

An adaptive design for case-driven vaccine efficacy study when incidence rate is unknown

XIAOMING LI, IVAN S. F. CHAN* AND KEAVEN M. ANDERSON

In many vaccine efficacy studies where the endpoint is a rare infection/disease event, an event-driven design is commonly used for testing the hypothesis that study vaccine lowers the risk of the event. Uncertainty of the incidence rate has a large impact on the sample size and study duration. To mitigate the risk of running a potentially large, long-duration efficacy trial with an uncertain event rate, we propose a two-stage adaptive design strategy with interim analyses to allow evaluation of study feasibility and sample size adaptation. During Stage I, a modest number of subjects will be enrolled and the feasibility of the study will be evaluated based on the incidence rate observed. If the feasibility of the study is established, at the end of Stage I a formal interim analysis will be performed, with a potential sample size adaptation based on the conditional rejection probability approach. The operating characteristics of this design are evaluated via simulation.

KEYWORDS AND PHRASES: Adaptive design, Vaccine efficacy, Event-driven, Unknown incidence rate.

1. INTRODUCTION

In many vaccine efficacy studies where the endpoint is a rare infection/disease event, an event-driven design (conditional on the total number of events) is commonly used for testing the hypothesis that study vaccine lowers the risk of the event. When there is good knowledge about the disease's natural history and incidence rate of the endpoint, the study sample size and duration can be accurately determined based on design parameters. However, when the incidence rate is not well-known, designing a trial to recruit an appropriate number of subjects in a reasonable timeframe with a reasonable up-front financial commitment is challenging.

In the clinical trial that motivates this research, investigators are interested in evaluating the effect of a vaccine (developed by Merck Research Laboratories) for prevention of a viral infection associated with a rare form of cancer. However, the natural history of the viral infection is not well-described, with large variability on the prevalence of infection in the general population ranging from 0 to 20%

reported in the medical literature. This uncertainty in disease natural history poses a great challenge in study planning as to how many subjects are required in the trial, how many study sites are needed, how long the study will last, and how much the trial may cost. In particular, the study team was charged to assess the feasibility of conducting an efficacy trial for this infection endpoint and, if feasible, to design a study to be completed in a reasonable time (up to 7 years) with a maximum of $\sim 8,000$ subjects.

There are several potential options for the design strategy that could be considered to deal with the uncertainty in the background infection rate. One option would be to design a large study up-front based on a conservative estimate of the incidence rate to obtain study results within the desired timeline. However, if the incidence rate is higher than the conservative estimate, the trial would be larger, of longer duration and more costly than needed. On the other hand, if the vaccine efficacy is lower than expected, the trial would be under-powered in spite of its great expense and duration, which represents the worst case scenario that the study team would try to avoid. A second option would be to conduct a pilot study first, from which a better estimate on the background incidence rate could be obtained and used subsequently to design an efficacy study. A major concern with this option would be a very long timeline since the lengthy follow-up for the first study would need to be completed before starting the pivotal efficacy study. In light of the large up-front investment of the first option and the long duration of the second option, we propose a two-stage adaptive design strategy built on the traditional fixed-number-of-events design to allow evaluation of feasibility and sample size adaptation in a single trial. During Stage I, a fixed number of subjects will be enrolled and the feasibility of the study will be evaluated based on the incidence rate observed in a blinded manner. If the feasibility of the study is established, at the end of Stage I an interim analysis will be performed with the following potential outcomes: 1) stop the study for either futility or overwhelming efficacy, 2) continue into Stage II to obtain the originally targeted number of events, or 3) adapt the design and increase the number of events in Stage II. The adaptation rule for the number of events is based on the conditional rejection probability approach (Muller and Schafer, 2001). In the latter two scenarios the number of subjects enrolled may also be increased in Stage II to ensure completion of the study in a

*Corresponding author.

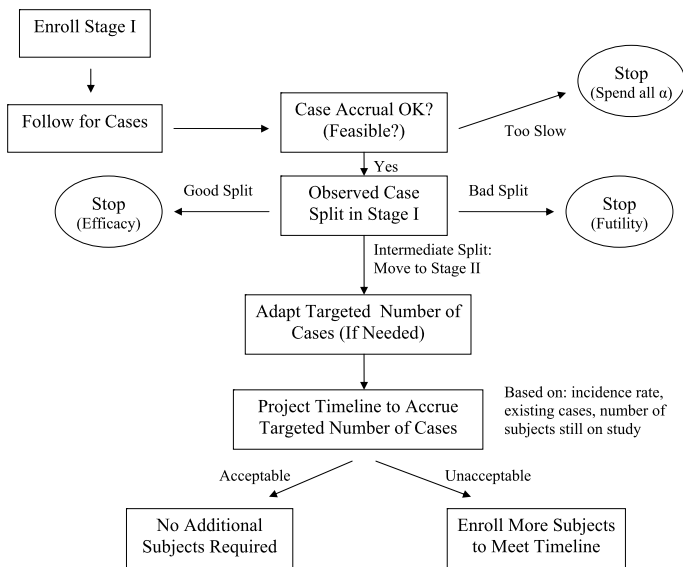


Figure 1. Design schema.

reasonable time-frame. In scenario 3), one or more interim analyses (IA) may be built in Stage II to provide additional opportunities to stop the trial early if the evidence of efficacy is either lacking or overwhelming. A schematic design of the study is illustrated in Fig. 1.

