STATISTICAL ANALYSIS PLAN

Study Title:	Comparison of High vs. Standard Dose Flu Vaccine in Pediatric Stem Cell Transplant Recipients
Name of Test Drug:	Fluzone Quadrivalent® and Fluzone® HD-TIV
Study Number:	DMID Protocol Number 16-0117
Protocol Version (Date):	Version 8.0 (12 September 2019)
Analysis Plan Version:	Version 1.0
Analysis Plan Date:	24 November 2020
Analysis Plan Author(s):	

1. INTRODUCTION

1.1 Overview

This phase II immunogenicity and safety trial compares a high-dose trivalent inactivated influenza vaccine (HD-TIV) to standard-dose quadrivalent inactivated influenza vaccine (SD-QIV) in pediatric hematopoietic stem cell transplant (HSCT) recipients. Participants were randomized to receive either two doses of 0.5 mL HD-TIV (60µg of each influenza antigen) or two doses of 0.5 mL SD-QIV (15µg of each influenza antigen) of the season-specific vaccine 28-42 days apart. This statistical analysis plan (SAP) describes the methods for the final analysis.

1.1.1 Study sites

Study subjects were recruited from the Outpatient Pediatric Oncology/HSCT Clinics or Clinical Trial Centers at the sites listed in Table 1.

1.1.2 Schedule of assessments

Study subjects were evaluated according to the following general schedule of assessments.

a. Screening: Subjects were screened by history and physical exam to ensure all eligibility criteria were met.
b. Study Visit 1: First blood sample collected; first vaccine administered.
c. Study Visit 2 (28-42 days following Study Visit 1): Second blood sample collected; second vaccine administered; visit window was 28-42 days following Visit 1.
d. Study Visit 3 (28-42 days following Study Visit 2): Third blood sample collected. Visit window was 28-42 days after administration of the second vaccine dose.
e. Study Visit 4 (180±42 days following Study Visit 2): Fourth blood sample collected. For the 2016/2017 enrollment season, visit window was 180±28 days following Visit 3. For the 2019/2020 enrollment season, this visit was optional.

1.1.3 Repeaters

A subset of subjects enrolled and vaccinated in 2016-17, 2017-18, or 2018-19 were re-enrolled for a maximum of one subsequent year. For example, subjects enrolled in 2016-17 could re-enroll in 2017-18 as repeaters. Subjects could only enroll as a repeater one time, the year following their original enrollment, and must have received at least one vaccine to be eligible as a repeater. Subjects enrolled as repeaters received two doses (separated by 28-42 days) of the same vaccine formulation (HD-TIV or SD-QIV) that they were randomized to receive in their first year of enrollment, again with investigators blinded to vaccine formulation.

1.1.4 Randomization

Randomization to HD-TIV or SD-QIV was frequency-matched within each site to ensure balance between the groups and was implemented using the Randomization Module within Vanderbilt's REDCap database. For subjects $\geq$ 12 months post-transplant, randomization was stratified on

treatment for graft versus host disease (GVHD) with systemic steroids. For subjects < 12 months post-transplant, randomization was stratified on meeting any of the following: (a) treatment for GVHD with systemic steroids, (b) history of alemtuzumab, (c) history of anti-thymocyte globulin, (d) receipt of a cord graft, (e) haploidentical transplant, or (f) post-transplant cyclophosphamide. Repeaters retained randomization from initial enrollment and were *not* re-randomized.

1.2 Study objectives

Immunogenicity and safety objectives are organized into subsections below. To avoid repetitive language, *all* objectives apply to the target study population (namely, pediatric HSCT recipients).

1.2.1 Objectives for HAI immunogenicity outcomes

The primary objective is to determine whether two doses of HD-TIV is associated with a higher geometric mean HAI titer to either influenza A antigen (H1N1 and H3N2, evaluated separately) compared to two doses of SD-QIV.

The secondary objectives pertaining to HAI immunogenicity outcomes are as follows:

- To determine whether two doses of HD-TIV is associated with a higher geometric mean HAI titer to influenza B antigens as compared to two doses of SD-QIV.
- To determine whether two doses of HD-TIV is associated with higher odds of seroconversion or seroprotection in *any* HAI antigen as compared to two doses of SD-QIV.
- To compare HAI immunogenicity for each antigen between the following groups: (a) one dose of HD-TIV vs. one dose of SD-QIV, (b) one dose of HD-TIV vs. two doses of HD-TIV, and (c) one dose of SD-QIV vs. two doses of SD-QIV.

