Supplementary Material

# Statistical Analysis Plan for Evaluating Baseline Modifiers of SARS-CoV-2 Vaccine Efficacy

Alex Luedtke, Lars van der Laan, and Sijia Li

## 1 Objectives

Using data from 4 Covid-19 vaccine efficacy trials testing the efficacy of 4 different vaccine regimens [1, 2, 3, 4], this study will evaluate whether there are any baseline covariates that are predictive of vaccine efficacy. The specific objectives are as follows:

1. For each vaccine, evaluate the extent to which each individual measured baseline covariate modifies the efficacy of the vaccine regimen under investigation during the blinded follow-up period.
2. In the setting of Objective 1, evaluate the extent to which the measured baseline covariates jointly modify vaccine efficacy.
3. If pursuing Objective 2 reveals evidence of heterogeneity of vaccine efficacy, then evaluate the importance of each individual baseline covariate in modifying vaccine efficacy when accounting for other measured covariates.
4. For each vaccine, evaluate whether incorporating information about how baseline covariates modify the efficacy of the other three vaccines under consideration yields an improved prediction of its efficacy.

For each objective and vaccine, efficacy will be evaluated in the same cohort and versus the same primary endpoint as were used in the primary study publication [1, 2, 3, 4]. The baseline covariates considered can be found in Table 1 in the appendix.

## 2 General Approach

### 2.1 Overview

Throughout this analysis plan, the marginal vaccine vaccine efficacy denotes the proportional reduction in the probability that a participant will experience the Covid-19 endpoint before a specified time $t_0$ on the vaccine arm versus on the placebo arm. The conditional vaccine efficacy is defined in the same way as its marginal counterpart but within strata defined by baseline covariates. For each

trial, the time point $t_0$ is selected to ensure stable estimation; in particular, it is selected to be the latest time point at which the risk set consists of at least 10% of the participants in both arms. All participants are right censored at time $t_0$ regardless of whether they were observed to experience event after $t_0$.

To provide an interpretable and unbiased analysis of how vaccine efficacy varies within covariate subgroups, the data are reweighted so that, within each stratum of baseline covariates and country, participants had an approximately equal probability of enrolling at any calendar time during the enrollment period (see Section 3 for details). These weights help improve the generalizability of our findings by removing differences in the times at which participants belonging to different covariate subgroups and country tended to enroll.

The analyses defined for each objective will be repeated a total of four times, once for each of the four vaccines under consideration. In the summary of these analyses that follows, we consider one such iteration of these analyses, that is, the evaluation of the stated objectives for one particular vaccine (e.g., Moderna).

Objective 1 will be repeated once for each baseline covariate under consideration. For each baseline covariate, the conditional vaccine efficacy will be estimated within each level of the baseline covariate via a targeted minimum loss-based estimator [5]. These estimates will be used to assess how the relative risk of a Covid-19 endpoint between vaccine and placebo recipients varies across levels of the baseline covariate. In Objective 1, the strength of univariate vaccine efficacy modification will be evaluated by comparing confidence intervals for the vaccine efficacy within each level of the covariates (all of which are categorical). In Objective 2, a machine learning method known as superlearning [6] will be used to more flexibly estimate the multivariate conditional vaccine efficacy functions. In this objective, the strength of vaccine efficacy modification will be evaluated by dividing participants into tertiles that define their vaccine efficacy conditional on multiple baseline covariates. The efficacy of the vaccine will then be estimated within each of these tertiles. If the vaccine efficacy across these tertiles varies, then there is evidence of vaccine efficacy modification. If there is evidence of modification, then variable importance will be evaluated in Objective 3. Variable importance will be assessed via the Shapley Population Variable Importance Measure (SPVIM) introduced in [7], which quantifies the marginal importance of each feature to the overall vaccine efficacy prediction. A rank-ordered list of covariates will be reported, where the list will be arranged in terms of the importance of each covariate for predicting the efficacy of the vaccine. Accompanying confidence intervals and hypothesis tests will also be reported. If there is no evidence of vaccine efficacy modification from Objective 2, then variable importance will not be assessed.

For Objective 4, the list of baseline covariates included will be restricted to contain only those that were measured consistently across the four trials. The text that follows focuses on the evaluation of this objective for the Moderna vaccine — repeating the objective for the other three vaccines is straightforward. As a first step towards completing this objective, the data from the efficacy trials for the three vaccines that are not under consideration (AstraZeneca, Johnson & Johnson, and Novavax) will be pooled together to obtain an estimate of the

conditional efficacy of an intervention that randomly assigns one of these three vaccines. For each participant in the COVE trial, the value of this pooled conditional vaccine efficacy estimate will then be used to define a new baseline covariate. The methods of Objectives 1 and 2 will then be repeated to determine whether this newly defined baseline covariate improves the estimate of the conditional efficacy of the Moderna vaccine. Objective 3 will also be repeated to assess whether the pooled vaccine efficacy estimate better predicts the efficacy of the vaccine than does any individual covariate.

## 2.2 Accounting for Variants

Because of the time and locations in which Janssen's ENSEMBLE trial was conducted, several variants were observed during the course of the blinded phase of this trial. To simplify interpretation, the analyses will focus on endpoints that were measured to be from the ancestral, alpha, or other variant, where, following [3], the 'other' variant is defined as viruses from the Wuhan ancestral lineage that have mutations other than E484K. Participants whose first Covid-19 event was from variant that was not ancestral, alpha, or other will be right-censored at the time of their event. Also, as the vast majority of endpoints in the South African sites were from the beta variant, all South African sites will be excluded from the analysis set.

## 2.3 Notation and a Useful Identity for the Conditional Vaccine efficacy

To introduce our notation, we begin by considering a simplified case where there is no right-censoring, the calendar time of participant enrollment is independent of their baseline covariates, and the analysis is performed in the intention to treat cohort, so that the vaccine arm indicator is also independent of baseline covariates. Later, we will imitate this idealized scenario by employing inverse probability weighting. For each participant, we observe $(W, A, Y)$, where $W = (W_1, W_2, \ldots, W_p)$ denotes a vector containing all covariates under consideration, $A$ denotes a randomization arm indicator, and $Y$ denotes an indicator of the event of interest (e.g., Covid-19 disease) by a particular time. We use $V$ denote a subset of the covariates in $W$ or the entire vector $W$, depending on the objective. In Objective 1, $V = W_j$ for a covariate $j$ and, in Objective 2, $V = W$.