The paper is organized as follows. In Section 2, the design considerations and proposed adaptive design strategy are described. The operating characteristics of the proposed adaptive design with application to the motivating trial are evaluated using simulations in Section 3. Discussion is presented in Section 4.

2. DESIGN CONSIDERATIONS AND THE PROPOSED ADAPTIVE DESIGN STRATEGY

For event-driven vaccine efficacy studies such as the motivating trial, one commonly used statistical method for the efficacy analysis is the conditional exact binomial method based on the Poisson model (Chan and Bohidar, 1998). Respectively for the control (C) and vaccine (V) groups (assuming a 1:1 randomization ratio as in the motivating trial for simplicity), let n_C and n_V denote the total number of events, F_C and F_V denote the total follow-up time (person-years), λ_C and λ_V denote the background risk per unit time, and ψ denotes the relative risk ($\psi = \lambda_V/\lambda_C$). Then the vaccine efficacy is measured as $VE = 1 - \psi$. The values of n_C and n_V are generally assumed to follow Poisson distributions with corresponding parameters $F_C\lambda_C$ and $F_V\psi\lambda_C$ as 1) the incidence rate is usually very low and 2) the events are independent of each other. Then, given the total number of events $n = n_V + n_C$, the number of events in the vaccine group n_v follows a binomial distribution with parameter

$\pi = \frac{\Psi F_V}{F_C + \Psi F_V}$, i.e., $f(n_v; \psi, n, F_V, F_C) = \binom{n}{n_v} \pi^{n_v} (1 - \pi)^{n - n_v}$. The hypothesis testing and statistical inference about the vaccine efficacy are based on this binomial distribution. Note that this general framework also works when the randomization ratio is different from 1:1, because the difference in sample sizes will be reflected in the differential follow-up times (F_C and F_V).

For the motivating trial, 17 events would be needed to have a 90% power to reject the null hypothesis on vaccine efficacy ($VE = 1 - \psi$) $H_0 : VE = 0\%$ assuming an expected VE of 85% at a 1-sided 0.025 level; thus, an initial target accrual of 17 events was proposed. The proposed duration for Stage I is approximately 4 years or, if the incidence rate is high, shorter; the desired Stage II duration is 3 years or less. In consideration of this timeline and resource factors, the initial study enrollment was proposed to be 2,000 subjects, with study feasibility assessed in a blinded manner at frequent intervals during Stage I. The number of additional subjects to be enrolled in Stage II will be determined such that the desired number of cases is expected to be accrued within the preferred timeline; Stage II would be eliminated if events accrue slowly.

For the purpose of study planning, design evaluation and simulation, the following assumptions on enrollment rate and loss-to-follow-up are considered: 1) 15% of enrolled subjects will be seropositive at baseline and hence not evaluable for the primary efficacy analysis, 2) the timeline for Stage I is 48 months, and the timeline for Stage II is 36 months, with a 3 month pause for analysis and decision-making between Stages I and II, 3) the initial 2,000 subjects in Stage I will take up to 6 months to enroll, and in Stage II 500 subjects can be enrolled each month, 4) 7%/6%/5% per year lost-to-follow-up rate for 1st year/2nd year/3rd year and beyond.

2.1 Feasibility evaluation

As described above, one of the main challenges in designing the motivating trial is the high uncertainty in background incidence rate, and it is important for the feasibility to be evaluated during the early portion of the trial. Hence the initial cohort enrolled into the study will be used to assess study feasibility based on the observed incidence rate of the event, and to inform the total number of subjects and duration necessary for successful completion of the study in a timely manner. This assessment would include an option for additional enrollment of subjects in order to remain within acceptable timelines, as well as an option to stop the trial early (for infeasibility) if a low incidence rate suggests an unreasonably large number of subjects would be needed to meet the maximum acceptable timeline.