The tertiary objectives pertaining to HAI immunogenicity outcomes are as follows:

- To compare HAI immunogenicity for each antigen between the one dose of HD-TIV and two doses of SD-QIV.
- To determine whether time post-transplant, subject age, or GVHD status (presence/absence) modifies the association between vaccination group and HAI response.
- To characterize the relationship between HAI titers, ex-vivo T and B cell phenotype, and ex-vivo influenza-specific T and B cell response in subjects receiving either HD-TIV or SD-QIV.
- To compare HAI immunogenicity outcomes between treatment groups (one dose or two doses of HD-TIV vs. one dose or two doses of SD-QIV, respectively) among those enrolled for a second year (repeaters).

1.2.2 Objectives for safety outcomes

The primary objectives for safety outcomes are to compare the frequency/severity of solicited local injection site adverse events (AEs) and systemic AEs between HD-TIV and SD-QIV following the first dose and following the second dose. The secondary objective is to compare safety outcomes between treatment groups among those enrolled for a second consecutive year (repeaters).

1.2.3 Further exploratory objectives

- To compare the persistence of HAI and microneutralization (MN) titers between HD-TIV and SD-QIV across all four antigens seven months after the final vaccine dose.
- To compare MN and NAI responses between subjects receiving HD-TIV and SD-QIV in the first year.
- To compare MN and NAI titers between treatment groups (two doses of HD-TIV vs. two doses of SD-QIV) among those enrolled for a second year, and compare them to HAI responses.
- To compare influenza virus detection during influenza season between treatment groups (two doses of HD-TIV vs. two doses of SD-QIV).
- To evaluate the correlation between HAI and baseline MN responses.

1.3 Planned analyses

Unblinded analyses corresponding to the primary and secondary objectives will be performed by the investigative team only after data are available from all participants and the SAP is finalized.

2. GENERAL CONSIDERATIONS FOR DATA ANALYSES

2.1 Estimands of interest

All primary analyses will be intention-to-treat (ITT) in nature, maintaining fidelity to randomized group. Per-protocol analyses may be performed as a sensitivity analysis.

2.2 Analysis sets

We characterize different study subsets for purposes of analysis in Table 2.

2.3 Missing data

- Subject visits might not occur on protocol-specified days. We will characterize the extent to which visit dates were faithful to the pre-specified protocol windows; for the purpose of analysis, observations will mapped to their "nearest" protocol-defined study visit window within a tolerance of at most ten days for Visits 1, 2, and 3. The study day as defined in Section 1.1.2 will be used when data are summarized by visit.
- A missing pre-treatment laboratory result would be treated as normal (i.e., no toxicity grade) for the laboratory abnormality summary.
- To account for other missing data (e.g., in immunogenicity data), we will employ multiple imputation with chained equations using at least fifty imputation iterations, aggregating via Rubin's rules; we will also compare these results to those of a complete-case analysis (the latter of which being a secondary approach for the purposes of a sensitivity analysis).

2.4 Robust methods

Where applicable, we will employ robust standard errors to allow model misspecification. Similarly, use of the *t*-test will allow for unequal variances between groups. Wald-based

confidence intervals and *p*-values will be formed on the basis of robust variance estimates.

2.5 Hypothesis testing

All tests will be conducted with a nominal significance level of $\alpha = 0.05$ (two-sided).

2.6 Outliers

Outliers will be identified during the data management and data analysis process for the purposes of error detection and reconciliation; all data deemed to be accurate will be included in the analyses. Sensitivity analyses may be performed to evaluate the impact of outliers.

2.7 Data transformations and lower limits of detection

Laboratory data (in particular, HAI titer values) with clearly defined lower and upper limits of detection (L and U, respectively) will be handled as follows:

1. Observations not meeting the lower limit of detection will be given a value of half the lower limit of detection [e.g., those not responding to a titer of 1:L will be given a titer of 1:(L/2)].
2. Observations meeting the upper limit of detection will be given a value of the upper limit of detection (e.g., those responding to a titer of 1:U will be given a titer of 1:U) in the primary analysis. A range of subsequent sensitivity analyses will be conducted to evaluate sensitivity to departures from this convention.

