Simulations on Sample Sizes and Powers Calculations for a Registry Vaccine Study

John Liang[1] PhD and Scott Patterson[1], PhD, PStat@

Affiliation: [1]Pfizer Inc., 500 Arcola Road, Collegeville, PA 19426, USA

Abstract

World Health Organization (WHO) guidance was issued in 2009 covering *Recommendations to assure the quality, safety and efficacy of pneumococcal conjugate vaccines.* From the new WHO guidance[1] (at section C2.2.3, page 27), two primary co-endpoints are recommended: *"The endpoints used in the primary analysis should be the percentage of subjects with IgG ≥ 0.35 μg/ml AND the serotype-specific IgG GMC (Geometric Mean Concentration) ratios".*

The purpose of this work is to show simulations to define sample size and power calculation for a study in support of WHO registration requirements accounting for multiplicity across serotypes and across endpoints.

Key Words: Multiplicity, across endpoints, across serotypes, Type-I error control

1. Study Design

A phase 3, parallel-group, randomized, active-controlled , double-blinded trial was to be designed to evaluate immunogenicity of Test Vaccine as compared with Control. There were 3 infant vaccinations and 1 toddler vaccination. The IgG (immunoglobulin G) data collected after 3 infant vaccinations are in interest for primary endpoints.

This study has two double-blind groups (*Control* and *Test*). Table 1 below shows the study flow chart for vaccinations and the bold draw.

Table 1. Schedule of Study Vaccine Administration and Blood Draw

Randomization Group	3-Month Visit	4-Month Visit	5-Month Visit	6-Month Visit
Control Group	X	X	X	
Test Group	X	X	X	
Blood sample for IgG				**X**

1.1 Study Objectives and Endpoints

The primary objective of the study is to demonstrate that the immune responses to the 13 pneumococcal serotypes induced by test vaccine in a 3-, 4-, and 5-month schedule (*Test Group*) are noninferior to the immune responses induced by control vaccine in a 3-, 4-, 5-month schedule (*Control Group*) when measured 1 month after the infant series.

Co-primary endpoints are as follows:

1) The percentage of subjects with an IgG concentration ≥ 0.35 μg/mL

2) The serotype-specific IgG GMC ratio.

There are 13 serotypes induced by the *test* vaccine and 7 serotypes induced by the *control* vaccine; however in both groups, all IgG antigen concentrations for all 13 serotypes were evaluated for both groups.

In this paper, we mainly show the work of simulations on sample size/power calculations based on these two primary endpoints for this study in support of registration.

2. Hypotheses and Acceptance Criterion

2.1 Statistical Hypotheses

Statistical inference will be made on the primary and co-primary endpoints of the primary immunological data of interest: the immunogenicity responses at 1 month after the infant series in subjects receiving *Test* vaccine at a schedule of 3, 4, 5 months (*Test Group*) relative to the responses in subjects receiving *Control* vaccine in a schedule of 3, 4, and 5 months (*Control Group*). The primary endpoint for each of the pneumococcal serotypes is the proportion of subjects in *Test Group* achieving a serotype-specific IgG concentration ≥0.35 μg/mL 1 month after the infant series and the proportion of subjects in *Control Group* achieving a serotype-specific IgG concentration ≥0.35 μg/mL 1 month after the infant series. The co-primary endpoint for each of the pneumococcal serotypes is the IgG geometric concentration (GMC) 1 month after infant series for *Test* and *Control Groups*.

The primary null hypothesis for each of the 13 pneumococcal serotypes from *Test Group* and *Control Group* is:

$$H_{01} : \pi_{test\_i} - \pi_{control\_i} \leq -0.10 ,$$

where i =1, 2, …, 13, π_{test_i} is the proportion of subjects achieving an IgG concentration ≥ 0.35 μg/mL in *Test Group* for the i-th serotype, and $\pi_{control\_i}$ is the proportion of subjects achieving an IgG concentration ≥ 0.35 μg/mL in *Control Group* for the i-th serotype.

The co-primary null hypothesis for each of the 13 pneumococcal serotypes for both *Test* and *Control Groups* is:

$$H_{02} : \mu_{test\_i} - \mu_{control\_i} \leq -0.693 ,$$

where i =1, 2, …, 13, μ_{test_i} is the log of the IgG GMC in *Test Group* for the i-th serotype, and $\mu_{control\_i}$ is the log of IgG antigen concentration in *Control Group* for the i-th serotype.

2.1 Statistical Decision Rules

For primary hypothesis, the noninferiority criterion for a given antibody serotype will be met if the lower bound of the 2-sided, 97.5% confidence interval, computed using the Chan and Zhang[2] procedure, for the difference in proportions (*Test* - *Control*) is greater than -0.10.

For co-primary hypothesis, the noninferiority criterion for a given antibody serotype will be met if the lower bound of the 2-sided, 97.5% confidence interval for the GMC ratio ($GMC_{test\_i}/GMC_{control\_i}$) greater than 0.5 (2-fold criterion).