For a distribution $P$ of $(W, A, Y)$, the conditional vaccine efficacy writes as $\mathrm{VE}_P(v) := 1 - \mathrm{RR}(v)$, where the conditional relative risk $\mathrm{RR}_P(v)$ is defined as

$$\mathrm{RR}_P(v) = \frac{P(Y = 1|A = 1, V = v)}{P(Y = 1|A = 0, V = v)}.$$

Many of our analyses make use of the following identity:

$$\begin{aligned}\mathrm{RR}_P(v) &= \frac{P(Y=1,A=1|V=v)}{P(Y=1,A=0|V=v)}\frac{P(A=0|V=v)}{P(A=1|V=v)} \\ &= \frac{P(A=1|Y=1,V=v)}{P(A=0|Y=1,V=v)}\frac{P(A=0|V=v)}{P(A=1|V=v)} \\ &= \frac{P(A=1|Y=1,V=v)}{1-P(A=1|Y=1,V=v)}\frac{P(A=0|V=v)}{P(A=1|V=v)}\end{aligned}$$

When $A$ is randomized in a 1:1 ratio – so that $P(A \mid W=w)=1/2$ for all $w$ – the above simplifies to

$$\mathrm{RR}_P(v) = \frac{P(A=1|Y=1,V=v)}{1-P(A=1|Y=1,V=v)}. \tag{1}$$

Thus, in the special case where $P(A \mid W=w)=1/2$ for all $w$, the conditional relative risk is simply equal to the odds of $A=1$ within strata of $V$ among cases (i.e., participants with $Y=1$). This ensures that the conditional relative risk can be estimated by simply running a logistic regression of $A$ against $V$ within strata of the covariates $V$ [8, 9], and also that variable importance for the conditional relative risk $\mathrm{RR}(v)$ can be assessed by evaluating variable importance for the conditional probability $P(A=1|Y=1,V=v)$.

Three simplifying assumptions were made in the discussion above. First, the validity of (1) relies on $A$ being randomized in a 1:1 ratio. Second, the entire preceding subsection supposed that the outcome $Y$ is fully observed (not censored). Third, participants did not necessarily enroll at calendar times that were independent of their baseline covariates. As we describe below, none these assumptions are either probably or certainly violated in all four trials. The next section details a unified approach that will be used to avoid needing any of these assumptions.

# 3 Accounting for possible measured confounders and right-censoring via inverse probability weighting

The data are reweighted to imitate a scenario where participants enrolled at calendar times that are independent of their baseline covariates. This is operationalized by fitting a proportional odds model (`MASS` package in R [10]) of enrolling among the first, second, third, and fourth quarters of study participants in each trial against main terms for each of the baseline covariates subsequently considered in the univariate analyses of vaccine efficacy. Data are then reweighted by the inverse of the probability that they would enroll in the calendar-time-quartile that they were observed to have enrolled in.

The simplifying assumption employed in Section 2.3 that $A$ is randomized in a 1:1 ratio is violated, to varying degrees, in all four efficacy trials considered. In

the AstraZeneca and Novavax trials, participants were randomized in a 2:1 ratio, rather than a 1:1 ratio. Hence, within each of these trials' intention to treat (ITT) cohort, an approximately 2:1 ratio of participants randomized to vaccine versus placebo is observed. In contrast, the 1:1 randomization ratio used in the Moderna and Johnson & Johnson trials makes it so that an approximate 1:1 ratio is observed in the ITT cohort in each of these trials. Another reason that this assumption may be violated in our case is that, for all four trials under consideration, the ITT cohort was not used as the primary analysis cohort. For each trial, membership to the primary analysis cohort is determined by post-randomization variables (e.g., status of having a confirmed SARS-CoV-2 infection before a certain time or missing a vaccination), there is no guarantee that a 1:1 randomization ratio will be seen in the primary analysis cohort even for trials where a 1:1 randomization ratio was used at baseline. In an effort to overcome both the fact that a 1:1 randomization ratio was not used in some trials and also that membership to the primary analysis conditions on post-randomization variables, inverse probability of vaccination weighting is used [11, 12]. In particular, observation $i$ in trial $j$ is reweighted an estimate of $1/\pi_j(A^i \mid W^i, Q^i)$, where $\pi_j(a \mid w, q)$ denotes the probability that $A = a$ given that covariates $W = w$ and enrollment quartile $Q = q$ in trial $j$ and $(A^i, W^i, Q^i)$ denotes the (randomization arm, baseline covariates, enrollment window) of observation $i$. For each trial, this estimate is obtained by fitting main terms logistic regression of $A$ against all baseline covariates in $W$ separately within each of the four enrollment windows $q$.

The simplifying assumption that there is no censoring, so that $Y$ is fully observed on all participants, is also violated. Indeed, some participants are right-censored before $Y$ can be assessed. To overcome this, right-censoring will be accounted for using inverse probability of censoring weighting [13]. These weights are defined separately by fitting a Kaplan-Meier estimator for the censoring time distribution within each stratum of participants defined by vaccine/placebo arm status, enrollment window, and age and/or risk strata that were predefined in the datasets. For Moderna these strata are [1] >=18 and <65 Years and Not at Risk [2] >=18 and <65 Years and at Risk [3] >=65 Years, for AstraZeneca they are [1] <65 and [2] >=65, for Janssen they are [1] 18-59 and not at risk, [2] >=60 and not at risk, [3] 18-59 and at risk, [4] >=60 and at risk, and for Novavax they are [1] <65 and [2] >=65.

The weight that each participant will receive in the analyses will correspond to the product of the inverse probability of enrollment weights, inverse probability of vaccination weights, and inverse probability of censoring weights.

# 4 Objective 1

Objective 1 will be evaluated separately for each vaccine and covariate under consideration. Here, we focus on one such vaccine-covariate pair, where we denote this covariate by $V$. In our analysis, all of our covariates are categorical, and so we focus on a categorical $V$ with levels $\{1, 2, \ldots, k\}$ here. A saturated

log-linear model is fitted via the targeted minimum loss-based estimator implemented in the `causalglm` R package [5]. The function is called with superlearner libraries consisting of the following algorithms, used for both estimating the propensity model and the outcome mean model: generalized additive model, random forest, gradient boosting with a maximum depths of 4-6, and highly adaptive lasso with maximum depth of 2, smoothness orders of 1 and number of knots of 20.

# 5 Objective 2

A nonparametric estimate of the relative risk will be reported for the vector of all baseline covariates, denoted by $V$ in this objective. This estimate will be obtained by estimating $E[A = 1 \mid Y = 1, V = v]$ using the `SuperLearner` R package [14]. Superlearner is an ensemble machine learning approach that uses cross-validation to select a weighted combination of predicted outcome scores across a collection of candidate algorithms to yield an optimal combination according to a prespecified criterion that performs at least as well as the best component algorithm. The candidate algorithms in SL can either be parametric, flexible machine learning algorithms, or a combination of both, making SL less prone to model misspecification than traditional parametric approaches. The guarantee that superlearner performs about as well as or better than the best candidate algorithm [15, 6] allows a rich library of parametric and flexible candidate algorithms to be included. Superlearner will be fitted on all participants using the following arguments:

- outcome: $A$
- predictors: $V$
- superlearner library: the library in Table 1.
- family: binomial
- observation weights: the product of the weights estimated in Section 3
- number of cross-validation folds: 5

All other arguments are set to their default values. In this analysis, all learners include, as a first step, excluding all observations for which it is not the case that $Y = 1$. The library that will be used in this analysis is listed in Table 1.