2.8 Software

All analyses will be performed in either R v. 4.0.2 or later, or Stata v. 15 or later (College Station, TX: StataCorp LLC).

3. SUBJECT DISPOSITION

These summary metrics are descriptive in nature (not inferential). These measures will be reported both by treatment group (HD-TIV and SD-QIV) and overall.

- Absolute and relative frequency of participants randomized at each investigator site.
- The summary of subject disposition for all screened participants. This summary will include the number of participants screened, screen failure participants who were not randomized, participants who met all eligibility criteria and were not randomized, participants randomized, participants randomized but never treated, participants FAS, and participants in the SAS.
- The number and percentage of the participants in the following categories:
    - Completed study
    - Repeaters
    - Prematurely discontinued from study prior to the data cut date

4. DESCRIPTIVE ANALYSIS OF BASELINE CHARACTERISTICS

Summaries of baseline subject characteristics (overall, and by treatment group) will be provided using the Full Analysis Set. Baseline subject characterisitcs include the following:

- Subject demographic data (e.g., sex, race/ethnicity, age, social history).
- General clinical characterisitcs (body weight, height, body mass index , baseline titers).
- Disease characteristics (e.g., type of transplant, GVHD status)

Descriptive statistics on continuous data will be reported as (N, sample mean, sample standard deviation, median, interquartiles, minimum, and maximum); absolute and relative frequency will be used to describe categorical data. Of note, we will not perform formal hypothesis testing to compare these variables between treatment groups.

5. STATISTICAL ANALYSES FOR IMMUNOGENICITY OUTCOMES

This section presents plans for analysis corresponding to each of the primary, secondary, and tertiary objectives listed in Section 1.2.1. We utilize the following general notation to describe models of interest:

- i : indicator of subject
- t : indicator of visit ($t = 2$, $t = 3$)
- X : randomized group (0 = SD, 1 = HD)
- Y : titer outcome
- $\boldsymbol{L}$: pre-treatment covariates
- γ_i : subject-specific random effect
- γ_S : study site random effect
- ε_{it} : error term

5.1 Analysis of continuous HAI titer responses

We will use mixed effect models to analyze continuous titers. All comparisons will be made on the geometric mean ratio scale. Consider the following model form:

$$\log(Y_{it}) = \beta_0 + \beta_1 \mathbf{1}(X = 1) + \beta_2 \mathbf{1}(t = 3) + \beta_3 \mathbf{1}(t = 3; X = 1) + \boldsymbol{\beta}_L \boldsymbol{L} + \gamma_i + \gamma_S + \varepsilon_{it},$$

under which all primary and secondary aims pertaining to continuous HAI titer outcomes can be addressed. The parameters of this model and the comparisons to which they correspond are summarized in Table 3.

Any continuous variables (e.g., time post-transplant and log-transformed baseline titer) used for adjustment will be flexibly modeled using natural cubic splines with knots at the 25th, 50th, and 75th percentiles.

5.2 Analysis of seroprotection

We will perform analyses similar to those of Section 5.1, instead using the binary indicator of *seroprotection* as the outcome, as defined by a titer of at least 1:40 to any of the four antigens. Correspondingly, we will utilize a logistic link instead of an identity link.

5.3 Analysis of seroconversion

We will perform analyses similar to those of Section 5.1, instead using the binary indicator of *seroconversion* as the outcome, as defined by at least a four-fold rise from baseline in titer within *any* of the four antigens. Correspondingly, we will utilize a logistic link instead of an identity link.

5.4 Effect modification

We will evaluate evidence of effect of vaccination on HAI response is modified by time post-transplant, age, or GVHD status by including (and subsequently testing) appropriate interaction terms in the model described in 5.1 (log-transformed continuous titer value).

5.5 Analysis of study repeaters

We will perform analyses similar to those described in Section 5.1, 5.2, and 5.3 in the RAS, using both their first and second year of data.

6. STATISTICAL ANALYSES FOR SAFETY OUTCOMES

6.1 Statistical analyses for safety outcomes

We will provide descriptive statistics (N, missing, mean, standard deviation, and quartiles) on solicited local and systemic AEs, overall and by randomized treatment group in the SAS.