If either noninferiority for the IgG responder objective or for the IgG GMC objective can be met for the all i = 1-13 serotypes, then the noninferiority of IgG responses for *Test Group* relative to *Control Group* for this study will be declared.

The alpha adjustment for primary comparisons (2-sided, 97.5% confidence interval is used) is based on WHO criteria[1], and that under that approach alpha is split between the GMC and responder hypotheses (97.5% CIs). The WHO criteria (Section C2.2.3 Primary analysis) indicates that:

> For the serotypes to the new vaccine and the licensed comparator, the endpoints used in the primary analysis should be:
>
> The percentage of subjects with an IgG concentration ≥ 0.35 μg/mL AND
> The serotype-specific IgG GMC ratios.

Therefore noninferiority for the study is demonstrated as below:

For each of all the 13 serotypes,

- the lower bound of 97.5% CIs for the difference in proportions is greater than -0.10 after the infant series

or

- the lower bound of 97.5% CIs for IgG GMC ratio is greater than 0.5 for 7 common serotypes after the infant series.

Noninferiority must be shown across all 13 serotypes to declare **success for the study**.

3. Type I Error Control

Experiment-wise type 1 error is therefore conservatively controlled across the 13 endpoints by application of the intersection-union testing procedure.

Multiplicity across endpoints was controlled by Bonferroni adjustment (i.e.. 5% type 1 error was divided equally across responder and GMC endpoints), therefore 97.5% confidence intervals will be used for inference . All serotypes must successfully meet either the test of proportions or GMCs for the study to succeed.

The null hypothesis of the test is as follows:

$$H_0 = \bigcup (H_{0k}^{(P)} \cap H_{0k}^{(CP)})$$

where U denotes the union and ∩ denotes intersection, *K = 1 -13;* $H_{0k}^{(P)}$ *and* $H_{0k}^{(CP)}$ *are* defined as follows*:*

The null hypothesis for primary endpoint (Responder):

$$H_{0k}^{(P)} : \pi_{test\_k} - \pi_{control\_k} \leq -0.10$$

where *k =1-13.*

The null hypothesis for co-primary endpoint (GMR):

$$H_{0k}^{(CP)} : \mu_{test\_k} - \mu_{control\_k} \leq -0.693$$

where *k =1-13.*

4. Simulations Results

Two previous similar studies were selected for simulations. IgG data from the 2 historical studies were logarithm-transformed first. Variance-covariance matrix and mean vector were calculated from the corresponding control group for each the 2 studies. Then multi-normally distributed random samples were generated (by a SAS macro).

It is noticed that it is quite time-consuming for calculating 97.5% CIs by using Chan and Zhang's[2] method; therefore it was decided to generate random samples for 1000 studies. After a several calculations and based we narrowed the scope of the number of subjects for each simulated study in a range 300 to 400 per group. Finally we ended up with 1000 simulated studies and 380 subjects per group for each simulated study.

After these random samples were generated, for the primary endpoint, we used mixed model to perform the comparison analysis and calculated GMRs and corresponding 97.5% CIs for each serotype in each study and for the co-primary endpoint, we calculated the difference (*Test – Control*) in proportion of subjects who achieving IgG concentration ≥ 0.35 μg/mL along with the 97.5% CIs of the difference in proportion. For difference in proportion and the corresponding 97.5% CIs, an exact method by Chan and Zhang[2] is used. In the appendix SAS code for calculation is provided.

4.1 Simulations Based the First Historical Study

Simulated power for the GMR endpoint regards to a specific serotype is define as follows:

Among the 1000 simulated studies, the 97.5% CIs of GMR (*Test/Control*) are calculated for the specific serotype. We look in to the lower limits of the 97.5% CIs of GMR to see if they are greater than 0.5. The proportion of such lower limits of 97.5% CIs of GMR that are greater than 0.5 among all lower limits of 97.5% CIs of GMR is defined as the simulated power for that specific serotype.

Similarly, we can define the simulated power for the co-primary endpoint. We look in to the lower limits of the 97.5% CIs of difference in proportion of subjects achieving IgG concentration ≥ 0.35 μg/mL to see if they are greater than -10%. The proportion of such lower limits of 97.5% CIs of difference in proportion that are greater than -10% among all lower limits of 97.5% CIs of difference is defined as the simulated power for that specific serotype.

Now we define a simulated power for the case **that either the lower limits of 97.5% CI of GMR> 0.5 or the lower limit of 97.5% CI of difference in portion of subjects achieving IgG concentration ≥ 0.35 μg/mL, are greater than -10%.** For a specific serotype, we look into lower limits of 97.5% CIs for both primary and co-primary endpoints.

Denote Y= 1, if either the lower limit of 97.5% CI of GMR> 0.5 or the lower limit of 97.5% CI of difference in portion (of subjects achieving IgG concentration ≥ 0.35 μg/mL) > -10% and 0, otherwise.

