## Statistical Analysis Plan

| | |
|---|---|
| Protocol Title: | A Phase 3, Randomized, Observer-blind, Placebo-Controlled, Group-Sequential Study to Determine the Immunogenicity and Safety of a Respiratory Syncytial Virus (RSV) F Nanoparticle Vaccine with Aluminum in Healthy Third-Trimester Pregnant, Women, and Safety and Efficacy of Maternally Transferred Antibodies in Preventing RSV Disease in Their Infants |
| Protocol Number: | RSV-M-301 |
| Authors: | Peter Gilbert, Ph.D.<br>Youyi Fong, Ph.D.<br><br>Statistical Center for HIV/AIDS Research and Prevention<br>Vaccine and Infectious Disease Division<br>Fred Hutchinson Cancer Research Center |
| Date: | 30 April 2020 |
| SAP Version: | v2.0 |

## *APPROVAL SIGNATURES*

Protocol: RSV-M-301

SAP Version: 2.0

Version Date: 30 April 2020

By signing below, I confirm that I have reviewed and approve the above version of the Statistical Analysis Plan.

Peter Gilbert, Ph.D.
Lead Statistician

5/4/2020
Date

Youyi Fong, Ph.D.
Co-Lead Statistician

May 4 20
Date

Louis F. Fries III, M.D.
Novavax Clinical Development

2 May 20
Date

Table of Contents

# 1 OVERVIEW

This statistical analysis plan focuses on the assessment of anti-F IgG ELISA, Palivizumab-competitive antibody (PCA), and Microneutralization (MN) 50% antibody titers to RSV/A and RSV/B, measured in mothers and their infants, as correlates of risk and as correlates of protection against RSV disease endpoints in infants in RSV-M-301.

The correlates analyses described in this SAP are conducted for the South Africa study sites and follow-up period of infants through to 90 days of age, for reasons including: (1) restricting to South Africa simplifies interpretation of results by restricting heterogeneity of factors related to RSV infection with medically significant or severe lower respiratory tract illnesses (LRTI); (2) Most RSV disease endpoints were in South Africa; (3) estimated vaccine efficacy was relatively high in South Africa; and (4) the protocol-specified vaccine efficacy analysis focused on follow-up through 90 days of age, and, as hypothesized, the attack rate in the placebo group and the corresponding estimated vaccine efficacy was highest when restricting to 90 days of follow-up due to the declining antibody in the infant subjects and by design to deliver babies during the peak of RSV season.

# 2 TWO-PHASE SAMPLING DESIGN FOR MEASURING MN TITERS

Anti-F IgG ELISA titers and PCA antibody titers were measured at all available time points from all RSV-M-301 study participants. A two-phase case-control sampling design is used for determining the subset of RSV-M-301 participants from whom to measure MN antibody titers.

Cases are defined as ITT-EFF-I infants in the analysis cohort of interest (defined in Section 3) with an illness started for any of the three efficacy endpoints (MS RSV LRTI, RSV LRTI with severe hypoxemia, RSV LRTI with hospitalization) through 90 days of age in the expanded data set.

A “cumulative case-control” design is used, which defines eligible controls as ITT-IMM-I infant and ITT-IMM-M mother-infant pairs in the analysis cohort of interest (defined in Section 3) who did not experience any of the three efficacy endpoints listed above through to 90 days of age and have available samples at Maternal Baseline, Maternal Day 14, and cord blood.

Immunogenicity data are measured in a two-phase sample, where, within each cell of a 2 by K table defined by (case, eligible control) cross-classified with K covariate strata, a without replacement sample is drawn. The K=20 covariate strata are defined by the cross-classification of treatment assignment, 5 location strata (site cluster NN, …, SS), and the indicator of whether the number of days between vaccination and birth is greater than or equal to 30 days.

For cases, the sampling probability in each cell is one such that immunogenicity data are measured for all cases. For eligible controls, we sample 2 controls for every case in each placebo arm stratum and sample 4 controls for every case in each vaccine arm stratum. Strata with no cases are treated as if they have a single case. Table 1 summarizes the number of cases and controls that are sampled for immunogenicity measurement across the K=20 covariate strata.

The sampling stratification variables are partly chosen based on association with the estimated VE (1-OR), where there was evidence that VE was greater for infants with >= 30 days between maternal vaccination and birth, and that VE varied over geographic locations. A study of such associations is summarized in Appendix A.

Table 1: Distribution of samples by location and indicator of vaccination-to-birth >= 30 days, separately for each treatment/case status cell, intersection of ITT-EFF-I, ITT-IMM-I cohorts (defined in Section 3).

| | Placebo | | | | Vaccine | | | |
|---|---|---|---|---|---|---|---|---|
| | control | | Cases | | Control | | Case | |
| | <30 Days | >=30 Days | <30 Days | >=30 Days | <30 Days | >=30 Days | <30 Days | >=30 Days |
| NN | 39 | 372 | 3 | 19 | 62 | 764 | 1 | 26 |
| PP | 12 | 79 | 1 | 11 | 30 | 155 | 2 | 7 |
| QQ | 15 | 82 | 2 | 6 | 39 | 170 | 5 | 5 |
| RR | 13 | 71 | 1 | 7 | 23 | 154 | 1 | 3 |
| SS | 11 | 40 | 0 | 1 | 18 | 89 | 0 | 2 |
| Total | 90 | 644 | 7 | 44 | 172 | 1332 | 9 | 43 |

# 3 COHORT DEFINITIONS

We define participant cohorts relevant to the correlates analyses, which are as defined in the RSV-M-301 protocol. (In this SAP we use both "population" and "cohort" to denote a cohort of participants enrolled in RSV-M-301.)

### Intent-to-Treat Populations

Intent-to-treat Efficacy (ITT-EFF) Population – defined as all infants (ITT-EFF-I) in the Safety Population for whom at least one post-partum, respectively, efficacy measurement is available as evidenced by collection of surveillance observations.

Intent-to-treat Immunogenicity (ITT-IMM) Population – defined as all maternal participants (ITT-IMM-M) and their infants (ITT-IMM-I) in the Safety Population for whom at least one post-treatment immunogenicity measurement is available.

### Per-Protocol Populations

### Per-Protocol Efficacy (PP-EFF) Populations

Per-Protocol Population for infant participants (PP-EFF-I) – defined as all infant participants who: a) are ≥ 37 weeks gestational age at birth, b) are born to maternal participants who received a study injection as randomized and ≥ 2 weeks prior to delivery, c) have not received prophylactic treatment with palivizumab between birth and Day 180 after delivery, d) have at least one documented post-partum contact during which active and/or passive surveillance activities for RSV-suspect illness can occur, and e) have no major protocol deviations affecting the primary efficacy outcomes as determined and documented by Novavax prior to database lock and unblinding.

### Per-Protocol Immunogenicity (PP-IMM) Populations

PP-IMM for Maternal Participants (PP-IMM-M) – defined as all maternal participants in the PP-EFF-M population and who receive the test article and regimen to which they were randomized, provide baseline and delivery (within 72 hours of delivery) serology data, and have no major

protocol deviations affecting the primary immunogenicity outcomes as determined and documented by Novavax prior to database lock and unblinding.

PP-IMM for Infant Participants (PP-IMM-I) – defined as all infant participants in the PP-EFF-I population and who: a) are ≥ 37 weeks gestational age at birth, b) are born to maternal participants who received a study injection as randomized and ≥ 2 weeks prior to delivery, c) have provided a cord blood specimen (or infant blood sample by venipuncture or heel stick within 72 hours of delivery as an acceptable substitute), d) have not received prophylactic treatment with palivizumab between birth and Day 180 after delivery, and e) have no major protocol deviations affecting the primary immunogenicity outcomes as determined and documented by Novavax prior to database lock and unblinding.

