SAMPLE SIZE ESTIMATION AND POWER CALCULATIONS FOR VACCINE EFFICACY TRIALS FOR EXCEEDINGLY RARE DISEASES

by

Matthew M. Loiacono

BA, Biology, Franklin & Marshall College, 2015

Submitted to the Graduate Faculty of

Department of Biostatistics

Graduate School of Public Health in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

GRADUATE SCHOOL OF PUBLIC HEALTH

This thesis was presented

by

Matthew M. Loiacono

It was defended on

April 11th, 2018

and approved by

Thesis Advisor:
Hanna Bandos, PhD
Research Assistant Professor, Biostatistics
Graduate School of Public Health
University of Pittsburgh

Committee Members:
Robert Krafty, PhD
Associate Professor, Department of Biostatistics
Graduate School of Public Health
University of Pittsburgh

Willem G. van Panhuis, MD PhD
Assistant Professor, Epidemiology
Assistant Professor, Biomedical Informatics
Affiliated Faculty, Public Health Dynamics Lab
School of Medicine and Graduate School of Public Health
University of Pittsburgh

SAMPLE SIZE ESTIMATION AND POWER CALCULATIONS FOR VACCINE EFFICACY TRIALS FOR EXCEEDINGLY RARE DISEASES

Matthew M. Loiacono, MS

University of Pittsburgh, 2018

ABSTRACT

The implementation of vaccines is one of the most profound advancements in the world of public health. While the scientific process underlying vaccine development is important, determining a vaccine's efficacy and safety profile would not be possible without vaccine efficacy trials. In terms of trial design, sample size estimation is one of the essential considerations to be made, ensuring that the trial achieves the targeted power, minimizing chances of a false-negative finding.

An often considered hypothesis for such trials is testing if a new vaccine is more efficacious than a placebo or the current standard-of-care. We investigated two relevant methods, the Z-Test Normal Approximation and the Exact Conditional Test under Poisson Assumption as presented by Chan and Bohidar. While the Normal Approximation is a generally acceptable method, the Exact Conditional approach may be better suited for trials with low disease incidence.

In this Thesis, we develop SAS code for estimating sample size via both methods, in addition to calculating exact power. Sample size and power calculations are performed under a range of different scenarios, stratified by high, intermediate, and low incidence of disease in the control group. Furthermore, under each scenario, simulations are utilized to evaluate the performance of the Exact Conditional approach.

Both methods are generally reasonable approaches to sample size estimation and power calculations for vaccine efficacy trials. In trials of higher disease incidence, the Normal Approximation is simpler to implement, and should adequately power the trial. For scenarios of low incidence (≤ 0.01), the Exact Conditional approach may be more favorable, as it may provide greater power to the trial. To minimize the chance of an underpowered trial, we propose implementing a strategic inflation of sample size. Lastly, we highlight the importance of congruency between methods used for sample size estimation, trial design, and data analysis.

Public Health Significance: Vaccine efficacy trials are required for vaccine licensure and distribution to the public. A robustly estimated sample size is crucial, as it ensures a trial is adequately powered, thereby improving the odds of obtaining conclusive results. The methods investigated here represent viable approaches to sample size estimation for such trials.

TABLE OF CONTENTS

PREFACE ..... **IX**

1.0 INTRODUCTION **1**

1.1 HISTORICAL OVERVIEW **1**

1.2 MODERN DAY AND CLINICAL TRIALS **3**

1.3 VACCINE EFFICACY TRIALS AND SAMPLE SIZE ESTIMATION **5**

2.0 METHODS **8**

2.1 POWER AND ALPHA LEVEL OF THE TEST **8**

2.2 HYPOTHESIS TESTING **8**

2.3 Z-TEST NORMAL APPROXIMATION **10**

2.4 EXACT CONDITIONAL TEST UNDER POISSON ASSUMPTION **12**

2.5 COMPUTATION AND SOFTWARE **14**

2.6 METHODOLOGY EVALUATION **15**

3.0 RESULTS **18**

4.0 DISCUSSION **26**

APPENDIX: SAS CODE **30**

BIBLIOGRAPHY **35**

LIST OF TABLES

Table 1. Sample size estimates and power calculations for scenarios of high incidence (P_1=0.15, π_1 = 0.65). 21

Table 2. Sample size estimates and power calculations for scenarios of intermediate incidence (P_1=0.05, π_1 = 0.65)..... 22

Table 3. Sample size estimates and power calculations for scenarios of low incidence (P_1=0.01, π_1 = 0.65). 23

LIST OF FIGURES

Figure 1. Sample size estimates versus pre-specified power (%). 24

Figure 2. Sample size estimates under varying levels of vaccine efficacy lower bound (a) and true vaccine efficacy (b). 25

PREFACE

Firstly, I would like to thank my Thesis advisor and mentor, Dr. Hanna Bandos, for her consistent guidance and support throughout this endeavor. The research and organizational skills I further developed under her guidance will certainly benefit my future studies and work. Similarly, I want to thank Dr. Wilbert van Panhuis and Dr. Robert Krafty for serving on my committee and for their feedback throughout this research. Additionally, I would like to thank my supervisors and mentors from my Summer 2017 Internship at Sanofi Pasteur, who inspired and helped direct the initial sample size work that ultimately led to this Thesis (Dr. Ayman Chit, Dr. Vitali Pool, Dr. Ed Thommes, Dr. Kerry Go, and Dr. Alex Selmani). Furthermore, I want to extend a warm thank you to my family, friends and loved ones for their constant support and motivation. On a lighter note, lastly, I would like to acknowledge both Constellation Coffee and Crazy Mocha Coffee Co., where a large portion of the code and writing for this Thesis were produced, for consistently providing excellent coffee and a work-friendly atmosphere.

1.0 INTRODUCTION

1.1 HISTORICAL OVERVIEW

Throughout the history of public health, there have been a number of particularly notable achievements that have not only altered the course of public health history for the better, but more importantly, have drastically improved health at an individual and population level. Some of these advancements and discoveries made in the 20th century include, but are not limited to, the fluoridation of drinking water, control of infectious diseases, standardization of motor-vehicle safety, and implementation of workplace safety protocols. These advancements, alongside many others, have made for an overall healthier population, with greater quality and quantity of life, all while dramatically reducing associated health care costs (Porter, 1999). However, aiding in the near-eradication of countless harmful diseases, the wide-scale implementation of vaccines is certainly one of the most profound advancements in the world of public health.

Although vaccines are frequently considered to be an advancement associated with the aforementioned 20th century public health advancements, the first known use of a vaccine, as we know them, dates back to the late 18th century. Edward Jenner, a country doctor in rural England, performed what we now regard as one of the first legitimate clinical trials. Jenner discovered that, by inoculating an individual with pus from a cowpox lesion, immunity to any subsequent exposures of small pox could be developed (Jenner, 1798; Stern & Markel, 2005). However, he

was not the first individual to attempt creating immunity via inoculation. The process of inoculation can be traced to the 15th century across parts of Asia, where subjects were inoculated with the live small pox virus. This particular method, though, was not only more dangerous, due to use of the live virus, but also less effective at producing immunity to repeated exposures (Plotkin, 2011). Jenner's method of inoculation, using a less dangerous variant of the virus, marked a turning point in the history of immunization, allowing us to create immunity well beyond the first or second exposure to a virus. His discovery, ultimately, laid the groundwork for the future of immunizations and vaccines.

Over the course of the following century, the small pox vaccine, as defined by Jenner, would go on to be used throughout Europe and the United States (Strassburg, 1982). Worthy of noting, the word vaccine, as Jenner had used it, was derived from the term *Variolae vaccinae*, or cowpox (Jenner, 1798). According to the World Health Organization, as of 2018, "a vaccine is a biological preparation that improves immunity to a particular disease… made from weakened or killed forms of the microbe, its toxins, or one of its surface proteins" (WHO, 2018). With that said, Jenner's smallpox vaccine would not meet the modern-day definition of a vaccine.

Nearly 100 years after the work of Jenner, it was the work of French chemist Louis Pasteur and his colleague, Emile Roux, that helped broaden and shape the definition of vaccines. Pasteur and Roux developed the first rabies vaccine by attenuating rabies grown in live rabbits in a lab setting. First tested on canines, and then on a small number of people over the course of the 19th century, their vaccine revealed its incredibly high success rate with minimal side effects (Geison, 2014). While this vaccine would now be more accurately described as an antitoxin, it nevertheless helped expand the definition of vaccines to include treatments beyond Jenner's cowpox vaccine.

Although many others contributed to the historical development of vaccines, Jenner, Pasteur, and Roux are worthy of highlighting, due to their particularly influential contributions.