The feasibility assessments during stage I of the study will be performed in a blinded fashion. Blinded assessments is recommended because it will lower the risk of perceived negative effects on study integrity that would result from

Table 1. Feasibility evaluation under different scenarios with 2,000 initial subjects enrolled

Time of look, year from FPE	Number of events	Point estimate (95% CI) of Incidence rate (/100 person-years)	Time to 11 infections: year from FPE	Time to 17 infections: year from FPE
2 (~1200 person-years)	1	0.08 (0.002, 0.46)	13	>20
	2	0.17 (0.02, 0.60)	6.5	10
	3	0.25 (0.05, 0.73)	4.5	6.5
	4	0.33 (0.09, 0.85)	4	5.5
3 (~2800 person-years)	3	0.11 (0.02, 0.32)	10	>15
	4	0.15 (0.04, 0.37)	7.5	11.5
	5	0.18 (0.06, 0.41)	6.3	9
	7	0.25 (0.10, 0.52)	4.5	6.5
	9	0.32 (0.15, 0.61)	4	5
3.5 (~3600 person-years)	5	0.14 (0.05, 0.32)	7.5	>15
	7	0.19 (0.08, 0.40)	6	6.5
	9	0.25 (0.11, 0.47)	4.5	5.5

repeated unblinded looks at the data. It will also reduce the risk of operational bias and avoid Type I error allocation to any unblinded data analysis.

During the feasibility assessment phase in Stage I, the likelihood of accruing the target number of events for the proposed interim and final analyses within the desired study timelines and sample size will be evaluated based on the initially enrolled subjects. Specifically, the incidence rate will be estimated based on the pooled information, i.e., all the observed cases and the total follow-up from the study will be used to calculate the incidence rate and predict the future case accrual. In fact, under the assumption that the vaccine efficacy is consistent over the course of the trial, the blinded assessment of the incidence rate would give essentially the same projection of future case accrual as the unblinded assessment.

After the blinded feasibility assessment, there is an option to stop the trial early for infeasibility. For example, if the data in Stage I indicate that the incidence is so low that it may take a very long time (e.g., 4+ years) to accrue even 6 cases, then only one (final) analysis will be performed for the study with 6 cases at 1-sided 0.025 level, in which case $H_0 : VE = 0\%$ will be rejected only if no event out of 6 is in the vaccine group. Six (6) is selected as it is the minimum number of cases with which it is possible to reject the null hypothesis with a 0/6 case split.

The first time point at which a formal blinded feasibility evaluation may be performed is 24 months after the study start. The reason that an earlier time point, say 18 months after first-patient-entered (FPE), is not selected is that the total follow-up time available for assessment of incidence would be extremely limited at that time. Table 1 shows the point estimate of the incidence rate and the corresponding 95% confidence interval at different time points for evaluation, and the time predicted to accrue 11 (target for interim analysis) and 17 cases.

An observed incidence rate that is approximately 0.15/100 person-years or lower will suggest an unreasonably long time to accrue the required number of cases for the interim (11) and final (17) analyses, in which case the study may be judged infeasible. On the other hand, if the observed incidence rate is approximately 0.25/100 person-years or higher, it may take as long as planned to accrue the required number of cases, in which case the study would be considered feasible. If an intermediate incidence rate were observed, then additional subjects could be enrolled to complete the trial within the desired timeline.

The desired decision may not be as clear if the observed incidence rate falls in between 0.15 and 0.25 per 100 person-years. For example, if an incidence rate of 0.17/100 person-years is observed at 2 years after FPE, it implies that it may take up to 6.5 years post FPE to accrue 11 cases (a total of 6,700 person years of follow-up time needed) if only the originally enrolled subjects are followed. Since the total follow-up time that can be accumulated from the 2,000 originally enrolled subjects is approximately 4,200 person-years by 4.5 years post FPE, an additional 2,500 person-years of follow-up time is required to expedite the time to accrue 11 cases to approximately 4.5 years from FPE. Based on a rate of 0.17/100 person-years, it is estimated that an additional 4,000 subjects need to be enrolled, which will yield the additional follow-up time required over the 2.5 years period (from 2 years to 4.5 years post FPE). Given the low incidence rate, it would not be an easy decision to make as to whether to enroll an additional 4,000 subjects or to declare infeasibility of running the study. However, if feasibility is declared and an additional 4,000 subjects are enrolled during Stage I, they will contribute substantially to the study follow-up time and the case accrual in Stage II; it may take less than a year of additional follow-up to accrue an additional 6 cases for a total of 17 cases as originally planned.

2.2 Timing of interim analysis (IA)

Given the uncertainty around vaccine efficacy and the need to build risk mitigation into the study design, the inclusion of an IA is strongly desired to provide an opportunity to stop the trial early either if the evidence of efficacy is lacking or overwhelming. If study feasibility is established during Stage I, the study will continue and an interim analysis will be performed before embarking on Stage II. A key consideration for the IA is when (at what information fraction) it should be done. Generally it is recommended to have the first interim analysis when approximately 40% or more of planned data is available (Gould, 2005), so an informative decision can be made with a decent amount of information, and at the same time provide an opportunity to save a substantial investment if indeed the treatment is futile. Since the motivating trial is event-driven, the information corresponds directly to the number of events. Adaptation scenarios with different numbers of events were considered for the Stage I IA and carefully compared with each other by simulations (details not shown). Factors considered for the evaluation include the overall study power, time to interim analyses, overall study duration, strength of interim evidence, the input from key scientific opinion leaders, and likelihood of requiring additional subject enrollment for Stage II of the study. After consideration of all of the above factors, the final proposal was for the IA to be performed when 11 events were observed. Again, inclusion of this IA assumes that the feasibility of the study was demonstrated in Stage I. The study will only be stopped early for overwhelming efficacy if 100% vaccine efficacy (0:11 case split between vaccine and placebo recipients) is observed at the interim analysis.