Participants will be divided into tertiles based on their conditional vaccine efficacy. Within each tertile, a stratified vaccine efficacy will be computed that corresponds to the proportional reduction in the Covid-19 endpoint probability among participants in that tertile when everyone receives the vaccine rather than placebo. Taking the first tertile as an example and under the simplifying assumptions employed in Section 2.3, this corresponds to estimating

$$1 - \frac{E_P[Y \mid A = 1, \mathrm{VE}_P(V) \leq \tau_P]}{E_P[Y \mid A = 0, \mathrm{VE}_P(V) \leq \tau_P]} = 1 - \frac{E_P[Y I\{\mathrm{VE}_P(V) \leq \tau_P\} \mid A = 1]}{E_P[Y I\{\mathrm{VE}_P(V) \leq \tau_P\} \mid A = 0]}.$$

Similar arguments to those used in [16] show that the above is a pathwise differentiable parameter of $P$, thereby facilitating consistent and asymptotically normal inference and the construction of Wald-type confidence intervals [17]. As in other parts of this SAP, the lack of validity of the simplifying assumptions from Section 2.3 will be accounted for using the inverse probability weights described in Section 3.

Fix $j \in \{0, 1, 2, 3\}$ and let $\tau_P(j)$ denote the $j/3$ quantile of $\mathrm{VE}_P(V)$ when $V \sim P$. The following algorithm can be used to obtain a point estimate and corresponding 95% confidence interval of the log-relative risk, $j \in \{0, 1, 2\}$,

$$\log \frac{E_P[Y \mid A = 1, \tau_P(j) \leq \mathrm{VE}_P(V) \leq \tau_P(j+1)]}{E_P[Y \mid A = 0, \tau_P(j) \leq \mathrm{VE}_P(V) \leq \tau_P(j+1)]}, \tag{2}$$

which can be mapped into a point estimate and 95% confidence interval for the VE in the subgroup of individuals with $\mathrm{VE}_P(V) \in [\tau_P(j), \tau_P(j+1)]$:

1. For each observation $i$, let $\mathrm{wgt}^i$ denote the inverse probability weight defined in Section 3.

2. Run for 20 random seeds. For each random seed $\ell$, use 5-fold cross-fitting as follows:

   (a) for each of the five random 4/5-1/5 training-validation splits of the data, $k \in \{1, 2, 3, 4, 5\}$:

   - train the model estimating VE on the cases in the training dataset and obtain predictions $\widehat{\mathrm{VE}}_{\ell,k}(V^i)$ for each individual $i$ in the validation dataset.
   - For $j \in \{0, 1, 2, 3\}$, let $\widehat{\tau}_{\ell,k}(j)$ denote the empirical $j/3$ quantile of $\widehat{\mathrm{VE}}_{\ell,k}(V^i)$ across observations $i$ in the validation set.
   - For each arm $a \in \{0, 1\}$ and $j \in \{0, 1, 2\}$, estimate the incidence in the $(j+1)$-th tertile of VE as follows:

   $$\hat{\mu}_{a,\ell,k}(j) := \frac{\sum_{i \in \mathcal{V}_{\ell,k}: A^i = a, \widehat{\tau}_{\ell,k}(j) \leq \widehat{\mathrm{VE}}_{\ell,k}(V^i) \leq \widehat{\tau}_{\ell,k}(j+1)} \mathrm{wgt}^i Y^i}{\sum_{i \in \mathcal{V}_{\ell,k}: A^i = a, \widehat{\tau}_{\ell,k}(j) \leq \widehat{\mathrm{VE}}_{\ell,k}(V^i) \leq \widehat{\tau}_{\ell,k}(j+1)} \mathrm{wgt}^i}.$$

   Above, $\mathcal{V}_{\ell,k}$ is the set of the indices of the observations that belong to the $k$-th validation set for the 5-fold split of the data defined via the $\ell$-th random seed considered.

   (b) To estimate the arm-specific incidence in the subgroup of individuals with $\mathrm{VE}_P(V) \in [\tau_P(j), \tau_P(j+1)]$ for the 5 splits defined by the given random seed $\ell$, average across these splits. In particular, for $a \in \{0, 1\}$ and $j \in \{0, 1, 2\}$, define

   $$\hat{\mu}_{a,\ell}(j) := \frac{1}{5} \sum_{k=1}^{5} \hat{\mu}_{a,\ell,k}(j).$$

3. The final arm-specific incidence estimate in the subgroup of individuals with $\text{VE}_P(V) \in [\tau_P(j), \tau_P(j+1)]$ is given by the average across the 20 choices of $\ell$ considered. In particular:

$$\hat{\mu}_a(j) := \frac{1}{20} \sum_{\ell=1}^{20} \hat{\mu}_{a,\ell}(j).$$

4. The estimate of the log-relative risk in (2) is given by:

$$\log \frac{\hat{\mu}_1(j)}{\hat{\mu}_0(j)}. \tag{3}$$

Standard errors for the log relative risk estimator in (3) are obtained via the nonparametric bootstrap, using 2000 bootstrap replications. Step 2 above is run over 20 random seeds in order to mitigate the dependence of the results on the particular folds selected during 5-fold cross-fitting.

# 6 Objective 3

If there is evidence of heterogeneity of vaccine efficacy in Objective 2, then the importance of individual variables for predicting vaccine efficacy will be assessed. Specifically, variable importance will be assessed for a vaccine if the two-sided 95% confidence interval for the vaccine efficacy in the tertile of individuals with the lowest vaccine efficacy overlaps with the two-sided 95% confidence interval for the highest tertile.

Variable importance for the relative risk functions considered in Objective 2 will be assessed nonparametrically via SPVIM [7]. The `vimp` R package will be used for this purpose. The variable importance analysis will be conducted separately for each trial. Hereafter we consider one such trial.

The function `sp_vim` in the `vimp` package will be used to assess variable importance for $v \mapsto P(A = 1 \mid Y = 1, V = v)$. By (1), this will also provide variable importance for $v \mapsto \text{RR}(v)$. The function `sp_vim` will be evaluated using case data (those with $Y = 1$) from a given trial and the following arguments

- outcome: $A$,
- covariates: $V$,
- number of folds for cross-fitting: 10,
- type of parameter: auc,
- superlearner library: the library in Table 2,
- inverse probability weights: product of the weights estimated in Section 3,
- stratified: True.