**Cohorts for Correlates Analyses**

The correlates analyses are conducted in the PP-IMM-M and PP-IMM-I Populations, which are the primary analysis populations for the immunogenicity analyses, intersected with the PP-EFF-M and PP-EFF-I populations, specified in the RSV-M-301 protocol. All correlates analyses are based on actual treatment received (RSV F Vaccine or Placebo).

As reported in Table 1, there are 51 placebo recipient cases and 52 vaccine recipient cases where antibody marker data are expected to be available for analysis.

The clinical data sets to be used for the analyses are based on the Day 180 unblinding and the adjudicated outcomes completed prior to unblinding by the independent Clinical Evaluation Adjudication Committee (CEAC).

# 4 ANTIBODY RESPONSE VARIABLES AND TIME POINTS

## 4.1 Antibody Assays

Four different assays are used to measure antibody responses:

- Anti-F IgG ELISA titer, reported in optical density (OD) units, with a negative response defined by a readout below Lower Limit Of Quantification (LLOQ) of 400; such readouts are set to 200 for analysis.
- Palivizumab-competitive antibody (PCA) concentration, reported as serum concentration corresponding to 50% binding of unlabeled palivizumab, with a negative response defined by a readout below the LLOQ of 12; such readouts are set to 6 for analysis.
- Microneutralization (MN) 50% antibody titer to RSV/A on HEp-2 cells, reported as the reciprocal of the serum dilution at which > 50% reduction in viral cytopathic effect (CPE) is observed. Titers will be reported in International Unit (IU) with the LLOQ of 13 for RSV/A. Titers below the LLOQ will also be set to 6.5 IU.
- MN 50% antibody titer to RSV/B on HEp-2 cells, defined and reported similarly as for RSV/A. Titers will be reported in International Unit (IU) with the LLOQ of 8 for RSV/B. Titers below the LLOQ will also be set to 4 IU.

As summarized in Section 7.2.1 of the protocol, anti-F IgG ELISA titer is measured using serial 3-fold dilutions of participant sera starting at 1:10. A four-parameter logistic (4-PL) curve fit is

applied to the reference standard and the reported titer in OD units is determined from the mean calculated from serial dilutions.

As summarized in Section 7.2.2 of the protocol, the PCA ELISA also uses serial dilutions of test sera, diluted minimally at 1:4, which are incubated with a fixed amount of biotin-labeled palivizumab in the RSV F-coated plates. Palivizumab-like antibodies in the test serum are quantitated based on OD values corresponding to 50% binding of unlabeled palivizumab. The final concentration is obtained by linear interpolation/extrapolation of the reference curve adjusted by dilution factor using a 4-PL curve fit analysis. The binding of biotinylated palivizumab to the antigen on the plate reflects the competition of palivizumab-like antibodies from human serum.

As summarized in Section 7.2.3 of the protocol, neutralizing antibody titer is defined as the reciprocal of the serum dilution at which > 50% reduction in viral cytopathic effect (CPE) is observed.

### 4.2 Antibody Assay Measurement Time Points

The antibody measurement sampling time points used for the correlates analyses specified in this SAP are as follows:

- Maternal: Baseline (screening/pre-immunization, -28 – 0 days), Day 14 (± 2 days) post-vaccination
- Infants: Birth (cord blood sampled up to 72 hours after birth)

In this SAP we refer to the sampling time points as maternal baseline, maternal Day 14, and infant birth. The Anti-F IgG and PCA antibody titers were also measured from maternal delivery samples, whereas the MN antibody titers were not measured from maternal delivery samples. Consistent with this sampling design, the correlates analyses focus on the three sampling time points maternal baseline, maternal Day 14, and infant birth.

### 4.3 Antibody Variable Sets

For each of the four assays, several distinct types of antibody variables are defined for analysis, which account for maternal and/or infant samples. In particular, the following sets are defined.

1. Maternal baseline antibody titer (log10 transformed)
2. Maternal Day 14 antibody titer (log10 transformed)
3. Maternal fold-rise in antibody titer from baseline to Day 14 (log10 transformed, i.e., log10 maternal Day 14 titer minus log10 maternal baseline titer)
4. Infant birth (cord blood) titer (log10 transformed)

## 5 UNSUPERVISED DESCRIPTIVE AND RISK SCORE ANALYSES FOR REFINING THE ANALYSIS PLAN

The correlates statistical analysis plan drafted below may be revised based on unsupervised analysis of immune response marker data, that is conducted based on the control group (free of RSV disease at 90 days of age) only, and therefore essentially studies the markers in the whole study population (given that RSV disease is a rare event). The fact that these unsupervised analyses are done independent of any information on case vs. control status makes it valid to refine the correlates analysis plan based on the analyses. The planned unsupervised analyses are described in Sections 5.1 and 5.2.

In addition, analyses are done of baseline demographic RSV risk scores, as described in Section 5.3.

## 5.1 Antibody Marker Variables for Unsupervised Descriptive Analysis

The descriptive analyses described in this section are done in controls only.

Table 2 lists the 16 antibody markers that are studied. The distribution of each of the markers is described with a boxplot, with stratification by treatment arm. Pairs plots will be used, with Spearman rank correlation estimates affixed to the plots, to describe intercorrelations of assay variables. Separate pairs plots will be made for each of the four assay classes (across the four types of titer variables), and for each of the four types of titer variables (across the four assays). These plots are all made separately for the vaccine and placebo groups.

Table 2: 16 antibody markers studied descriptively and in correlates analyses (all plotted and analyzed on the log10 transformed scale).

| |
|---|
| Anti-F IgG titer<br>a. Maternal baseline antibody titer<br>b. Maternal Day 14 antibody titer<br>c. Maternal fold-rise antibody titer (baseline to Day 14)<br>d. Infant birth antibody titer |
| PCA titer<br>a. Maternal baseline antibody titer<br>b. Maternal Day 14 antibody titer<br>c. Maternal fold-rise antibody titer (baseline to Day 14)<br>d. Infant birth antibody titer |
| MN titer RSV/A |

| |
|---|
| a. Maternal baseline antibody titer<br>b. Maternal Day 14 antibody titer<br>c. Maternal fold-rise antibody titer (baseline to Day 14)<br>d. Infant birth antibody titer |
| MN titer RSV/B<br>a. Maternal baseline antibody titer<br>b. Maternal Day 14 antibody titer<br>c. Maternal fold-rise antibody titer (baseline to Day 14)<br>d. Infant birth antibody titer |