1.2 MODERN DAY AND CLINICAL TRIALS

As of 2018, there are nearly 100 vaccines approved for use in the United States, with an estimated 21 million hospitalizations and 700,000+ deaths having been prevented between 1994-2014, due to vaccines (CDC, 2014; FDA, 2018). Similarly, an estimated 2-3 million deaths are prevented annually across the globe as a result of vaccine implementation. The realized benefits of vaccination as a form of preventive care, in terms of overall population health as well as associated health care costs, are not to be underestimated (WHO, 2017). Although the scientific discovery process that underlies vaccine discovery and development is certainly fundamental to bringing these vaccines to the public, the rigorous testing protocol to determine their effectiveness as well as any potential side effects is paramount to their licensure, distribution, and successful implementation.

In order for any new pharmaceutical product to be released to the general public, each product must be assessed over a series of tests, or clinical trial phases. Prior to the initiation of a clinical trial for a new product, however, certain preclinical studies must first be performed. Primarily relying on animal testing, preclinical studies are intended to determine a safe baseline dosage for the first phases of the clinical trials, while also providing a preliminary assessment of the product's safety. Species are chosen to ensure maximum relatability to humans and in doing so, safe baseline dosages can be estimated allometrically (Atanasov et al., 2015). Following preclinical studies, phases 0-IV of the clinical trial introduce human subjects and allow for further

assessment of product safety, dosing, and efficacy. Phase 0, introduced by the FDA in 2006, is an exploratory investigational study in which a small number of subjects (usually less than 20) receive a micro-dose of the product to demonstrate possible therapeutic benefits. This phase serves as a way of quickly separating effective from ineffective new products, and is not required for product licensure (Fromer, 2006). More traditionally, clinical trials begin with phase I, a study intended to establish the dose and schedule of a drug, conducted on a small cohort of volunteers. Design for phase I trials varies based on the drug and the results from the preclinical studies, but one of the most common methods is the "3 + 3" dose-escalation design. Differences in particular designs aside, phase I trial designs expose all subjects to the drug and aim to implement escalating doses and varied scheduling to ultimately determine the recommended phase II dose, in addition to the drug's safety (Hansen, Graham, Pond, & Siu, 2014; Le Tourneau, Lee, & Siu, 2009).

While phase 0 and I trials establish the baseline dosing, safety, and scheduling, a phase II trial, conducted in up to several hundred individuals, serves to further investigate the new drug's efficacy, dosing, and safety profile, in addition to potentially adverse effects. In general, the most commonly used design in phase II trials is a single-arm trial. However, in some cases, a randomized control design may also be implemented, where the control group receives the current standard-of-care, if it exists (Mandrekar & Sargent, 2010).

After a successful phase II trial, a large scale phase III trial is conducted to more precisely assess these same characteristics across a larger group of patients. Phase III trials follow a randomized control design, and generally include anywhere from 250-3,000 individuals, with varying levels of risk. It is possible that the size of these trials can even reach upwards of 10,000 individuals. In terms of treatment, in some special cases, individuals may continue to receive treatment outside of the trial to maintain the therapeutic benefits while the drug is undergoing

licensing. This particular type of trial is referred to as phase IIIB trial (Friedman, Furberg, DeMets, Reboussin, & Granger, 2015). More generally, a phase III trial helps investigators to get a much more accurate estimate of the drug's true efficacy. Following a successful phase III trial, a drug receives its licensure and is then ready to be released to its target market or the general public.

Once a drug has received licensing and complete the phase III trial, a phase IV trial can take place. As opposed to the experimental nature of phase 0-III trials, a phase IV trial is an observational trial in nature. Although not all phase IV trials are the same, the majority take shape as post-marketing surveillance. Phases 0-III trials help to identify most adverse effects, but in the case of excessively rare adverse effects, the phase IV trial is crucial to identifying these events (Mitchell, Philipose, & Sanford, 1993; Suvarna, 2010). Semantics aside, this generalized outline of clinical trials is applicable to nearly all new pharmaceutical products, including the vast majority of vaccines.

1.3 VACCINE EFFICACY TRIALS AND SAMPLE SIZE ESTIMATION

While assessing the safety profile and dosing schedule of a new vaccine are both important aspects of its development, the primary objective in conducting clinical trials is to establish the efficacy of the vaccine in question. In establishing the efficacy, more specifically, investigators aim to demonstrate that the test vaccine reduces the incidence of a particular disease, relative to a placebo or the standard-of-care. In terms of the physiological components, preclinical studies and phase I trials are useful to establish these characteristics, therefore setting the stage for phases II-IV. However, from a statistical standpoint, one of the vital calculations to be made for a vaccine

efficacy trial is the sample size estimation, that helps ensure the pre-specified statistical power is achieved, thus minimizing the chance of a false negative finding.

Prior to sample size estimation, the desired minimum power for the trial must first be specified, usually by the lead investigator(s) and/or statisticians involved in the trial design. Similarly, the significance level of the test must also be specified during the earlier trial design stages, where the level of the test is indicative of the maximum allowed chance of a Type I error rate, or false-positive rate. After establishing the desired level of the test, power for the trial, and anticipated effect size, the expected required sample size can be estimated. This calculation is crucial for a number of reasons, and over or underestimation can have a range of negative consequences. An incorrectly calculated sample size, where the sample size is underestimated, can result in the specified power never being achieved, potentially leading to inconclusive results. As phase II-IV clinical trials become larger and more costly, an overestimated sample size, conversely, could incur unnecessarily high expenses for the organization leading the clinical trial (Sakpal, 2010).

Different approaches exist to calculating a sample size for a vaccine efficacy trial, depending on the hypothesis being tested. In regards to the classical test of comparing two proportions or incidences, appropriate sample size formulas have been derived (Breslow & Day, 1987). However, these methods cannot account for a non-zero bound for the value of the null vaccine efficacy, and are thus limiting in terms of the scope of their applicability. Different test statistics have been proposed for this precise situation, and corresponding asymptotic sample size calculations have also been derived (Chan & Bohidar, 1998; Chow, Shao, & Wang, 2003; Farrington & Manning, 1990; Gart & Nam, 1988; Miettinen & Nurminen, 1985). Although these methods take into account a non-zero lower bound, they may not be optimal for trials with a large

sample size and excessively low disease incidence. For this type of scenario, an exact conditional test where the number of cases is assumed to follow a Poisson distribution has been previously presented by Breslow and Day, and further modified by Chan and Bohidar to account for hypotheses with a non-zero lower bound (Bohidar & Chan, 1996; Breslow & Day, 1987; Chan & Bohidar, 1998).

In this paper, we investigated the sample size estimation methods as presented by Chan and Bohidar, including the Z-Test Normal Approximation and the Exact Conditional Test under Poisson Assumption (Chan & Bohidar, 1998). In Section 2.1, the concepts of statistical power and level of the test are reviewed. In Section 2.2, the specific hypothesis of interest for a vaccine efficacy trial is established. Two approaches to power and sample size calculation are presented in Sections 2.3 and 2.4. In Section 2.5, a description is provided of the software used to program both of these methods, as well as our algorithm to determine the correct required number of cases for the Exact Conditional approach. In Section 2.6, the chosen scenarios for projecting sample size are discussed. Our results are presented in Section 3. Lastly, in Section 4, our findings are summarized, the importance of continuity in trial design and analysis is highlighted, and a potential strategy to help avoid under-powering a trial is proposed.

2.0 METHODS

2.1 POWER AND ALPHA LEVEL OF THE TEST

Two vital parameters that must be specified, prior to sample size estimation, are the significance level of the test and the minimum desired statistical power for the trial. In specifying the level of the test, the investigator is establishing the maximum chance of a Type I Error, or false-positive, which is denoted as (α). Generally, the level of the test is designated within the range of 2.5% to 5%. The power of the study is the likelihood to detect a protective effect of the vaccine, if the vaccine does in fact have a protective effect. We denote power as (1-β), where (β) is the probability of making a Type II Error, or a false-negative. Clinical trials are generally designed to achieve a minimum of 80% power (Chow et al., 2003). Studies can be designed to have a targeted power greater than 80%, but this will usually result in a substantial increase in the sample size.

2.2 HYPOTHESIS TESTING

For the simplicity of the presentation, all of the formulas here are presented for a case of equal sample sizes between the test and control groups. Let us define n as the number of patients in each group, P_1 as the incidence of disease in the control group, and P_2 as the incidence of disease in the test vaccine group, where $P_2 \leq P_1$. We then denote vaccine efficacy (π) as (1 – relative risk) of test subjects having the disease relative to control subjects:

(1) $$\pi = 1 - (P_2 / P_1)$$

In the case of a vaccine efficacy trial, where a pre-existing vaccine is already the standard-of-care, we generally refer to the new vaccine's expected efficacy as the relative vaccine efficacy, a term that is frequently used in vaccine efficacy studies of this nature (Ashkenazi et al., 2006; Comstock, 1994; DiazGranados et al., 2014; Storsaeter & Olin, 1992). Therefore, the relative vaccine efficacy value is not the test vaccine's true efficacy, but rather, it is the test vaccine's efficacy in relation to the control vaccine's efficacy (Shim & Galvani, 2012). In either case, whether we are working with a relative vaccine efficacy of the new test vaccine and standard-of-care, or simple vaccine efficacy compared to a placebo, the general process of hypothesis testing, power calculation, and sample size estimation remain the same.