Further, we will use a mixed model to separately compare the odds of local/systemic events at each visit between treatment groups.

6.2 Analysis of study repeaters

We will perform an analysis similar to that described in Section 6.1, using subjects in the RAS set who have completed at least one memory aid in their second year.

7. STATISTICAL ANALYSES FOR EXPLORATORY OBJECTIVES

In this section, we outline possible exploratory analyses of interest.

7.1 MN and NAI titers

We will examine MN and NAI titers in a fashion analogous to that described in Section 5.1 in the FAS. We will perform an analysis similar to that described in Section 7.1 in the RAS.

7.2 Persistence/durability of HAI and MN titers

We will perform analyses similar to those described in Section 5.1 in the FAS, except using information from Visit 4 as the outcome in order to assess long-term durability of HAI and MN titers.

7.3 Influenza detection

We will use a mixed model to compare the odds of symptomatic influenza, detected by PCR, during flu season between randomized treatment groups using the FAS. Further, we will evaluate whether there is evidence that baseline covariates (e.g., GVHD) predict odds of symptomatic influenza.

7.4 HAI and MN responses

We will determine whether there is evidence of correlation between HAI and MN responses.

7.5 HAI titers and ex-vivo T and B cell phenotypes

We will perform an analysis in the FAS to evaluate whether there is evidence that the baseline phenotype of T or B cells predicts HAI titers among study groups after vaccination. In addition to baseline percentages of CD4+ T cells, CD8+ T cells, and CD19+ B cells, a cell clustering algorithm (FlowSOM) will be used to delineate clusters of each parent cell population. Changes in cluster frequencies will be evaluated cross-sectionally in relation to time since transplantation. For study repeater subjects, the same individual will be evaluated at the baseline vaccination visit for each year. We hypothesize that cell clusters that increase in frequency with time after transplantation represent evidence of immune reconstitution, and the presence of these cell clusters at high frequency at the time of vaccination will be associated with higher subsequent magnitudes of HAI titers.

8. TABLES

TABLE 1: STUDY SITES	
#	**Site**
1	Vanderbilt University
2	Seattle Children's Hospital
3	Children's Hospital of Philadelphia (CHOP)
4	Children's Mercy Hospital
5	Cincinnati Children's Hospital
6	Nationwide Children's Hospital
7	St. Jude Children's Research Hospital
8	UCSF (University of California San Francisco) Benioff Children's Hospital
9	Baylor School of Medicine, Texas Children's Hospital

TABLE 2: DEFINITION OF ANALYSIS SETS		
#	**Analysis set**	**Description**
1	All randomized (ARAS)	All participants randomized into the study, for use in by-subject listings.
2	Full (FAS)	All participants who (1) are randomized into the study and (2) have received at least 1 dose of either vaccine.
3	Safety (SAS)	All participants who (1) are randomized into the study and (2) have completed at least one memory aid.
4	Repeaters (RAS)	All participants who are (1) randomized into the study and (2) received at least one vaccination as a repeater.

TABLE 3: PARAMETERS FOR MIXED MODEL OF INTEREST						
#	**Comparison**	**Visit**	**Comparison**	**Parameter**	**Null hypothesis**	**Classification**
1	2 doses HD-TIV 2 doses SD-QIV	3 3	Between-group	$\exp(\beta_3)$	H_0: $\beta_3 = 0$	Primary
2	1 dose HD-TIV 1 dose SD-QIV	2 2	Between-group	$\exp(\beta_1)$	H_0: $\beta_1 = 0$	Secondary
3	1 dose SD-QIV 2 doses SD-QIV	2 3	Within SD-QIV	$\exp(\beta_2)$	H_0: $\beta_2 = 0$	Secondary
4	1 dose HD-TIV 2 doses HD-TIV	2 3	Within HD-TIV	$\exp(\beta_2 + \beta_3)$	H_0: $\beta_2 + \beta_3 = 0$	Secondary
5	1 dose HD-TIV 2 doses SD-QIV	2 3	Between-group	$\exp(\beta_2 - \beta_1)$	H_0: $\beta_2 - \beta_1 = 0$	Tertiary