## 5.2 Decisions for Correlates Analyses Based on the Unsupervised Descriptive Analyses

1. If the MN titer RSV/A and MN titer RSV/B responses tend to be highly correlated, e.g., with Spearman rank correlation coefficients typically greater than 0.9, then the two variables will be combined into a single antibody variable for the analyses. For each variable type maternal baseline, maternal Day 14, and infant birth, the single variable would be the geometric mean of the two titer readouts.
2. Studying post-baseline marker correlates of VE (a type of a correlate of protection, Qin et al., 2007; Plotkin and Gilbert, 2012; Moodie et al., 2018) is only expected to be successful if a baseline variable is available that is reasonably predictive of the post-baseline marker that is being assessed as a correlate of VE. This statement is based on the set-up of the correlates of VE framework within the field of principal stratification causal inference (Frangakis and Rubin, 2002), which conducts estimation and hypothesis testing about a parameter that has been called the "VE curve," which is defined by a contrast in two conditional risks: the risk of disease in vaccine recipients conditional on the value of an immunological marker in vaccine recipients vs. the risk of disease in placebo recipients conditional on the value of the counterfactual value of the same immunological marker that would have occurred had the placebo recipient, counter to fact, been assigned to the vaccine arm. The logic of this framework is that the "VE curve" measures vaccine efficacy in subgroups defined by the immunological marker in vaccine recipients, with an interpretation similar to vaccine efficacy for the overall cohort except applied to subgroups of vaccine recipients. Gilbert et al. (2014) illustrates use of the VE curve framework (for a herpes zoster randomized vaccine efficacy trial), which discusses the interpretation and value of the framework. In addition, the statement that a baseline predictor is needed to effectively make inference about the VE curve is based on a series of simulation studies reported in statistical methods papers [e.g., spanning from Gilbert and Hudgens (2008) to Juraska et al. (2018)]. Therefore, for each of the 12 post-baseline antibody markers defined by the four assays and marker types 2. through 4. defined

in Section 4.3, the marker will only be studied as a correlate of VE if the Spearman rank correlation between the marker and the same marker at baseline exceeds 0.5. An alternative type of correlate of protection is a mediator of VE, as discussed in Section 8. Mediators of VE will be assessed regardless of whether any baseline variables are available for predicting the post-baseline marker under evaluation.

## 5.3 Demographic Covariates and Development of Risk Scores for Confounding Control

Table 3 lists "demographic and potentially RSV prognostic" covariates of interest that are considered in various correlates analyses. These covariates are grouped into two categories: maternal enrollment variables and delivery/birth variables. In addition to these two categories, an important variable for the analysis is time from vaccination to birth, which we treat as a separate variable in its own right to adjust for, and not considered as part of either risk score. The variable is encoded numerically as the number of days between vaccination and birth.

Table 3: "Demographic putative RSV exposure" covariates that are adjusted for in various correlates analyses.

| |
|---|
| **Maternal Enrollment Variables** |
| Indicator of > 28, corresponding to the randomization of pregnant women within <= 28 years-old and > 28 years-old age strata |
| Age as a continuous variable, for more refined confounding control in some analyses |
| Number of previous children (count variable) |
| BMI (continuous) |
| HIV status |
| Smoker status |
| Asthma status |
| **Delivery/Birth Variables** |
| RSV season intensity score at the time of birth (This score will be developed based on the RSV infection data, which will depend on each infant participant's geographic location and birth date/calendar time. The score will be developed using both treatment arms pooled to maximize |

| |
|---|
| the information. After the score is defined, it will be evaluated for its correlation with the RSV disease endpoints.) |
| Indicator of evidence for GBS colonization |
| Exposure to intrapartum antibiotic prophylaxis: defined as intravenous penicillin, ampicillin, cefazolin, clindamycin or vancomycin, for ≤2 hours before delivery |
| Indicator infant lives with a smoker |
| Infant sex |
| Indicator other children < 5 years of age in home |
| Indicator of day care, or infant lives in home with day care attendee |
| Birthweight (continuous and low vs. not low) [low birthweight is a defined adverse event of special interest (AESI)] |
| Ratio of length to birthweight (a marker of being a 'physiologically young infant' at birth) |
| Frontal Occipital Head Circumference (FOC) (continuous) |
| Estimated gestational age at birth in days (count variable) |
| Small for gestational age [defined adverse event of special interest (AESI)] |
| Intrauterine growth retardation [defined adverse event of special interest (AESI)] |

For each of the two covariate categories in Table 3, a risk score is developed, where one or more of these risk scores are controlled for in correlates analyses to adjust for potential confounding. The risk scores are developed using placebo group data only.

Each risk score is defined as the logit of the predicted RSV disease risk from a superlearner estimated regression model. The same settings of superlearner (e.g., loss function, cross-validation technique, library of learners) are used as are used for implementation of superlearner for antibody marker correlates analyses, as described in Section 7.8.1. The number of days between vaccination and birth is one of the input variables for the superlearner model for the Delivery/Birth risk score. The indicator of the days between vaccination and birth being >= 30 days (vs. 14-29 days) is adjusted for as a separate variable not part of either risk score.

The risk score based on the Maternal Enrollment variables is selected as follows, based on the n=6 covariates. Six superlearner models are fit, with model k (among k=1, …, 6) only allowing models with at most k covariates in the model. The six superlearner models are compared by CV-AUC. The most parsimonious model (i.e., the model with smallest value of k) with estimated

CV-AUC no more than 0.01 less than the superlearner model with highest estimated CV-AUC will be selected. The risk score based on Delivery/Birth variables will be selected in the same way, with models k=1, …, 8 considered, such that the largest considered models include 8 covariates. Larger models are not considered based on the available sample size.

***Mapping of the set of baseline risk scores included in covariate-adjustment for the various correlates analyses.*** All of the correlates analyses adjust for the baseline risk score; the fact it is measured at baseline (before randomization) ensures that it is straightforward to adjust for the variable and retain the benefits of randomization. Because the risk score defined from delivery/birth samples is measured after randomization, it is less clear whether and when to adjust for the score, due to either post-randomization selection bias or because the score is on the causal pathway of the vaccine effect on RSV disease (i.e., partially mediates the vaccine effect on RSV disease). Our general strategy is to only allow delivery/birth variables to be included in covariate adjustment if they are thought to be highly unlikely in the causal pathway of the vaccine effect on RSV disease. For example, delivery/birth variables such as the indicator of living with a smoker are thought not to be on the causal pathway, such that analyses generally can safely adjust for these variables. Breastfeeding is judged to be on the causal pathway, and thus it is not adjusted for.

Among delivery/birth variables, the number of days between vaccination and delivery is particularly interesting because it is potentially entangled with the level of immune responses of infants. Thus, it may need to be treated differently when examining different time points. Specifications for this issue are detailed in Section 7.3.1.

We consider all of the Delivery/Birth variables listed in Table 3 to be unlikely on the causal VE pathway such that they are included in risk score building analyses.

# 6 RSV DISEASE ENDPOINTS

The correlates analyses study the following three RSV disease endpoints, only counting endpoints occurring by the Day 90 visit (if an infant has a Day 90 visit) or by 90 days after birth (if an infant does not have a Day 90 visit):

1. ["RSV Disease"] Composite endpoint defined as the first occurrence of any of the three protocol-specified endpoints MS RSV LRTI, RSV LRTI with severe hypoxemia, RSV LRTI with hospitalization in the expanded data set.
2. RSV LRTI with Severe Hypoxemia
3. RSV LRTI with Severe Hypoxemia without cough

# 7 CORRELATES OF RISK (COR) ASSESSMENT

## 7.1 CoR Objectives

The following CoR objectives are assessed in the vaccine and placebo groups separately.

**Objective 1 [Univariable CoRs]** To assess each antibody marker (defined by one of the four assays and marker type 1. through 4. defined in Section 4.3) as a correlate of risk of the RSV disease endpoint (16 total markers)

**Objective 2 [Multivariable CoRs Focused]**

a. **[By sample/time point across assay types]** To assess four sets of antibody markers as multivariable correlates of risk of the RSV disease endpoint; these sets are (1) Baseline titers from the four assays; (2) Day 14 titers from the four assays; (3) Day 14 fold-rise titers from the four assays; and (4) Infant birth titers from the four assays.

b. **[By assay type across sample/time points]** To assess each of the four assays measured across the multiple sample/time points (baseline, Day 14, Day 14 fold-rise, infant birth) as multivariable correlates of risk of the RSV disease endpoint.