2.3 Method for potential adaptation on total number of events

After the IA at the end of Stage I, one of the key decisions needed to be made is whether to adapt the total number of events required for the final analysis based on the observed case split between the vaccine and placebo groups, or equivalently the observed vaccine efficacy at the IA. Table 2 summarizes the potential outcomes of the IA. Futility will be declared if 5 or more events are observed in the vaccine group (corresponding to an observed vaccine efficacy of <17%). If the futility bound is not crossed (i.e., observing 4 or less cases in the vaccine group), then the total number of events may be increased such that the conditional power is at least 80% assuming a true vaccine efficacy of 75%. This 75% VE assumption used in adaptation, although lower than the 85% VE assumption expected for the vaccine used in the initial design, is considered the minimum vaccine efficacy profile that will be acceptable to the scientific community for the vaccine to be administered to a large population for prevention of the disease.

The adaptation of the number of events is based on the conditional rejection probability principle (CRP) by Muller

and Schafer, (2001). The basic concept is that for a group sequential trial with K looks at the information fractions t_1, \dots, t_K , at any interim look L , design parameters for the rest of the study may be changed (including the maximum sample size, the number of future interim looks, etc.), providing that the remainder of the trial preserves the CRP of rejecting H_0 , at look L . The method is further described below:

- Let z_L be the observed Wald statistic at look L . Then the CRP (ε_0) is defined as $\varepsilon_0 = \Pr(H_0 \text{ is rejected at any look } j(> L)|z_L, H_0)$.
- With any data dependent changes at the interim look L , the overall unconditional type-I error of the entire trial will be preserved, as long as the CRP for the modified trial beyond the change point remains ε_0 or lower, under the null hypothesis.

It is straightforward to extend this approach to an event-driven efficacy study, by using a single binomial distribution to model the number of events in the treatment group. The default design is assumed to have analyses after 11 and 17 endpoints, with a positive trial resulting if 0/11 or $\leq 4/17$ cases are in the vaccine group. This design has a Type I error rate of $p_1 = (0.5)^{11} = 0.0005$ for the first analysis and 0.0241 for the final analysis. Probabilities calculations for this design or more complex exact binomial group sequential designs can be computed using the `gsBinomialExact()` function from the `gsDesign` R package (<http://cran.r-project.org/>).

The corresponding adaptation strategy for the number of events is summarized in Table 2.

The additional number of events required is based on a search such that the conditional power exceeds 80% when the stopping boundaries are chosen to maintain the conditional error rate. Note that all the calculations below are based on equal follow-up times in the two groups. However, in reality the follow-up times for the two groups probably will not be exactly equal, but should be very close to equal due to the 1:1 randomization ratio and the low incidence rate of the event. The actual follow-up time will be accounted for in the analysis; the subtle difference in follow-up time between two groups is not expected to have a material impact on the analyses.

As shown in Table 2, at the first interim analysis, if $n_{v1} = 0$, then stop the trial for efficacy, i.e., $H_0 : VE = 0\%$ is rejected; else if $n_{v1} \geq 5$, then stop the trial for futility, i.e., $H_0 : VE = 0\%$ will not be rejected; or else if $n_{v1} = 1, 2, 3$ or 4 then continue the trial into stage II.

Conditional rejection probabilities are calculated based on the probability of ≤ 4 vaccine group events out of 17 total events given the number of vaccine group events observed after 11 total events. If there is no vaccine efficacy, the probability (conditional upon the total number of events) that an event occurs in the vaccine group will be 0.5, and this probability will reduce to 0.2 if the true vaccine efficacy is 75%. As an example, if $n_{v1} = 1$ (or 2) out of the first 11

Table 2. Summary of the adaption strategy for number of events based on 11 events at the first efficacy analysis

No. of vaccine cases in Stage I (CRP)	Additional cases targeted for Stage II		Total cases in vaccine group	Decision
0			-	Stop for efficacy
1 or 2 (0.6563 or 0.3438)	6		≤ 4 (out of $11 + 6 = 17$ cases)	Stop for efficacy
			> 4 (out of $11 + 6 = 17$ cases)	Stop for futility
3 (0.1094)	12		≤ 6 (out of $11 + 12 = 23$ cases)	H_0 rejected
			> 6 (out of $11 + 12 = 23$ cases)	H_0 not rejected
4 (0.0156)	24	Interim analysis at 12/24 cases available	≤ 5 (out of $11 + 12 = 23$ cases)	Stop for efficacy
			≥ 9 (out of $11 + 12 = 23$ cases)	Stop for futility
		Final analysis at 24/24 available cases	6, 7 or 8 (out of $11 + 12 = 23$ cases)	Continue for additional 12 cases
			≤ 10 (out of $11 + 12 + 12 = 35$ cases)	H_0 rejected
≥ 5			≥ 11 (out of $11 + 12 + 12 = 35$ cases)	H_0 not rejected
			-	Stop for futility