In a vaccine efficacy trial, we are interested in determining whether or not a new vaccine is more efficacious than the control. Specifying a minimum vaccine efficacy level of the new vaccine, denoted as π_0, the hypothesis used to test that the new vaccine is more efficacious than the control is represented as follows:

(2) $$H_0\colon \pi \leq \pi_0 \quad versus \quad H_1\colon \pi > \pi_0$$

For a trial with measurements of relative vaccine efficacy, π_0 would be set equal to zero, as the new vaccine's efficacy is relative to that of the control. If the control group is entirely unvaccinated, then π_0 would be set to a value greater than zero, to test whether or not the new vaccine's efficacy is greater than π_0, thus demonstrating some level of efficacy.

2.3 Z-TEST NORMAL APPROXIMATION

For the Z-Test Normal Approximation method, we used the test statistic that has been previously established, for the given hypothesis (2) (Chan & Bohidar, 1998; Farrington & Manning, 1990; Miettinen & Nurminen, 1985):

(3) $$Z = \frac{\hat{P}_2 - (1-\pi_0)\hat{P}_1}{\hat{\sigma}_0/\sqrt{n}}$$

(4) $$\hat{\sigma}_0 = [\widetilde{P_2}(1-\widetilde{P_2}) + (1-\pi_0)^2\,\widetilde{P_1}(1-\widetilde{P_1})]^{1/2}$$

Here, small values of the Z statistics provide evidence against the null hypothesis. In this case, $\hat{P}_1$ and $\hat{P}_2$ are the observed incidence in the control and test groups, respectively, and $\widetilde{P_1}$ and $\widetilde{P_2}$ are the maximum likelihood estimates of P_1 and P_2, within the confines of the null hypothesis (Chan & Bohidar, 1998). In this test statistic, *n* represents the sample size of either the control or test group, in accordance with our assumption of equal sample sizes between the test and control groups.

Based on this test statistic, asymptotic formulas for power and sample size have also been previously derived (Chan & Bohidar, 1998). Firstly, the limiting values of $\widetilde{P_1}$ and $\widetilde{P_2}$ can be denoted as $\overline{P_1}$ and $\overline{P_2}$. Using the planned incidence in the test and control vaccine groups, P_1 and P_2, the respective values of $\overline{P_1}$ and $\overline{P_2}$ can be determined as follows:

(5) $$\overline{P_2} = \{a - [a^2 - 8b]^{1/2}\}/4 \quad \textit{and} \quad \overline{P_1} = \overline{P_2}/(1-\pi_0)$$

where, the values for *a* and *b* in this expression can be calculated as such:

(6) $$a = (1-\pi_0)(1+P_1)+1+P_2 \quad \textit{and} \quad b = (1-\pi_0)(P_1+P_2)$$

Consequently, $\overline{P_1}$ and $\overline{P_2}$ can then be used in place of $\widetilde{P_1}$ and $\widetilde{P_2}$ in (4) to determine the limiting value of $\hat{\sigma}_0$, denoted as $\overline{\sigma_0}$:

(7) $$\overline{\sigma_0} = [\overline{P_2}(1-\overline{P_2}) + (1-\pi_0)^2\,\overline{P_1}(1-\overline{P_1})]^{1/2}$$

Furthermore, σ_1 can be calculated by replacing $\widetilde{P_1}$ and $\widetilde{P_2}$ with P_1 and P_2 in (4):

(8) $$\sigma_1 = [P_2(1-P_2) + (1-\pi_0)^2\,P_1(1-P_1)]^{1/2}$$

With $\overline{\sigma_0}$ and σ_1 defined, the asymptotic power for a Z-Test of α level to test our given null hypothesis (2), can be determined, where Φ is the standard Normal distribution function:

(9) $$1-\beta = \Phi\{\frac{1}{\sigma_1}\left[Z_{1-\alpha}\overline{\sigma_0} + n^{1/2}P_1(\pi_1-\pi_0)\right]\}$$

Based on this formula for asymptotic power of the test, the total sample size, *N*, can then be determined as follows:

(10) $$N = 2[\left(Z_{1-\alpha}\overline{\sigma_0} + Z_{1-\beta}\sigma_1\right)^2 / \left(P_1(\pi_1-\pi_0)\right)^2]$$

2.4 EXACT CONDITIONAL TEST UNDER POISSON ASSUMPTION

If the disease of interest is rare and has a low incidence rate, a large scale trial may be required. In this particular situation, an exact conditional test, where the number of cases in the study can be approximated by a Poisson distribution, has been previously proposed (Breslow & Day, 1987; Gail, 1AD). Slightly modified by Chan and Bohidar to account for a null hypothesis with a non-zero lower bound, this Exact Conditional Test procedure, is presented below (Bohidar & Chan, 1996; Chan & Bohidar, 1998).

Let X and Y represent the number of cases in the control and test vaccine groups, where they independently follow Binomial distributions, BIN(n, P_1) and BIN(n, P_2), respectively. In the case of a rare disease, where n is adequately large and P_1 and P_2 are adequately small, we can assume that $nP_1 \rightarrow \lambda_1$ and $nP_2 \rightarrow \lambda_2$. Therefore, X and Y can be approximated as POI(λ_1) and POI(λ_2), respectively. The total accumulated number of cases observed during the course of the trial, $X+Y$, is denoted as T. Consequently, the appropriate distribution of Y (number of cases in the test vaccine group), conditional upon $T = t$, for any $t \in \{0, 1, 2, 3\ldots, T\}$, can be derived as follows:

$$X \sim POI(\lambda_1) \text{ and } Y \sim POI(\lambda_2), \text{where } X \text{ and } Y \text{ are independent}$$

$$Therefore, T = X + Y \sim POI(\lambda_1 + \lambda_2)$$

$$P(Y = k | T = t) = \frac{P(Y = k, T = t)}{P(T = t)}$$

$$\frac{P(Y = k, X + Y = t)}{P(T = t)} = \frac{P(Y = k, X = t - k)}{P(T = t)}$$

$$= \frac{\frac{\lambda_2^k}{k!} * \frac{\lambda_1^{t-k}}{(t-k)!}}{\frac{(\lambda_1 + \lambda_2)^t}{n!}} = \binom{t}{k} \frac{\lambda_2^k \, \lambda_1^{t-k}}{(\lambda_1 + \lambda_2)^n} = \binom{t}{k} \left(\frac{\lambda_2}{\lambda_1 + \lambda_2}\right)^k \left(\frac{\lambda_1}{\lambda_1 + \lambda_2}\right)^{t-k}$$

$$Let\ \Theta = \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{nP_2}{nP_1 + nP_2} = \frac{1-\pi}{2-\pi}$$

Substituting Θ into the final step of the derivation above, the conditional distribution of Y given T is then shown to follow a Binomial distribution:

(11) $$Y \mid T \sim BIN(T, \Theta) = \binom{T}{k}\Theta^k(1-\Theta)^{T-k}$$

To accommodate for the decreasing nature of Θ, our hypothesis (2) can then be represented accordingly:

(12) $$H_0\colon\ \Theta \geq \Theta_0 \quad versus \quad H_1\colon\ \Theta < \Theta_0$$

where $\Theta_0 = (1-\pi_0)/(2-\pi_0)$. If Y_{obs} is the number of cases observed in the test vaccine group, the exact conditional p-value can then be calculated as:

(13) $$p = \Pr[Y \leq Y_{obs} \mid Y \sim \mathrm{BIN}(T, \Theta_0)] = \sum_{k=0}^{Yobs}\binom{T}{k}\Theta_0^k(1-\Theta_0)^{T-k}$$

Similarly, given our hypothesis (2) against a specified alternative, the exact power and critical value (Y_c) at a pre-specified level (α) can be calculated as such:

(14) $$1-\beta = \Pr[Y \leq Y_c \mid Y \sim \mathrm{BIN}(T, \Theta_1)] = \sum_{k=0}^{Yc}\binom{T}{k}\Theta_1^k(1-\Theta_1)^{T-k}$$

$$where\ \ \Theta_1 = (1-\pi_1)/(2-\pi_1)$$

With a minimum value of *T* determined for the level of the test that will also achieve the desired minimum power, a vaccine efficacy trial can then be designed around accruing a specified total number of cases. The corresponding total sample size (*N*) for the vaccine efficacy trial can be calculated by the following formula:

(15) $$N \approx 2 * \frac{T}{(2-\pi_1)P_1}$$

2.5 COMPUTATION AND SOFTWARE

All formulas and methodology, as described here, were coded with SAS® software (Version 9.4, Cary, NC). Aside from assuming equal sample sizes between the control and test groups, all formulas were coded exactly as presented by Chan and Bohidar, and are provided in the Appendix (Chan & Bohidar, 1998). Using what we refer to as a "brute force" type algorithm, we were able to determine the correct minimum number of cases to be accrued over the trial (*T*), as required for the Exact Conditional approach, and then the projected sample size. The logic of this procedure is described below:

