. The model was then used to generate a predicted probability of death (PDeathDx) for each subject.

PDeathLabs refers to the predicted probability of death based upon 49 parameters derived from complete value sets for 4 vital signs (systolic blood pressure, diastolic blood pressure, O2 saturation, and body mass index) and 7 routine laboratory tests (estimated glomerular filtration rate, ALT, hematocrit, serum albumin, low-density lipoprotein cholesterol, high-density lipoprotein cholesterol, and hemoglobin A1c) (Murata GH, doi pending). Entries for these 11 clinical measurements were retrieved if their recorded dates were ≥ 14 days prior to the diagnosis of COVID. 13 parameters were derived for each type of measurement to reflect criteria used by practitioners to assess metabolic control (total = 13 * 11 = 143). Logistic modeling showed that 49 of these parameters were independently predictive of death (Murata GH, doi pending). The model was used to assign a predicted probability of death (PDeathLabs) based on clinical measurements to each subject.

Current treatment was identified by reviewing all outpatient medications active on the 14$^{th}$ day prior to the COVID diagnosis. A patient was considered on treatment if (s)he still had a supply of medications from their most recent “fill” on the cutoff date. The VA system assigns each formulation to one or more drug classes. A process identical to the one above was used to assign a RR and CI to each of the 343 VA drug classes. AggRiskRx refers to the protective effect of 8 VA drug classes with an upper boundary for CI ≤ 0.80. This definition presumes that

a protective effect goes beyond neutralizing the underlying condition and is therefore likely to be independent of its initial indication. An aggregate effect for all 8 classes was derived by log transforming the relative risk for each and adding the transformed values. This approach assumes that their effect was independent and that the aggregate effect was the product of the individual RR. We did not examine high-risk drugs because the RR of pre-existing conditions reflects the underlying disease as well as the drugs used to treat the condition.

Age at diagnosis, gender, race, ethnicity, veteran status, smoking history and use of supplemental oxygen were retrieved from the CSDR. The CSDR was also interrogated for the 2-year (Charl2Yrs) and lifetime Charlson Comorbidity Index (CharlEver) and the 2-year (Elix2Yrs) and lifetime Elixhauser (ElixEver) scores.

***Statistical methods*** – Univariate analysis was used to compare the attributes of patients who died and survived. Group differences in nominal variables were tested by chi-square analysis. Group differences in continuous variables were examined by the student's t-test or Mann-Whitney U-test.

Main model - Stepwise logistic regression was used to construct a multivariate model for COVID death in the entire sample. The dependent variable was death within 60 days of the diagnosis. The predictor of interest was prior vaccination for COVID-19. Covariates included age, gender, race, ethnicity, veteran status, current smoking, use of supplemental oxygen, PDeathDx, PDeathLabs, AggRiskRx, Charl2Yrs, CharlEver, Elix2Yrs, ElixEver and delta variant virus infection. Variables were entered in a stepwise fashion with a P-to-enter of 0.01 and to remove of 0.05. The model was used to derive an overall predicted probability of death (PDeath) for each patient. The ability of PDeath to discriminate between the two groups was assessed by the area under its receiver operator characteristic (ROC) curve. An adjusted odds ratio (OR) and its 95% CI was derived for the vaccination term. A standard on-line calculator was used to convert the adjusted OR to an equivalent RR. The identical procedure was used to evaluate the delta term.

Early variants versus delta – Separate models were developed for early variants (pre-July) and for delta (July - September) using the methods described above. The objective was to determine if predictors of death had changed significantly and if the effectiveness of prior vaccination differed for the two groups.

Split-sample validation – The purpose of this analysis was to determine if the estimates for vaccine and delta virus effects were stable across independent samples. Recall that our models retrospectively test the effect of vaccination, while controlling for as many confounders as possible. They were not designed as prediction models and not intended for use in future patients. Nevertheless, it is important to show that they perform well in different subsets and over time. A computer-generated random number was assigned to each subject. Patients with values ≤ 0.6 were assigned to a derivation set, while the remainder were assigned to a validation set. The model was re-derived for the derivation set. The predictors of the model were fitted to the validation sample to determine if the predictors remained significant. The derivation model was also used to assign a PDeath to each member of the validation set. The area under its ROC curve was used to assess discriminating power. Patients in the validation set were assigned to deciles of risk based upon PDeath. Chi-square analysis was used to determine if the frequency of observed cases varied across the deciles of risk.