CRP – Conditional rejection probabilities.

events at Stage I, then the probability of observing ≤ 3 (or ≤ 2) events in the vaccine group out of the next 6 events in Stage II is 0.983 (or 0.901). Thus, the conditional power is $> 80\%$ in this case and no adaptation of the number of events would be needed to accrue $n_2 = n_{v2} + n_{p2} = 6$ additional cases in Stage II. The conditional rejection probability for Stage II is controlled (at 0.6563 for $n_{v1} = 1$ or 0.3438 for $n_{v1} = 2$) under the null hypothesis so that the unconditional type I error rate is controlled at the one-sided 0.025 level.

If $n_{v1} = 3$, then with the original design at most 1 of the next 6 cases could be in the vaccine group for the trial to be positive, corresponding to a conditional error probability under the null hypothesis of 0.1094. In order to have $> 80\%$ conditional power with 75% vaccine efficacy, $n_2 = n_{v2} + n_{p2} = 12$ additional cases will need to be accrued in Stage II. At the final analysis if $n_{v1} + n_{v2} \leq 6$ then $H_0 : VE = 0\%$ will be rejected. In this case ($n_{v1} = 3$, $n_2 = 12$) the conditional rejection probability for Stage II of 0.073 is less than the required 0.1094 under the null hypothesis, and the conditional power (based on $VE = 75\%$) of 79.5% is approximately 80%.

If $n_{v1} = 4$, then $n_2 = n_{v2} + n_{p2} = 24$ additional cases will need to be accrued in Stage II. Another interim analysis will be performed when $n_{2I} = n_{v2I} + n_{p2I} = 12$ out of $n_2 = 24$ cases are available, with the final analysis to be conducted (if necessary) when all $n_2 = 24$ cases are available, as the following:

- if $n_{v2I} \leq 1$ then stop the trial for efficacy, i.e., $H_0 : VE = 0\%$ will be rejected,
- else if $n_{v2I} \geq 5$, then stop the trial for futility, i.e., $H_0 : VE = 0\%$ will not be rejected,
- else if $n_{v2I} = 2, 3$ or 4, then continue the trial for another $n_{2F} = n_{v2F} + n_{p2F} = 12$ cases,

in which case at the end of Stage II if $n_{v2I} + n_{v2F} \leq 6$ out of $n_2 = 24$ cases then reject $H_0 : VE = 0\%$, otherwise, do not reject $H_0 : VE = 0\%$. Here the conditional rejection

probability for Stage II is controlled (at 0.0156) under the null hypothesis, and the conditional power (based on $VE = 75\%$) would be $\sim 80\%$.

Note that the precise design above was chosen from many possible candidate designs of the same type based on simulation studies presented below, after considering the following criteria: the timing to interim analysis, the strength of evidence at the interim analysis, overall study duration, and overall design characteristics in terms of power under different scenarios and the maximum number of subjects that may be required.

2.4 Sample size re-estimation for Stage II

As mentioned before, the proposed duration for Stage II is approximately 3 years. The number of additional subjects to be enrolled in Stage II will be determined such that the desired number of cases is expected to be accrued within the preferred timeline; this number would be 0 if no sample size increase is needed.

In terms of the maximum study enrollment, we evaluated a range from 7,500 to 10,000 subjects. The comparison indicates that there is unlikely to be any significant incremental value of increasing the maximum sample size to 10,000 from 7,500.

Based on the observed incidence rate in Stage I and the required number of cases for Stage II (with or without adaptation), and assuming a vaccine efficacy of 75%, it can be projected, based on the initially enrolled (2,000) subjects, as to how long it will take to accrue the additionally required cases in Stage II from the end of Stage I. If it is expected to take at most approximately 3 years, then there will be no adaptation (increase) in sample size, and only the originally enrolled 2,000 subjects will be followed for further case accrual. However, if it is predicted to take more than 3 years, then additional subjects may be enrolled into Stage II (again in a 1:1 randomization ratio) to meet the desired timeline.