For a specific serotype, the proportion (Y=1) among all cases (Y=1 or 0) is defined as the simulated power for the "Either Endpoint" case.

The following Table 2 shows the simulation results based the IgG data from first historical study.

For serotypes 8 and 13, the simulated for GMR endpoint are high (both are greater than 99.4%) but the simulated powers for responder endpoint are low (36.1% and 50.6% respectively). With no surprise, the powers for either one (still maintains relatively high (greater than 99.5%)

For Serotype 6, for both endpoints, the powers are less than 83% but the power for either one is 90.4%, which meets the minimal requirement for power from regulatory authority.

Table 2. Powers for responder endpoint (IgG ≥ 0.35 µg/mL) and for the GMR endpoint from simulations for 380 subjects per arm by serotypes (1000 studies) - Based on Study 1

Serotype	Power for proportion endpoint	Power for GMR endpoint	Power for either endpoint
Serotypes 1, 2, 4, 5, 7, 11, and 12	≥99.9%	≥99.9%	≥99.9%
Serotype 3	98.9%	≥99.9%	≥99.9%
Serotype 6	82.8%	74.8%	90.4%
Serotype 8	36.1%	≥99.9%	≥99.9%
Serotype 9	≥99.9%	99.7%	≥99.9%
Serotype 10	99.6%	≥99.9%	≥99.9%
Serotype 13	50.6%	99.4%	99.5%

Mean vectors and variance-covariance matrix from Study 1; IgG data used for the multi-normal sample generations. When simulating, we assume zero difference of mean IgG concentration between Group 2 and Group 1 for each serotype.

If the lower limit of the 97.5% confidence interval of difference in proportions (Test – Control) > - 10%, then non-inferiority is claimed.

2-fold criterion (i.e. the lower limit of 97.5% confidence interval of GMR > 0.5) used for non-inferiority claim.

4.1 Simulations Based the Second Historical Study

Simulations were performed based on IgG data from the second historical study. The simulation results are shown in Table 3. All simulated powers for the primary endpoint, the co-primary endpoint, and the "Either Endpoint" are greater than 99.9%.

Table 3. Powers for responder endpoint (IgG ≥ 0.35 µg/mL) and for the GMR endpoint from simulations for 380 subjects per arm by serotypes (1000 studies) – Based on Study 2

Serotype	Power for proportion endpoint	Power for GMR endpoint	Power for either endpoint
All serotypes 1 to 13	≥99.9%	≥99.9%	≥99.9%

Mean vectors and variance-covariance matrix from Study 2; IgG data used for the multi-normal sample generations. When simulating, we assume zero difference of mean IgG concentration between Group 2 and Group 1 for each serotype.

If the lower limit of the 97.5% confidence interval of difference in proportions (Test – Control) > - 10%, then non-inferiority is claimed.

2-fold criterion (i.e. the lower limit of 97.5% confidence interval of GMR > 0.5) used for non-inferiority claim.

5. Overall Simulated Power

In any simulated study, if for a given serotype, the lower limit of the 97.5% CI the difference in proportion (*Test* − *Control*) > −0.10 OR the limit of the 97.5% CI of GMR (*Test* relative to *Control*) > 0.5, then non-inferiority can be declared for that serotype.

Non-inferiority must be shown across all 13 serotypes to declare success for the study. The proportion of successes of the studies is calculated among the 1000 studies and for both history data and the simulated powers (success proportions) are **90.3% and 99.9%** based on the IgG data from the first and second historical studies, respectively.

Acknowledgements

The authors of this paper would like to thank the clinical team, especially thank Dr. Daniel A. Scott and Dr. Alejandra Gurtman, of the registry study for their valuable inputs in the simulation work.

References

1. WHO Recommendations to assure the quality, safety and efficacy of pneumococcal conjugate vaccines Proposed replacement of: TRS 927, Annex 2, EXPERT OMMITTEE ON BIOLOGICAL STANDARDIZATION Geneva, 19 to 23 October 2009
2. Chan ISF, Zhang Z. Test-based exact confidence intervals for the difference of two binomial proportions. Biometrics. 1999; 55:1202-1209.

Appendix for SAS Code

Mixed model is used to perform the analysis for GMR and corresponding 97.5% C.Is.

The SAS code for the mixed model is as follows:

```
ods listing close;
proc mixed data=trialf method=reml   ITDETAILS   maxiter=200;
     by s serotype;
     class group;
     model lnigg=group;
     estimate 'Group 2:1' group -1 1/cl alpha=0.025;
     ods output Estimates=test;
run;
```

An exact method by Chang and Zhang[2] is used by executing SAS StatXact procedure
SAS Code:

```
proc binomial data = sum_resp1 max_time=720 gamma = 0.000001 alpha=0.975
       out = ex_diff noprint;
  riskdiff / ex one;
  by s serotype;
  po group;
  ou outcome;
  weight count;
 run;
```