## RESULTS

On September 30, 2021, there were 347,220 COVID patients in VA's COVID Shared Data Resource. 339,772 (or 97.9%) had at least one pre-existing condition and form the basis for this report. The mean age at the time of diagnosis was 58.6 ± 16.7 years; 84.1% were male; 22.9% were members of a racial minority; 9.0% were Hispanic; 95.8% were veterans; 0.7% were on supplemental oxygen; and 11.8% were current smokers. 9.1% had been fully vaccinated at least 14 days prior to the COVID diagnosis. The median interval between vaccination and diagnosis was 154 days (interquartile range 111 to 185). 21.5% acquired their

infections after July 1, 2021 and were presumed to have the delta variant. Overall, 18,120 patients (5.33%) died within 60 days of their diagnosis.

**Table 1** shows the results of univariate analysis comparing non-survivors and survivors. Non-survivors were older and more likely to be male, white, and on supplemental oxygen, but less likely to be Hispanic or current smokers. Vaccinated patients were less likely to die than the unvaccinated (3.95% vs 5.47%, respectively; $P < 0.001$). The case fatality rate was lower for those acquiring delta than earlier variants (4.64% vs 5.52%; $P < 0.001$). This finding persisted even when vaccinated patients were removed from the analysis (5.06% vs 5.55%; $P < 0.001$).

**Figure 1** shows the raw CFR for 8 cohorts defined by the time since completed vaccination in 4-week blocks. This analysis was restricted to 25,818 patients vaccinated prior to infection with the delta variant. Each cohort was comprised of at least 2,500 patients. Cohort 1 had their vaccinations 2-6 weeks prior to diagnosis, while cohort 8 had vaccinations 30-34 weeks before diagnosis. The lowest CFR was 2.19% at 10-14 weeks (cohort 3) but rose to 5.24% for cohort 8 with longest time since their completed vaccinations.

The main multivariate model is presented in **Table 2**. 239,393 patients (70.5%) had complete data sets available for multivariate modeling. 13 variables were identified as statistically significant and independent determinants of death at 60 days. Highlights include, a poorer prognosis for the elderly, males, and Hispanics, while being White was protective. PDeathDx, PDeathLabs, AggRiskRx, and 3 of 4 comorbidity measures were all significant predictors of death. The adjusted odds ratio for delta infections were 1.87 ± 0.05, which corresponds to a relative risk of 1.78. The adjusted odds ratio for prior vaccination was 0.280 ± 0.011 and a relative risk of 0.291. This observation suggests that the delta variant is substantially more lethal than earlier variants – an effect that is largely be offset by prior vaccination.

**Tables 3** and **4** show the multivariate models for early COVID variants and delta, respectively. Of 11 variables identified as predictors before July 1, 2021, 8 were still significant after the emergence of delta. The adjusted odds ratio for vaccination prior to July 2021 was 0.404 ± 0.033, while the odds ratio thereafter was 0.259 ± 0.012. The relative risk of death from delta associated with vaccination was 0.269 corresponding to an effectiveness of 73.1%. This observation suggests that prior vaccination was more effective in reducing the case fatality rate for delta than earlier variants. However, only 4,649 (or 15.1%) of 18,120 breakthrough infections occurred before July 1. The earlier odds ratios were therefore based on a relatively small number of deaths in the vaccinated group.

**Table 5** presents the model re-derived on the derivation set and then applied to the validation set (**Table 6**). The objective was to determine if the estimates of the vaccine and delta effects were stable across independent samples. 13 variables were identified as significant predictors of death in the derivation set, of which 12 remaining significant in the validation set with the last cofactor (Native American ancestry) showed a trend towards significance. Although not designed as a prediction model, the ROC area under the curve in the validation set was 0.826, demonstrating that the model was able to differentiate between dying and surviving patients in a different set of data.

The observed versus predicted probability of death in the validation set over deciles of predicted risk is shown in **Table 7**. There is a reasonable correlation between expected and observed rates until predicted probabilities > 0.5. At that point, the predicted risk exceeds the observed in the small number so classified. 16,955 validation patients (17.9%) had a predicted probability of death ≥ 0.1 for which the CFR in this group was 20.9%. There were only minor differences in the adjusted odds ratios for prior vaccination (0.284 versus 0.275) and for a delta infection (1.86 versus 1.88) between the two samples.