Table 3. Sample size re-estimation for Stage II per outcomes from Stage I

Timing of Stage I analysis (yr from FPI)	Placebo follow-up time in Stage I (pyrs)	Number of placebo cases in Stage I*	Viral infection incidence rate (/100 pyrs)	Additional number of cases required, per design	Time to targeted number of cases w/o increasing sample size (yr from Stage I)	Additional no. of subjects needed to complete the study within 3 years from Stage I	Additional no. of subjects needed to complete the study within 4 years from Stage I
3	1050	7	0.67	24	4.5 (>3)	2000	500
		8, 9, or 10	≥ 0.76	≤ 12	≤ 2 (<3)	0	0
3.5	1400	7	0.5	24	6.5 (>3)	4000	2000
		8, 9, or 10	≥ 0.57	≤ 12	≤ 2.5 (<3)	0	0
4	1750	7	0.4	24	9 (>3)	6000 [†]	3000
		8	0.46	12	3.5 (>3)	500 [‡]	0
		9 or 10	≥ 0.51	6	≤ 1.5 (<3)	0	0
4.5	2100	7	0.33	24	12 (>3)	8000 [†]	4000
		8	0.38	12	4.5 (>3)	1500	500
		9 or 10	≥ 0.43	6	≤ 2 (<3)	0	0

*If the number of cases observed in the placebo group during Stage I is 11 or fewer than 7, then the study may be stopped due to efficacy or futility and not move into Stage II, per design.

[†]Exceeds the maximum allowance of 7500, hence may not increase enrollment or extend the timeline for Stage II.

[‡]It may not be worth it or efficient to enroll additional subjects.

The number of additional subjects to be enrolled will be determined such that it will take approximately 3 years from the end of Stage I to accrue the additionally required events in Stage II.

The details on how the recommendation would be made are summarized in Table 3 based on various outcomes from Stage I under a range of scenarios (timing of Stage I analysis, Stage I outcome, preferred timelines for Stage II). The assumptions made for the calculations include: 1) 15% of enrolled will be ‘event’ positive at baseline and hence excluded from the primary efficacy analysis, 2) 7%/6%/5% per year lost-to-follow-up rate for 1st year/2nd year/3rd year and beyond, 3) a vaccine efficacy of 75%.

Note that when sample size needs to be increased in Stage II, the number of additional subjects needed would depend on how long the study team is willing to wait for the end of study from the end of Stage I. A preferred timeline of 3 years from Stage I for study completion was assumed in the table, based on previous decisions. For illustration purposes, a timeline of 4 years from Stage I for completion of the study was also included. It is shown that in the event that the sample size needs to be increased, if one is willing to wait one additional year (4 years instead of 3 years for Stage II), the additional sample size required could be substantially reduced (by at least 50%). The additional number of subjects to complete the study within shorter timelines (no less than 2 years due to time for enrollment and criteria for negativity at Day 1 and Month 7) can also be calculated if desired, but the number of additional subjects required, if any, to achieve that goal would be much larger than with a longer timeline, since it would not be efficient to enroll addi-

tional subjects who would only be followed for a short period of time (due to the time required for enrollment, criteria for negativity testing, and evaluation of the viral infection over an extended duration for case confirmation, etc.).

3. OPERATING CHARACTERISTICS OF THE PROPOSED ADAPTIVE DESIGN STRATEGY

To evaluate the characteristics of the proposed adaptive design, simulations are used due to the complexity of the design. Factors considered in the simulation studies include the feasibility evaluation, different incidence rates and vaccine efficacies; and the results are summarized in Table 4. The assumptions on enrollment and loss-to-follow-up summarized in Section 2 were used in the simulation.

As shown in Table 4, the overall study type I error rate is well controlled below the nominal level of 1-sided 2.5%, regardless of the incidence rate. This is aligned with the fact that the exact conditional method is generally conservative. The study power increases with increase in true vaccine efficacy or incidence rate. The main reason that the study power increases with incidence rate can be explained as follows: when the incidence rate is lower, there is a bigger chance for infeasibility, or even when feasibility is established a bigger chance for the trial not to proceed with a potentially prohibitive large sample size, which would result in lower power to reject the null efficacy hypothesis.

Also shown in Table 4 is the probability that the sample size is increased during the trial. It is of interest to note that the sample size is unlikely to be increased in

Table 4. Overall study power of the proposed design accounting for feasibility evaluation (2,000 initial enrollment, 7,500 maximum enrollment), based on simulations (5,000 iterations)

VE	Incidence rate (/100 pyrs)	Power (%) (1-sided 2.5% level)	Expected total number of subjects	% the sample size is increased	Expected number of extra ¹ subjects	Average study duration (months)	Percentage ² of stopping at (percentage of stopping for efficacy)		
							Stage I ³	Stage II interim ⁴	Stage II
0%	0.2	1.7	2200	7%	2850	67	90% (0%)	9% (1%)	0% (0%)
	0.5	1.9	2260	12%	2100	43	76% (0%)	21% (2%)	3% (0%)
	0.8	1.7	2050	4%	1070	34	76% (0%)	22% (2%)	3% (0%)
80%	0.2	45	2100	3%	3310	82	82% (28%)	18% (17%)	0% (0%)
	0.3	64	2270	10%	2680	81	53% (20%)	46% (43%)	0% (0%)
	0.4	79	2300	14%	2200	77	33% (16%)	66% (62%)	1% (1%)
	0.5	87	2300	14%	2140	69	22% (14%)	75% (71%)	3% (2%)
	0.8	91	2160	8%	2080	51	16% (13%)	79% (74%)	5% (4%)
90%	0.2	63	2040	1%	4130	84	88% (51%)	12% (12%)	0% (0%)
	0.3	78	2080	3%	2840	77	67% (45%)	33% (33%)	0% (0%)
	0.4	89	2090	4%	2300	72	49% (39%)	50% (50%)	0% (0%)
	0.5	96	2080	4%	2060	65	42% (38%)	58% (58%)	0% (0%)
	0.8	99	2040	2%	1630	49	35% (35%)	65% (64%)	0% (0%)