1) Consider all values of *T* from 1 to 1000
2) Determine a critical value (Y_c) for each respective *T*, if it exists
    a) Using equation (14) and considering all Y values $\in$ {0, 1 ,2, 3…, T}, Y_c is the minimum value at which power is $\geq$ the pre-specified minimum power
    b) Any values of *T* without a determined Y_c value are no longer considered

3) Level of the test, for each combination of T and its respective Y_c value is considered
    a) Only values of T where exact level, using equation (13), is < the pre-specified *alpha* are considered
4) Determine the minimum required value of T from this reduced set of T's
    a) The minimum T at which all following T's increase by a value of 1 without skipping
        i) Due to the discrete nature of the Binomial distribution, we want to ensure that if an extra case is obtained, power does not drop below the pre-specified minimum power
5) Determine the expected sample size for the trial using equation (15)

2.6 METHODOLOGY EVALUATION

In their work, Chan and Bohidar compared the statistical power and size of the Exact Conditional and Normal Approximation approaches, as well as compared their performance, in terms of the sample size requirements. In order to further assess the performance of these methods, we expanded upon the set of vaccine efficacy trial scenarios considered in their paper. A scenario, as we denoted it, can be described by three primary variables: the incidence of disease in the control group (P_1), true vaccine efficacy (π_1), and the vaccine efficacy lower bound (π_0).

In real-world scenarios, disease incidence can range drastically, depending on a large number of factors. In their paper, Chan and Bohidar suggested that the Exact Test has substantially less power than the Z-Test when the incidence rate in the control group is ≥ 0.3. Therefore, we focused on the rates less than 0.3, stratifying by high, intermediate, and low incidence scenarios. The three values of disease incidence of the control group used in this paper were based upon what

we considered to be realistic values that might be observed today, for a range of diseases. The high incidence value was baselined off of the world-wide incidence of malaria in 2000 (~0.155), an incidence much higher than observed today, as this was expected to adequately represent a realistic upper bound for incidence (World Bank, 2018). For the low incidence scenario, an incidence of 0.01 was used, as this was low enough to demonstrate differences between the two methods, while still producing realistically viable sample size estimates.

Similarly, vaccine efficacy values range widely, from 0% up to nearly 100% in real-world trials. After reviewing a number of different vaccine efficacy trials, we decided to fix the true efficacy at an intermediate value of 65%. In addition to a zero lower-bound for the null efficacy, a lower-bound of 15% was also considered, as this represented what we found to be a realistic lower-bound in a vaccine efficacy trial design (Alonso et al., 1994; Aronson et al., 2004; Cassio de Moraes et al., 1992; Coursaget et al., 1986; Fine et al., 1994; Gross, et al., 1995; Kulkarni et al., 2017; Rerks-Ngarm et al., 2009).

With the scenarios defined as such, the changes in sample size requirements were investigated, while simultaneously increasing the minimum pre-specified power, from 80% to 95%, in increments of 5%, as well as two values for the significance level of the test, 2.5% and 5.0%. The minimum sample size required under these scenarios was evaluated by the Exact Conditional and Normal Approximation approaches, as described in Sections 2.3 and 2.4.

Furthermore, the validity of the Exact Conditional approach to sample size estimation was investigated via simulations. Empirical power was calculated under the same set of scenarios described above, allowing us to evaluate the performance of the Exact Conditional approach, in terms of power. Briefly, the simulations were performed as follows. The number of cases in the control and treatment groups was simulated independently, based on BIN(n, P_1) for the control

subjects and BIN(n, P_2) for the test subjects, where n is 1/2 of the sample size for the scenario, as estimated via the Exact Conditional approach. For each simulated trial, a p-value was calculated using formula (13). Simulating 2,000,000 trials per scenario, statistical power was then calculated as the proportion of those trials in which the null hypothesis was rejected at the pre-specified level for that particular scenario.

3.0 RESULTS

Estimates of the sample sizes under a range of hypothetical vaccine efficacy trial scenarios are presented in Tables 1, 2, and 3. Within these tables, we presented identical sets of scenarios, except for different disease incidence in the control group (P_1), stratified by high, intermediate, and low disease incidence.

Regardless of the specific incidence in the control group, there were a number of overarching patterns that arose across the three scenarios. First, the exact power, number of total required cases, and exact level remained the same across Tables 1, 2, and 3. As explained in Section 2.5, these components were reliant upon the null and true vaccine efficacy values, which were identical across the three primary incidence scenarios, thus explaining this particular pattern.

When specifying a smaller alpha level under the same scenario, estimated sample size increased. Similarly, when specifying a greater minimum power for two otherwise identical scenarios, estimated sample size again increased. This observation was logical, as both decreasing alpha and increasing power serve to provide one with greater certainty in the findings of a trial, consequently, requiring a greater number of subjects to be accrued. To confirm the trend in the relationship between power and sample size, additional scenarios were considered (Figure 1).

In terms of the vaccine efficacy values, an increase in the efficacy lower bound resulted in an increase in sample size for both methods, across all three disease incidence scenarios. To further investigate this relationship, as well as the effect of varying the true vaccine efficacy value (π_1), an additional set of scenarios was considered, where either the efficacy lower bound or true vaccine efficacy were changing, while keeping all other parameters fixed (Figure 2). In Figure 2a, with the

true efficacy fixed, increasing the efficacy lower bound translated to a decreased effect size and therefore, increased sample size. On the contrary, in Figure 2b, with the efficacy lower bound fixed, increasing the true efficacy translated to an increased effect size and therefore, decreased sample size. These two observations succinctly demonstrated the generally expected relationship between effect size and sample size, where sample size estimates increase as effect size decreases.

The results of the simulations are also presented in Tables 1-3. Within this identified set of scenarios, the empirical power was almost always less than the exact power for the targeted number of cases. This can be attributed to the fact that, when using the two Binomial distributions to simulate our data, many of the "trials" resulted in fewer than the targeted number of cases (T). In each of those trials, the exact power could be substantially lower than the targeted power. Although theoretically it is possible that there are scenarios in which the empirical (unconditional) power is higher than the targeted power, we can generally expect that in regular scenarios when the targeted range of statistical power is 70-90%, unconditional power would be lower than exact.

Furthermore, there are some particular patterns that arose as a result of changes in the disease incidence of the control group. For the case of high incidence, sample size estimates were smaller overall. More specifically, comparing a fixed scenario across any two tables, the sample size estimates of the Exact Conditional approach increased proportional to the factor decrease in the control incidence. Referencing the Exact Conditional sample size formula (15), the sample size only relies upon the control incidence via P_1 in the denominator, as neither the minimum required number of cases (T) nor the true efficacy (π_1) rely upon control incidence. On the other hand, while a similarly inverse relationship was observed between the Normal Approximation sample size and control incidence, this relationship was not exactly proportional as observed with

the Exact Conditional approach. Likewise, in referencing equations (6) through (10), the relationship between sample size and P_1 cannot be as easily disentangled.

The differences in sample size estimates of the Exact Conditional approach and the Normal Approximation method were similarly reflected by the efficiency measure in Tables 1-3. Within our designated set of scenarios, efficiency values were always less than 100%, indicating that the estimated sample sizes via the Normal Approximation were predominantly less than those from the Exact Conditional approach. On average, efficiency was overall lowest when incidence was highest. In comparing Table 1 and Table 2, there were some individual scenarios where this does not hold true, but this is likely attributed to the discrete nature of the Binomial distribution used for the Exact Conditional approach. The higher efficiency observed in Table 3 suggests that, under a scenario of low incidence, sample size estimates using the Exact Conditional approach should be relatively close to those when using the Normal Approximation approach.

Table 1. Sample size estimates and power calculations for scenarios of high incidence (P_1=0.15, π_1 = 0.65).

Efficacy Lower Bound	Pre-specified Alpha	Pre-specified Power	Exact Level	Exact Power	Simulated Power	Exact Conditional Test		Z-Test	Efficiency (Z-Test/Exact)
						Total # of Cases	Expected Sample Size	Asymptotic Sample Size	
(π_0)	(%)	(%)	(%)	(%)	(%)	(T)	(N)	(N)	(%)
0	2.5	80.0	2.35	86.2	82.4	37	366	300	82.0
0	2.5	85.0	2.18	89.6	87.8	42	416	342	82.2
0	2.5	90.0	2.00	92.1	91.6	47	466	400	85.8
0	2.5	95.0	2.20	96.2	95.8	56	554	492	88.8
0.15	2.5	80.0	2.45	85.2	82.1	51	504	422	83.7
0.15	2.5	85.0	1.94	86.9	86.4	57	564	482	85.5
0.15	2.5	90.0	2.02	92.1	91.8	67	662	562	84.9
0.15	2.5	95.0	2.02	96.3	96.4	82	810	690	85.2
0	5.0	80.0	4.36	83.4	80.3	28	278	236	84.9
0	5.0	85.0	4.01	87.7	86.5	33	326	274	84.1
0	5.0	90.0	4.03	92.8	92.3	40	396	326	82.3
0	5.0	95.0	4.27	96.7	96.3	49	484	410	84.7
0.15	5.0	80.0	4.05	85.6	84.4	44	436	332	76.2
0.15	5.0	85.0	4.11	89.0	88.1	49	484	384	79.3
0.15	5.0	90.0	4.13	91.6	91.1	54	534	456	85.4
0.15	5.0	95.0	4.49	96.0	95.6	66	652	572	87.7

Table 2. Sample size estimates and power calculations for scenarios of intermediate incidence (P_1=0.05, π_1 = 0.65).

