The Design of a Phase 2 Malaria Vaccine Trial Based on a Cohort Study

Michael P. Fay[1], M. Elizabeth Halloran[2], Dean A. Follmann[1], David Diemert[1], Alassane Dicko[3], Mahamadou Thera[3], Issaka Sagara[3], Amy Klion[1], Ogobara Doumbo[3], Allan Saul[1]

1: National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA; 2: Emory University, Atlanta, GA, USA; 3: Malaria Research and Training Center, University of Bamako, Bamako, Mali.

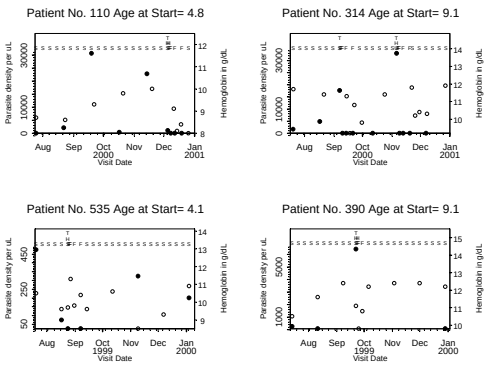
Abstract

INTRODUCTION:
With multiple vaccines targeting blood stage antigens of malaria under development, we need efficient methods for Phase 2 trials aimed at demonstrating a biological impact of the vaccine. There are two important factors: 1. picking endpoints that will predict final efficacy in reducing death and severe morbidity; 2. trial designs that minimize the number of volunteers required. Using longitudinal data from an endemic population in Mali, we studied several possible clinical, parasitologic and combination endpoints.
METHODS:
We assume that the differences in endpoints between 4 year old control subjects and 4 year old vaccinated subjects are similar to the observed differences between non-vaccinated 4 and 8 year olds from a longitudinal study in Donéguébougou, Mali. Children had weekly visits with monthly blood smears, with additional unscheduled visits when malaria symptoms were present. Children with malaria symptoms had smears read immediately and if positive, were treated. Sample size calculations were done by: 1) only using data from children close to the target ages and using the associated means and variances in standard sample size formulas, or 2) fitting the whole data set to a mathematical model of the age related response, then using the calculated mean response at the target ages. Sample sizes could be adjusted if the vaccine was efficacious in only a proportion of those vaccinated.
RESULTS:
During the 1999 (and 2000) malaria transmission seasons, 184 (193) children aged 3 months to 20 years were followed for at least 140 days, of which 23 (18) were 3-5 years old and 25 (39) were 7-9 years old at enrollment. 67 children were followed in both seasons. Both maximum and average parasitemia over all scheduled visits during a season produced large sample size estimates, while all but treatment follow-up (ABTFU) visits gave smaller ones. For the ABTFU visits, maximum parasitemia gave smaller sample sizes than average parasitemia. Parasite densities above 1000 or above 10,000 per μl, with or without fever, performed similarly to maximum parasitemia, although there was considerable variability in the estimates. For example, by method (1), for a 100% efficacious vaccine, using maximum parasitemia over ABTFU visits resulted in an estimated sample size per vaccination group of 61 for the 1999 season (80% CI 18 to 524, by bootstrap) vs. 27 for the 2000 season (80% CI 14 to 65), while the endpoint of parasitemia over 10,000 with fever resulted in an estimate of 15 (80% CI 8 to 35) for 1999 vs. 101 (80% CI 29 to 1650) for 2000. A vaccine that was efficacious in only half the subjects required a greater than 4-fold increase in sample sizes.

Methods

Preliminary Cohort Study Design

•**Location**: The data come from a longitudinal study in Donéguébougou, Mali, which is a rural village 30 km northeast of Bamako with approximately 1200 inhabitants.
•**Seasonality:** Malaria transmission occurs mainly during the rainy season from June to November. The data for this poster covers the 1999 season (July 13, 1999 – January 17, 2000) and 2000 season (July 19, 2000 – December 31, 2000).
•**Study Design:** Children were seen at weekly scheduled visits at the study clinic and at unscheduled self-referring visits. At all visits they were questioned for symptoms of malaria and had their axillary temperature taken. Malaria thick blood films were prepared from capillary blood obtained by fingerstick at the first visit of each season, at every fourth weekly visit, and at intervening visits if symptoms of malaria were present. Smears were read contemporaneously only if symptoms of malaria were present, and children were treated with either sulfadoxine-pyrimethamine (uncomplicated malaria) or parenteral quinine (severe malaria) if smear-positive. Children were followed after treatment to ensure an adequate response, and thick films were performed at some – but not all – follow-up visits. Each time that capillary blood was obtained for a malaria film, hemoglobin concentration (g/dl) was also determined.