All other arguments are set to their default values. Variable importance values will be reported in a forest plot, where each variable importance measure will be accompanied by the 95% confidence interval returned by the `sp_vim` package. Variables will be listed from top to bottom in order of decreasing magnitude of the variable importance measure.

# 7 Objective 4

The text that follows focuses on the evaluation of this objective for the Moderna vaccine — repeating the objective for the other three vaccines is straightforward. As a first step towards completing this objective, the data from the efficacy trials for the three vaccines that are not under consideration (AstraZeneca, Johnson & Johnson, and Novavax) will be pooled together to obtain an estimate of the conditional efficacy of an intervention that randomly assigns one of these three vaccines. This is done by pooling the data from the AstraZeneca, Johnson & Johnson, and Novavax trials and subsequently running the methods of Objective 2 to estimate the relative risk on the pooled vaccine arm versus on the pooled placebo arm, conditional on baseline covariates that were measured across all four trials and for which there were at least 25 events per randomization arm. These covariates are: age, sex, ethnicity, BMI, obesity and living condition. In this analysis and this analysis only, continuous (rather than categorical) age and BMI variables are used. When estimating this pooled relative risk, each participant receives the same inverse probability of vaccination and censoring weight as they received when completing the earlier objectives.

Once this conditional pooled relative risk function has been estimated, its value is evaluated for each participant in the Moderna trial and used to define a new covariate. The methods of Objective 2 are then repeated with this additional covariate included.

# A Tables

# References


[1] Lindsey R Baden, Hana M El Sahly, Brandon Essink, Karen Kotloff, Sharon Frey, Rick Novak, David Diemert, Stephen A Spector, Nadine Rouphael, C Buddy Creech, et al. Efficacy and safety of the mrna-1273 sars-cov-2 vaccine. *New England Journal of Medicine*, 384(5):403–416, 2021.

[2] Ann R Falsey, Magdalena E Sobieszczyk, Ian Hirsch, Stephanie Sproule, Merlin L Robb, Lawrence Corey, Kathleen M Neuzil, William Hahn, Julie Hunt, Mark J Mulligan, et al. Phase 3 safety and efficacy of azd1222 (chadox1 ncov-19) covid-19 vaccine. *New England Journal of Medicine*, 385(25):2348–2360, 2021.

[3] Jerald Sadoff, Glenda Gray, An Vandebosch, Vicky Cárdenas, Georgi Shukarev, Beatriz Grinsztejn, Paul A Goepfert, Carla Truyers, Hein Fennema, Bart Spiessens, et al. Safety and efficacy of single-dose ad26. cov2. s vaccine against covid-19. *New England Journal of Medicine*, 384(23): 2187–2201, 2021.

[4] Lisa M Dunkle, Karen L Kotloff, Cynthia L Gay, Germán Áñez, Jeffrey M Adelglass, Alejandro Q Barrat Hernández, Wayne L Harper, Daniel M Duncanson, Monica A McArthur, Diana F Florescu, et al. Efficacy and safety of nvx-cov2373 in adults in the united states and mexico. *New England Journal of Medicine*, 2021.

[5] causalglm : interpretable and model-robust causal inference for heterogeneous treatment effects. `https://tlverse.org/causalglm/`. Accessed: 2022-03-03.

[6] Mark J Van der Laan, Eric C Polley, and Alan E Hubbard. Super learner. *Statistical applications in genetics and molecular biology*, 6(1), 2007.

[7] Brian Williamson and Jean Feng. Efficient nonparametric statistical inference on population feature importance using shapley values. In *International Conference on Machine Learning*, pages 10282–10291. PMLR, 2020.

[8] Binyan Jiang, Rui Song, Jialiang Li, and Donglin Zeng. Entropy learning for dynamic treatment regimes. *Statistica Sinica*, 29(4):1633, 2019.

[9] Hongxiang Qiu, Alex Luedtke, and Mark J van der Laan. Discussion of "entropy learning for dynamic treatment regimes". *Statistica Sinica*, 29(4): 1666–1677, 2019.

[10] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S.* Springer, New York, fourth edition, 2002. URL `https://www.stats.ox.ac.uk/pub/MASS4/`. ISBN 0-387-95457-0.

Table 1: List of baseline covariates considered, and whether they were available in each trial (denoted by an X).

| | Moderna | AstraZeneca | Janssen | Novavax |
|---|---|---|---|---|
| **Personal Characteristics** | | | | |
| Age | | | | |
| 18-29 yr | X | X | X | X |
| 30-39 yr | X | X | X | X |
| 40-49 yr | X | X | X | X |
| 50-59 yr | X | X | X | X |
| 60-69 yr | X | X | X | X |
| $\geq$ 70 yr | X | X | X | X |
| Ethnicity | | | | |
| Hispanic | X | X | X | X |
| Not Hispanic | X | X | X | X |
| Race | | | | |
| American Indian/Alaska Native | X | X | X | X |
| Black or African American | X | X | X | X |
| Multiple | X | X | X | X |
| Other | X | X | X | X |
| White | X | X | X | X |
| Sex | | | | |
| Female | X | X | X | X |
| Male | X | X | X | X |
| Intersex/Unknown | | | X | |
| **Health Characteristics** | | | | |
| Body mass index (kg/m$^2$) | | | | |
| Healthy weight or Underweight ($< 25$) | X | X | X | X |
| Overweight ($\geq 25$, $< 30$) | X | X | X | X |
| Obese ($\geq 30$) | X | X | X | X |
| Class 3 Obese ($\geq 40$) | X | X | X | X |
| Cardiovascular Disease | X | X | X | X |
| Diabetes | X | X | X | X |
| HIV | X | X | X | X |
| Kidney Disease | X | X | X | X |
| Liver disease | X | X | X | X |
| Chronic Lung Disease | X | X | X | X |
| **Risk Characteristics** | | | | |
| Risk of Exposure | | | | |
| Low | X | X | X | X |
| Medium | X | X | X | X |
| High | X | X | X | X |
| Risk from Living Condition | | | | |
| Low | X | X | X | X |
| Medium | X | X | X | X |
| High | X | X | X | X |
| Very High | X | X | X | X |
| Smoke history | X | X | X | X |
| **Geographic Location** | | | | |
| Argentina | | | X | |
| Brazil | | | X | |
| Chile | | X | X | |
| Colombia | | | X | |
| Mexico | | | X | X |
| Peru | | X | X | |
| USA | X | X | X | X |

Table 2: SuperLearner [14] library used to estimate the vaccine efficacy conditional on multiple baseline covariates. The family is set to `binomial()`.

| Algorithms | Screens[2]/ Tuning Parameters |
|---|---|
| SL.ranger | All |
| SL.glm | All |
| SL.gam | All |
| SL.polymars | All |
| SL.bayesglm | All |
| SL.nnet | All |


[11] Daniel G Horvitz and Donovan J Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

[12] Miguel A Hernán and James M Robins. *Causal Inference: What If*. Chapman & Hall/CRC, Boca Raton, 2020.