Efficacy Lower Bound	Pre-specified Alpha	Pre-specified Power	Exact Level	Exact Power	Simulated Power	Exact Conditional Test		Z-Test	Efficiency (Z-Test/Exact)
						Total # of Cases	Expected Sample Size	Asymptotic Sample Size	
(π_0)	(%)	(%)	(%)	(%)	(%)	(T)	(N)	(N)	(%)
0	2.5	80.0	2.35	86.2	81.4	37	1098	968	81.4
0	2.5	85.0	2.18	89.6	86.8	42	1246	1106	86.8
0	2.5	90.0	2.00	92.1	90.7	47	1394	1294	90.7
0	2.5	95.0	2.20	96.2	95.1	56	1660	1600	95.1
0.15	2.5	80.0	2.45	85.2	81.1	51	1512	1364	81.1
0.15	2.5	85.0	1.94	86.9	85.5	57	1690	1556	85.5
0.15	2.5	90.0	2.02	92.1	91.0	67	1986	1814	91.0
0.15	2.5	95.0	2.02	96.3	95.8	82	2430	2232	95.8
0	5.0	80.0	4.36	83.4	79.0	28	830	762	79.0
0	5.0	85.0	4.01	87.7	85.5	33	978	886	85.5
0	5.0	90.0	4.03	92.8	91.4	40	1186	1054	91.4
0	5.0	95.0	4.27	96.7	95.7	49	1452	1332	95.7
0.15	5.0	80.0	4.05	85.6	83.3	44	1304	1072	83.3
0.15	5.0	85.0	4.11	89.0	87.2	49	1452	1242	87.2
0.15	5.0	90.0	4.13	91.6	90.2	54	1600	1474	90.2
0.15	5.0	95.0	4.49	96.0	95.0	66	1956	1854	95.0

Table 3. Sample size estimates and power calculations for scenarios of low incidence (P_1=0.01, π_1 = 0.65).

Efficacy Lower Bound	Pre-specified Alpha	Pre-specified Power	Exact Level	Exact Power	Simulated Power	Exact Conditional Test		Z-Test	Efficiency (Z-Test/Exact)
						Total # of Cases	Expected Sample Size	Asymptotic Sample Size	
(π_0)	(%)	(%)	(%)	(%)	(%)	(T)	(N)	(N)	(%)
0	2.5	80.0	2.35	86.2	80.9	37	5482	4980	90.8
0	2.5	85.0	2.18	89.6	86.4	42	6224	5696	91.5
0	2.5	90.0	2.00	92.1	90.3	47	6964	6666	95.7
0	2.5	95.0	2.20	96.2	94.8	56	8298	8244	99.4
0.15	2.5	80.0	2.45	85.2	80.7	51	7556	7014	92.8
0.15	2.5	85.0	1.94	86.9	85.1	57	8446	7998	94.7
0.15	2.5	90.0	2.02	92.1	90.6	67	9926	9330	94.0
0.15	2.5	95.0	2.02	96.3	95.6	82	12150	11488	94.6
0	5.0	80.0	4.36	83.4	78.6	28	4150	3924	94.6
0	5.0	85.0	4.01	87.7	85.1	33	4890	4562	93.3
0	5.0	90.0	4.03	92.8	91.1	40	5926	5434	91.7
0	5.0	95.0	4.27	96.7	95.5	49	7260	6864	94.6
0.15	5.0	80.0	4.05	85.6	82.9	44	6520	5512	84.5
0.15	5.0	85.0	4.11	89.0	86.9	49	7260	6388	88.0
0.15	5.0	90.0	4.13	91.6	89.9	54	8000	7582	94.8
0.15	5.0	95.0	4.49	96.0	94.8	66	9778	9540	97.6

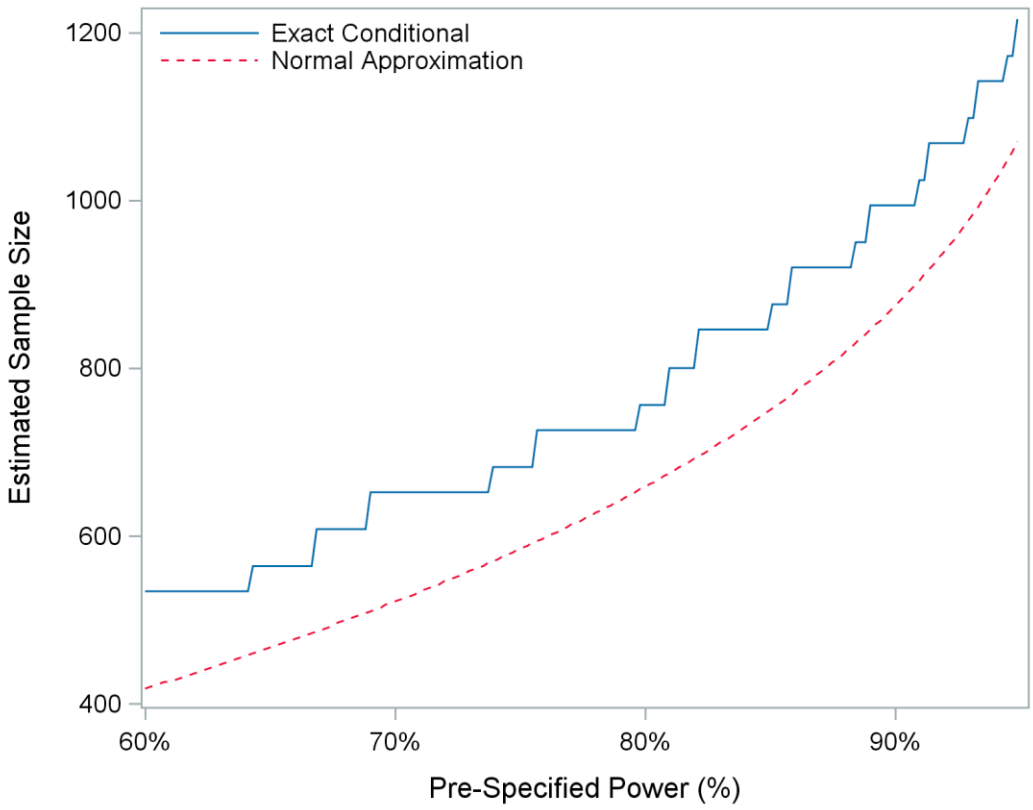

Figure 1. Sample size estimates versus pre-specified power (%).

Scenario parameters fixed as follows: *P1=0.01*, π_0*=0.15* π_1*=0.65*, *Alpha=5%*

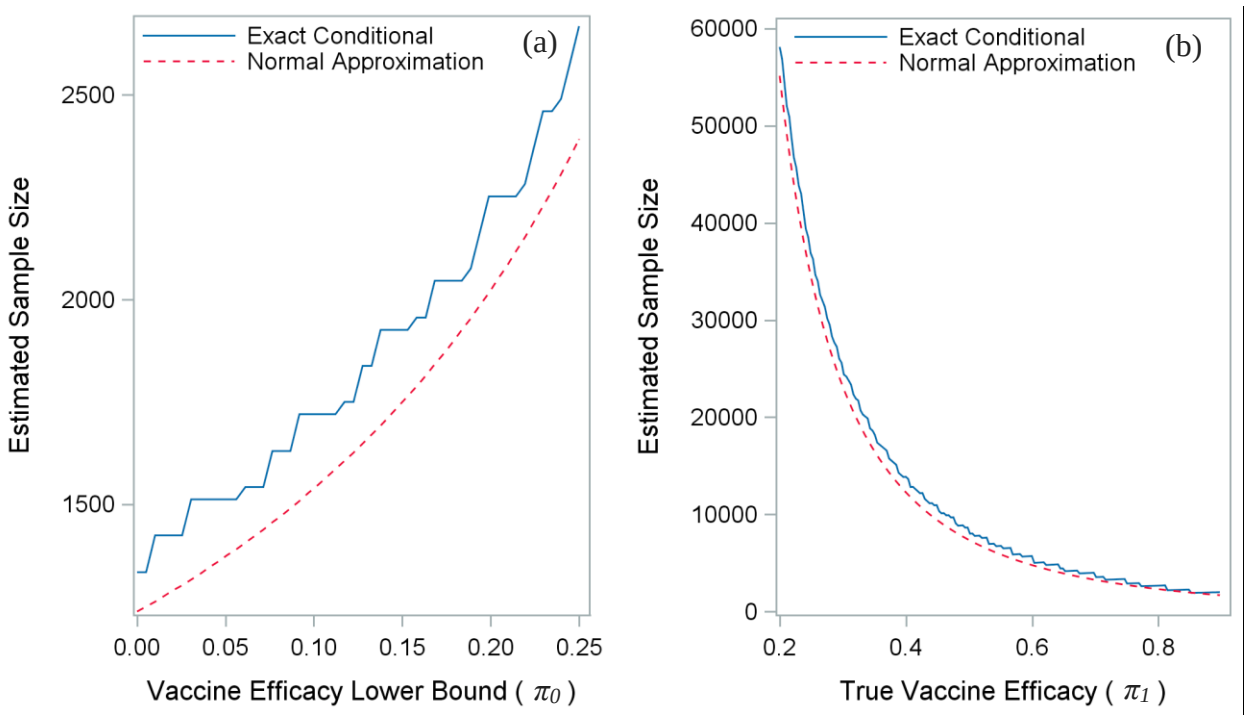

Figure 2. Sample size estimates under varying levels of vaccine efficacy lower bound (a) and true vaccine efficacy (b).