Example Data: The above graphs represent sample data from 4 individuals. Solid dots by the left axis are parasitemia levels, open circles by the right axis are hemoglobin levels. T=treated, H=high temperature (>=37.5º C). Visit types: S=scheduled, U=unscheduled, F=treatment follow-up.

Laboratory Methods: Thick blood films were stained with Giemsa and parasitemia was assessed by counting the number of asexual *P. falciparum* parasites until 300 leukocytes were observed. Parasite densities were converted to parasites per microliter of blood, assuming an average leukocyte count of 7500/ml. Hemoglobin was determined using a Hemacue® analyzer (Labnet International, Buckley, WA).

Methods for Sample Size Calculation

Motivational Overview

We wish to estimate a sample size needed to show an effect of vaccine even if it provides only partial protection in some subjects and no effect in others. We assume the partial protection effect (the difference in mean endpoint values between vaccinated 4 year olds who respond to the vaccine and unvaccinated 4 year olds) will be similar in magnitude to the difference in means between 4 year olds and 8 year olds, and use the cohort data to estimate the latter differences on several different possible endpoints. We additionally modify our sample size calculation to allow for the possibility that only a proportion of the population responds to the vaccine.

Sample Size Formula
Suppose that

- **u(4)** and **v(4)** are the means and variance estimates of a response for the 4 year olds, and **u(8)** and **v(8)** are similarly defined for the 8 year olds.
- vaccinated 4 year olds respond with means and variances similar to unvaccinated 8 year olds.
- **r** is the proportion of vaccinated subjects who respond as if unvaccinated
- **Zq** is the (1-q)th quantile of the standard normal distribution (e.g., $Z_{0.05/2}$ =1.96)

Then using the central limit theorem on the means, Fay, Halloran and Follmann (2005) showed that the sample size for each group which gives 80% power for a two-sided 0.05 test of a difference in means is approximately

$$N = \frac{[v(4)+v(8)+r\{v(4)+v(8)\}](Z_{.025}+Z_{.2})^2}{(1-r)^2\{u(4)-u(8)\}^2} + \frac{r(1-r)(Z_{.025}+Z_{.2})^2}{(1-r)^2}$$

Estimating Means and Variances:

Coarse Approach

In order to estimate the means and variances for the 4 and 8 year olds, one simple approach (the coarse approach) is to simply use sample means and variances from the cohort study of the possible response variables. Because the study was not large enough to give enough patients of those exact ages, we enlarged the age range and used 3-5 year olds to represent 4 year olds, and 7-9 year olds to represent 8 year olds.

Smooth Approach

In the smooth approach, we parametrically estimated the means as a function of age, and estimated the variance by parametric assumptions, either a Poisson model with overdispersion, or a least-squares model on the log of the means where the error has constant variance. We modelled the mean at age a, u(a), with the following function using 3 parameters (b,c, and, d)

u(a;b,c,d) = c exp(- ½ {[log(a) – log(d)]/b}²)

The function has the following properties (regardless of the values of b,c, and d):

- u(0)=0, i.e., mean response is 0 at birth.
- The function u(a) is uni-modal. It starts at zero, has one peak and then decreases back to zero as $a \to \infty$
- The age at the peak is d.
- The scaling factor is b. This is similar to the standard error of a log-normal distribution. It determines how pointy the peak is.
- The mean response at the peak is c.

Poisson Model with Overdispersion: The Poisson model with overdispersion has mean u(a;b,c,d) for any age a, and associated variance f * u(a;b,c,d), where f is the overdispersion parameter. We used that model when the response variable had small counts. We estimated the mean by maximum likelihood assuming Poisson responses (equivalent to maximum quasi-likelihood with mean and variance as described). We estimated f by taking the maximum of 1 or the generalized Pearson statistic as recommended by McCullagh and Nelder (1989, p. 2000).
Least-Squares Model: We used the least-squares model with continuous or approximately continuous responses. Taking the log of both sides of the mean expression above gives

$$\log(u) = \log(c) - .5\,(\log(a)-\log(d))^2 \,/\, b^2$$

For the i^{th} child with response y_i and age a_i, the equation can be rewritten in linear model form as

$$\log(y_i) = \beta_0 + \beta_1 \log(a_i) + \beta_2 \log^2(a_i) + \varepsilon_i$$

where ε_i is an error term, and we assume all the error terms are independent and have the same variance, σ^2, and

$$\beta_0 = \log(c) - \tfrac{1}{2}\log^2(d)/b^2$$
$$\beta_1 = \log(d)/b^2$$
$$\beta_2 = -1/b^2$$

We estimated the β parameters by least-squares methods, then solved for the b,c, and d parameters. We estimated σ^2 by

$$\frac{1}{M-3}\sum_{i=1}^{M}\{\log(\ y_i\ ) - \log(\ u(a_i;\hat{b},\hat{c},\hat{d})\ )\}^2$$

where M is the number of patients in the cohort, and the b,c, and d values with hats are the estimates. For the least squares method, we calculated the sample size from the differences in the log(u) values instead of the differences in the means. To be specific, in the sample size equation above, we inserted our estimate of log(u(4)) for u(4), our estimate of log(u(8)) for u(8), and our estimate of σ^2 for v(4) and v(8).