[13] James M Robins and Andrea Rotnitzky. Recovery of information and adjustment for dependent censoring using surrogate markers. In *AIDS epidemiology*, pages 297–331. Springer, 1992.

[14] Eric Polley, Erin LeDell, Chris Kennedy, and Mark van der Laan. *SuperLearner: Super Learner Prediction*, 2021. URL `https://CRAN.R-project.org/package=SuperLearner`. R package version 2.0-28.

[15] Mark J van der Laan, Sandrine Dudoit, and Aad W van der Vaart. The cross-validated adaptive epsilon-net estimator. *Statistics & Decisions*, 24 (3):373–395, 2006.

[16] Alexander R Luedtke and Mark J van der Laan. Optimal individualized treatments in resource-limited settings. *The international journal of biostatistics*, 12(1):283–303, 2016.

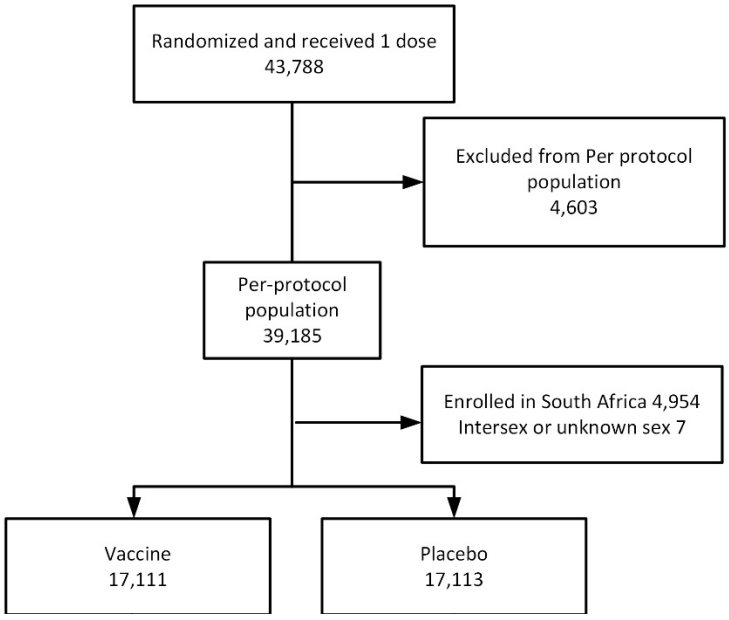
[17] Peter J Bickel, Chris AJ Klaassen, Peter J Bickel, Ya'acov Ritov, J Klaassen, Jon A Wellner, and YA'Acov Ritov. *Efficient and adaptive estimation for semiparametric models*, volume 4. Springer, 1993.

Supplementary Methods

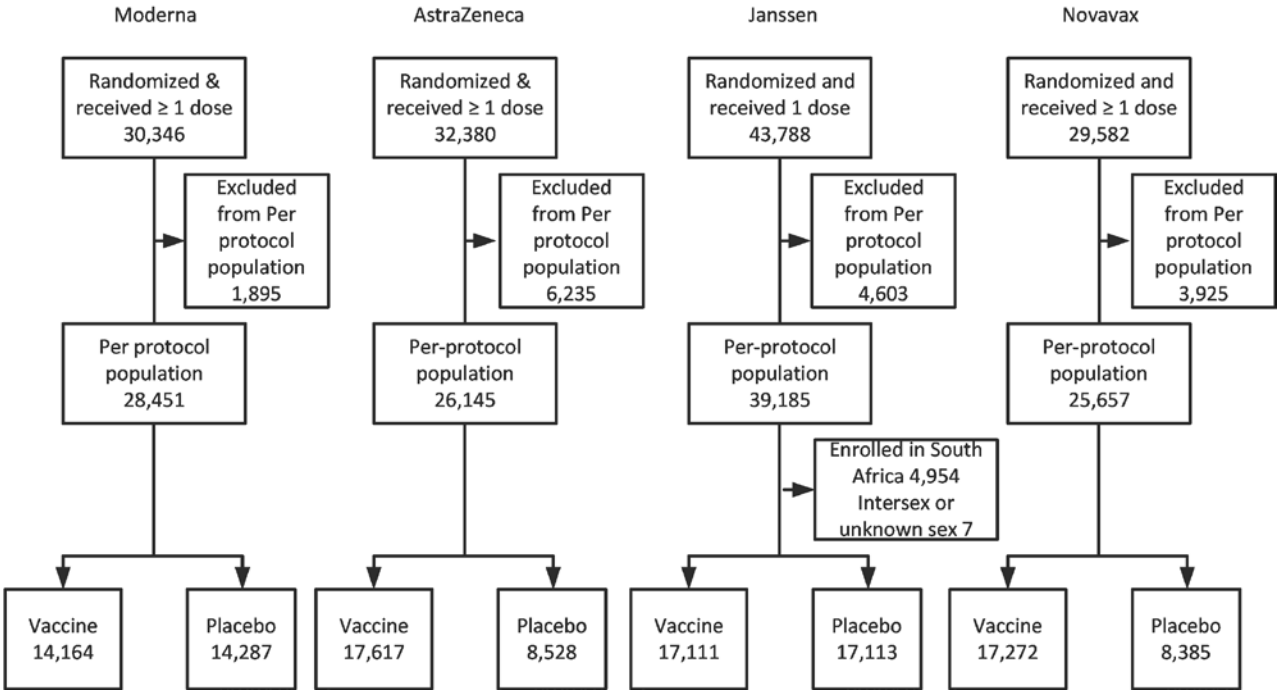
*Predictors of study outcomes*

Occupational risk was determined by attributing OSHA-defined hazard recognition scores on a scale of low, medium, high, or very high risk to self-reported workplace information provided by participants. If individuals selected more than one category, the highest risk score was used. Living situation risk was scored on a scale of low, medium, high, or very high risk based on number of co-habitants for AstraZeneca, Novavax, and Janssen, and on housing type for COVE Moderna (in which a specific number of cohabitants was not collected). For Moderna, if more than one housing type was selected, the highest risk score was used.

Supplementary Figure 1. Enrollment and allocation of participants in the ENSEMBLE trial showing exclusion of participants with COVID-19 events that were not from the ancestral/alpha/other lineages.