Scenario parameters fixed as follows: (a): *P1=0.01*, π_0*=0.00, Alpha=5%, Power=80%*
(b): *P1=0.01*, π_1*=0.65, Alpha=5%, Power=80%*

4.0 DISCUSSION

In planning and designing a vaccine efficacy trial, both the Exact Conditional approach and Normal Approximation approach are viable methods to sample size estimation, each with their own advantages and disadvantages. The Normal Approximation procedure represents a reasonable and less specialized approach, requiring minimal programming and computational complexity. This technique is also conveniently available in some statistical packages, such as PASS (NCSS LLC, Kaysville, Utah). Therefore, the simplicity of implementing it is one of its strong advantages. The Exact Conditional method, on the other hand, requires more initial programming to be performed, but is still relatively fast to execute on a modern-day computer.

In terms of sample size requirements, we found that the Exact Conditional method provided estimates consistently larger than those of the Normal Approximation, particularly, up to 25% larger in scenarios of higher incidence. These differences, however, became substantially smaller under low incidence scenarios, with most being under 10%, but no greater than ~16%. Our observations agreed with those by Chan and Bohidar. They concluded that, under scenarios of low incidence and fixed large sample sizes (~4,000+), the Exact Conditional Test performed similarly to the Z-Test, in terms of power, and for some cases, tended to be even more powerful (Chan & Bohidar, 1998).

Despite the involved complexity of the Exact Conditional approach, there are some considerations to be made that may lend it some favorability. From a practicality and a trial design standpoint, this approach may be more favorable, due to its conditionality upon a required number of cases. With such an approach, a trial would be designed to accrue a fixed number of events, and a situation where the pre-specified statistical power is not achieved would therefore be avoided.

While a given trial is in motion, these total number of cases can be easily monitored, effectively serving as an indicator of the progress of the trial. Depending on the timeline and trial schedule, decisions to end the trial early or extend it, in order to accrue more cases, can be supported with a quick reference to this easily accessible value. On the contrary, under the fixed sample size/duration trial design, there is the possibility that the trial may be underpowered, due to a lower incidence rate in the control group than anticipated. An important and sometimes overlooked aspect of a biostatistician's job lies in his or her ability to clearly communicate the rationale of their design and analysis plan to non-statistician investigators on the trial. Correspondingly, the required total number of cases condition and its utility as an indicator of trial's progress may be a more intuitive and easily understood measure.

While either of the methods we have investigated are generally reasonable for determining a sample size that should adequately power a trial, careful considerations should be given to the congruency between the methods used for planning, design, and final analysis of the collected data. As illustrated by the results of our simulations, inconsistencies between these major stages of a clinical trial could lead to underpowered trials. After evaluating the required sample size using the conditional approach, we simulated our trial data unconditionally via two independent Binomial distributions, with probabilities equal to the incidence rates in the control and test groups. Our approach to the simulation was more representative of the real-world, where trials are designed to accrue and observe a fixed number of subjects, and then run for a fixed amount of time. Such inconsistence between our method of sample size estimation and the design of the simulations led to a notable number of "trials" under each scenario accruing less than the required number of cases. This then resulted in nearly all of the scenarios we presented having a simulated power that was less than the exact power. This suggests that in a real-world situation, given the Exact Conditional

sample size estimate, a fixed sample size/duration trial may likely be underpowered, due to the minimum number of cases condition not being satisfied.

In practice, it is also possible that sample size may be estimated using the Normal Approximation, yet, in the analysis stage, the Exact Conditional Test may be used. This might present a problem, similar to what was observed in our simulations, with differences in exact versus empirical power. Considering that the Normal Approximation sample size estimates, within the scenarios we defined, were consistently less than those of the Exact Conditional approach, these trials would be significantly less likely to accrue the necessary number of cases. Thus, if analyzed with the Exact Conditional Test, there is a high probability that the trial will be underpowered. It is therefore crucial that the method used for sample size estimation matches that which is used for the data analysis.

As discussed earlier, prolonging a trial is a possible solution to accruing more cases when the required number of cases condition is not met, but this may not always be a logistically or financially viable option. In an effort to avoid under-powering a trial, we propose a potential inflation of the sample size estimate be used during the trial design stage of a fixed number of events trial to ensure that, with a given probability, the required number of cases will be observed (Flahault, Cadilhac, & Thomas, 2005). The specific probability used, however, should be further explored by investigators, in order to best determine a sample size that is manageable and reasonable, given the budget and other constraints of the trial. In combination with careful considerations of congruency in the trial design and analysis methodology, this will help to ensure that a trial achieves its desired power, allowing investigators to confidently draw conclusions.

In designing a clinical trial, one of the essential considerations to be made is the sample size estimation, which targets the pre-specified statistical power, minimizing the chance of a false-

negative conclusion. Incorrectly estimated sample sizes could lead to unnecessary costs or inconclusive study results. In this paper, we assessed the practicality of the Exact Conditional Test under Poisson Assumption and Z-Test Normal Approximation approaches to sample size estimation and power calculation for vaccine efficacy trials (Chan & Bohidar, 1998). In trials where incidence in the control group is high, the Normal Approximation is a reasonable and easy to implement approach that should adequately power the trial. For trials with low incidence ($\leq$ 0.01), the Exact Conditional approach may be more favorable, as it may provide greater power to the trial, at the cost of a marginal increase in sample size. To increase the likelihood that the required number of cases is accrued in a trial, we advise the use of a strategic sample size inflation, thereby decreasing the chances of an underpowered trial. Lastly, to better ensure the trial's targeted power is achieved, it is important that the methodology used for sample size estimation, study design, and data analysis are congruent.