**Objective 3 [Machine Learning Multivariable CoRs]** To build 'estimated optimal surrogates' (Price et al., 2018) based on different specified antibody variable sets, i.e., models that best predict the RSV disease endpoint.

All of the objectives are assessed in a manner adjusting for potential confounding covariates, as described in Sections 7.3, 7.4, and 7.5.

We elaborate the questions being addressed by the different analyses.

Objective 1 studies each of the 16 individual markers as a correlate of RSV disease within each of the vaccine and placebo groups, controlling for baseline exposure covariates, which asks questions about the prognostic value of a single measurement (one marker at one time point). If these analyses identify a strong CoR that can be validated as a surrogate endpoint, then it would provide a practical surrogate endpoint in only requiring a single assay measurement. Ideally the surrogate endpoint would be a Day 14 maternal readout, such that studies would not need to study infant samples.

Objective 2 asks, in a structured way, whether improved CoRs can be developed by either (a) using multiple assays at a given sample/time point; or (2) using multiple sampling time points for a given assay. If it showed stronger CoRs than Objective 1, it could lend support to a valid surrogate endpoint that would require multiple assays and/or sample/time points.

Objective 3 is similar to Objective 2, except that it fits a larger set of regression models, as a more comprehensive search for the best possible CoR, attempting to be maximally data driven. The analysis is similar to the machine learning CoR analysis conducted in Neidich et al. (2019, Figure 3). Objectives 1 and 2, on the other hand, ask specific, focused questions, such that Objectives 1 and 2 are complementary to Objective 3.

Objectives 1 and 2 use formal hypothesis testing (p-values) for whether antibody markers or sets of antibody markers are correlated with RSV disease; whereas Objective 3 is based on quantification of individual-level classification accuracy. To generate additional hypotheses, we add a fourth objective, an exploratory objective, which, unlike Objectives 1, 2, and 3, is not fully pre-specified, and thus it is only hypothesis generating.

**Objective 4 [Multivariable CoRs Exploratory]** For each of the eight antibody marker sets analyzed in Objective 2, to conduct logistic regression modeling, controlling for baseline demographic covariates, to explore best-fitting models of RSV disease that include elements such as polynomials, hinges/thresholds, and interactions.

## 7.2 Antibody Markers/Variable Sets for the CoR Objectives

***Objective 1.*** The 16 antibody markers defined in Table 2 (excluding the maternal delivery and TPP markers) are assessed as univariable CoRs.

***Objective 2a.*** The same 16 antibody markers as assessed in Objective 1 are assessed, arranged into four separate analyses of (1) the maternal baseline titer variables; (2) the maternal Day 14 titer variables; (3) the Day 14 fold-rise titer variables; and (4) the infant birth titer variables.

***Objective 2b.*** Again, the same 16 antibody markers as assessed in Objective 1 are assessed, arranged into four separate analyses of (1) the anti-F IgG titer variables; (2) the PCA titer variables; (3) the MN titer RSV/A variables; and (4) the MN titer RSV/B variables.

***Objective 3.*** An estimated optimal surrogate is built, by methods in Price et al. (2018), based on each of several sets of variables listed below. Note that for maternal fold-rise markers, three ways of inputting the marker are always considered in the regression models: the log10 fold-rise as a quantitative variable, the indicator of at least a 2-fold rise, and the indicator of at least a 4-fold rise. The latter two are included based on the specification in the RSV-M-301 protocol.

1. Maternal baseline demographic covariates only (reference model 1 – 1 variable set)
2. Delivery/birth baseline demographic covariates only (reference model 2 – 1 variable set)
3. Both sets of baseline variables in 1. and 2. (reference model 3 – 1 variable set)
4. Baseline demographic covariates AND each of 16 variable sets defined by one of the 16 individual markers listed in Table 2 (Whether the maternal baseline, delivery/birth baseline, or both sets of baseline variables are included depends on the time point of the individual marker: for baseline and Day 14 markers only maternal baseline variables are included, whereas for infant birth markers both sets are included.) (16 variable sets)
5. Baseline demographic covariates AND each of the 8 multivariable sets of markers assessed in Objective 2 (one or both baseline variable sets are chosen as in 4.) (8 variable sets)
6. Baseline demographic covariates AND Maternal baseline markers AND one of the following three sets:
    a. Maternal Day 14 markers (all 4 assays) (1 variable set)
    b. Maternal fold-rise markers (all 4 assays) (1 variable set)
    c. Infant birth markers (all 4 assays) (1 variable set)
7. Baseline demographic covariates and all 16 individual marker variables (full model) (1 variable set)

In total, 31 variable sets are studied. The reason to include reference sets 1, 2, and 3 is it allows asking how much incremental improvement in predicting RSV disease is obtained by adding antibody marker variables on top of baseline demographic/exposure covariates. The reason to

include the sets described in 4. and 5. is that it will allow systematic/objective quantification and comparison of classification accuracy of each CoR model studied in Objectives 1 and 2. Sets 6a through 6c allow studying whether and how including baseline (reflecting prior exposure or immunity) antibody titers improve prediction compared to post-baseline antibody markers. These analyses complement other variable sets that are the same except they do not include baseline markers. A reason to include models that do not include baseline variables is that, if best CoR models are found to not include baseline variables, then it could support use of a practical surrogate endpoint that does not require assay measurement of baseline samples. Set 7 is included as the full model that considers all variables together, which serves as another reference model.

***Objective 4 (Exploratory).*** Objective 4 studies the same variables as investigated in Objective 2.

## 7.3 Data Analysis for Objective 1

### 7.3.1 Logistic Regression Modeling

Logistic regression modeling is applied with adjustment for one or both of the RSV disease risk scores developed based on the covariates listed in Table 3 and may also adjust for the number of days from vaccination to birth. Of the two risk scores, the timing of the antibody markers measured for a given analysis determines which are included in the analysis. In particular, the study of maternal baseline and maternal Day 14 antibody markers only include the Maternal Baseline Variable risk score; the time from vaccination to birth will also be included as a covariate so that the comparison between different marker levels is done for mother-infant pairs with similar times from vaccination to birth. Although birth happens at a post-randomization time point, including time from vaccination to birth is justified in part because it was specified in the protocol and the cohort for analysis is live-born infants. The study of infant birth markers includes both risk scores, but not the indicator of time from vaccination to birth. Excluding time from vaccination to birth allows study of the full impact of the infant birth markers because time from vaccination to birth may affect risk by impacting the level of infant birth markers.

The logistic regression models are applied with weighting to account for the two-phase sampling of antibody markers, where the weight for each disease case stratum is 1 and the weight for each un-diseased control stratum is the number of participants in the stratum divided by the number of participants in the stratum that also have MN antibody marker data measured. This inverse probability weighting corrects for bias that can occur from the biased/oversampling of certain RSV disease-free control strata based on their covariate information. It also enables estimation of absolute conditional risk (i.e., estimation of the intercept coefficient in the logistic regression model).

For analyses of the anti-F IgG and PCA antibody titers, ordinary logistic regression modeling is used fit by the *glm* R package, which means that the full data sets are used. For analyses of MN antibody titers, fitting of the logistic regression models accommodates the outcome-dependent stratified biomarker sampling design via maximum likelihood estimation in a logistic regression model (Breslow and Holubkov, 1997) and inverse probability weighted maximum partial likelihood estimation in a Cox proportional hazards model [Lin and Ying (1993)], respectively. A version of the Breslow and Holubkov method designed specifically to accommodate the 2-phase sampling design is applied, as implemented in the R package *osDesign*. While the method actually estimates odds ratios, the low event rate implies the odds ratios closely approximate relative risks. Moreover, the low event rate implies that this dichotomous-endpoint method has negligible power loss compared to a time-to-event method. The advantage of the 2-phase logistic

regression method is that maximum likelihood estimation is fully efficient (providing maximum precision and statistical power in large samples), whereas the inverse probability weighted partial likelihood method used in the Cox proportional hazards model (as implemented in the R package cch with LinYing method) is not efficient (albeit in practice in the rare event setting the methods often provide comparable efficiency).

### 7.3.2 Nonparametric Threshold Modeling

The nonparametric CoR threshold estimation method of Donovan et al. (2019) is applied, using the version accounting for right-censoring of some follow-up times. It will be applied to each of the 16 markers listed in Table 2, separately for the vaccine and placebo groups. The identical method as that used to produce the output of Figure 2 of Donovan et al. (2019) will be used, with results presented in the same way as Figure 2. The method uses the same empirical two-phase sampling estimated weights as used for the logistic regression analyses.

### 7.3.3 Descriptive Plots of Antibody Titers and RSV Disease Endpoint Rates

Boxplots describe the distribution of each of the 16 antibody variables described in Table 2 for subgroups of participants defined by treatment group (vaccine vs. placebo) cross-classified with RSV disease endpoint status (case vs. control).

For each of the 16 antibody variables, estimated cumulative RSV disease endpoint incidence curves over time are plotted for subgroups of vaccine recipients defined by the lower, middle, and upper third of antibody titer response values at Month 13 (Low, Medium, High subgroups), and similarly for the three subgroups of placebo recipients. For each variable, the tertile cut-points are defined based on controls only and pooling over vaccine and placebo recipients, such that the same cut-points are used for analyses of vaccine recipients and analyses of placebo recipients. For the MN titer RSV/A and MN titer RSV/B variables, if supported by the descriptive analyses, common tertile cut-points for the RSV/A and RSV/B antigens are used (averaging across the antigenic specific cut-points on the log10 scale).

The cumulative incidence curves are estimated via the Kaplan-Meier method with inverse probability weighting (empirical weight estimates) that account for the two-phase case-control sampling design.

### 7.3.4 Multiple Hypothesis Testing Adjustment (Objective 1)

For the univariable CoR analyses of the 16 antibody marker variables, family-wise error rate (Holm-Bonferroni) and false-discovery rate (q-values; Benjamini-Hochberg) adjustment will be applied, separately for the vaccine and placebo groups. The multiplicity adjustment is applied separately for each of the three time points maternal baseline, maternal Day 14, and delivery/birth, where the adjustment is over all studied antibody markers at the given time point. The reason for this approach is that assessment of CoRs at each of these time points can be viewed as separate study question. For each analysis the unadjusted p-value, the FWER-adjusted p-value, and the q-value is reported for a Wald test of whether the odds ratio differs from 1. All p-values and q-values are 2-sided. As a guideline for interpreting CoR findings (but not meant to be a rigid gateway), results with unadjusted p-value <= 0.05 and q-value <= 0.10 are flagged as having statistical evidence for being a CoR. Results with FWER-adjusted p-value <= 0.05 are flagged as having more robust statistical evidence for being a CoR. CoP Objectives 2 and 3 listed below will only be done for a given RSV disease endpoint and marker if the marker is demonstrated to be a CoR according to the criterion of unadjusted p-value <= 0.05 and q-value <= 0.10. A potential advantage of this filtering may be reduction in penalty for multiple hypothesis testing adjustment.

## 7.4 Data Analysis for Objective 2

Logistic regression modeling similar to that conducted for Objective 1 is used, except that for analysis of each marker set (1), (2), (3), (4) specified in Objective 2a and of each marker set (1), (2), (3), (4) specified in Objective 2b, all antibody markers specified in the set are included in the logistic regression model fit. A generalized Wald test for the vector of coefficients corresponding to the 4 markers all equaling zero is applied to test for an overall multivariable CoR. Models without any MN antibody markers are fit with *glm* and models with at least one MN antibody marker are fit with *osDesign*. Confounding control is done similarly as for the logistic regression modeling for Objective 1, which means that baseline demographic/putative exposure risk score or scores are adjusted for in the analyses. In addition, given the potential importance of pre-existing maternal exposure to RSV, all analyses except the ones that already include one or more baseline antibody markers are repeated adding in adjustment for baseline antibody titers. Specifically, all multivariable marker analyses adjust for baseline demographic covariates, and all multivariable analyses of post-baseline antibody markers are repeated adjusting for both baseline demographic covariates and baseline antibody markers, where for Objective 2b, this baseline antibody marker is simply the baseline antibody titer of the same kind of marker measured post-baseline. For Objective 2a, ideally the baseline marker for all four assays would be adjusted for, but this cannot be supported by the sample size of the study. Therefore, only the baseline MN titer is adjusted for (an individual's geometric mean of MN RSV/A and MN RSV/B titers).

### 7.4.1 Multiple Hypothesis Testing Adjustment (Objective 2)

Objective 2 fits 16 multivariable logistic regression models (eight antibody marker sets, vaccine and placebo). For the vaccine and placebo groups separately, Holm-Bonferroni FWER-adjusted p-values and Benjamini-Hochberg q-values are reported along with unadjusted p-values for the 4 generalized Wald tests. In addition, for the vaccine and placebo groups separately, Benjamini-Hochberg q-values are reported for the set of 16 individual coefficients for the 16 antibody marker variables. As a guideline for interpreting CoR findings, results with unadjusted p-value <= 0.05 and q-value <= 0.10 are flagged as having statistical evidence for being a CoR. Results with FWER-adjusted p-value <= 0.05 are flagged as having more robust statistical evidence for being a CoR.

## 7.5 Data Analysis for Objective 3

### 7.5.1 Implementation of Superlearner Regression for Calculating the Estimated Optimal Surrogate

The following details are used in the implementation of superlearner:

- The analysis is done separately for the vaccine and placebo arms.
- Pre-scale each input variable to have mean 0 and standard deviation 1 in the treatment arm under study (for all variables including binary, count, and continuous variables).
- Given the limited number of RSV disease cases, only allow learning algorithms to have a maximum of 4 antibody response variables. Also use leave-one-out cross-validation and

negative log-likelihood loss, which some studies have shown tend to perform well in small sample size settings.

- Include learning algorithms with and without lasso pre-screening (with default tuning parameter selection), and with and without logistic regression univariate 2-sided p-value screening (at level $p < 0.10$).
- Include high-correlation variable screening, not allowing any pair of input variables to have Spearman rank correlation $r > 0.9$. When a pair of variables has $r > 0.9$, one of the variables is randomly selected.
- The superlearner is conducted averaging over 10 random seeds, to make results less dependent on random number generator seed.
- All of the learners are implemented with weighting to account for the two-phase sampling design (with weights defined the same as for Objectives 1 and 2).
- Results for comparing classification accuracy of different models are based on point and 95% confidence interval estimates of cross-validated area under the ROC curve (CV-AUC) and difference in CV-AUC as a predictiveness metric (Hubbard et al., 2016; Williamson et al., in press). Results are presented as forest plots of point and 95% confidence interval estimates similar to those used in Neidich et al. (2019, Figure 3) and Magaret et al. (2019). Implementation of this CV-AUC estimation is done using the R package *vimp* available on CRAN, including the same empirical inverse probability weights that are used for other purposes.
- Results are obtained using the Superlearner R package available at CRAN.

Table 4 lists the learning algorithms that are applied to estimate the conditional probability of RSV disease based on the input variable sets considered above. Most of the algorithms are "non-data-adaptive" type learning algorithms, such as parametric regression models (e.g., glms), which are simple and stable and advantageous for an application with limited number of endpoint events. Two more data-adaptive type algorithms are included, for increasing diversity of the types of learners: SL.randomForest (to include a tree-based method) and SL.xgboost. All of the selected learners are coded into the SuperLearner R package.