Supplementary Figure 2. CONSORT diagram



Supplementary Table 1. SARS-CoV-2 Vaccines investigated in Phase 3 trials coordinated by the COVID-19 Prevention Network (CoVPN) and analyzed in this study.

| Trial | Location of study sites | Vaccine product | Primary dose regimen | Study dates (before blinded cutoff/ crossover) | N for each study* | | VE symptomatic infection (+/-95%CI)*** | VE severe disease | Reference |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | Placebo | Vaccine | | | |
| Moderna (COVE) | United States | mRNA | 2 IM doses of mRNA-1273 or placebo 28 days apart | July 27, 2020 – March 26, 2021 | 14,073 | 14,134 | 94.1 (89.3-96.8) | 100** | El Sahly et al. 2021 PMC8482810 NCT04470427 |
| AstraZeneca/Oxford (AZD1222) | Chile, Peru, United States | Non-replicating adenovirus | 2 IM doses of AZD1222 or placebo 28 days apart | August 28, 2020 – July 30, 2021 | 8,550 | 17,662 | 73.98 (65.34-80.47) | 84.97 (58.97-94.5) | Falsey et al. 2021 PMC8522798 NCT04516746 |
| Janssen/ JNJ (ENSEMBLE) | Argentina, Brazil, Chile, Colombia, Mexico, Peru, South Africa, United States | Non-replicating adenovirus | Single IM dose of Ad26.COV2.S or placebo | September 21, 2020 – July 9, 2021 | 19,691 | 19,630 | 66.9 (59-73.4) | 76.7 (54.6-89.1) | Sadoff et al. 2021 PMC8220996 NCT04505722<br><br>Sadoff et al. 2022 PMC8849184 |
| Novavax (PREVENT-19) | Mexico, United States | Protein nanoparticle and Matrix-M1 adjuvant | 2 IM doses of NVX-CoV2373 or placebo 21 days apart | December 27, 2020 – June 1, 2021 | 9,868 | 19,714 | 90.4(82.9-94.6) | 100 (87-100) | Dunkle et al. 2022 PMC8693692 NCT04611802 |

*Population for primary efficacy analysis
**VE could not be estimated to 1.0
***For JNJ, Moderna, and AstraZeneca, the efficacy listed was calculated for the period from two weeks to two months after completion of primary dose regimen. For Novavax, the period was one week to three months.

Supplementary Table 2**.** Summary of Symptomatic Covid-19 endpoints excluded due to their occurrence after $t_0$ by randomization arm in the per-protocol cohort of each trial excluding intersex participants (# endpoints excluded/ total # endpoints).[†] The first row lists the number of any type of variants excluded, and the second row restricts to the number of ancestral/alpha/other variants excluded from the analysis.

| | Moderna $t_0$ = 179 days | | AstraZeneca $t_0$ = 191 days | | Janssen $t_0$ = 188 days | | Novavax $t_0$ = 115 days | |
|---|---|---|---|---|---|---|---|---|
| | Placebo | Vaccine | Placebo | Vaccine | Placebo | Vaccine | Placebo | Vaccine |
| All variants | 3/744 | 0/55 | 3/184 | 7/141 | 53/1082 | 37/495 | 2/79 | 0/17 |
| Analysis | 3/744 | 0/55 | 3/184 | 7/141 | 2/540 | 1/174 | 2/79 | 0/17 |

[†]: Separate time points $t0$ were chosen for each trial such that about 10% participants are at risk in the vaccine arm. $t0$ is the number of days from time of event or censoring date since randomization.

Supplementary Table 3**.** Summary of follow-up time since randomization in the per-protocol cohort of each trial excluding intersex participants and South African participants.

| | Moderna | | AstraZeneca | | Janssen | | Novavax | |
|---|---|---|---|---|---|---|---|---|
| | Median | [Min, Max] | Median | [Min, Max] | Median | [Min, Max] | Median | [Min, Max] |
| Follow-up time | 146 | [28, 243] | 104 | [39, 334] | 113 | [1, 284] | 92 | [23, 275] |

Supplementary Table 4**.** Estimates of vaccine efficacy (VE) against COVID-19 within subgroups defined by categorical baseline covariates, along with corresponding unadjusted 95% confidence intervals. Missing entries are due to the subgroup having fewer than 25 endpoints across the vaccine and placebo arms.

| | **Moderna** | **AstraZeneca** | **Janssen** | **Novavax** |
|---|---|---|---|---|
| Overall | 0.93 [0.90,0.95] | 0.65 [0.52,0.74] | 0.71 [0.64,0.77] | 0.91 [0.86,0.95] |
| **Socio-demographic** | | | | |
| Ethnicity | | | | |
| Hispanic/Latino | 0.96 [0.92,0.98] | 0.70 [0.02,0.91] | 0.66 [0.50,0.77] | 0.89 [0.39,0.98] |
| Not Hispanic/Latino | 0.92 [0.87,0.95] | 0.76 [0.51,0.89] | 0.76 [0.38,0.91] | 0.96 [0.87,0.99] |
| Race | | | | |
| American Indian/Alaska native | | 0.47 [-0.83,0.85] | 0.77 [0.67,0.84] | |
| Asian | 0.99 [0.97,1.00] | | | |
| Black/African American | 0.94 [0.90,0.96] | | 0.82 [-0.13,0.97] | |
| Other | 0.99 [0.99,1.00] | | 0.65 [-4.01,0.98] | |
| White | 0.92 [0.87,0.95] | 0.77 [0.44,0.90] | 0.70 [0.43,0.85] | |
| Multiple races | | | 0.66 [0.41,0.80] | |
| Sex | | | | |
| Female | 0.93 [0.88,0.96] | 0.75 [0.26,0.92] | 0.75 [0.56,0.85] | 0.93 [0.83,0.97] |
| Male | 0.92 [0.87,0.95] | 0.71 [0.45,0.85] | 0.67 [0.46,0.80] | 0.87 [0.74,0.94] |
| **Health characteristics** | | | | |
| Cardiovascular disease | | | | |
| Yes | 0.89 [0.76,0.95] | 0.88 [0.16,0.98] | 0.77 [0.02,0.94] | |
| No | 0.94 [0.91,0.96] | 0.70 [0.37,0.86] | 0.69 [0.55,0.79] | |
| Diabetes | | | | |
| Yes | 0.93 [0.77,0.98] | 0.82 [0.04,0.97] | 0.54 [-3.03,0.95] | |
| No | 0.93 [0.89,0.95] | 0.74 [0.42,0.88] | 0.72 [0.60,0.81] | |
| Chronic Lung Disease | | | | |
| Yes | 0.92 [0.67,0.98] | 0.35 [-1.67,0.84] | 0.72 [-0.55,0.95] | |
| No | 0.93 [0.89,0.95] | 0.75 [0.51,0.88] | 0.70 [0.57, 0.80] | |
| **Risk characteristics** | | | | |
| Workplace risk of Exposure€ | | | | |
| Low | | 0.78 [-0.27,0.96] | 0.70 [0.56,0.80] | 0.94 [0.79,0.98] |
| Medium | 0.93 [0.90,0.96] | 0.77 [0.42,0.91] | | 0.92 [0.79,0.97] |
| High | 0.92 [0.84,0.96] | 0.64 [-0.12,0.89] | 0.61 [0.25,0.79] | |
| Risk from living condition¥ | | | | |
| Low | 0.92 [0.82,0.97] | 0.84 [0.56,0.94] | 0.75 [0.52,0.87] | |
| Medium | 0.93 [0.89,0.96] | 0.65 [-0.14,0.89] | 0.74 [0.58,0.84] | |
| High | 0.88 [0.71,0.95] | 0.80 [-0.25,0.97] | 0.60 [0.09,0.83] | |
| Very high | 0.93 [0.28,0.99] | 0.61 [-1.13,0.93] | 0.37 [-1.12,0.81] | |

| | Moderna | AstraZeneca | Janssen | Novavax |
|---|---|---|---|---|
| Smoking history | | | | |
| Yes | | 0.61 [-0.31,0.89] | | 0.89 [0.65,0.97] |
| No | | 0.76 [0.50,0.88] | | 0.95 [0.83,0.98] |
| **Geographic location** | | | | |
| USA | | 0.78 [0.55,0.89] | 0.72 [0.30,0.89] | 0.93 [0.83,0.97] |
| Argentina | | | 0.44 [-0.33,0.76] | |
| Brazil | | | 0.73 [<-100.17,1.00] | |
| Colombia | | | 0.82 [0.73,0.88] | |
| Peru | | 0.45 [-0.47,0.79] | 0.46 [0.28,0.6] | |

[€]Detailed derivation of exposure risk based on OSHA categories is provided in Supplementary Methods

[¥]Living condition encompasses housing type and household size, detailed derivation provided in Supplementary Methods

Supplementary Table 5 Estimated tertiles of vaccine efficacy (VE) against the symptomatic Covid-19 endpoint reported for subgroups and corresponding unadjusted 95% confidence intervals. In each column, the newly reported subgroups are (i) subgroups where vaccine efficacy is predicted using only an estimate of VE based on data from the other three trials (pooled VE), and (ii) subgroups where vaccine efficacy is predicted using both an estimate of VE based on data from the other three trials and all available baseline covariates (All baseline covariates and pooled VE). If needed, estimates were projected to satisfy the population-level constraint that efficacy should be nondecreasing when moving from the first to the third tertile of participants.

| | **Moderna** | **AstraZeneca** | **Janssen** | **Novavax** |
|---|---|---|---|---|
| **Pooled VE** | | | | |
| Lowest VE | 0.93 [0.88,0.96] | 0.53 [0.27,0.79] | 0.68 [0.55,0.81] | 0.90 [0.75,0.96] |
| Middle VE | 0.93 [0.88,0.98] | 0.73 [0.55,0.89] | 0.68 [0.55,0.81] | 0.90 [0.75,1.00] |
| Highest VE | 0.93 [0.88,0.98] | 0.73 [0.55,0.90] | 0.77 [0.67,0.87] | 0.90 [0.79,1.00] |
| **All baseline covariates** | | | | |
| Lowest VE | 0.93 [0.89,0.97] | 0.62 [0.38,0.81] | 0.68 [0.55,0.80] | 0.89 [0.75,1.00] |
| Middle VE | 0.93 [0.89,0.97] | 0.62 [0.38,0.86] | 0.69 [0.57,0.82] | 0.89 [0.75,1.00] |
| Highest VE | 0.93 [0.89,0.97] | 0.74 [0.56,0.93] | 0.75 [0.63,0.86] | 0.94 [0.86,1.00] |
| **All baseline covariates and pooled VE** | | | | |
| Lowest VE | 0.93 [0.88,0.97] | 0.53 [0.30,0.76] | 0.67 [0.54,0.80] | 0.90 [0.77,1.00] |
| Middle VE | 0.93 [0.88,0.97] | 0.71 [0.52,0.91] | 0.67 [0.54,0.80] | 0.90 [0.77,1.00] |
| Highest VE | 0.93 [0.89,0.97] | 0.76 [0.59,0.93] | 0.77 [0.66,0.88] | 0.93 [0.86,1.00] |

Supplementary Table 6. Severe Covid-19 endpoint† of ancestral/alpha/other strains summaries by subgroup and randomization arm in the per-protocol cohort of each trial excluding South African participants and intersex participants.

| | **Moderna** | | **AstraZeneca** | | **Janssen** | | **Novavax** | |
|---|---|---|---|---|---|---|---|---|
| | Placebo | Vaccine | Placebo | Vaccine | Placebo | Vaccine | Placebo | Vaccine |
| Overall | 106/14164 | 2/14287 | 10/8528 | 1/17617 | 75/17113 | 19/17111 | 4/8385 | 0/17272 |
| **Socio-demographic** | | | | | | | | |
| Age (years) | | | | | | | | |
| ≥ 18 to < 65 | 76/10569 | 1/10661 | 8/6692 | 1/13878 | 65/13480 | 16/13505 | 4/7417 | 0/15228 |
| ≥ 65 | 30/3595 | 1/3626 | 2/1836 | 0/3739 | 10/3633 | 3/3606 | 0/968 | 0/2044 |
| Ethnicity | | | | | | | | |
| Hispanic/ Latino | 21/2787 | 1/2831 | 3/2064 | 0/4032 | 45/8889 | 10/8767 | 0/1801 | 0/3707 |
| Not Hispanic/ Latino | 85/11377 | 1/11456 | 7/6464 | 1/13585 | 30/8224 | 9/8344 | 4/6584 | 0/13565 |
| Race | | | | | | | | |
| American Indian/Alaska Native | 0/113 | 0/109 | 2/372 | 0/747 | 20/1616 | 7/1641 | 0/522 | 0/1068 |
| Asian | 4/700 | 0/628 | 0/355 | 0/738 | 0/639 | 1/699 | 0/375 | 0/757 |
| Black or African American | 6/1352 | 0/1395 | 1/699 | 0/1401 | 9/1451 | 1/1416 | 0/947 | 0/1881 |
| Multiple | 3/304 | 0/300 | 0/203 | 0/421 | 12/951 | 0/928 | 0/137 | 0/296 |
| White | 86/11273 | 2/11391 | 7/6735 | 1/13972 | 33/11835 | 10/11838 | 4/6350 | 0/13124 |
| Other | 7/422 | 0/464 | 0/164 | 0/338 | 1/621 | 0/589 | 0/54 | 0/146 |
| Sex | | | | | | | | |
| Male | 56/6670 | 1/6848 | 4/3714 | 1/7732 | 32/7629 | 6/7617 | 2/4158 | 0/8283 |
| Female | 50/7494 | 1/7439 | 6/4814 | 0/9885 | 43/9484 | 13/9494 | 2/4227 | 0/8989 |
| **Health Characteristics** | | | | | | | | |
| Body mass index ($kg/m^2$) | | | | | | | | |