APPENDIX: SAS CODE

```
*************************************************************************
MACRO: Vaccine Efficacy Trial Sample Size using the Exact Conditional Test
*************************************************************************

-Calculates the exact power and sample size for a vaccine efficacy trial,
using an Exact Conditional Test under Poisson Assumption
-All formulas are based upon (Chan & Bohidar, 1998)
-Assumes equal sample size between test and control groups
-Vaccine efficacy here is defined as (1-relative risk) of participants in the
test group having the disease, relative to the control group
-This method defines the minimum # of cases required to achieve the pre-
specified power, where the trial is then designed around accruing >= this #
of total cases

*Required inputs are as follows (only values between 0 and 1 are accepted):

      P1     = incidence in the control group
      VE1    = true efficacy of test vaccine, under the alternative
      VE0    = minimum efficacy of test vaccine, under the null (NOTE: must
be less than VE1)
      Power  = pre-specified power desired to achieve
      Alpha  = pre-specified maximum one-sided level of the test
      Output = name of temporary dataset to where the results will be output

When this MACRO is run, a temporary dataset will be output, along with a
printout of this dataset;

%MACRO SampleSize_ExactConditional(P1=, VE1=, VE0=, Power=, Alpha=, Output=
);
%LET c = 1;
OPTIONS SPOOL;
/*Calculates and assigns theta values as macro variables for the power
calculation based on input VE0 VE1 */
DATA _NULL_;
      teta0=(1-&VE0)/(1+&c-&VE0);
      teta1=(1-&VE1)/(1+&c-&VE1);
      CALL SYMPUT("theta0", teta0);
      CALL SYMPUT("theta1", teta1);
RUN;

/*Loops through all possible T's (where T is total # of cases accumulated) 1-
1000, calculates the power and level at ALL y(obs) values*/
DATA temp;
      RETAIN VE1 P1 T Yc power level  n2;
      VE1=&VE1;
      P1=&P1;
      VE0=&VE0;
      alpha=&alpha;           *Alpha is the nominal significance level;
      powerreq=&power;  *powerreq is the required power as specified;
            DO T=1 TO 1000;
```

```
                Yc=0;
                level=0;
                power=0;
                DO x=1 TO T;
                        Yc=x;
                        power = PROBBNML(&theta1,T,x);      *power is
the exact power of the test;
                        level = PROBBNML(&theta0,T,x);      *level is
the exact level of the test;
                        n2=T/((&c+1-&VE1)*(&P1));
                        N=ceil(n2)*2;
                        OUTPUT;
                END;
                OUTPUT;
        END;
DROP x;
RUN;

/*Removes any observations were level is > pre-specified ALPHA */
DATA temp;
      SET temp;
      IF (level <= &alpha AND power >= &power);
RUN;

/*Keeps 1st observation PER T where minimum power is reached (this is the
critical value Yc) */
DATA temp;
      SET temp;
      BY t;
      IF first.t=1;
RUN;

/* Creates FLAG1 variable where flag1=1 if current T is previous T+1 */
DATA temp;
      SET temp;
      IF T = (LAG1(T)+1) THEN flag1=1;
      ELSE flag1=0;
RUN;

/* Creates 20 lag variables (F1-F20) based on the flag1 variable */
DATA temp;
      SET temp;
      F1 = LAG1(FLAG1); F2 = LAG2(FLAG1); F3 = LAG3(FLAG1); F4 = LAG4(FLAG1);
F5 = LAG5(FLAG1);
      F6 = LAG6(FLAG1); F7 = LAG7(FLAG1); F8 = LAG8(FLAG1); F9 = LAG9(FLAG1);
F10 = LAG10(FLAG1);
      F11 = LAG11(FLAG1); F12 = LAG12(FLAG1); F13 = LAG13(FLAG1); F14 =
LAG14(FLAG1); F15 = LAG15(FLAG1);
      F16 = LAG16(FLAG1); F17 = LAG17(FLAG1); F18 = LAG18(FLAG1); F19 =
LAG19(FLAG1); F20 = LAG20(FLAG1);
RUN;


/* Creates new FLAG2 variable where FLAG2=1 if all F1-F20=1, else FLAG2=0 */
/* logic: if we count back 20 from the last FLAG2=0, then we will end up at
the first T when power stabilizes and T's become consecutive, increasing
1 at a time without skipping*/
```

```
DATA temp;
      SET temp;
      IF F1=1 AND F2=1 AND F3=1 AND F4=1 AND F5=1 AND F6=1 AND F7=1
      AND F8=1 AND F9=1 AND F10=1 AND F11=1 AND F12=1 AND F13=1 AND
      F14=1 AND F15=1 AND F16=1 AND F17=1 AND F18=1 AND F19=1 AND F20=1 THEN
FLAG2=1;
      ELSE FLAG2=0;
RUN;

/* Keeps only FLAG2=0 */
DATA temp;
      SET temp;
      WHERE FLAG2=0;
      DROP F1-F20 FLAG1;
RUN;

/*Assigns the maximum value of T in the previous datastep to MACRO variable
"maxT", then subtracts 20 */
DATA _NULL_;
      SET temp;
      CALL SYMPUT("maxT", MAX(T)-20);
RUN;

/*Identifies the lowest minimum T for which power will always be >= pre-
specified power based, on maxT in previous step */
DATA &output LABEL;
      SET temp;
      IF T=&maxT;
      DROP FLAG2;
      FORMAT power level alpha powerreq PERCENT10.2;
      LABEL       P1 = "Incidence Rate in Control"
                  VE1 = "True Efficacy of Test Vaccine"
                  VE0 = "Efficacy Lower Bount"
                  alpha = "Pre-Specified Level"
                  powerreq = "Required Power"
                  level = "Exact Level"
                  power = "Exact Power"
                  T = "Total # of cases (Exact Conditional)"
                  N = "Expected Sample Size (Exact Conditional)";
RUN;

PROC DELETE DATA=temp;
RUN;

/* Prints output dataset */
PROC PRINT DATA=&output LABEL;
      VAR P1 VE1 VE0 powerreq alpha power level N;
RUN;

%MEND;
```

```
*************************************************************************
MACRO: Vaccine Efficacy Trial Sample Size using the Z Test Normal
Approximation
*************************************************************************

-Calculates an estimated asymptotic sample size for a vaccine efficacy trial
using a normal approximation approach
-All formulas are based upon (Chan & Bohidar, 1998)

*Required inputs are as follows (only values between 0 and 1 are accepted):

      P1     = incidence in the control group
      VE1    = true efficacy of test vaccine, under the alternative
      VE0    = minimum efficacy of test vaccine, under the null (NOTE: must
be less than VE1)
      Power  = pre-specified power desired to achieve
      Alpha  = pre-specified maximum one-sided level of the test
      Output = name of temporary dataset to where the results will be output

When this MACRO is run, a temporary dataset will be output, along with a
printout of this dataset;

%MACRO SampleSize_NormalApprox(P1=, VE1=, VE0=, Power=, Alpha=, Output= );
%LET c = 1;
OPTIONS SPOOL;

/*Calculates and assigns p2 (incidence in test group) */
DATA _NULL_;
      p2=(1-&VE1)*&P1;
      CALL SYMPUT("p2", p2);
RUN;

/* Calculates all the necessary parameters (sigma0 and sigma1) for the Z-test
via Normal Approximation and assigns them to macro variables */
DATA _NULL_;
      sigma1=(&P2*(1-&P2)+(1/&c)*((1-&VE0)**2)*(&P1)*(1-&P1))**(1/2);
      a=(1-&VE0)*(1+&c*&P1)+&c+&p2;
      b=(1-&VE0)*(&c*&P1+&p2);
      P2z=((a-((a**2)-4*b*(1+&c))**(1/2))/(2*(1+&c)));
      p1z=p2z/(1-&VE0);
      sigma0=(p2z*(1-p2z)+(1/&c)*((1-&VE0)**2)*(p1z)*(1-p1z))**(1/2);
      CALL SYMPUT("sigma0", sigma0);
      CALL SYMPUT("sigma1", sigma1);
RUN;

/* Calculates the Normal Approximation Sample Size */
DATA &output LABEL;
      P1=&p1;
      VE1=&VE1;
      VE0=&VE0;
      powerreq=&power;
      alpha=&alpha;
      zalpha=QUANTILE("norm", &alpha);
      zbeta=QUANTILE("norm", 1-&power);
      N2z=((zalpha*&sigma0+zbeta*&sigma1)**2)/((&P1*(&VE1-&VE0))**2); *sample
size for vaccine group;
```

```
      Nz=ceil(n2z)*2;   *total sample size (when c=1, this is simply
doubling;
      DROP zalpha zbeta N2z;
      FORMAT power level alpha powerreq PERCENT10.2;
      LABEL       P1 = "Incidence Rate in Control"
                  VE1 = "True Efficacy of Test Vaccine"
                  VE0 = "Efficacy Lower Bount"
                  alpha = "Pre-Specificed Level"
                  powerreq = "Required Power"
                  Nz = "Expected Sample Size (Normal Approximation)";
RUN;

/* Prints output dataset */
PROC PRINT DATA=&output LABEL;
      VAR P1 VE1 VE0 powerreq alpha Nz;
RUN;

%MEND;


/* Example execution of macros */
%SampleSize_ExactConditional(P1=0.05, VE1=0.65, VE0=0.15, Power=0.85,
Alpha=0.05, Output=sampleExact);
%SampleSize_NormalApprox(P1=0.05, VE1=0.65, VE0=0.15, Power=0.85, Alpha=0.05,
Output=sampleNormal);
```

BIBLIOGRAPHY

Alonso, P., Smith, T., Armstrong Schellenberg, J. R., Masanja, H., Mwankusye, S., Urassa, H., Tanner, M. (1994). Randomised trial of efficacy of SPf66 vaccine against Plasmodium falciparum malaria in children in southern Tanzania. *The Lancet*, *344*(8931), 1175–1181. https://doi.org/10.1016/S0140-6736(94)90505-3

Aronson, N. E., Santosham, M., Comstock, G. W., Howard, R. S., Moulton, L. H., Rhoades, E. R., & Harrison, L. H. (2004). Long-term Efficacy of BCG Vaccine in American Indians and Alaska Natives. *JAMA*, *291*(17), 2086. https://doi.org/10.1001/jama.291.17.2086

Ashkenazi, S., Vertruyen, et al. (2006). Superior Relative Efficacy of Live Attenuated Influenza Vaccine Compared With Inactivated Influenza Vaccine in Young Children With Recurrent Respiratory Tract Infections. *The Pediatric Infectious Disease Journal*, *25*(10), 870–879. https://doi.org/10.1097/01.inf.0000237829.66310.85

Atanasov, A. G., Waltenberger, B., Pferschy-Wenzig, E. M., Linder, T., Wawrosch, C., Uhrin, P., … Stuppner, H. (2015, December 1). Discovery and resupply of pharmacologically active plant-derived natural products: A review. *Biotechnology Advances*. Elsevier. https://doi.org/10.1016/j.biotechadv.2015.08.001