Table 4: Learning algorithms in the superlearner library of estimators of the conditional probability of RSV disease.

| Algorithm Type | Input Variable Sets (Listed in Section 7.2) | Screens* |
|---|---|---|
| SL.mean (Base model: No covariates) | None | None |
| SL.glm | All 31 variable sets | All, Lasso, univariate logistic regression p-value < 0.10, Low-collinear. |

| | | |
|---|---|---|
| SL.glm.interaction | All 31 variable sets | All, Lasso, univariate logistic regression p-value < 0.10, Low-collinear. |
| SL.bayesglm | All 31 variable sets | All, Lasso, univariate logistic regression p-value < 0.10, Low-collinear. |
| SL.step (stepwise) | All 31 variable sets | Lasso, univariate logistic regression p-value < 0.10, Low-collinear. |
| SL.glmnet (lasso) | All 31 variable sets | Lasso, univariate logistic regression p-value < 0.10, Low-collinear. |
| SL.gam | All 31 variable sets | Lasso, univariate logistic regression p-value < 0.10, Low-collinear. |
| SL.randomForest | All 31 variable sets | Lasso, univariate logistic regression p-value < 0.10, Low-collinear. |
| SL.xgboost | All 31 variable sets | Lasso, univariate logistic regression p-value < 0.10, Low-collinear. |

*All = include all variables; Lasso = include variables with non-zero coefficients in the standard implementation of SL.glmnet that optimizes the lasso tuning parameter via cross-validation; Low-collinearity = do not allow any pairs of quantitative variables with Spearman rank correlation > 0.90; Univariate logistic regression p-value < 0.10 = Wald test 2-sided p-value in a logistic regression model < 0.10.

In order to evaluate the relative performance of the superlearner estimated models for each treatment arm and variable set (31 variable sets defined in Section 7.2), for each learning algorithm specified in Table 4, the CV-AUC is estimated with a 95% confidence interval (Hubbard, Kherad-Pajouh, and van der Laan, 2016). For each of the vaccine and placebo groups, the point and 95% confidence interval estimates of CV-AUC are reported in a forest plot, in two panels (one for the vaccine group, one for the placebo group). These plots provide a way to discern which antibody variables and sampling time points/sources provide the most information in predicting RSV disease.

In addition, for selected variable sets, similar forest plots will be made comparing performance of the various estimated models (e.g., by individual learning algorithm types such as lasso), including discrete superlearner and superlearner models. The plot will be examined to determine which individual learning algorithm types are performing the best. If there is an interpretable

algorithm that has performance close to the best-performing algorithm (which is most likely to be the superlearner), then it will be fit on the entire data set (for a given treatment arm) and the estimated model presented in a table.

For each of the vaccine and placebo groups, cross-validated ROC curves are plotted for the superlearner estimated models for each of the input variable sets. In addition, for each treatment group, boxplots of cross-validated estimated probabilities of RSV disease by case-control status (as estimated from the superlearner models) are plotted.

# 8 CORRELATES OF PROTECTION ASSESSMENT

While multiple formal types of "correlates of protection" have been defined and statistical techniques developed for their assessment [e.g., Qin et al. (2007); Gilbert et al. (2008); Plotkin and Gilbert (2012, 2016)], key concepts relevant to meeting objectives of this SAP and consistent with the above-cited work include: (1) broadly speaking a CoP is a marker that can somehow be used to reliably predict clinical vaccine efficacy (VE) for some population without needing to measure the clinical endpoint; (2) correlates of risk are an important component of developing CoPs, but because they do not assess vaccine effects directly, additional more direct analyses of CoPs are warranted; (3) CoPs can either be non-mechanistic or mechanistic, either way potentially meeting the statistical goal of predicting VE, where this SAP focuses on assessment of CoPs without the ability to empirically distinguish non-mechanistic vs. mechanistic (other research outside of this correlates study would be needed); (4) two formal types of CoPs that we seek to assess are correlates of VE (within the principal stratification statistical framework, Frangakis and Rubin, 2002) and mediators of VE. For examples, correlates of VE were assessed for herpes zoster vaccination (Gilbert et al., 2014), dengue vaccination (Moodie et al., 2018), and influenza vaccination (Gilbert et al., 2019), and mediators of VE were assessed for influenza vaccination (Cowling et al., 2018). While the Prentice framework for surrogate endpoint evaluation has long been used as another approach to evaluation of CoPs, for this SAP we opt to use the mediator of VE framework, which is similar except that it uses more formal assessment of mediation using advances in causal inference since the seminal paper Prentice (1989). An essential concept in Prentice (1989) is that a marker is a valid surrogate endpoint if it fully mediates the vaccine effect on the clinical endpoint, but this paper did not access causal inference language that allowed formal assessment of mediation; thus Cowling et al. (2018) represents one translation of the original Prentice (1989) concept to causal inference.

## 8.1 CoP Objectives

We assess the following three CoP objectives:

1. [Baseline Correlates of VE] To assess each of the four baseline antibody markers as a correlate of VE against RSV disease
2. [Post-baseline Correlates of VE] To assess each qualifying post-baseline antibody marker (among the 12 listed in Table 2) in vaccine recipients as a correlate of VE against RSV disease
3. [Mediators of VE] To assess each of the 12 post-baseline antibody markers listed in Table 2 as a mediator of VE against RSV disease

Section 5.3 and the end of Section 7.3 defines the criterion for a post-baseline antibody marker to qualify for correlate of VE assessment. This criterion is applied to each of the 12 post-baseline antibody markers listed in Table 2, to decide which markers are included for correlates of VE assessment.

In contrast to post-baseline antibody markers, no screening is needed for baseline antibody markers, because baseline markers can be studied as correlates of VE using regression methods under the same assumptions that are made for correlates of risk analyses (not requiring additional assumptions that are needed for post-baseline correlate of VE and mediation analysis).

In addition, no screening is done for deciding whether to conduct mediators of VE analysis.

## 8.2 Correlates of VE Data Analysis

Baseline correlates of VE models will be built by combining the CoR models developed in Objectives 1 and 2 using only the baseline covariates. We will first test interaction between vaccination and each of the 4 baseline antibody titer markers separately, adjusting for the baseline demographic variables. For each baseline marker, if there is evidence of interaction (based on unadjusted p <= 0.10), we will plot point and 95% pointwise confidence interval estimates of VE as a function of the baseline antibody marker, based on the logistic regression model. Thus, a 2-sided interaction p-value <= 0.10 is used as a criterion for baseline marker modification of VE. If there is no evidence of interaction (based on unadjusted p > 0.10), we will report this finding. Secondly, in multivariable logistic regression models including all 4 baseline antibody titer markers, we will explore whether any of four vaccination status × antibody titer variables have a significant interaction (based on unadjusted p <= 0.10). Focusing on a final model including each significant interaction term, we will plot point and 95% pointwise confidence interval estimates of VE as a function of the relevant baseline antibody markers.

Model fitting will be performed using the R osDesign or *glm* as appropriate (the former appropriate if inverse probability weighting is needed) unless the baseline CoVE model contains interactions between treatment and baseline MN marker. In that scenario calibration weighted estimation of the Cox model (Breslow et al. 2009, Fong and Gilbert 2015) may bring enhanced efficiency. For example, if the baseline anti-F IgG titer is not in the baseline CoVE model and can be used to predict the baseline MN level with $R^2 > 0.5$, we will use the sptm package in R to fit a calibration weighted Cox proportion hazards model.