| | Moderna | | AstraZeneca | | Janssen | | Novavax | |
|---|---|---|---|---|---|---|---|---|
| Healthy weight or Underweight (< 25) | 15/3861 | 0/3970 | 1/2482 | 0/5275 | 20/5379 | 5/5501 | 0/2463 | 0/5220 |
| Overweight (≥ 25, <30) | 31/4938 | 1/4857 | 3/3079 | 0/6406 | 33/7055 | 5/6901 | 0/2719 | 0/5610 |
| Obese (≥ 30) | 60/5365 | 1/5460 | 6/2967 | 1/5936 | 22/4679 | 9/4709 | 4/3203 | 0/6442 |
| Class 3 Obese (≥ 40) | 16/995 | 0/1015 | 1/473 | 1/976 | 4/601 | 1/643 | 1/572 | 0/1267 |
| Cardiovascular Disease | 37/4472 | 1/4468 | 7/2392 | 1/5057 | 22/4550 | 2/4422 | 3/2021 | 0/4059 |
| Diabetes | 23/1468 | 0/1484 | 1/873 | 0/1627 | 11/1587 | 1/1608 | 3/858 | 0/1622 |
| HIV | 1/87 | 0/93 | 2/134 | 0/278 | 2/274 | 0/240 | 0/49 | 0/135 |
| Kidney Disease | 2/74 | 0/73 | 1/45 | 0/131 | 0/113 | 0/106 | 1/56 | 0/125 |
| Liver disease | 1/101 | 0/113 | 1/160 | 0/308 | 2/168 | 0/169 | 0/62 | 0/134 |
| Chronic Lung Disease | 11/808 | 1/808 | 1/1055 | 1/2027 | 4/1150 | 3/1114 | 2/1264 | 0/2461 |
| **Risk Characteristics** | | | | | | | | |
| Workplace Risk of Exposure[€] | | | | | | | | |
| Low | | | 3/2666 | 0/5204 | 70/16355 | 17/16390 | 2/4951 | 0/9956 |
| Medium | 53/6735 | 1/6760 | 6/3765 | 1/7734 | 1/229 | 0/249 | 2/2670 | 0/5607 |
| High | 53/7429 | 1/7527 | 1/2097 | 0/4679 | 4/529 | 2/472 | 0/764 | 0/1709 |
| Risk from Living[¥] Condition | | | | | | | | |
| Low | 18/2354 | 1/2322 | 4/4365 | 0/9044 | 39/8676 | 8/8739 | 3/6808 | 0/14002 |
| Medium | 81/10229 | 1/10348 | 4/1593 | 1/3227 | 24/4978 | 5/4923 | 1/1117 | 0/2274 |
| High | 2/1179 | 0/1165 | 2/1336 | 0/2744 | 10/2694 | 5/2683 | 0/339 | 0/733 |
| Very High | 5/402 | 0/452 | 0/1234 | 0/2602 | 2/765 | 1/766 | 0/121 | 0/263 |
| Smoke history | 1/253 | 0/236 | 4/1705 | 0/3471 | 1/369 | 2/374 | 3/2609 | 0/5330 |
| **Geographic Location** | | | | | | | | |
| USA | 106/14164 | 2/14287 | 8/7423 | 1/15389 | 34/9121 | 11/9156 | 4/7887 | 0/16261 |
| Argentina | | | | | 2/1414 | 0/1400 | | |

| | Moderna | | AstraZeneca | | Janssen | | Novavax | |
|---|---|---|---|---|---|---|---|---|
| Brazil | | | | | 4/3385 | 0/3394 | | |
| Chile | | | 0/670 | 0/1358 | 0/539 | 0/528 | | |
| Colombia | | | | | 31/1862 | 5/1856 | | |
| Mexico | | | | | 1/218 | 2/207 | 0/498 | 0/1011 |
| Peru | | | 2/435 | 0/870 | 3/574 | 1/570 | | |

*†*: All participants were right censored at time *t*0 regardless of whether they were observed to experience event

after *t*0. Separate time points *t*0 were chosen for each trial such that about 10% participants are at risk in the vaccine arm.

[§]Category is defined across all clinical sites. Indigenous people from South America were classified together with the American Indian or Alaska Native United States and Mexico demographic according to the FDA definition (American Indian or Alaska Native: A person having origins in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or community attachment). In this analysis, the Moderna, AstraZeneca, Janssen and Novavax trials included 222, 237, 212, and 1590 participants, respectively, who identified as American Indian or Alaskan Native from North America.

[€]Detailed derivation of exposure risk based on OSHA categories is provided in Supplementary Methods

[¥]Living condition encompasses housing type and household size, detailed derivation provided in Supplementary Methods

Supplementary Table 7. Estimates of vaccine efficacy (VE) of the Janssen vaccine against severe Covid-19 within subgroups defined by categorical baseline covariates, along with corresponding 95% confidence intervals (CI).

| | Janssen | |
|---|---|---|
| | VE | 95% CI |
| Overall | 0.85 | [0.77, 0.90] |
| **Socio-demographic** | | |
| Ethnicity | | |
| Hispanic/Latino | 0.85 | [0.71,0.92] |
| Not Hispanic/Latino | 0.85 | [-0.79,0.99] |
| Sex | | |
| Female | 0.90 | [0.63,0.97] |
| Male | 0.80 | [0.54,0.92] |
| **Health characteristics** | | |
| Obesity | | |
| Yes | 0.71 | [0.18,0.9] |
| No | 0.89 | [0.7,0.96] |
| **Risk characteristics** | | |
| Risk from living condition | | |
| Low | 0.85 | [0.74,0.91] |
| Medium | 0.97 | [0.84,1.00] |
| **Geographic Location** | | |
| USA | 0.81 | [-0.56,0.98] |
| Colombia | 0.93 | [0.88,0.96] |

Missing entries are due to the subgroup having fewer than 25 endpoints across the vaccine and placebo arms.

Supplementary Table 8. Estimated tertiles of vaccine efficacy (VE) against the severe Covid-19 endpoint. The ENSEMBLE trial, which evaluated the Janssen vaccine, was the only trial with sufficient severe endpoints on the vaccine arm to evaluate heterogeneity of vaccine efficacy against severe Covid-19. If needed, estimates were projected to satisfy the population-level constraint that efficacy should be nondecreasing when moving from the first to the third tertile of participants.

| | **Janssen** | |
|---|---|---|
| | VE | 95% CI |
| Lowest VE | 0.75 | [0.48,1.00] |
| Middle VE | 0.76 | [0.48,1.00] |
| Highest VE | 0.97 | [0.91,1.00] |