2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Computation of probability of study success based on data accumulated until an Interim Analysis

Promotor :
Prof. dr. Roel BRAEKERS

Promotor :
Ms. NANCY FRANCOIS

Edith Milanzi

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

universiteit hasselt
KNOWLEDGE IN ACTION

Universiteit Hasselt | Campus Hasselt | Martelarenlaan 42 | BE-3500 Hasselt
Universiteit Hasselt | Campus Diepenbeek | Agoralaan Gebouw D | BE-3590 Diepenbeek

universiteit hasselt | Maastricht University

2013•2014
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Computation of probability of study success based on data accumulated until an Interim Analysis

Promotor :
Prof. dr. Roel BRAEKERS

Promotor :
Ms. NANCY FRANCOIS

Edith Milanzi

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics

universiteit hasselt | Maastricht University

Declaration of Authorship

I, Edith .B Milanzi, declare that this thesis titled, 'Computation of probability of study success based on data accumulated until an Interim Analysis' and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: _____

Date: _____

Acknowledgements

First and foremost i would like to thank God, Only His everlasting blessings have brought me this far.

Many thanks to my supervisors Nancy Francois and Prof Dr.Roel Braekers for their guidance throughout my project. I am sincerely grateful to them for sharing their expertise and illuminating views on a number of issues related to the project.

I would also like to express my appreciation to my family for their unconditional support, throughout my degree. Elasma in particular, your patience, understanding and mostly your selflessness i will never take for granted.

Contents

List of Figures

List of Tables

UNIVERSITEIT HASSELT

Abstract

Faculty of Science

Master in Biostatistics

Computation of probability of study success based on data accumulated until an Interim Analysis

by Edith .B Milanzi

The primary objective of this study was to determine the probability of success of a vaccine efficacy trial at study end. In randomised clinical trials, often times, sponsors and investigators are interested to know whether the trial will demonstrate a successful result at the end. Interim analyses are planned for these purposes. Conditional power is one of the tools used to assess the probability that a study will succeed in the end given data observed at a given point in time. Based on simulations, a two-arm randomised Phase II vaccine efficacy trial was simulated and based on data observed until an interim analysis, the probability of success was estimated. The trial was replicated 1000 times depicting the design of the actual trial and the conditional power for every trial was calculated and an average estimate was obtained. The estimate of probability of success at study end based on the data observed was 63.9 %. Therefore it was concluded that there was about 64 % chance that the trial would run to completion given the observed data. More elegant alternatives were however recommended such as the predictive power which accounts for variability of the treatment effect to assess the probability of success.

Chapter 1

Introduction

The pharmaceutical industry has grown exponentially over the years, and the most logical expectation would be that the amount of drugs to treat various diseases would also grow at approximately the same rate. However this is not currently the case. An ongoing and serious challenge facing the pharmaceutical industry is the high failure rate in the latter stages of drug development, resulting in low research and development (R&D) productivity [1]. As such resources are deployed in producing drugs that are eventually rendered useless.

Gan et al [2], reviewed 235 recently published phase III randomized clinical trials. The authors report that 62% of the trials did not achieve results with statistical significance. Trying to explain the high failure rate, they note the actual magnitude of benefit achieved in a clinical trial is nearly always less than what was predicted at the time the trial was designed (designated δ) and conclude, "investigators consistently make overly-optimistic assumptions regarding treatment benefits when designing randomised clinical trials.[3]. Such instances have been evidenced in oncology trial but this does not mean the problem is unique to oncology trials but also vaccine efficacy trials which this report is focused on. This problem of high failure rate indeed universally affects any drug development process. Randomized clinical trials utilize enormous resources, most importantly money but also human resources. Realizing that this is a huge risk when it comes to sponsoring such trials, many sponsors often want to know whether their investment in the drug process is worth it. In order to accommodate such concerns, during the design of the study, in between analyses are planned at a specified time during the trial. These analyses, called interim analyses help to shed light on the progress of the study in terms of the effect of the treatment (or indeed the relevant hypothesis in

question). Often the interim analyses will be able to tell whether the drug under study is effective (at the point of analysis), or is worse than the standard/comparison. Apart from that, these analyses also help to address safety concerns such that if the drug under study displays more harm than good, the safety of the patients has to respected and the trial must be stopped to uphold ethical considerations, as they are a vital aspect in clinical trials. In cases where the drug is found to have higher efficacy before the intended trial duration, it can be terminated and conclusions that the drug is effective can be made. Equivalently, the trial can be terminated if the drug is performing worse and threatening the lives of the patients.

An important and critical aspect of the clinical trial is the power of the study. It is calculated during the design of the clinical trial. This statistical power is the probability that the study will succeed by attaining statistical significance,at a designated effect size. This however does not provide a reliable answer to the question of successful completion of the study. This is so because traditional power is the probability of success (achieving statistical significance) at an assumed effect size where the assumed effect size is often based on regulatory, payer, and/or marketing requirements or needs, and may not be supported by available evidence or reflect the true treatment effect [4]. In addition this power does not make use of any available data regarding the treatment effect.
On the other hand, Conditional power is the probability of concluding a positive study at the end of trial, given the interim results of treatment effect. It conditions on the current data, and assumes a particular point estimate (usually the maximum likelihood estimate) for the treatment effect [5]. It differs from traditional statistical power in that an observed amount of data is used to calculate conditional power. The conditional power approach rather than the traditional power is more appealing in determining whether a study will be successful on completion. The conditional power function is a useful device for communicating with clinical investigators as it can be used to illustrate the effects of low accrual or to aid the decision of continuing the study [6]. If the conditional power is very small then the investigation can be stopped in order to channel resources to more promising investments. However both the traditional and conditional power consider the treatment effect to be fixed which is usually a restrictive assumption.
Alternatively the probability of success can be determined by considering that the effect size varies(not fixed). This can be done by computing the predictive power of the study. The predictive power therefore accounts for the variability of the

treatment effect.

Both Wang [4] and Trzaskoma [7]]focus on the hybrid of bayesian and frequentists concepts in determining the probability of success. By simulating studies under different scenarios and accounting for the variability of the treatment effect size, they separately showed that both conditional power and predictive power can help optimize individual study designs, for example, if the power is high, it makes sense to employ an aggressive study design such as allowing stopping early for efficacy; on the other hand, futility interim analysis may be needed if the available data indicate a low power. Furthermore, conditional power approach can also be used for sample size re-estimation purposes to create one single interim analysis rule. This allows interim results to be used for adjusting the original sample size if necessary to ensure adequate power in the final analysis, or to terminate the trial early on the basis of such interim findings.

In the same line of thought, Carroll [1] approached the probability of success problem in a context of moving form a phase II trial to a Phase III trial also from a bayesian point of view. In this case, the prior is based on data collected form the Phase II study to inform decision on the Phase III study. The author is however quick to mention that other choices of prior distributions can also be used as long as they are plausible.

1.1 Vaccine Efficacy

Vaccine efficacy (VE) is defined as the degree in which the vaccine offers protection against the target infection or disease.[8]. Several effects of vaccine can be studied depending on the interest of the study. These effects as described by Halloran et al [9] include: vaccine efficacy of susceptibility where there is measurement of how protective vaccination is against infection, and vaccine susceptibility to disease whose distinction with the former lies in the case definition of the disease but are usually considered as one in literature. Normally the interest in vaccine efficacy trials has been to determine how well vaccination protects the vaccinated individual i.e the vaccine efficacy of susceptibility to disease [8].

In conducting vaccine efficacy trials there is need to quantify the occurrence of new cases of disease in the population. This is often done by determining the incidence of the disease/infection in the population and it is associated with individuals who

were initially healthy and then got infected. There are three types of incidence measures commonly used: the cumulative incidence, the incidence rate and the hazard rate. If in a vaccine field efficacy trial the endpoint is infection then the three incidence measures are usually termed the attack rate, the infection rate and the force of infection respectively.[8]
However the vaccine efficacy does not depend on which incidence measure is used.

1.2 Problem Description

Acute otitis media (AOM) is the most common bacterial infectious disease among children and often a painful experience for the affected child. Prior to the advent of an effective vaccine against the pneumococcal serotypes associated with otitis media, approximately two thirds of children in the USA, for example, experienced at least 1 episode of otitis media, and almost 1 in 6 children experienced 3 or more episodes of otitis media in their first years of life [10]. Characteristic symptoms of AOM are effusion in the middle ear accompanied by signs of acute illness as earache, otorrhoea ,ear tugging, fever, irritability, anorexia, vomiting or diarrhoea. In the long run these may lead to delayed speech development as well as imparied hearing apart from resistance of pneumococcal strains to common antibiotics. As such there is a growing and justified need for effective vaccines [11].
Despite the introduction of the first licensed 7-valent pneumococcal conjugate vaccine (Prevnar, Pfizer) in the childhood routine immunization schedule, Navajo and White Mountain Apache children continue to suffer from pneumococcal diseases and have carriage rates, therefore the development of protein-based pneumococcal vaccines that have the potential to provide protection across all pneumococcal serotypes, in addition to pneumococcal conjugate vaccine, would be of great public health interest to AI/AN populations.
A Vaccine Efficacy phase II Trial was designed by GlaxoSmithKline(GSK) Vaccines to demonstrate the VE of the pneumococcal dPly and PhtD proteins in preventing AOM in Native American infants aged less than 24 months, living in the southwestern US in and around the Navajo and White Mountain Apache reservations, as well as to evaluate the impact on Acute Lower Respiratory Tract infections (ALRI) and on nasopharyngeal carriages up to the second year of life. The efficacy of a three-dose primary course at 2, 4 and 6 months of age followed by a booster dose at 12-15 months of age with GSK Vaccines pneumococcal protein

vaccine co-administered with Prevnar 13 in preventing clinical AOM diagnosed and verified against American Academy of Pediatrics (AAP) criteria will be established. In view of this trial,the following objective were developed for this project;

1.3 Objectives

- Computation of an integrated conditional probability of success at study end (based on accumulated data up to an Interim Analysis).

The organization of the report is as follows, the next chapter gives the brief outline of methods implemented in this project. Results are presented in chapter 3, followed by discussion and conclusions in chapter 4.

Chapter 2

Methodology

2.1 Trial Setting Description

This trial was designed as a Phase II, double-blind, randomized, placebo-controlled, multi-centric, single-country study with two parallel groups. The two study groups,the treatment group dPly/PhtD group was expected to enroll approximately 900 subjects receiving GSK Vaccines' pneumococcal protein vaccine (dPly-PhtD vaccine) co-administered with Pfizer's Prevnar 13 vaccine. The control group was also expected to enroll approximately 900 subjects receiving GSK Vaccine' placebo co-administered with Pfizer's Prevnar 13 vaccine representing a 1:1 randomization ratio. Subjects were expected to be followed up for a maximum of 22 months with an accrual period of 16 months. During the design phase of the study there was no licensed pediatric pneumococcal vaccine containing pneumococcal proteins available to act as an active comparator justifying the use of the placebo.
The primary endpoint of the trial is to demonstrate the efficacy of a three-dose primary course at 2, 4 and 6 months of age followed by a booster dose at 12-15 months of age with GSK Biologicals' pneumococcal protein vaccine co-administered with Prevnar 13 in preventing clinical AOM diagnosed and verified against American Academy of Pediatrics (AAP) criteria. As such the endpoint is time to episodes of acute otitis media.
With a time to event endpoint the Vaccine Efficacy is essentially defined as 1 minus the hazard ratio denoted

$$VE = (1 - \frac{\lambda_1}{\lambda_0}) \times 100$$

where λ_0 is the hazard rate in the placebo group and λ_1 is the hazard rate of the vaccinated group. As such the hypothesis was formulated as follows:

$$Ho : VE \leq 0\%$$

$$Ha : VE > 0\%$$

Or equivalently in terms of the hazard ratio(HR):

$$Ho : HR = 1\%$$

$$Ha : HR < 1\%$$

The above hypotheses correspond to the hypothesis that pneumococcal proteins in the vaccine will reduce the incidence of clinical AOM compared to control group if the vaccine efficacy is greater than 0%.

2.2 Statistical Methodology

2.2.1 Cox proportional Hazards Model

Cox's proportional hazards model is the most commonly used model for clinical trial data and provides reliable estimates of survival times, as well as the relative risk associated with time-to-event occurrence[12]. Considering that we were dealing with time to event data, first attempt of analysis was done using the Cox model which only uses the time to the first event. In time to first event data analysis,if T is a non-negative continuous random variable representing the waiting time until the occurrence of an event, then the survival function which gives the probability of not experiencing an event just before duration t, or more generally, the probability that the event of interest has not occurred by duration t is given by:

$$S(t) = Pr(T \geq t.)$$

The Cox proportional hazards model is given by

$$\lambda_x t = \lambda_0 exp(\beta x)$$

where λ_0 is the baseline hazard function and is assumed to depend on time t and β an unknown vector of regression coefficients for covariate x that does not depend on time . This builds up to the important assumption of proportional hazards. As such hazard ratios between two individuals of different covariate information are used to compare survival between two individuals. The Cox proportional hazards model relates the hazard rate for individuals at the value $x(\lambda_x(t))$, to the hazard rate for individuals at the baseline value $\lambda_0(t)$. This produces an estimate for the hazard ratio

$$HR = \lambda_x(t)/\lambda_0(t)$$

The hazard ratio represents the chance of experiencing events in the treatment arm as a ratio of the chance of events occurring in the control arm. The Cox proportional hazards model was fitted considering the time to first occurence of Acute otitis media. However, the common feature with the standard Cox regression model is that it only considers the time to the occurrence of the first event. In this study there were expected cases of recurrent episodes of AOM such that the cox proportional model would not be appropriate to analyse such data. This is because the model ignores any recurrent episodes. An extension of the Cox model ,the Andersen and Gill model(A-G) was considered for analysis to account for recurrence and is briefly described below.

2.2.2 Andersen-Gill Model

Recurrent events, are events that are experienced a number of instances times from the same individual. These can include, myocardial infactions, and tumour recurrence. The analyses of recurrence time to event data are prone to inappropriate analyses using the cox model because of its simplicity.
The A-G model is one of the models used to analyze data with repeated occurrences. Also known as the counting process model, this approach models the repeated episodes for each person as separate observations, with the risk set not constrained by the number of events occurring within an individual, and it makes a strong assumption of independence among multiple observations per person over time and also between individuals.[13]. For the *kth* event of the *ith* subject at time t, in the counting process model, the hazard function is given by

$$\lambda_{ik}(t) = \lambda_0(t)exp(x_{ik}\beta)$$

where β represents the covariate vector (p fixed effects) for the *ith* subject with respect to the *kth* event and x_{ik} is the covariate matrix and λ_0 represents the common baseline hazard for all events. This model is an extension to the standard Cox Proportional Hazards model [14]. The A-G model uses the counting process formulation. As pointed out by Castaneda and Gerritse [15], the primary difference between the cox proportional hazards model and the A-G model is the definition of the individuals at risk. For the Cox model, an individual ceases to be at risk when an event occurs while for the A-G model the individual still remains at risk after an event. It is also known as a conditional model,distinct from a marginal model where subject is at risk from the start of treatment and does not depend on any previous events. The partial likelihood for the A-G model is given by

$$L(\beta) = \prod_{i=1}^{n} \prod_{k=1}^{Ki} \left(\frac{exp(\beta^T x_i)}{\sum_{i \in R(tik)}^{n} exp(\beta^T xi)} \right)^{\delta_i k}$$

where $\delta_i k$=0 if censoring intervent time and 1 otherwise. [16]

2.3 Probability of Success

Suppose we wish to test our null hypothesis using a reference test Z, then according Jennison and Turnbull [**?**], the conditional power at stage k which represents the probability that a trial will succeed at the end given data(D_k) observed up until a given point in time is given by

$$P(\theta)_k = Pr_\theta(Z \ will \ reject \ H_0 | D(k))$$

A very low value of this probability will suggest that continuation of the trial is futile and the primary endpoint of efficacy will be no longer pursued. In this project, any probability values less than 15 % will lead to the trial stopping for the primary outcome.

For successive analyses, assuming Z_k is a sufficient statistic for θ at stage k, and for an information level I_k, then it can replace D_k such that the conditional distribution of Z_K given Z_k can be written as

$$Z_K | Z_k \sim \mathcal{N}(Z_k \sqrt{(I_k / I_K)} + (\theta(I_K - I_k)/\sqrt{I_K}), 1 - I_k/I_K)$$

and there on sided conditional power at analysis k is as presented by Jennison and Turnbull [**?**] is also given by

$$P(\theta)_k = \Phi\{\frac{Z_k\sqrt{I_k} - z_\alpha\sqrt{I_K} + (I_K - I_k)\theta}{\sqrt{(I_K - I_k)}}\}$$

Where Z_K is the test statistics at final analysis and I_K represents the information (number of events) required at final analysis and θ represented by $ln(HR)$. Zk is calculated as($\theta \times \sqrt{Ik}$). Since we were using a one sided test at 0.074 significance level, $Z\alpha$ is the critical value for significance level of 0.074.
The above formula was applied to each simulated dataset to calculate the conditional power for each trial. The resulting conditional powers were averaged to obtain one estimate of probability of success.

2.4 Data Simulation

Simulation studies use computer intensive procedures to test particular hypotheses and assess the appropriateness and accuracy of a variety of statistical methods in relation to the known truth. [17] To estimate probability of success, 1000 datasets were simulated closely depicting the design of the trial. The Interim Analysis was planned to be conducted at 50% of the information, thus after at least 633 reported clinical AOM episodes were diagnosed and verified against AAP criteria since 1266 reported cases were expected at final analysis. Inorder to accumulate 633 events,the trial was simulated and cut off 17 months into the trial where the approximately, the postulated number of events were obtained. As such only reported cases till 17 months were used for analysis.

A hypothesized hazard rate of 0.6 event /child /year was used for the placebo group which corresponded to a hazard rate of 0.465 in the treatment group for a true VE of 17 %. These rates assumed a constant event rate over time, which is consistent with an exponential distribution of the time to event. The complete follow-up for each subject was set to 22 months and accrual of the individuals was set to 16 months such that 633 events were obtained shortly after accrual period ended. With an expected loss to followup of 10 % , about 1537 subjects were expected to be evaluable as such this was the number of generated subjects. Based on an almost constant accrual rate of 96 individuals per month, a total of approximately

633 events would be accrued by aprroximately 17 months into the trial. The simulation was done according to Metcalfe et al [16]. For each subject,inter-event times were simulated as independent realization of an exponential distribution of rate

$$\lambda = exp(x_i\beta)$$

obtained from the inversion method

$$t_{ik} = -\frac{lnu_i}{exp(x_i\beta)} = -\frac{lnu_i}{\lambda}$$

where i=1, ..., n individuals; k=0, ..., ki events observed from an individual; t_{ik} denoted as the interval time from $ti_k - 1$ until ti_k and $u_i \sim$ U[0,1], a uniform distribution. By transforming u_i ,interval-event times t_{ik} for each subject were generated. All generated times greater than 17 months were censored. Time to occurrence of primary endpoint during the defined efficacy follow-up period was compared between groups by estimating Vaccine efficacy(VE) and its 92.6% confidence interval (one sided, 0.074 significance level)using the Cox model and A-G model. All analyses were done in SAS statistical package version 9.3

Chapter 3

Results

3.1 Exploratory Data Analysis

3.1.1 Descriptive Statistics

From the 1000 datasets simulated and based on approximately 1540 subjects. The number of occurrences that were determined as failed or censored observations were tabulated. There were more events experienced in the placebo group than the vaccinated group up to follow up time as shown below.

Table 3.1: Table of Status by Group

Status	Group		
	Placebo	Vaccinated	Total
Censored	1360	1226	2586
	41.34	37.26	78.6
Failed	406	298	704
	12.34	9.06	21.4

On average, 21.4 % of the children experienced atleast one event at the time of analysis. Events that were censored were those that were experienced after 17 months of the trial which was the cutoff point for the analysis. About 70 % of the children had not yet experienced an event until this point in time. An overview of the recurrences by treatment group also seemed appropriate inorder to gain insight in the differences between the two groups if any.

TABLE 3.2: Recurrent Events by Group

	Group		
Events	Placebo	Vaccinated	Total
0	405	307	710
	12.31	9.33	21.82
1	578	520	1098
	17.57	15.81	33.37
2	450	398	848
	13.68	12.1	25.78
3	222	196	418
	6.75	5.96	12.71
4	80	80	160
	2.43	2.43	4.86
5	20	15	35
	0.61	0.46	1.66
6	10	7	17
	0.3	0.21	0.52
7	1	1	1
	0.03	0.03	0.06
Total	1766	1524	3290
	53.8	46.32	100

Recurrent events were experienced in both groups with a maximum of 7 recurrent events experienced by two subjects,one in each group. Slightly more recurrent events per child were observed in the placebo group rather than the vaccinated group. For example 13.68 % of the placebo experienced at least 2 recurrent events while a slightly lower percentage of 12.1% experienced the same in the vaccinated group.

3.1.2 Proportional Hazards Assumption

Both the cox and the A-G model assume the proportionality of the hazards. In this regard the assumption was investigated graphically using Kaplan Meier Curves and the cumulative hazard functions were also plotted.

Figure 3.1, shows the two curves do not seem to cross, reflecting proportional hazards,and the treatment group is seen to survive longer until an episode of AOM as compared to the placebo group (Log rank test pvalue= 0.02). However,

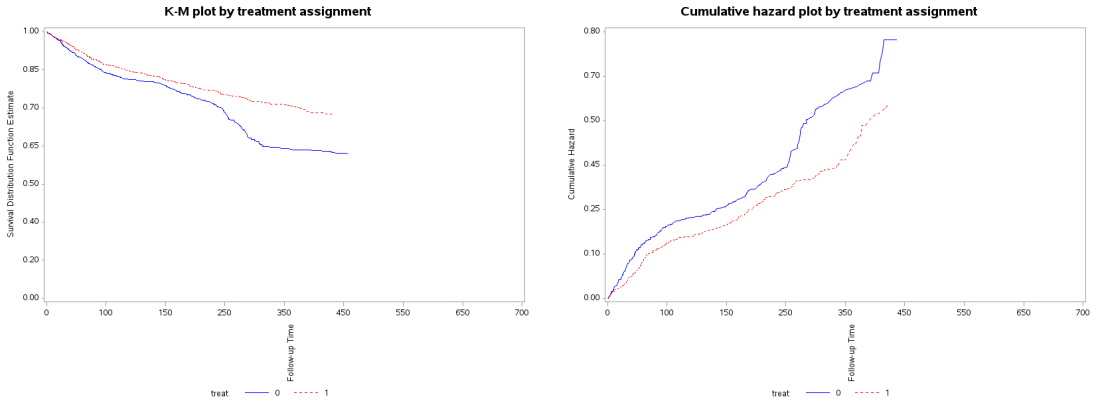

FIGURE 3.1: Kaplan Meier Curves and Cumulative Hazard curves for Cox Model

in the Cox model,the standard Kaplan Meier analysis assumes that after an event has been observed, the individual is no longer at risk [18]. This might not be appropriate for recurrent events as the individual is still in the risk set even if the first event occurred. As such there is an enormous amount of information that is lost in displaying such curves for recurrent data. In accounting for that, survival curves for the recurrent events were plotted as seen in Fig 3.3. .

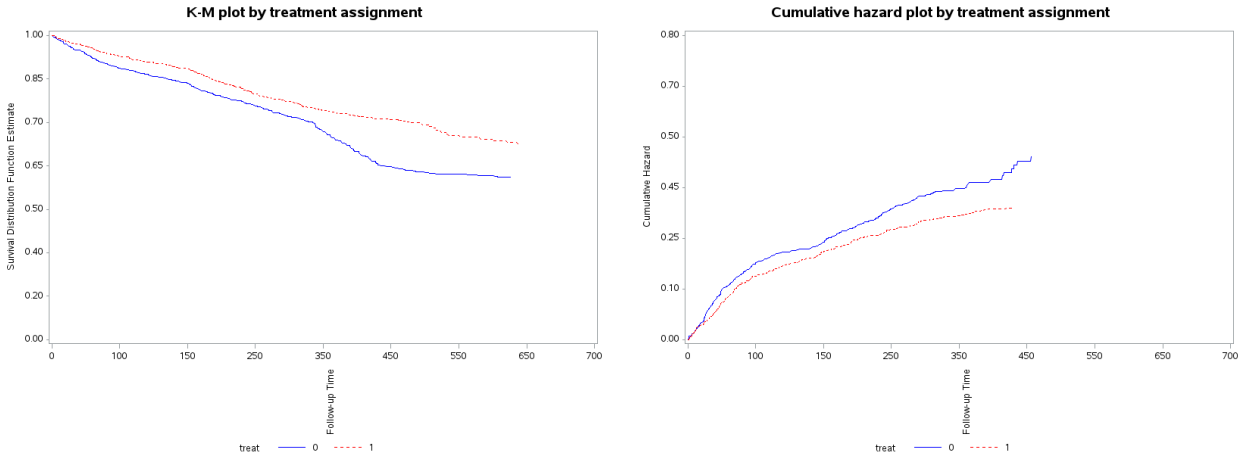

FIGURE 3.2: Kaplan Meier Curves and Cumulative Hazard curves for A-G Model

It can be observed that the rate at which the Kaplan meier curve of the recurrent events is dropping is slower than the rate at which the same curve is dropping for the cox model. By the middle of the follow up time(330 days), the probability of surviving without an event for the cox model is approximately 63 % for the control group and close to 70 % for the treatment group. As for the curves that

account for the recurrent events, by middle of follow up time, the probability of surviving is slightly above 70 % for the control group while it is about 80 % for the treatment group. It seems however that probability of experiencing an AOM episode is higher after half the time of follow up for the control group than the treatment.

3.2 Statistical Analysis

Based on the accumulated data until 17 months into the beginning of the study, the Cox model and A-G model were fitted for each dataset. The covariate considered here was the treatment alone where 0 represents the placebo group and 1 represents the vaccinated group. The covariate vector of fixed effects therefore represents the treatment effect. For further statistical analysis of the A-G model the data was manipulated into the A-G data structure. Based on this structure of the data, intervals in which events occur are split between start times and stop times, such that for a second episode of AOM, the starting time is the previous time(stop time) which the first event occurred. These starting and stopping times are modelled independent variables in the model. If an individual has 1 event, there are two rows representing data for that subject,i.e 1 row indicating time of entry until the event, and the last row,indicating time of event until censorship and it follows for more subsequent events.
Each event from the same individual is assumed to be independent and a subject

Table 3.3: A-G Data Structure

ID	Tstart	Tstop	Event	Status
1	0	12	1	1
1	12	17	2	1
1	17	25	2	0
2	0	12	0	0
3	0	23	1	1
3	23	28	2	1
3	23	40	2	0

contributes to the risk set for an event as long as the subject is under observation at the time the event occurs. A sample data structure of the A-G model has been presented in table3.3.

The hazard ratio of the two groups and the treatment effect of the vaccine was estimated from an assumed true VE of 17% using the Proc phreg procedure in SAS.

Table 3.4 presents the average estimates based on the simulated datasets for both the Cox model and the A-G model. Level of significance was 0.074 in order to account for the multiple analyses to be done during the trial.

The estimates for the Cox model were obtained after deleting any recurrent events

Table 3.4: Parameter Estimates

Model	Parameter	Estimate(Standard Error)	Hazard Ratio	92.6% HR CI
Cox Model	Treatment	-0.210(0.091)	0.814	[0.713;0.928]
A-G	Treatment	-0.218(0.078)	0.806	[0.717;0.906]

from the same individual such that there was only one record per individual. This was done assuming that the interest was in the time to first event. The A-G model estimates were obtained as a result of a robust sandwich estimate for the covariance matrix which produces robust standard errors as a way of accounting for the correlation of events from one subject as opposed to the model based standard errors.

Table 3.5: Vaccine Efficacy at 17 Months into the trial

Model	Vaccine Efficacy	92.6% VE CI
Cox Model	18.6%	[9 %;28.9 %]
A-G	19.4%	[10%;28.3 %]

From the estimated values of the hazard ratios the vaccine efficacy for each model was calculated with the A-G model displaying a higher vaccine efficacy than the cox model. This is expected because information from recurrent episodes is utilized in the A-G model. In addition the cox model has got slightly wider confidence limits than the A-G and model. A true VE of 17% was assumed at the design of the trial representing a bias of 9.4 % for the cox model and 14.11 % for the A-G model. These percentages are considered high suggesting the estimates could be biased.

3.3 Probability of Success

The traditional power of the trial during design phase,where 1800 subjects were targeted, was calculated prior to the trial for different scenarios corresponding to

different incidence rates and the target vaccine efficacy. A high significance level was chosen because of the uncertainity around the incidence of AOM in the studied population. For example with a 0.6 incidence rate, and VE equals 11%, the power was calculated as 81% and around 92.1 % for a vaccine efficacy of 17%.

Considering that these calculations were made before the data was obtained, there was need to calculate the conditional power now based on the accumulated data. By applying the formula in chapter 2. The following estimate was obtained for a true VE of 17 %.
For the standard Cox model, the conditional power and therefore the probability

Table 3.6: Conditional power

Model	Conditional power
Standard Cox Model	58.3
A-G	63.9

to observe a successful trial at the end was 58.3% while the A-G model it is about 63.9. % The A-G model indicates more power for the study to run to completion than the cox model.

Chapter 4

Discussion and conclusion

Data accumulated until an interim analysis gives an insight into the progress of the trial and sometimes a decision can be made to stop the trial or proceed with it. Statistical techniques are used to aid this decision. Estimation of the conditional power is one of the tools used to help in making this kind of decision. Conditional power expresses the probability of the trial running to completion given that some data has already been accumulated up until a given point in time. This is usually of interest for all the concerned parties conducting the trial and most importantly the sponsor whose resources are invested for a significant treatment effect. In this trial, with a fixed estimate of the Vaccine efficacy of 17% i.e to detect a 17 % reduction in the hazard ratio of experiencing episodes of Acute otitis media, and with the null hypothesis of VE being less than 0% (no treatment effect),it is expected that there will be insufficient evidence to reject it if the conditional power is less than 15 % . The probability of the trial being successful at the end given the data observed until an interim point was calculated as 58.3 % based on the cox model and 63.9% based on the A-G model which was more appropriate and therefore preferable in this case. Thus there is about 63.9 % chance that the trial will reject the null hypothesis of hazard ratio being equal to 1 and consequently, the vaccine efficacy being less than 0%. This is considerably further from the stopping rule of $\leq$ 15% power that was hypothesized to stop the trial. This estimate is also considerably lower than the 92.1 % that was postulated at the beggining of the trial. However the conditional power suggests that the trial will succeed if the current trend in the treatment effect is maintained.

It should be noted however that, this decision to either proceed or terminate the trial should not entirely be driven by the statistical considerations. There are

other aspects of the trial that also play part in making this decision. For example in cohort studies,slow case accrual, costly laboratory tests, or tests that require destruction of unique biological samples can make it desirable to stop early when the data indicate no relevant effect [19]. The A-G estimates were a result of a robust sandwich estimator for the covariance matrix as opposed to the model based matrix and this provides a simple and valid approach to analyze recurrent events. A one sided 92.6 %confidence interval of [0.717;0.906] suggests significance of the treatment for the A-G model. From a hazard ratio of 0.806, the vaccine efficacy is obtained as 19.4 % and is contained in the VE confidence interval of [10 %;28.3%] .
We can therefore say that the treatment is shown to delay occurrence of the first episode and the subsequent episodes where applicable. The treatment effect -0.21 translates to a hazard ratio of 0.846 with a standard error of 0.09 for the cox model. The hazard ratio 92.6 % confidence interval does not contain 1 suggesting significance. However this model was used for illustrative purposes and deemed inappropriate for recurrent data.

According to a review by Fletcher et al [10], clinical trials during the development of AOM vaccines found that their efficacies against a microbiological outcome, vaccine-serotype pneumococcus in middle ear fluid isolates(MEF), were quite similar (about 60%), yet vaccine efficacies against a clinical outcome, "clinical episodes" of otitis media, varied considerably ranging from -1 to 34 % . This is inline with the results obtained.

By considering all AOM episodes, a slight increase in vaccine attributable benefit could be detected compared to analytical methods that only take the first episode of AOM experienced by each child into account. [11]. The total number of episodes of acute otitis media in a population and also most importantly the effects of recurrence of the infection are better off in indicating burden of the disease. Thus, the reduction of the total number of episodes might be a more relevant measure of vaccine efficacy than the reduction of the number of children with at least one episode. This is why, statistical methods which account for recurrence in the data and therefore utilising this information are more appropriate then the methods that take into account only time to first episode.It has never been shown that there is a good correlation between the time to first recurrence and the course of illness as a whole. Therefore, although time to first recurrence is indeed an

important measure of survival in many fields of medical research (e.g. cancer research), it may only provide a rough approximation of the true course of illness.[18]

In comparison with the Cox proportional hazards model, it can be noted that there were not pronounced differences between the results of the two approaches interms of the Vaccine efficacy though the A-G model displayed slightly higher efficacy. This was expected because more information is used in the A-G model to estimate the efficacy of the vaccine. This also expresses why the conditional power computed is also higher than the cox model. It should be mentioned however, that differences between the two approaches could be quite substantial in most cases for example when there is complete follow up for each individual.

Finally, the conditional power has got a number of shortfalls. As described by Tweel et al [19],First, a choice must be made for a plausible parameter value to calculate the Conditional Power: for example, the parameter value as specified in the design phase under the alternative hypothesis, or the parameter value based on the data obtained so far. Second, a choice must be made for the critical value for the Conditional Power to decide for early stopping or continuing. Third, a choice must be made for the "optimal" information fraction to estimate the Conditional Power. Further more, conditional power assumes that the treatment effect does not vary which in most cases is not realistic. There are several statistical approaches to circumvent around these shortfalls. The most common alternative is use of group sequential analysis and use of predictive power in place of conditional power. The use of predictive power addresses the shortfall of fixed treatment effect by accounting for the variability of the true treatment effect. In this approach, the observed data is used to update some prior distribution for the treatment effect, and then the predictive distribution of the result at the final analysis is obtained by integrating over the posterior distribution of the treatment effect parameter [4]. These predictive probabilities have a distinct advantage over the conditional probabilities in that the predictive probabilities take into account both prior notions of the likely values for the true treatment probability and the evidence in the data for the true value.[20].

In this paper, we investigated the conditional power of a phase II vaccine efficacy trial via simulations. Two models of analysis that handle survival data were used, but with different approaches. The cox model and A-G model. The cox model was used in order to illustrate the incompetence of the model to handle recurrent data which the trial is characterized with. Conditionally, there is 63.9 % probability

that the trial will succeed at the end to reject the null hypothesis that the vaccine efficacy is less than 0%. However we would recommend using group sequential analysis or the predictive power to reliably estimate the probability of success at study end.

Bibliography

[1] Kevin J Carroll. Decision making from Phase II to Phase III and the probability of success: reassured by "assurance"? *Journal of biopharmaceutical statistics*, 23(5):1188–200, January 2013. ISSN 1520-5711. doi: 10.1080/10543406.2013.813527.

[2] Hui K Gan, Benoit You, Gregory R Pond, and Eric X Chen. Assumptions of expected benefits in randomized phase III trials evaluating systemic treatments for cancer. *Journal of the National Cancer Institute*, 104 (8):590–8, April 2012. ISSN 1460-2105. doi: 10.1093/jnci/djs141. URL `http://www.ncbi.nlm.nih.gov/pubmed/22491345`.

[3] Laleh Amiri-Kordestani and Tito Fojo. Why do phase III clinical trials in oncology fail so often? *Journal of the National Cancer Institute*, 104(8): 568–9, April 2012. ISSN 1460-2105. doi: 10.1093/jnci/djs180.

[4] Yanping Wang, Haoda Fu, Pandurang Kulkarni, and Christopher Kaiser. Evaluating and utilizing probability of study success in clinical development. *Clinical trials (London, England)*, 10(March):407–13, January 2013. ISSN 1740-7753. doi: 10.1177/1740774513478229.

[5] Kristine R Broglio, David N Stivers, and Donald a Berry. Predicting clinical trial results based on announcements of interim analyses. *Trials*, 15(1):73, January 2014. ISSN 1745-6215. doi: 10.1186/1745-6215-15-73.

[6] Bruce Jennison, Christopher and Turnbull W. *Group Sequential Methods with Applications to Clinical Trials*. Chapman & Hall/CRC, 2000. ISBN 0-8493-0316-8.

[7] Benjamin Trzaskoma and Andreas Sashegyi. Predictive probability of success and the assessment of futility in large outcomes trials. *Journal of biopharmaceutical statistics*, 17(1):45–63, January 2007. ISSN 1054-3406. doi: 10.1080/10543400601001485.

[8] Nauta Jozef. *Statistics in Clinical Vaccine Trials*. Springer New York, 2011. ISBN 9783642146909. doi: 10.1007/978-3-642-14691-6.

[9] M.Elizabeth Halloran, Longini Jr M.Ira, and Struchiner J. Claudio. *Design And Analysis of Vaccine Studies*. Springer New York, 2010. ISBN 9780387403137.

[10] Mark a Fletcher and Bernard Fritzell. Pneumococcal conjugate vaccines and otitis media: an appraisal of the clinical trials. *International journal of otolaryngology*, 2012:312935, January 2012. ISSN 1687-921X. doi: 10.1155/2012/312935.

[11] Antje Jahn-Eimermacher, Jean-Baptist du Prel, and Heinz-Josef Schmitt. Assessing vaccine efficacy for the prevention of acute otitis media by pneumococcal vaccination in children: a methodological overview of statistical practice in randomized controlled clinical trials. *Vaccine*, 25(33):6237–44, August 2007. ISSN 0264-410X. doi: 10.1016/j.vaccine.2007.05.062. URL `http://www.sciencedirect.com/science/article/pii/S0264410X07006548#`.

[12] Rodrigo Villegas, Olga Julià, and Jordi Ocaña. Empirical study of correlated survival times for recurrent events with proportional hazards margins and the effect of correlation and censoring. *BMC medical research methodology*, 13(1): 95, January 2013. ISSN 1471-2288. doi: 10.1186/1471-2288-13-95.

[13] Nirmala Pandeya, David M Purdie, Adèle Green, and Gail Williams. Repeated occurrence of basal cell carcinoma of the skin and multifailure survival analysis: follow-up data from the Nambour Skin Cancer Prevention Trial. *American journal of epidemiology*, 161(8):748–54, April 2005. ISSN 0002-9262. doi: 10.1093/aje/kwi098. URL `http://aje.oxfordjournals.org/content/161/8/748.full.pdf`.

[14] Zhenchao Guo, Thomas M Gill, and Heather G Allore. NIH Public Access. 47(2):107–116, 2009.

[15] Bart Gerritse. Appraisal of Several Methods to Model Time to Multiple Events per Subject : Modelling Time to Hospitalizations and Death. *Revista Colombiana de Estadistica*, 33(1):43–61, 2010.

[16] Chris Metcalfe and Simon G Thompson. The importance of varying the event generation process in simulation studies of statistical methods for recurrent events. *Statistics in medicine*, 25(1):165–79, January 2006. ISSN 0277-6715. doi: 10.1002/sim.2310.

[17] Andrea Burton, Douglas G Altman, