STATISTICAL ANALYSIS PLAN

Assessment of Correlates of Risk and Protection in the Phase 3 CYD14 and CYD15 Dengue Vaccine Efficacy Trials

Date:
17 September 2014
SAP version 1

Prepared by

Peter Gilbert, Zoe Moodie, Michal Juraska, Ying Huang

SAP Modification History

The version history of and modifications to this statistical analysis plan are described below.

Date: 17 September, 2014
SAP version: 1
Modifications:

Minor editorial updates were made in February of 2016 prior to submission.

Table of Contents

1 OVERVIEW

This statistical analysis plan focuses on the assessment of neutralizing antibody titer readouts measured at Month 13 as correlates of risk and as correlates of protection in CYD14 and CYD15.

2 Case-Cohort Sampling of Immune Responses

The sampling plans for CYD14 and CYD15 are described in the protocols and in the primary publications of these studies (Capeding et al, 2014; Villar et al., 2015). The "immunogenicity subset" refers to the participants that had immune responses measured at months 0, 7, 13, and 25. Immune responses were also measured from all participants that experienced the primary dengue endpoint after Month 13 and up to Month 25, the end of the active surveillance phase.

3 COHORT DEFINITIONS

- **Month 13 cases:** Post-Month 13 cases that have not experienced a primary dengue endpoint (DENV-ANY) before Month 13 and have a primary dengue endpoint between Month 13 and the end of the Active phase (i.e., up to Month 25) with Month 13 samples available
- **Month 13 immunogenicity subset cases:** The subset of Month 13 cases in the immunogenicity subset
- **Month 13 controls:** Participants never registering a primary study endpoint (DENV-ANY) during the Active Phase and at risk at Month 13 (i.e., date of last contact occurred after Month 13)
- **Month 13 immunogenicity subset controls:** The subset of Month 13 controls in the immunogenicity subset

* Controls for the serotype-specific analysis also include those infected with another serotype between Month 13 and 25 as they are considered at risk for infection with the serotype being analyzed.

We refer to the vaccine cases and sampled controls as the "vaccine case-cohort" and to the placebo cases and sampled controls as the "placebo case-cohort."

4 Immune Response Variables and Time Points

The analyses assess the four dengue serotype-specific neutralizing antibody titers, plus the area under the magnitude-breadth (AUC-MB) curve (Huang et al., 2009) based on the four serotypes, as correlates of risk and correlates of protection. The neutralizing antibody titers are continuous values analyzed on a $\log_{10}$ scale. For each serotype DENV-1, DENV-2, DENV-3, DENV-4, neutralizing antibody titer is defined via the plaque reduction neutralization test assay as the highest dilution of serum at which $\geq 50\%$ of dengue challenge virus in plaque counts was neutralized compared to the negative control wells, which represent a virus load of 100%. The lower limit of quantitation is $\log_{10}(10)$ and values below this limit are set to $\log_{10}(5)$. The AUC-MB is the geometric mean of the available serotype-specific neutralizing antibody titers and is referred to as the "average titer". The average titer provides a single summary measure for the assessment of correlates of risk and protection against the primary dengue endpoint (DENV-ANY). These analyses have greater precision than the serotype-specific dengue endpoint analyses due to the larger number of available dengue endpoints and decreased technical measurement error. Precision is improved by using the average of the four correlated neutralizing antibody titer variables, instead of a single titer variable. Principal components analyses will be performed to compare the variability explained by the primary principal components relative to the variability explained by the average titer and by each of the four individual titer variables.

In addition, baseline-subtracted versions of the above five variables may be assessed as correlates of risk and protection in the CYD14 and CYD15 trials where the immunogenicity subsets may be large enough to permit this (baseline samples are only available for the immunogenicity subsets).

The primary immunogenicity time-point for measuring dengue neutralization titers is Month 13 (Visit 10, 28 days after the third vaccination).

5 Correlates of Risk Assessment

5.1 Primary Objectives

1. To assess in vaccine and placebo recipients the average titer measured at Month 13 as a predictor of the rate of the subsequent overall dengue primary endpoint (DENV-ANY).
2. To assess in vaccine and placebo recipients the four dengue serotype-specific neutralizing antibody titers measured at Month 13 as predictors of the rate of the subsequent homologous serotype-specific dengue endpoint (DENV-1, DENV-2, DENV-3, DENV-4).

5.2 Secondary Objectives

1. To identify models based on the four dengue serotype-specific neutralization variables and the average titer measured at Month 13 that best-classify vaccine and placebo recipients by status of whether they subsequently experience the overall dengue primary endpoint (DENV-ANY)

2. To assess whether and how baseline characteristics (including demographics and where available, baseline neutralization variables) modify the correlates of risk in vaccine and placebo recipients.

5.3 Descriptive Plots of Neutralization Titers and Dengue Endpoint Rates

Boxplots describe the distribution of each immune variable for subgroups of participants defined by treatment group (vaccine vs. placebo) cross-classified with primary dengue endpoint status.

Cumulative dengue incidence curves are plotted for the three subgroups of vaccine recipients defined by the lower, middle, and upper third of neutralization response values at Month 13 (Low, Medium, High subgroups), and similarly for the three subgroups of placebo recipients defined by the response value tertiles at Month 13. (The tertiles are defined using common cut-points for the vaccine and placebo groups, based on treatment arm pooled data. In addition, the cut-points are defined using inverse probability weighting, in order that the tertiles are defined for the population of participants at-risk at the Month 13 visit.) Depending on the rates of neutralization above the lower limit of quantification, these subgroup categories may be modified to Positive vs. Negative response or Negative response vs. the lower and upper half of readouts among positive responders. Depending on descriptive analyses, the Low, Medium, and High categories may be defined differently than these default choices; however, in all instances the categories are determined without respect to case/control status and are fixed before conducting any descriptive or inferential analyses that are unblinded to case/control status.

The plots are done for the primary endpoint regardless of dengue serotype (DENV-ANY) as well as for the primary endpoint with each of the four dengue serotypes separately. The curves are estimated via the Kaplan-Meier method with inverse probability weighting that accounts for the case-cohort sampling design.

5.4 Relative Risk Estimation

Logistic regression and Cox proportional hazards models are used for evaluating whether and how the Month 13 immune response variables affect the subsequent rate of the primary dengue endpoint overall and for each of the serotypes. The fitting methods accommodate 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 (i.e., that assesses the time from the Month 13 sampling date to the date of dengue endpoint diagnosis). 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).

Two regression methods are used to assess consistency of results, with relative advantages of theoretical efficiency and accommodation of the failure time, respectively. The right-censored failure time is defined as the time between the Month 13 (Visit 10) sample date and the date of diagnosis of the primary dengue endpoint or the last contact date, as defined in the primary SAP.

All regression analyses adjust for treatment, protocol-specified categories (2−5, 6−11, 12−14 years in CYD14 and 9−11, 12−16 years in CYD15), gender and country. It the models do not support this many adjustment factors, then only the variables significantly associated with the dengue endpoint under consideration will be retained. If any baseline covariates measured from the whole cohort predict dengue serotype-specific neutralization titers reasonably well (e.g., $R^2 > 0.50$), then for univariate models of those titers, the Cox model will be fit using the more-efficient method of Breslow et al. (2009a, 2009b) instead of the Lin and Ying estimator. The correlates of risk analysis will be conducted by pooling over vaccine and placebo groups and considering interaction term(s) between treatment and the neutralization titer variable(s).

Given lack of knowledge about what threshold levels of neutralization response could be predictive of the dengue endpoints, in exploratory analyses the method of Fong et al. (2015) will be used for hypothesis testing of whether dichotomized neutralization responses predict risk for any possible threshold value. These methods may provide insights on "change-point" thresholds that are predictive of reduced dengue risk.

5.5 Univariate and Multivariate Models

The analysis applies the models with the five dengue neutralization immune variables (four serotypes plus average titer) entered as quantitative readouts or as categorical variables (e.g., Low, Medium, High) as defined above. In the quantitative variable analyses, values below the lower quantification limit of the assay are set to half the quantification limit. For the primary dengue endpoint, models are fit that include all four serotype-specific variables simultaneously and that include each of the variables separately, considered without and with baseline subtraction where possible (i.e., for those in the immunogenicity subset who have baseline data available). In the multivariable models a generalized Wald statistic is used to test the overall null hypothesis that none of the variables associate with the overall primary endpoint.

For each serotype-specific dengue endpoint, the models will be fit with only the serotype-specific neutralization variable (without or with baseline subtraction where possible) included in the model (primary objective 2).

P-values for testing each serotype-specific titer were adjusted over the four serotypes using family-wise error rate (Holm-Bonferroni) and false-discovery rate (q-values; Benjamini-Hochberg) adjustment, separately for the vaccine and placebo groups. All p-values and q-values are 2-sided.

5.6 Missing immunological measurements

The following methods are used if (and only if) more than 10% of participants in the vaccine or placebo case-cohort are missing any immune response data and will be applied on final models, after model selection is done based on the "complete-case" analysis, whereby participants missing relevant immunological variables are excluded from the analyses. When the probability of missing immunological measurements depends on either the unobserved missing value or other unobserved covariates or observed covariates not included in the modeling, ignoring these missing data in the methods described below could yield biased results. Multiple imputation is used to fill in the missing data. The Mice package in R is used to perform linear-regression-based imputation with outcome status and any participant variables predictive of the immune response as independent variables. A total of 20 imputed datasets are generated and results are combined across imputed datasets using standard multiple imputation rules (Rubin, 1996).

6 Correlates of Protection Assessment

6.1 Primary Objectives

1. To assess the average titer measured at Month 13 as a predictor of the level of VE against the overall dengue primary endpoint (DENV-ANY).
2. To assess each of the four dengue serotype-specific neutralizing antibody titers measured at Month 13 as a predictor of the level of VE against the homologous serotype-specific dengue endpoint (DENV-1, DENV-2, DENV-3, DENV-4).