For CoP Objective 2, the vaccine efficacy curve VE($s$) is the vaccine efficacy for subgroups defined by the level $s$ of a given antibody response marker in vaccinees, varying over a range of $s$ values. Methods for estimation and inference of the VE curve are aided by including baseline variables that predict the given antibody response marker in vaccinees and/or by closeout placebo vaccination (Follmann, 2006); both strategies provide means for predicting the counterfactual antibody response marker (if assigned vaccine) for placebo recipients.

We estimate VE($s$) curves (point estimates, 95% pointwise and simultaneous confidence intervals, and 2-sided p-values for whether VE($s$) varies in $s$) using the same Juraska et al. (2018) method used in Moodie et al. (2018) and Gilbert et al. (2019), implemented using the *pssmooth* R package available at CRAN. The same weights are used as in the correlates of risk analyses. Also as for correlates of risk analyses, the method is first applied adjusting for the appropriate

risk score listed in Table 3, where for analysis of Day 14 maternal antibody markers, the Maternal Enrollment Variable risk score is adjusted for as the potential confounder; and for analysis of infant delivery markers, both the Maternal Enrollment Variable risk score and the Delivery/Birth Variable risk score are adjusted for. Then, a second set of analyses are done that also adjust for the same/matched antibody marker measured at baseline. As for the analyses of baseline markers as VE-modifiers, a 2-sided p-value <= 0.10 for VE($s$) varying in $s$ is taken as a criterion for post-baseline marker modification of VE.

### 8.3 Mediator of VE Data Analysis

The method described in Cowling et al. (2018) is applied to assess each of the 12 post-baseline antibody markers listed in Table 2 as mediators of vaccine efficacy. The method is implemented as in Cowling et al. (2018), based on modifying their publicly-supplied R code, with modification that weights are used in all model fits to account for the two-phase sampling design (the same as used for correlates of risk analyses, as described in Section 7.3.1). The method is implemented using as the failure time the days since birth until RSV disease, with right-censoring by the date of the last study visit. Similar to the correlates of VE analyses, the method is first applied adjusting for the appropriate risk score listed in Table 3, and secondly is applied also adjusting for the same/matched antibody marker measured at baseline.

The outputs of the mediation analysis for each antibody marker are point and 95% confidence interval estimates of the total effect (equivalent to VE), the direct effect (the vaccine effect on RSV disease if the vaccine effect on the antibody response is deactivated), the indirect effect (the vaccine effect on RSV disease via other immunological variables/mechanisms not captured by the antibody marker under investigation), and the proportion mediated by the antibody marker.

In addition, the mediation analyses will be applied to a selected set of estimated optimal surrogates for vaccine recipients developed through Objective 3, to estimate the proportion of VE mediated by the estimated optimal surrogates. These analyses need to be interpreted with caution given that the estimated optimal surrogates are selected as best synthesis variables for predicting RSV disease; this prior modeling to select the marker can potentially bias inferences.

For both the correlate of VE and mediator of VE analyses, the same multiplicity adjustments as used in the correlates of risk analyses (described in Section 7.3.4) are used.

## 9 REFERENCES


1. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J Roy Stat Soc B Met 1995;57:289-300.
2. Breslow NE, Holubkov R. Maximum likelihood estimation for logistic regression parameters under two-phase, outcome dependent sampling. Journal of the Royal Statistical Society Series B, Statistical methodology 1997; 59:447-61.
3. Breslow NE, Lumley T, Ballantyne CM, Chambless LE, Kulich M. Using the whole cohort in the analysis of case-cohort data. American journal of epidemiology. 2009 Apr 8;169(11):1398-405.
4. Cowling BJ, Lim WW, Perera RA, Fang VJ, Leung GM, Peiris JM, Tchetgen Tchetgen EJ. Influenza hemagglutination-inhibition antibody titer as a mediator of vaccine-induced protection for influenza B. Clinical Infectious Diseases. 2018 Sep 8;68(10):1713-7.

5. Donovan K, Hudgens M, Gilbert PB. Nonparametric inference for immune response thresholds of risk in vaccine studies. Annals of Applied Statistics. 2019 13(2):1147−1165.
6. Follmann D. Augmented designs to assess immune response in vaccine trials. Biometrics. 2006 Dec;62(4):1161-9.
7. Frangakis CE, Rubin DB. Principal stratification in causal inference. Biometrics 2002 58(1):21–29.
8. Fong Y, Gilbert P. Calibration weighted estimation of semiparametric transformation models for two-phase sampling. Statistics in medicine. 2015 May 10;34(10):1695-707.
9. Gilbert PB, Fong Y, Carpp LN, Monto AS, Martin ET, Petrie JG. HAI and NAI titer correlates of inactivated and live attenuated influenza vaccine efficacy. BMC Infectious Diseases. 2019 May 22;19(1):453. doi: 10.1186/s12879-019-4049-5.
10. Gilbert PB, Gabriel EE, Miao X, Li X, Su SC, Parrino J, Chan IS. Fold rise in antibody titers by gpELISA is an excellent correlate of protection for a herpes zoster vaccine, demonstrated via the vaccine efficacy curve. J. Infect. Dis. 2014;210:1573-81.
11. Gilbert PB, Qin L, Self SG. Evaluating a surrogate endpoint at three levels, with application to vaccine development. Statistics in Medicine 2008; 27(23):4758-78.
12. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 1979;6:65-70.
13. Hubbard AE, Kherad-Pajouh S, van der Laan MJ. Statistical inference for data adaptive target parameters. The international journal of biostatistics. 2016 May 1;12(1):3-19.
14. Juraska MJ, Huang Y, Gilbert PB. Inference on treatment effect modification by biomarker response in a three-phase sampling design. Biostatistics. 2018; doi:10.1093/biostatistics/kxy074
15. Lin, DY and Ying, Z. Cox regression with incomplete covariate measurements. Journal of the American Statistical Association 1993; 88: 1341–1349.
16. Liu Q, Shepherd BE, Wanga V, Chun L. Covariate-adjusted Spearman's rank correlation with probability-scale residuals. Doctoral dissertation, Vanderbilt University, implemented in the R package PResiduals.
17. Magaret CA, Benkeser D, Williamson BD, Borate BR, Carpp LN, Georgiev IS, Setliff I, Dingens AS, Simon N, Carone M, Simpkins C, Montefiori DC, Alter G, Yu W-H, Juraska MJ, Edlefsen PT, Karuna S, Mgodi NM, Edugupanti S, Gilbert PB. Prediction of VRC01 neutralization sensitivity by HIV-1 gp160 sequence features. PLoS Computational Biology. 2019 Apr 1;15(4):e1006952.
18. Moodie Z, Juraska MJ, Huang Y, Zhuang Y, Fong Y, Carpp LN, Self SG, Chambonneau L, Small R, Jackson N, Noriega F, Gilbert PB . Neutralizing antibody correlates analysis of tetravalent dengue vaccine efficacy trials in Asia and Latin America. Journal of Infectious Diseases. 2018 Feb 14;217(5):742-753. doi: 10.1093/infdis/jix609.
19. Neidich SD, Fong Y, Li S, Geraghty D, Williamson BD*, Young WC, Goodman D, Seaton KE, Shen X, Sawant S, Zhang L, deCamp A, Blette BS, Shao M, Yates NL, Feely F, Pyo C-W, Ferrari G, Frank I, Karuna ST, Swann E, Mascola JM, Graham BS, Hammer SM, Sobieszczyk M, Corey L, Janes HE, McElrath MJ, Gottardo R, Gilbert PB, Tomaras GD. Antibody Fc effector functions and IgG3 associate with decreased HIV-1 risk. Journal of Clinical Investigation. 2019;10-1172.
20. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. Statistics in medicine. 1989 Apr;8(4):431-40.
21. Price BL, Gilbert PB, van der Laan MJ. Estimation of the optimal surrogate based on a randomized trial. Biometrics. 2018 74(4), 1271-1281.
22. Qin L, Gilbert PB, Corey L, McElrath MJ, Self SG. A framework for assessing immunological correlates of protection in vaccine trials. J Infect Dis. 2007;196(9):1304-12.
23. Williamson B, Gilbert PB, Simon N, Carone M. Nonparametric variable importance assessment using machine learning techniques. Biometrics, in press.