Accounting for Variability Through Bootstrapping

We calculated bootstrap confidence intervals on the sample size estimates using the percentile confidence intervals using 1000 bootstrap replications. We used linear interpolations between order statistics for the percentile intervals instead of defining them as rounded up (for the upper limit) or down (for the lower limit) as in Efron and Tibshirani (1993). This is the default method in the Splus software which we used. For some of the bootstrap replications for the smooth method using the Poisson model, the algorithm did not converge, and the percentile intervals were calculated only using the converging replicates.

Results

Data Used for Analysis

For our data analysis, we used subjects who were followed for the whole season, and who were not missing any of the summary variables. For the 1999 season data from 184 subjects were included (range of 168-182 days of follow-up), while for the 2000 season data from 193 subjects were included (140-165 days of follow-up). 67 subjects were included in the data sets from both seasons.

There were 23 children aged 3-5 years in the 1999 season and 25 aged 7-9 years. In the 2000 season there were 18 aged 3-5 years and 39 aged 7-9 years.

Sample Sizes (with 80% confidence intervals)

Below we list the estimated sample sizes needed in each group to demonstrate a significant difference in response using different endpoints with 80% power and using a two-sided 0.05 level test. Note that r is the proportion of vaccinated subjects that act as if they received the control regimen. We also list 80% bootstrap confidence intervals on the sample size estimates.

	Coarse Approach		Smooth Approach	
	1999	2000	1999	2000
Number of Times the Patient is Treated During the Season				
r=0	12 (7, 24)	199 (38, 4515)	86 (60, 125)	67 (50, 95)
r=.5	61 (35, 116)	803 (165, 14713)	378 (268, 539)	299 (226, 415)
Number of Times Hemoglobin<10 During the Season				
r=0	66 (17, 910)	67 (22, 749)	27 (17, 37)	15 (11, 20)
r=.5	275 (85, 3520)	301 (101, 3214)	137 (87, 181)	79 (50, 101)
Minimum Hemoglobin During the Season				
r=0	106 (26, 1342)	224 (42, 6474)	35 (25, 47)	25 (19, 30)
r=.5	403 (95, 6366)	818 (158, 22869)	145 (108, 194)	105 (84, 127)
Cumulative Parasitemia for the Season (using all visits)				
r=0	49 (18, 524)	35 (17, 76)	49 (35, 68)	25 (19, 32)
r=.5	238 (95, 976)	201 (96, 366)	202 (147, 280)	108 (81, 135)
Average Parasitemia per Season over all Scheduled Visits				
r=0	1541713 (66, 10796)	82 (28, 1030)	21935 (735, 160976)	86 (51, 146)
r=.5	6542053 (275, 61795)	448 (162, 3510)	87746 (2942, 643910)	351 (212, 591)
Average Parasitemia per Season over all but Treatment Follow-up Visits				
r=0	147 (27, 2861)	32 (16, 64)	65 (45, 98)	23 (18, 29)
r=.5	569 (125, 10272)	185 (88, 318)	268 (186, 397)	100 (77, 123)
Average Log Parasitemia per Season over all Scheduled Visits				
r=0	69 (23, 866)	1717 (53, 10845)	3375 (652, 69902)	132 (82, 228)
r=.5	266 (77, 3952)	7561 (247, 58346)	13505 (2613, 279615)	536 (333, 919)
Average Log Parasitemia per Season over all but Treatment Follow-up Visits				
r=0	860 (57, 7939)	80 (24, 892)	168 (99, 321)	37 (27, 50)
r=.5	3032 (202, 40859)	295 (91, 2994)	679 (403, 1293)	154 (116, 205)
Maximum Parasitemia per Season Over all Scheduled Visits				
r=0	573 (61, 7593)	87 (31, 825)	12820 (640, 30032)	91 (54, 154)
r=.5	2799 (299, 27633)	475 (179, 3339)	51288 (2566, 520133)	370 (222, 623)
Maximum Parasitemia per Season over all but Treatment Follow-up Visits				
r=0	61 (18, 524)	27 (14, 65)	59 (42, 85)	26 (19, 34)
r=.5	274 (96, 1709)	155 (85, 314)	244 (172, 346)	112 (84, 141)
Maximum Log Parasitemia per Season over all Scheduled Visits				
r=0	5290 (60, 10963)	1151 (41, 7790)	1091 (304, 12967)	75 (46, 124)
r=.5	22179 (298, 52466)	5446 (183, 42299)	4370 (1221, 51871)	306 (190, 502)
Maximum Log Parasitemia per Season over all but Treatment Follow-up Visits				
r=0	46 (14, 290)	21 (10, 55)	61 (42, 88)	24 (19, 30)
r=.5	187 (58, 1276)	86 (42, 211)	249 (176, 360)	102 (80, 126)
Number of Episodes with Parasitemia>0 during the Season (using all visits)				
r=0	35 (16, 112)	1158 (70, 13724)	78 (53, 120)	258 (152, 526)
r=.5	180 (85, 569)	4121 (252, 30905)	332 (229, 503)	1069 (635, 2154)
Number of Episodes with Parasitemia>=1000 during the Season (using all visits)				
r=0	49 (17, 272)	52 (20, 254)	56 (38, 88)	22 (18, 28)
r=.5	228 (85, 1304)	182 (72, 1136)	249 (174, 384)	106 (88, 132)
Number of Episodes with Parasitemia>=3000 during the Season (using all visits)				
r=0	106(19, 3454)	24(10, 88)	48 (34, 75)	23(18, 29)
r=.5	424(73, 12929)	87(41, 279)	221 (161, 335)	110(91, 137)
Number of Episodes with Parasitemia>=10000 during the Season (using all visits)				
r=0	34 (14, 114)	88 (23, 1780)	49 (992, 74)	24 (18, 32)
r=.5	162 (70, 480)	350 (99, 4239)	228 (164, 337)	119 (96, 155)
No. of Episodes with Parasitemia>0 and Temperature>37.5 C (using all visits)				
r=0	14 (7, 31)	65 (20, 702)	79 (57, 113)	26 (20, 34)
r=.5	70 (39, 137)	272 (84, 2804)	349 (255, 490)	126 (102, 163)
No. of Episodes with Parasitemia>=1000 and Temperature>37.5 C (using all visits)				
r=0	31 (13, 102)	68 (19, 681)	72 (49, 101)	29 (22, 39)
r=.5	158 (74, 513)	279 (89, 2289)	326 (228, 450)	142 (112, 187)
No. of Episodes with Parasitemia>=3000 and Temperature>37.5 C (using all visits)				
r=0	30 (10, 165)	63(18, 701)	56(39, 79)	29(22, 41)
r=.5	149 (51, 753)	271(83, 2763)	258(188, 358)	145(112, 200)
No. of Episodes with Parasitemia>=10000 and Temperature>37.5 C (using all visits)				
r=0	15 (8, 35)	101 (29, 1650)	48 (33, 68)	35 (25, 52)
r=.5	82 (45, 171)	471 (140, 7406)	227 (163, 317)	172 (129, 250)

