STATISTICAL ANALYSIS PLAN

TITLE: The impact of vaccination on severity of illness in COVID-19: A multicenter cohort study

NCT04912700

Approval date: 05/13/2021

Statistical Analysis Plan:
This is a low-risk retrospective cohort study via the secondary use of electronic medical records. There is no formal determination of sample size. This study will include all patients meeting the inclusion/exclusion criteria between 12/15/20 to 4/30/21 (see Appendix A). For the primary outcome of *Aim 1*, there are no previous studies with data for designing power analysis. We assume that 15% of unvaccinated patients will experience severe COVID-19 infection (composite ICU admission, mechanical ventilation, or in-hospital mortality). A two-sided test on comparisons of proportions for partially and fully vaccinated versus unvaccinated patients, respectively, with a type I error rate of 0.05 for each comparison, was used. A representing cohort of greater than 5,500 patients (assuming ~ 92% unvaccinated, 7% partially vaccinated, and 1% fully vaccinated patients) achieves a power of at least 80% to detect the reduction of 5% and 10% on rate of severe COVID-19 infection for partially and fully vaccinated patients, respectively, compared to unvaccinated patients.

Patient characteristics and outcomes of interest stratified by defined vaccinated groups (unvaccinated, partially vaccinated, fully vaccinated) will be summarized using appropriate descriptive statistics and testing procedures on the comparisons for continuous and categorical variables. Tests of bivariate analyses will be performed using analysis of variance (or equivalent nonparametric test) and Chi-squared test (or equivalent exact test) depending on continuous and categorical variables, respectively.

For *Aim 1*, we will quantify the crude rate of severe COVID-19 infection (composite ICU admission, mechanical ventilation, or in-hospital mortality) among defined vaccinated groups. In addition to assessment of the difference of rates of severe infection in univariate analysis, multivariable logistic regression or Cox regression controlling for measured patient characteristics, will be used to assess the association between vaccinated status and occurrence of severe COVID-19 infection. Propensity score approaches on matching and/or weighting will be also employed to evaluate the effect of vaccinated status in regression analyses. With regards to matching procedure, three-way matching, pairwise approach, or common-referent (unvaccinated) group approach will be considered. For *Aim 2*, we will quantify the crude rate of each additional clinical outcomes among defined vaccinated groups, respectively. To account for the influence of selection bias in study cohort, if applicable, the effect of vaccinated status on each of additional outcomes will be evaluated using propensity score weighting.
For *Aim 3*, the crude incidence of ED presentations among different vaccinated groups will be reported using frequencies or rates, accounting for population size of the local community in each of vaccinated groups. Poisson regression will be used to assess the effect of vaccinated groups on ED presentations.
For *Aim 4*, for patients presenting to ED, the crude rate of hospitalization among different vaccinated groups will be reported. Generalized linear model will be applied to compare the effect of vaccinated groups on hospitalization.
For *Exploratory of Aim 1*, for fully vaccinated group (patients), descriptive statistics of clinical outcomes of interest stratified by vaccination types will be summarized; any difference will be tested via analysis of variance (or equivalent nonparametric test) and Chi-squared test (or equivalent exact test) depending on continuous and categorical variables, respectively. For *Exploratory of Aim 2,* for fully vaccinated group (patients), logistic regression model will be used to identify predictors, including demographic, clinical and lab measures, for hospital admission. For missing data, the multiple imputation or missing treated as a category will be considered depending on appropriateness. Validation of the built model based on the identified variables will be examined.