¹Additional subjects to be enrolled, given the sample size is increased for stage II.

²Percentages are rounded to the closest single digit.

³Either at the end of Stage I or when 6 cases are accrued if infeasibility is indicated.

⁴This will be the end of Stage II unless 24 more cases are required such that a second interim analysis is needed as shown in Table 2.

the following situations: 1) when incidence rate is very low, the trial may be terminated early due to lack of feasibility, i.e., the required sample size is prohibitively large to achieve the targeted conditional power; 2) when incidence rate is high, fast accrual of events is possible so that the trial can be completed without sample size increase; 3) when the true underlying vaccine efficacy is high, no adaptation (increase) in the targeted number of events would be required.

The furthestmost right panel in Table 4 shows the probability of stopping the trial at each stage, and the corresponding probability of stopping the trial for efficacy (in parenthesis). With the adaptive design, it can be seen that in almost every scenario, the trial did not continue to the end of stage II. When there is no vaccine efficacy ($VE = 0\%$), the trial is most likely to be stopped at the end of Stage I. When the vaccine efficacy is high ($VE = 80\%$ or 90%), the trial will be more likely to be stopped at the end of Stage I or during the additional interim analysis in Stage II.

4. DISCUSSION

Due to the high uncertainty in the incidence rate of the event of interest, it was challenging to design the trial discussed here in an informative way based on a more traditional approach. Two such approaches considered, namely 1) committing a large sample size up-front so it is guaranteed that the trial will be completed within a desired time frame, and 2) having a pilot study first to better estimate

the incidence rate to inform the design of a subsequent efficacy trial, are not efficient in addressing the issue. Hence, an adaptive design strategy was proposed, which allows for feasibility evaluation, interim analysis and potential adaptation on the number of events to be accrued and/or number of subjects to be enrolled, while controlling the overall study type I error rate. Simulations demonstrate that the proposed design works well by stopping early if indeed the trial is ‘infeasible’ due to either a very low incidence rate, or a combination of low incidence rate and vaccine efficacy such that a prohibitively large number of subjects may be required for the trial to be completed within a desired time frame. In addition, if the feasibility is established and the trial does move forward into Stage II, it often limits the number of subjects in the trial, sparing resources and time.

Note it is important to point out again that the feasibility evaluation during the Stage I be done in a blinded manner; this evaluation can be performed continuously, if needed. In practice, it will not be preferred to perform such an evaluation too early during the trial before adequate data are available to make an informative decision. Also it may not be efficient and necessary to perform such an evaluation on an on-going basis. As shown in Table 1, it would be sufficient to perform such an evaluation at a few specified time points.

As in most, if not all, adaptively designed trials, there are many operational challenges, probably more so for the design strategy proposed here. First, the study team needs to have a well-designed strategy to collect data, particularly the number of events, during Stage I and an efficient channel

for communication for a timely evaluation/decision. Knowing that any interim analysis has been performed and the trial is continued provides some partial information on the observed vaccine efficacy due to the extreme results that would have required halting the trial. The sample size for the remainder of the trial also provides information. One potential strategy to mask as much information as possible is not to communicate the detailed decision strategy to the study sites so that study personnel and subjects are not aware of the interim outcome. In addition, the sample size adaptation for Stage II could be done in increments of 500 (as implemented in the simulations), and the final number of subjects to be enrolled needs not to be communicated.

In development of this adaptive design, the authors have made comparisons with group sequential designs. Details are not presented here to save space. Such comparisons are recommended in the draft FDA guidance on adaptive design (<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf>) in order to justify adaptations based on unblinded interim results. While the group sequential options considered were reasonably efficient, the 2-stage adaptive design presented here was selected after considering tradeoffs in statistical and operational efficiency to balance the upfront resource requirement (number of subjects enrolled, site selection, financial commitment, etc) and a high likelihood to complete the pivotal phase of efficacy trial in a reasonable timeframe.

Another important aspect to consider for an adaptively designed trial is how to estimate and infer the treatment effect accounting for the adaptation. The point estimate of vaccine efficacy can be obtained as the observed value, which is straightforward to interpret and communicate with the scientific communities. Although this 'naïve' estimator may be biased due to interim analysis and/or adaptation, it is shown analytically (Li et al., 2010) that the bias is minimum based on the current study design when the true vaccine efficacy lies in the region of interest (>60%).

To obtain confidence intervals following an adaptive design, one can use the stage-wise ordering method (Jennison and Turnbull, 2000). First, the outcome space is ordered by whether the null hypothesis is rejected, the stage the trial is stopped, and the observed vaccine efficacy; inferences are then based on the sufficient statistic triplet (K, S_K, N_K) , where $K \in \{1, 2, 3\}$ is the stage at which the study stops, S_K is the total number of cases in the vaccine group, and N_K is the total number of cases. More specifically, $K = 1$ if the study stops at the end of Stage I; $K = 2$ if the study stops at the end of Stage II when there is no 2nd interim analysis or stops at the interim analysis of Stage II; and $K = 3$ if the study stops at the

end of Stage II when there is a 2nd interim analysis (in Stage II). Let (k, s, n) and (k', s', n') are two potential realizations of (K, S_K, N_K) , then $(k', s', n') \succ (k, s, n)$ if any of the following conditions holds: 1) the hypothesis is rejected with (k', s', n') but not (k, s, n) ; 2) the hypothesis is rejected with both (k', s', n') and (k, s, n) , and $k' < k$; 3) the hypothesis is not rejected with either (k', s', n') or (k, s, n) , and $k' > k$; or 4) $k' = k$ and $s'/n' < s/n$ (i.e., ties can be broken by observed vaccine efficacy). Based on the above ordering of the study outcome, a one-sided p-value is given as $\Pr_{p=p_0}\{(K, S_K, N_K) \geq (k^*, s^*, n^*)\}$, where (k^*, s^*, n^*) is the observed outcome, i.e., the trial stops at stage k^* with s^* cases in the vaccine group and a total of n^* cases. And the confidence interval (p_L, p_U) can be obtained as $\Pr_{p=p_L}\{(K, S_K, N_K) \geq (k^*, s^*, n^*)\} = \alpha/2$, and $\Pr_{p=p_U}\{(K, S_K, N_K) \leq (k^*, s^*, n^*)\} = \alpha/2$. It can be shown that whenever the null hypothesis is rejected based on the design, the p-value is < 0.025 and the lower bound of the CI is > 0 , and the confidence interval obtained based on this approach for the design has the correct (conservative) coverage (Li et al., 2010).

Received 08 November 2011

REFERENCES

- CHAN, I. S. F. and BOHIDAR, N. R. (1998). Exact power and sample size for vaccine efficacy studies. *Communications in Statistics, Theory and Methods* **27** 1305–1322.
- DUFFY, D. E. and SANTNER, T. J. (1987). Confidence intervals for a binomial parameter based on multistage tests. *Biometrics* **43** 81–93.
- FDA Draft Guidance for Industry: Adaptive design clinical trials for drugs and biologics. February 2010.
- FURTH, A. F. and SARGENT, D. J. (2006). Duffy-Santner confidence interval for the two-stage three-outcome design. *Journal of Biopharmaceutical Statistics* **16** 875–880. [MR2297411](#)
- GAYDOS, B., ANDERSON, K. M., BERRY, D., BURNHAM, N., CHUANG-STEIN, C., DUDINAK, J., FARDIPOUR, P., GALLO, P., GIVENS, S., LEWIS, R., MACA, J., PINHEIRO, J., PRITCHETT, Y. and KRAMS, M. (2009). Good practices for adaptive clinical trials in pharmaceutical product development. *Biostatistics* **43** 539–556.
- GOULD, A. L. (2005). Timing of futility analysis for 'proof of concept' trials. *Statistics in Medicine* **24** 1815–1835. [MR2140943](#)
- JENNISON, C. and TURNBULL, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*. Chapman and Hall/CRC, Boca Raton. [MR1710781](#)
- LI, X., CHAN, I. S. F. and ANDERSON, K. M. (2010). Statistical Inference Following an Adaptive Design for Case-Driven Vaccine Efficacy Study. In *Proc. of ASA JSM*, Vancouver, Canada.
- MULLER, H. H. and SCHAFER, H. (2001). Adaptive group sequential designs for clinical trials: Combining the advantages of adaptive and of classical group sequential approaches. *Biometrics* **57** 886–891. [MR1859823](#)
- Xiaoming Li
 Gilead Sciences
 199 E. Blaine Street
 Seattle, WA 98102
 USA
 E-mail address: xiaoming.li@gilead.com

Ivan S. F. Chan
Merck Research Laboratories
351 N. Sumneytown Pike
North Wales, PA 19454
USA
E-mail address: ivan_chan@merck.com

Keaven M. Anderson
Merck Research Laboratories
351 N. Sumneytown Pike
North Wales, PA 19454
USA
E-mail address: keaven_anderson@merck.com