6.2 Secondary Objectives

3. To identify models based on the four dengue serotype-specific neutralization variables and the average titer measured at Month 13 that provide the most accurate models for predicting VE against the primary dengue endpoint (DENV-ANY).
4. To assess whether and how baseline characteristics (including demographics and where available, baseline neutralization variables) modify the correlates of protection.

Without trial design augmentations, efficacy trials permit assessment of correlates of risk (CoRs), but additional assumptions are needed to infer that a CoR is also a CoP (a CoP is a reliable statistical predictor of VE) (Qin et al., 2007; Gilbert, Qin, and Self, 2008). CoPs will be assessed via the principal stratification "vaccine efficacy curve" approach.

6.3 Vaccine efficacy curve approach

Statistical methods have been developed to directly assess CoPs via estimation of the vaccine efficacy curve (several publications including Huang and Gilbert (2011), Huang, Gilbert, and Wolfson, 2013, and Gabriel and Gilbert (2014)). The vaccine efficacy curve VE(s) is interpreted as the vaccine efficacy for subgroups with level s of the vaccine-induced immune response under consideration. Methods for inference about the VE curve are aided by either a baseline immunogenicity predictor (BIP) or closeout placebo vaccination (CPV) to predict the Month 13 dengue neutralization titers of placebo recipients. An exploratory analysis will be conducted to develop the best baseline variable or set of variables that predict the dengue neutralization titers. Based on the best available BIP model, the methods of Juraska and Gilbert (2015) and of Huang, Gilbert, and Wolfson (2013) will be applied to estimate the vaccine efficacy curve. The curve will be estimated separately for each of the neutralization variables, treated as a quantitative variable as considered above. Each of the estimated vaccine efficacy curves will be plotted together with 95% confidence intervals, and the testing procedures of Juraska and Gilbert (2015) and of Huang, Gilbert, and Wolfson (2013) will be applied to assess whether the vaccine efficacy curve changes with the immune response (this tests for some value as a CoP). The estimated vaccine efficacy curves will be examined to assess if the data support thresholds of protection. In addition, the procedures of Huang, Gilbert, and Wolfson (2013) will be used to assess if baseline-subtracted readouts provide superior CoPs to unsubtracted readouts. The method of Gabriel and Gilbert (2014), which is similar to Huang and Gilbert (2011) except it account for a right-censored time-to-event outcome via Weibull regression modeling, may also be used if the data support that vaccine efficacy and/or the potential correlation of immune responses with outcomes changes with time.

The same multiplicity adjustments as used in the correlates of risk analyses (described in Section 5.5) are used for the correlates of protection analyses.

7 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 N, Lumley T, Ballantyne C, Chambless L, Kulich M. Improved Horvitz-Thompson estimation of model parameters from two-phase stratified samples: Applications in Epidemiology. Statistical Biosciences 2009a; 1:32-49.
4. Breslow N, Lumley T, Ballantyne C, Chambless L, Kulich M. Using the whole cohort in the analysis of case-cohort data. American Journal of Epidemiology 2009b;169: 1398-1405.
5. Capeding MR, Tran NH, Hadinegoro SR, et al. Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. Lancet 2014;384:1358-65.
6. Fong Y, Di Chongzhi D, Permar S. Change-point testing in logistic regression models with interaction term. Statistics in Medicine 2015; 34(9):1483-94.
7. Fong Y, Huang Y, Gilbert PB. Estimation and hypothesis testing under four types of change point models using chnpt. Department of Biostatistics, University of Washington Technical Report 2015.
8. Gabriel E, Gilbert PB. Evaluating principle surrogate endpoints with time-to-event data accounting for time-varying treatment efficacy. Biostatistics. 2014;15(2):251-65.
9. 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.
10. Holm S. A Simple Sequentially Rejective Multiple Test Procedure. Scandinavian Journal of Statistics 1979;6:65-70.
11. Huang Y, Gilbert PB. Comparing biomarkers as principal surrogate endpoints. Biometrics. 2011;67(4):1442-51. PMCID: 3163011.
12. Huang Y, Gilbert PB, Wolfson J. Design and estimation for evaluating principal surrogate markers in vaccine trials. Biometrics 2013; 69:301-309.
13. Huang Y, Gilbert PB, Montefiori DC, Self SG. Simultaneous evaluation of the magnitude and breadth of a left and right censored multivariate response, with application to HIV vaccine development. Statistics in Biopharmaceutical Research 2009;1(1):81-91.
14. Juraska M, Gilbert PB. Nonparametric smoothing method for assessing effect modification by a post-randomization biomarker with application to vaccine efficacy trials. Technical report, Fred Hutch 2015.
15. Lin, DY and Ying, Z. Cox regression with incomplete covariate measurements. Journal of the American Statistical Association 1993; 88: 1341–1349.
16. 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.
17. Rubin DB. Multiple imputation after 18+ years. Journal of the American Statistical Association 1996;434:473-89.
18. Villar L, Dayan GH, Arredondo-Garcia JL, et al. Efficacy of a tetravalent dengue vaccine in children in Latin America. N Engl J Med 2015;372:113-23.