Breslow, N. E., & Day, N. E. (1987). Statistical methods in cancer research. Volume II--The design and analysis of cohort studies. *IARC Scientific Publications*, (82), 1–406. Retrieved from http://www.iarc.fr/en/publications/pdfs-online/stat/sp82/

Cassio de Moraes, J., Camargo, M. C. C., Rossetto Hidalgo, N. T., Aparecida Barbosa, H., Gattas, V. C., Vasconcelos, H. de. G., … Broome, C. V. (1992). Protective efficacy of a serogroup B meningococcal vaccine in Sao Paulo, Brazil. *The Lancet*, *340*(8827), 1074–1078. https://doi.org/10.1016/0140-6736(92)93086-3

Chan, I. S. F., & Bohidar, N. R. (1998). Exact power and sample size for vaccine efficacy studies. *Commun Stat Theory Meth*, *27*(6), 1305–1322. https://doi.org/10.1080/03610929808832160

Chow, S. C., Shao, J., & Wang, H. (2003). *Sample size calculations in clinical research*. Marcel Dekker. Retrieved from https://www.crcpress.com/Sample-Size-Calculations-in-Clinical-Research/Chow-Wang-Shao/p/book/9780203911341

Comstock, G. W. (1994). Evaluating vaccination effectiveness and vaccine efficacy by means of case-control studies. *Epidemiologic Reviews*, *16*(1), 77–89. https://doi.org/10.1093/OXFORDJOURNALS.EPIREV.A036147

Coursaget, P., Chotard, J., Vincelot, P., Diop-Mar, I., Yvonnet, B., Sarr, M., … Chiron, J. P. (1986). SEVEN-YEAR STUDY OF HEPATITIS B VACCINE EFFICACY IN INFANTS FROM AN ENDEMIC AREA (SENEGAL). *The Lancet*, *328*(8516), 1143–1145. https://doi.org/10.1016/S0140-6736(86)90543-X

DiazGranados, C. A., Dunning, A. J., Kimmel, M., Kirby, D., Treanor, J., Collins, A., … Talbot, H. K. (2014). Efficacy of High-Dose versus Standard-Dose Influenza Vaccine in Older Adults. *New England Journal of Medicine*, *371*(7), 635–645. https://doi.org/10.1056/NEJMoa1315727

Farrington, C. P., & Manning, G. (1990). Test statistics and sample size formulae for comparative binomial trials with null hypothesis of non-zero risk difference or non-unity relative risk. *Statistics in Medicine*, *9*(12), 1447–1454. https://doi.org/10.1002/sim.4780091208

Fine, M. J., Smith, M. A., Carson, C. A., Meffe, F., Sankey, S. S., Weissfeld, L. A., … Kapoor, W. N. (1994). Efficacy of Pneumococcal Vaccination in Adults: A meta-analysis of randomized controlled trials. *Arch Intern Med*, *154*(23), 2666–2677. https://doi.org/10.1001/archinte.1994.00420230051007

Flahault, A., Cadilhac, M., & Thomas, G. (2005). Sample size calculation should be performed for design accuracy in diagnostic test studies. *Journal of Clinical Epidemiology*, *58*(8), 859–862. https://doi.org/10.1016/j.jclinepi.2004.12.009

Friedman, L. M., Furberg, C. D., DeMets, D. L., Reboussin, D. M., & Granger, C. B. (2015). *Fundamentals of Clinical Trials*. *Fundamentals of Clinical Trials*. https://doi.org/10.1007/978-3-319-18539-2_11

Fromer, M. J. (2006). FDA Introduces New Phase 0 for Clinical Trials. *Oncology Times*, *28*(15), 18–19. https://doi.org/10.1097/01.COT.0000293387.34124.a8

Gail, M. (1974). Power Computations for Designing Comparative Poisson Trials Author (s): Mitchell Gail Published by : International Biometric Society Stable URL : http://www.jstor.org/stable/2529645. *BIOMETRIC*, *30*(2), 231–237.

Gart, J. J., & Nam, J. (1988). Approximate Interval Estimation of the Ratio of Binomial Parameters: A Review and Corrections for Skewness. *Biometrics*, *44*(2), 323. https://doi.org/10.2307/2531848

Geison, G. L. (2014). *The private science of Louis Pasteur*. Princeton University Press.

Gross, P. A., Hermogenes, A. W., Sacks, H. S., Lau, J., & Levandowski, R. A. (1995). The Efficacy of Influenza Vaccine in Elderly Persons. *Annals of Internal Medicine*, *123*(7), 518. https://doi.org/10.7326/0003-4819-123-7-199510010-00008

Hansen, A. R., Graham, D. M., Pond, G. R., & Siu, L. L. (2014). Phase 1 trial design: is 3 + 3 the best? *Cancer Control : Journal of the Moffitt Cancer Center*, *21*(3), 200–8. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/24955703

Jenner, E. (1798). An inquiry into the causes and effects of the variolae vaccinae.

Kulkarni, P. S., Desai, S., Tewari, T., Kawade, A., Goyal, N., Garg, B. S., … Raj, C. V. (2017). A randomized Phase III clinical trial to assess the efficacy of a bovine-human reassortant pentavalent rotavirus vaccine in Indian infants. *Vaccine*, *35*(45), 6228–6237. https://doi.org/10.1016/j.vaccine.2017.09.014

Le Tourneau, C., Lee, J. J., & Siu, L. L. (2009). Dose escalation methods in phase I cancer clinical trials. *Journal of the National Cancer Institute*, *101*(10), 708–20. https://doi.org/10.1093/jnci/djp079

Mandrekar, S. J., & Sargent, D. J. (2010). Randomized phase II trials: time for a new era in clinical trial design. *Journal of Thoracic Oncology : Official Publication of the International Association for the Study of Lung Cancer*, *5*(7), 932–4. https://doi.org/10.1097/JTO.0b013e3181e2eadf

Miettinen, O., & Nurminen, M. (1985). Comparative analysis of two rates. *Statistics in Medicine*, *4*(2), 213–226. https://doi.org/10.1002/sim.4780040211

Mitchell, V., Philipose, N., & Sanford, J. (1993). Stages of Vaccine Development. *The Children's Vaccine Initiative: Achieving the Vision*, 109–120. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK236428/

Plotkin, S. A. (ed.). 2011. *History of Vaccine Development* (2011 Ed.). Springer Science & Business Media. https://doi.org/10.1007/978-1-4419-1339-5
Porter, D. (1999). *Health , Civilization and the State. A history of public health from ancient to modern time*. *Health Civilization and the State*. https://doi.org/10.1353/bhm.2000.0059
Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., … Kim, J. H. (2009). Vaccination with ALVAC and AIDSVAX to Prevent HIV-1 Infection in Thailand. *New England Journal of Medicine*, *361*(23), 2209–2220. https://doi.org/10.1056/NEJMoa0908492
Sakpal, T. V. (2010). Sample size estimation in clinical trial. *Perspectives in Clinical Research*, *1*(2), 67–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21829786
Shim, E., & Galvani, A. P. (2012). Distinguishing vaccine efficacy and effectiveness. *Vaccine*, *30*(47), 6700–5. https://doi.org/10.1016/j.vaccine.2012.08.045
Stern, A. M., & Markel, H. (2005). The History Of Vaccines And Immunization: Familiar Patterns, New Challenges. *Health Affairs*, *24*(3), 611–621. https://doi.org/10.1377/hlthaff.24.3.611
Storsaeter, J., & Olin, P. (1992). Relative efficacy of two acellular pertussis vaccines during three years of passive surveillance. *Vaccine*, *10*(3), 142–144. https://doi.org/1(Test, n.d.)0.1016/0264-410X(92)90002-2
Strassburg, M. A. (1982). The global eradication of smallpox. *AJIC: American Journal of Infection Control*, *10*(2), 53–59. https://doi.org/10.1016/0196-6553(82)90003-7
Suvarna, V. (2010). Phase IV of Drug Development. *Perspectives in Clinical Research*, *1*(2), 57–60. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21829783
U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, CDC Newsroom. (CDC 2014) Report shows 20-year US immunization program spares millions of children from diseases. Retrieved October 16, 2017, from http://www.cdc.gov/media/releases/2014/p0424-immunization-program.html
U.S. Food and Drug Administration. (FDA, 2018). Approved Products - Vaccines Licensed for Use in the United States. Retrieved February 23, 2018, from https://www.fda.gov/BiologicsBloodVaccines/Vaccines/ApprovedProducts/UCM093833
World Health Organization. (WHO, 2017). Immunization coverage. Retrieved from http://www.who.int/mediacentre/factsheets/fs378/en/
World Health Organization. (WHO, 2018). Vaccines. Retrieved from http://www.who.int/topics/vaccines/en/
World Bank. Incidence of malaria (per 1,000 population at risk). *World Bank*. Retrieved from https://data.worldbank.org/indicator/SH.MLR.INCD.P3.