Discussion

We have explored the use of different endpoints to assess the efficacy of candidate malaria vaccines in a Phase 2 study, calculating sample sizes as if the difference between the vaccine group and the control group will be similar to the observed difference between 4 year olds and 8 year olds. When the sample sizes are smaller, this means that the standardized differences in means between the age groups are larger. In general it is harder to see differences between age groups when parasitemia is summarized only using data from scheduled visits since some of the highest parasitemia values (those from unscheduled sick visits) would not be included. Furthermore, it seems better to summarize parasitemia using either the maximum parasitemia, the maximum log parasitemia, or the number of episodes exceeding a given parasitemia value, rather than averaging parasitemia over the season. This analysis shows that if the motivating assumption is approximately correct then there are many possible endpoints which result in similar sample size requirements. The choice of endpoint should therefore not be made solely upon the basis of the estimated sample size but rather on other factors such as the plausibility that the endpoint is a marker of a biological effect.

It is important to stress the assumption these sample size calculations use - that differences in a given efficacy endpoint between vaccinated and unvaccinated 4 year olds will be similar to historically-observed differences between unvaccinated 4 year olds and unvaccinated 8 year olds. This simple assumption does not account for the possibility that because 4 year olds have a smaller surface area than 8 year olds, they will likely be challenged by infectious bites less often. If the variables are corrected for the estimated number of infectious bites, then the effect sizes of these corrected variables will be larger and hence the necessary sample sizes will be smaller.

References

•Efron, B. and Tibshirani, R.J. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, London.
•Fay, M.P., Halloran, M.E., and Follmann, D.A. (2005). Accounting for variability in sample size estimation with applications to nonadherence, missing data, and estimation of variance and effect size. (unpublished manuscript).
•McCullagh, P., and Nelder, J.A. (1989). *Generalized Linear Models, second edition.* Chapman and Hall, London.