## 10 SUMMARY OF CHANGES

| Date | Version | Summary of Changes |
|---|---|---|
| 30-Apr-2020 | 2.0 | • Revised composite endpoint definition to exclude cough to better align with case definition used for sampling. |
| 10-Apr-2020 | 1.0 | • New |

# 11 APPENDIX A

This appendix presents analyses supporting the selection of the baseline stratification variables used for defining the two-phase sampling design. The stratification variables are partly chosen based on association with the estimated VE (1-OR). This issue is studied in Tables A1 through A5 below.

- Different control populations. VE appears similar (Table A1).
- Season. VE appears similar (Table A2).
- Location. VE appears *different* (Table A3).
- PP (per-protocol) status. VE appears similar (Table A4).
- Time from vaccination to Birth >= 30 Days. VE appears *different* (Table A5).

Table A6 lists participant counts by RSV season and location, within each treatment arm. All tables are based on the Intent-To-Treat (ITT) efficacy population.

Table A1: Odds for different control subpopulations are different, but odds ratios (OR) between vaccine and placebo are similar.

| | no_swap | swab_rsvneg | swab_rsvpos |
|---|---|---|---|
| vacc | 0.13 | 0.06 | 0.26 |
| plac | 0.24 | 0.12 | 0.5 |
| OR | 0.55 | 0.48 | 0.53 |

Table A2: Odds for different seasons (NN location only) are different, but odds ratios (OR) between vaccine and placebo are similar.

| | S1 | s2 | s3 |
|---|---|---|---|
| vacc | 0.031 | 0.022 | 0.043 |
| plac | 0 | 0.041 | 0.073 |
| OR | | 0.546 | 0.584 |

Table A3: Odds for different locations (season 3 only) are somewhat different. For example, Barnard's two-sided test for comparing the case:control distribution between PP and SS in the placebo arm (1,7,47,54) is 0.06.

| | NN | PP | QQ | RR | SS |
|---|---|---|---|---|---|
| vacc case | 17 | 7 | 10 | 2 | 2 |
| vacc control | 399 | 113 | 191 | 128 | 95 |
| vacc odd | 0.043 | 0.062 | 0.052 | 0.016 | 0.021 |
| plac case | 14 | 7 | 7 | 6 | 1 |
| plac control | 192 | 54 | 90 | 60 | 47 |
| plac odd | 0.073 | 0.13 | 0.078 | 0.1 | 0.021 |
| OR | 0.584 | 0.478 | 0.673 | 0.156 | 0.989 |

Table A4: Attack rates (for any of the 3 efficacy endpoints) are similar between PP and not PP in both vaccine and placebo.

| | Placebo | | | vaccine | | | VE |
|---|---|---|---|---|---|---|---|
| | Control | Case | attack rate | Control | Case | attack rate | |
| PP = No | 62 | 5 | 0.075 | 121 | 4 | 0.032 | 0.57 |
| PP = Yes | 686 | 46 | 0.063 | 1399 | 48 | 0.033 | 0.48 |

Table A5: Attack rates are different by time from vaccination to birth in the vaccine group but not in the placebo group.

| | Placebo | | | vaccine | | | VE |
|---|---|---|---|---|---|---|---|
| | Control | Case | attack rate | Control | Case | attack rate | |
| Vax to Birth <= 30 Days | 91 | 7 | 0.071 | 174 | 9 | 0.049 | 0.31 |
| Vax to Birth > 30 Days | 657 | 44 | 0.063 | 1346 | 43 | 0.031 | 0.51 |

Table A6: Participant counts by RSV season and location.

| | | Placebo | | | Vaccine | | |
|---|---|---|---|---|---|---|---|
| | region | s1 | s2 | s3 | s1 | s2 | s3 |
| Case# | NN | 0 | 8 | 14 | 1 | 9 | 17 |
| | PP | 2 | 3 | 7 | 0 | 2 | 7 |
| | QQ | 0 | 1 | 7 | 0 | 0 | 10 |
| | RR | 0 | 2 | 6 | 0 | 2 | 2 |
| | SS | 0 | 0 | 1 | 0 | 0 | 2 |
| Control# | NN | 31 | 197 | 192 | 32 | 406 | 399 |
| | PP | 10 | 27 | 54 | 9 | 63 | 113 |
| | QQ | 0 | 8 | 90 | 0 | 19 | 191 |
| | RR | 0 | 24 | 60 | 0 | 49 | 128 |
| | SS | 0 | 8 | 47 | 0 | 16 | 95 |
| Odds | NN | 0 | 0.041 | 0.073 | 0.031 | 0.022 | 0.043 |
| | PP | 0.2 | 0.111 | 0.13 | 0 | 0.032 | 0.062 |
| | QQ | | 0.125 | 0.078 | | 0 | 0.052 |
| | RR | | 0.083 | 0.1 | | 0.041 | 0.016 |
| | SS | | 0 | 0.021 | | 0 | 0.021 |
| VE (1-OR) | NN | | | | | 0.45 | 0.42 |
| | PP | | | | 1 | 0.71 | 0.52 |
| | QQ | | | | | 1 | 0.33 |
| | RR | | | | | 0.51 | 0.84 |
| | SS | | | | | | 0.01 |

Palivizumab Competitive ELISA (PCA) measurements suggest different levels of antibody response exist between locations (results not shown).

Table A7: Breastfeeding Status Summary (These data are not used because breastfeeding may be in the causal pathway of the vaccine effect on RSV disease.)

| Country Group | Cat | D+14 | D+35 | D+60 | D+90 | D+112 | D+120 | D+180 |
|---|---|---|---|---|---|---|---|---|
| World | 1 | 3123 | 2738 | 2476 | 2160 | 74 | 1776 | 912 |
| World | 2 | 902 | 1074 | 1119 | 1207 | 43 | 1128 | 1055 |
| World | 3 | 397 | 604 | 802 | 1001 | 51 | 1255 | 2379 |
| US | 1 | 450 | 344 | 298 | 236 | 7 | 201 | 109 |
| US | 2 | 346 | 340 | 298 | 296 | 4 | 279 | 292 |
| US | 3 | 247 | 361 | 441 | 488 | 9 | 514 | 610 |
| ZA | 1 | 1884 | 1651 | 1472 | 1261 | 42 | 979 | 468 |
| ZA | 2 | 280 | 436 | 520 | 601 | 21 | 565 | 380 |
| ZA | 3 | 118 | 184 | 275 | 403 | 18 | 617 | 1418 |
| ROW | 1 | 789 | 743 | 706 | 663 | 25 | 596 | 335 |
| ROW | 2 | 276 | 298 | 301 | 310 | 18 | 284 | 383 |
| ROW | 3 | 32 | 59 | 86 | 110 | 24 | 124 | 351 |

cat=1: fed breast milk from mother only

cat=2: fed breast milk from mother and other food

cat=3: fed not from breast milk from mother