Eight variables were highly significant determinants of COVID death across all models: age at diagnosis, PDeathDx, PDeathLabs, AggRiskRx, vaccination status, male, Charl2Yrs, and

ElixEver. Because these factors were influential across independent samples of varying sizes, were stable over time, and predicted survival for viral variants, they could represent a core set of patient attributes for future studies of vaccine effectiveness.

## DISCUSSION

In this paper, we stress the importance of a robust system for classifying vaccine effects. The reason is that the usual methods for evaluating effectiveness are composite measures reflecting 3 underlying risks (exposure, infection, death) and may not precisely define the mechanisms by which vaccines have failed. For example, the mortality rate is a function of the probabilities of exposure, developing an infection once exposed, and dying once infected. Improvements in the last 2 outcomes (i.e., infection and death) may be offset by increases in the first (i.e., exposure) as patients abandon COVID precautions. On the other hand, CFR is a function of only the last risk: death. In this study, we determine the extent to which vaccination provides protection against mortality as a metric not affected by masking, social distancing, handwashing, early testing, and quarantines.

Observational studies of vaccine effectiveness are heavily biased. Patients at the highest risk of death are more likely to receive the vaccine for a number of reasons, including personal choice, concern of their physicians, and/or national policies driven by vaccine shortages and stressed delivery systems. This prioritization confounds the relationship between the intervention and outcome because the benefits of vaccination are offset by their preferential use in patients with the poorest prognosis. The accuracy of estimates for vaccine effect depends upon the extent to which this bias is removed. Our study is unique in that we performed a systematic review of major domains in the medical record, identified observations associated with a poor outcome, and used summary measures of the findings to adjust the effects of vaccination. The review included clinical measurements that were not tested as candidate variables in other models but are critical determinants of survival (such as oxygen saturation). This approach is innovative because it represents intensive computer processing of hundreds of

millions of observations to identify hundreds of potential confounders. While not perfect, our estimate of vaccine effect likely represents the most accurate metric reported to date.

We found that the adjusted odds ratio for vaccination was 0.280. This value corresponds to a 71% reduction in the risk of death. This benefit was observed at a median of 5.1 months after vaccination. Importantly, substantial benefits of vaccination were observed before and after the emergence of delta, although the former was significantly less. A validation study showed that the components of the original model were still significant in an independent sample, that it retained its ability to differentiate between those who died or survived, and that the estimate of vaccine effect remained stable. CFR declined with time to a nadir at 10-14 weeks after vaccination and then rose thereafter. Since CFR is not affected by patient behaviors or environmental factors, this pattern is consistent with the acquisition, and subsequent loss, of a physiological factor that promoted recovery from an established infection. The timing observed for the data also suggests that a booster shot at 6 months may miss a significant proportion of patients who might benefit. Our study differs from that of Cohn et al (1) who studied vaccine effectiveness over time. Their measures of effectiveness included infection and mortality rates instead of case fatality rates as in this study. Nevertheless, data presented here confirm the vital importance of vaccination and support the use of a booster shot to mitigate risk for those affected by the delta variant.

Prior vaccination has been included in other prediction models for COVID mortality. For example, Hippisley-Cox et al (6) published a multivariate model for COVID death in a large, vaccinated cohort in England. The investigators found that the adjusted hazard ratio for full versus partial vaccination was 0.17 (0.13 to 0.22), suggesting a major effect. Differences in vaccine effectiveness between our study and theirs may be explained by the populations studied. To be eligible for military service (and thus inclusion in the present study), individuals cannot have a debilitating congenital abnormality. It is also possible that Veterans may have diminished immune responses to vaccination that may not be fully explained by covariates,

even in a highly specified model. The effect of vaccination on the case fatality rate may therefore depend upon the population studied.

Our definition of a delta infection relied on information provided by the Centers for Disease Control (CDC). CDC identifies and tracks COVID variants through genomic surveillance. It collects specimens for sequencing through the National SARS-CoV-2 Strain Surveillance program as well as data from commercial and academic laboratories under contract to state and local public health agencies. Nowcast is their model for reporting proportions of the variants by HHS region and week. CDC announced that delta became the dominant variant on July 7 and that the variant accounted for 83% of isolates by July 20. These findings corresponded to an alarming increase in the 7-day moving average for reported cases from 12,000 in late June to over 60,000 by July 27, 2021. This increase prompted CDC to issue new guidance on vaccinations and personal precautions for delta on that date. For the week of October 3, 2021, delta accounted for 99.2% of reported cases – a date which corresponds to the end of our sampling period. It was not overtaken by omicron as the dominant variant until the week of December 18, 2021. These data suggest that the great majority of cases presenting between July - September 2021 had delta. Our study showed that these patients had a worse prognosis compared to preceding cases and that the difference could not be explained by patient attributes or vaccination. A different approach was used by groups studying delta in England and Canada (10, 11). These groups studied patients with variants confirmed by viral genome sequencing and likewise found a poor prognosis for delta. Another difference between our study and theirs relates to the control of confounders. Poor outcomes can be attributed to the virulence of the variant, the degree to which the host response is compromised by pre-existing conditions, or both. Measures of virulence are valid to the extent that the latter is controlled. Although outcomes in their models were adjusted for demographic characteristics, one model (10) did not include comorbidities while the other included the presence of “any documented major comorbidity” (11). Our novel approach to pre-existing conditions, vital signs, laboratory

tests, and outpatient medications enable more rigorous control of cofounders than previously reported methods. In fact, we could not demonstrate that delta had a poorer outcome until these factors and vaccination status were controlled. This issue becomes more and more important as each surge culls the most vulnerable patients in a population – leading to a gradual change of patient characteristics in the susceptible population. Although they differ in design, all 3 studies confirm that the outcomes of delta infection are poorer and that the variant is more virulent than early variants.

Our programming has all the elements of a surveillance system for COVID outcomes. For example, surveys can be done over a moving time window in which the most recent quarter is compared to the prior 2 or 3. Results can be reported on a national level, by region, or for each facility. The system can track the outcomes of a dominant variant, monitor the transition from one variant to another, or generate aggregate results when multiple variants become established in the community. Similarly, newer anti-virals can be included in the model so that their effects on CFR can be adjusted for multiple patient characteristics, vaccination status, and the dominant variant. Because each patient has a predicted probability of death, an expected CFR can be derived for each facility. Differences between the observed and expected number of deaths should prompt an investigation into the cause. This investigation may be extended to include viral genome sequencing on stored samples. This process would add targeted testing on a national level to random sampling to identify variants of concern. The approach is practical and uses data already collected through electronic medical records. Finally, we have already run the routines repeatedly to resolve problems and identify the resources required to establish the workflow and framework.

Our validation studies show that our models can be used to predict COVID deaths even though not originally generated for that purpose. Given adequate resources, such an assessment could be performed for every veteran before (s)he becomes ill. The risks can then be used to allocate scarce resources such as ICU beds, monoclonal antibody infusions, and/or

anti-virals at the time of presentation. Prescreening is particularly important for Paxlovid® because its use is indicated for patients with mild to moderate symptoms but who are at risk for severe complications. The same information can be also be used as a data-driven approach for counselling patients who are unvaccinated. Based upon our validation study, about 70% of veterans will have enough data to derive a predicted probability of death if they become infected. Of these patients, 1 of every 6 will belong to a group with a CFR of > 20%. We propose that members of this group be targeted for counseling if they are not vaccinated. Counseling would include a personalized assessment of risk, the patient's findings that contribute to that risk, and the degree to which the risk is mitigated by vaccination. This individualized approach is more likely to overcome vaccine hesitancy than a general discussion of risks and benefits which patients may not find relevant. The proposed strategies represent a personalized approach to COVID care.

Our models do not include the use of anti-virals, dexamethasone, anticoagulation, or monoclonal antibodies. The reason for such is that the most commonly used anti-viral agent – remdesivir – has little effect on patient mortality (7, 8). In addition, the effectiveness of dexamethasone treatment is not definitive, guidelines for anticoagulation therapies have not been fully formulated, and the use of monoclonal antibodies remains typically restricted. However, as effective anti-virals and other therapies are utilized on future cases, our models will be modified to evaluate their effects.

Although our findings show the effectiveness of vaccination, one cannot definitively prove that vaccination improved survival. For example, while we controlled for as many clinical variables as possible, there is no way of measuring all relevant patient attributes. Vaccination could be a marker for many traits that affect recovery from serious illness such as physical fitness, nutrition, medication compliance, preventive care and so forth. This possibility is suggested by the observation that vaccinated patients have a lower risk of death than unvaccinated persons even when COVID deaths are excluded (9).

Our conclusions are limited to patients with characteristics of the veteran population. Because our approach requires an extensive amount of baseline data, the results are also biased towards patients with chronic conditions that require periodic evaluation. Our methods represent a new approach to evaluating the effectiveness of interventions in observational studies. If validated by others for COVID and other diseases, findings presented here represent an alternative and perhaps more robust method for handling confounders.

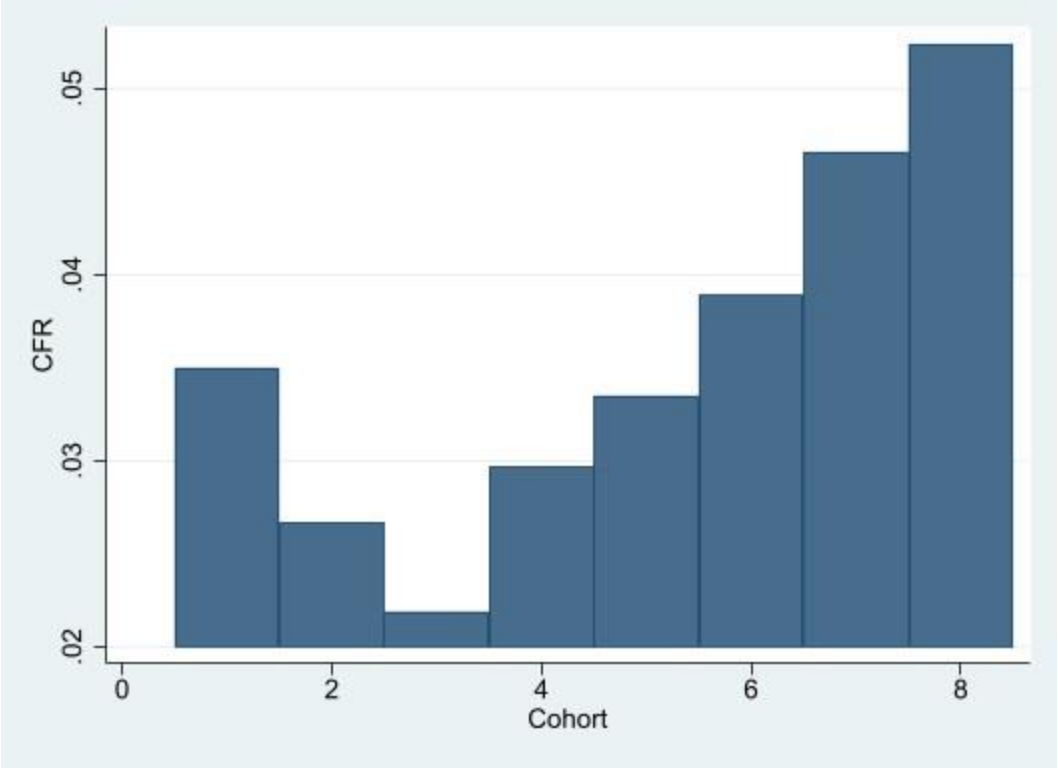


**FIGURE 1:** Case fatality rate by time since vaccination (in 4-week blocks). Cohort 1 was comprised of patients vaccinated 2-6 weeks prior to presentation while cohort 8 was vaccinated 30-34 weeks previously. All cohorts had at least 2,500 patients.

**TABLE 1: CHARACTERISTICS OF NON-SURVIVORS AND SURVIVORS**

| Attribute | Non-Survivors | Survivors | P_Value |
|---|---|---|---|
| Age at diagnosis (years) | 76.1 ± 11.2 | 57.6 ± 16.4 | < 0.001 |
| Male | 97.5% | 83.4% | <0.001 |
| White | 72.5% | 63.1% | <0.001 |
| Hispanic | 7.0% | 9.1% | <0.001 |
| O2 supplementation | 1.8% | 0.7% | <0.001 |
| Current smoker | 8.9% | 12.0% | <0.001 |
| Charl2Yrs | 3.23 ± 2.65 | 1.41 ± 2.00 | <0.001* |
| CharlEver | 4.99 ± 3.31 | 2.28 ± 2.77 | <0.001* |
| Elix2Yrs | 11.87 ± 12.75 | 4.77 ± 8.45 | <0.001* |
| ElixEver | 21.28 ± 16.62 | 9.47 ± 12.76 | <0.001* |
| PDeathDx | 0.139 ± 0.128 | 0.049 ± 0.064 | <0.001* |
| PDeathLabs | 0.136 ± 0.106 | 0.058 ± 0.065 | <0.001* |
| AggRiskRx | -0.0854 ± 0.2720 | -0.2025 ± 0.5424 | <0.001* |

* Mann-Whitney U-test

**TABLE 2: MAIN MULTIVARIATE MODEL**

| Variable | Odds Ratio | Std Error | Z_Score | P_Value | Lower CI | Higher CI |
|---|---|---|---|---|---|---|
| AgeAtDx | 1.0656 | 0.0010 | 65.8100 | 0.0000 | 1.0636 | 1.0676 |
| PDeathDx | 5.6789 | 0.5630 | 17.5200 | 0.0000 | 4.6760 | 6.8970 |
| PDeathLabs | 15.5174 | 1.7617 | 24.1500 | 0.0000 | 12.4217 | 19.3845 |
| VaccFlag | 0.2802 | 0.0110 | -32.5100 | 0.0000 | 0.2595 | 0.3026 |
| Delta | 1.8656 | 0.0494 | 23.5400 | 0.0000 | 1.7712 | 1.9651 |
| CharlEver | 1.0216 | 0.0060 | 3.6200 | 0.0000 | 1.0098 | 1.0334 |
| Male | 1.8282 | 0.1037 | 10.6400 | 0.0000 | 1.6359 | 2.0431 |
| AggRiskRx | 1.2671 | 0.0397 | 7.5700 | 0.0000 | 1.1918 | 1.3473 |
| White | 0.8599 | 0.0176 | -7.3800 | 0.0000 | 0.8261 | 0.8951 |
| ElixEver | 1.0065 | 0.0009 | 6.9700 | 0.0000 | 1.0047 | 1.0083 |
| Hispanic | 1.2367 | 0.0433 | 6.0700 | 0.0000 | 1.1548 | 1.3245 |
| Charl2Yrs | 1.0323 | 0.0061 | 5.3500 | 0.0000 | 1.0203 | 1.0443 |
| Veteran | 2.9615 | 1.5124 | 2.1300 | 0.0340 | 1.0884 | 8.0579 |
| Constant | 0.0001 | 0.0000 | -18.2200 | 0.0000 | 0.0000 | 0.0002 |

**n = 239,393; ROC = 0.824**

**TABLE 3: MULTIVARIATE MODEL FOR EARLY VARIANTS**

| Variable | Odds Ratio | Std Error | Z_Score | P_Value | Lower CI | Higher CI |
|---|---|---|---|---|---|---|
| AgeAtDx | 1.0681 | 0.0012 | 60.0600 | 0.0000 | 1.0658 | 1.0704 |
| PDeathDx | 5.7727 | 0.6256 | 16.1800 | 0.0000 | 4.6680 | 7.1388 |
| PDeathLabs | 14.6853 | 1.8320 | 21.5400 | 0.0000 | 11.5000 | 18.7530 |
| CharlEver | 1.0255 | 0.0067 | 3.8400 | 0.0000 | 1.0124 | 1.0388 |
| VaccFlag | 0.4039 | 0.0332 | -11.0400 | 0.0000 | 0.3438 | 0.4744 |
| Male | 1.8551 | 0.1233 | 9.2900 | 0.0000 | 1.6285 | 2.1133 |
| White | 0.8241 | 0.0187 | -8.5500 | 0.0000 | 0.7883 | 0.8614 |
| AggRiskRx | 1.3161 | 0.0485 | 7.4500 | 0.0000 | 1.2244 | 1.4147 |
| Hispanic | 1.2593 | 0.0488 | 5.9500 | 0.0000 | 1.1672 | 1.3585 |
| ElixEver | 1.0059 | 0.0010 | 5.7100 | 0.0000 | 1.0039 | 1.0080 |
| Charl2Yrs | 1.0250 | 0.0067 | 3.7600 | 0.0000 | 1.0119 | 1.0383 |
| Constant | 0.0002 | 0.0000 | -87.3900 | 0.0000 | 0.0002 | 0.0003 |

**n = 186,505; ROC = 0.826**

**TABLE 4: MULTIVARIATE MODEL FOR DELTA VARIANT**

| Variable | Odds Ratio | Std Error | Z_Score | P_Value | Lower CI | Higher CI |
|---|---|---|---|---|---|---|
| AgeAtDx | 1.0570 | 0.0021 | 27.5400 | 0.0000 | 1.0529 | 1.0612 |
| VaccFlag | 0.2593 | 0.0119 | -29.4700 | 0.0000 | 0.2370 | 0.2836 |
| Charl2Yrs | 1.0704 | 0.0119 | 6.1100 | 0.0000 | 1.0473 | 1.0941 |
| PDeathLabs | 20.0041 | 5.5087 | 10.8800 | 0.0000 | 11.6604 | 34.3181 |
| PDeathDx | 4.5774 | 1.1316 | 6.1500 | 0.0000 | 2.8196 | 7.4311 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Male | 1.9115 | 0.2051 | 6.0400 | 0.0000 | 1.5490 | 2.3589 |
| ElixEver | 1.0094 | 0.0018 | 5.2800 | 0.0000 | 1.0059 | 1.0129 |
| O2Supp | 1.8408 | 0.2729 | 4.1200 | 0.0000 | 1.3766 | 2.4616 |
| AggRiskRx | 1.1373 | 0.0664 | 2.2000 | 0.0270 | 1.0144 | 1.2751 |
| Constant | 0.0007 | 0.0001 | -46.2300 | 0.0000 | 0.0005 | 0.0010 |

**n = 52,888; ROC = 0.818**

**TABLE 5: MULTIVARIATE MODEL FOR DERIVATION SET**

| Variable | Odds Ratio | Std Error | Z_Score | P_Value | Lower CI | Higher CI |
|---|---|---|---|---|---|---|
| AgeAtDx | 1.0642 | 0.0013 | 50.1500 | 0.0000 | 1.0616 | 1.0668 |
| PDeathLabs | 15.9648 | 2.3263 | 19.0100 | 0.0000 | 11.9986 | 21.2421 |
| PDeathDx | 5.8042 | 0.7409 | 13.7800 | 0.0000 | 4.5195 | 7.4542 |
| VaccFlag | 0.2836 | 0.0143 | -25.0400 | 0.0000 | 0.2570 | 0.3130 |
| Delta | 1.8565 | 0.0632 | 18.1900 | 0.0000 | 1.7368 | 1.9845 |
| CharlEver | 1.0176 | 0.0078 | 2.2800 | 0.0230 | 1.0024 | 1.0329 |
| Male | 1.9171 | 0.1414 | 8.8200 | 0.0000 | 1.6591 | 2.2153 |
| ElixEver | 1.0072 | 0.0012 | 5.9600 | 0.0000 | 1.0048 | 1.0096 |
| AggRiskRx | 1.2760 | 0.0515 | 6.0400 | 0.0000 | 1.1790 | 1.3811 |
| White | 0.8752 | 0.0234 | -4.9900 | 0.0000 | 0.8305 | 0.9222 |
| Hispanic | 1.2690 | 0.0569 | 5.3200 | 0.0000 | 1.1623 | 1.3855 |
| Charl2Yrs | 1.0336 | 0.0079 | 4.3300 | 0.0000 | 1.0183 | 1.0492 |
| NativeAm | 1.5546 | 0.1721 | 3.9900 | 0.0000 | 1.2514 | 1.9312 |
| Constant | 0.0003 | 0.0000 | -76.7700 | 0.0000 | 0.0002 | 0.0003 |

**n = 144,427; ROC = 0.823**

**TABLE 6: MULTIVARIATE MODEL FOR THE VALIDATION SET**

| Variable | Odds Ratio | Std Error | Z_Score | P_Value | Lower CI | Higher CI |
|---|---|---|---|---|---|---|
| AgeAtDx | 1.0678 | 0.0016 | 42.6100 | 0.0000 | 1.0646 | 1.0710 |
| PDeathDx | 5.5345 | 0.8713 | 10.8700 | 0.0000 | 4.0652 | 7.5349 |
| PDeathLabs | 14.4228 | 2.6136 | 14.7300 | 0.0000 | 10.1112 | 20.5730 |
| VaccFlag | 0.2750 | 0.0171 | -20.7300 | 0.0000 | 0.2434 | 0.3107 |
| Delta | 1.8780 | 0.0793 | 14.9200 | 0.0000 | 1.7288 | 2.0401 |
| CharlEver | 1.0276 | 0.0095 | 2.9300 | 0.0030 | 1.0090 | 1.0464 |
| Male | 1.7797 | 0.1566 | 6.5500 | 0.0000 | 1.4978 | 2.1145 |
| ElixEver | 1.0056 | 0.0015 | 3.8000 | 0.0000 | 1.0027 | 1.0085 |
| AggRiskRx | 1.2492 | 0.0618 | 4.4900 | 0.0000 | 1.1337 | 1.3765 |
| White | 0.8618 | 0.0284 | -4.5200 | 0.0000 | 0.8080 | 0.9193 |
| Hispanic | 1.1781 | 0.0660 | 2.9200 | 0.0030 | 1.0555 | 1.3149 |
| Charl2Yrs | 1.0308 | 0.0097 | 3.2200 | 0.0010 | 1.0119 | 1.0501 |
| NativeAm | 1.1253 | 0.1687 | 0.7900 | 0.4310 | 0.8387 | 1.5097 |
| Constant | 0.0002 | 0.0000 | -63.5800 | 0.0000 | 0.0002 | 0.0003 |

**n = 94,966; ROC = 0.826**

**TABLE 7: OBSERVED VERSUS PREDICTED DEATHS IN THE VALIDATION SET**

| Decile | **0** | **.1** | **.2** | **.3** | **.4** | **.5** | **.6** | **.7** | **.8** | **Total** |
|---|---|---|---|---|---|---|---|---|---|---|
| Alive | 75537 | 9152 | 2533 | 992 | 423 | 203 | 82 | 24 | 4 | 88950 |
| % | 96.83 | 83.96 | 74.79 | 68.09 | 63.90 | 60.24 | 57.34 | 42.11 | 36.36 | 93.67 |
| Deceased | 2474 | 1749 | 854 | 465 | 239 | 134 | 61 | 33 | 7 | 6016 |
| % | ***3.17*** | ***16.04*** | ***25.21*** | ***31.91*** | ***36.10*** | ***39.76*** | ***42.66*** | ***57.89*** | ***63.64*** | 6.33 |
| Total | 78011 | 10901 | 3387 | 1457 | 662 | 337 | 143 | 57 | 11 | 94966 |
| % | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 | 100.00 |

## REFERENCES

1. Cohn BA, Cirillo PM, Murphy CC, et al. SARS-CoV-2 vaccine protection and deaths among US veterans during 2021. Science 2021. DOI: 10.1126/science.abm0620.

2. Tartof SY, Slezak JM, Fischer H, et a. Effectiveness of mRNA BNT162b2 COVID-19 vaccine up to 6 months in a large integrated health system in the USA: a retrospective cohort study. Lancet. 2021 Oct 16-22;398(10309);1407-16.

3. Yair Goldberg Y, Mandel M, Bar-On YM, et al. Waning immunity after the BNT162b2 vaccine in Israel. N Engl J Med. 2021 Dec 9; 385:e85.

4. Moghadas SM, Vilches TN, Zhang K, et al. The impact of vaccination on COVID-19 outbreaks in the United States. medRxiv. Preprint. 2020 Nov 30 [revised 2021 Jan 2]. DOI: 10.1101/2020.11.27.20240051.

5. Arbel R, Hammerman A, Sergienko R, et al. BNT162b2 vaccine booster and mortality due to Covid-19. N Engl J Med. 2021 Dec 23;385:2413-2420.

6. Hippisley-Cox J**,** Coupland CAC, Mehta N, et al. Risk prediction of covid-19 related death and hospital admission in adults after covid-19 vaccination: national prospective cohort study. BMJ 2021;374:n2244.

7. Wilt TJ, Kaka AS, MacDonald R, et al. COVID-19: Remdesivir for Adults – A Living Review. Updated August 2021. Washington, DC: Evidence Synthesis Program, Health Services Research and Development Service, Office of Research and Development, Department of Veterans Affairs. VA ESP Project #09-009; Aug 2021.

8. Ohl M, Miller D, Lund B, et al. Association of remdesivir treatment with survival and length of hospital stay among US veterans hospitalized with COVID-19. *JAMA Network Open*. July 15, 2021;4(7):e2114741.

9. Xu S, Huang R, Sy LS, et al. COVID-19 vaccination and non-COVID-19 mortality risk – seven integrated health care organizations, United States, December 14, 2020 – July 31, 2021. MMWR Morb Mortal Wkly Rep. 2021;70:1520-24.

10. Twohig KA, Nyberg T, Zaidi A, et al. Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. Lancet Infect Dis. 2022 Jan;22(1):35-42. doi: 10.1016/S1473-3099(21)00475-8. Epub 2021 Aug 27.

11. Fisman DN, Tuite AR. Evaluation of the relative virulence of novel SARS-CoV-2 variants: a retrospective cohort study in Ontario, Canada. CMAJ 2021 October 25;193:E1619-25. doi: 10.1503/cmaj.211248.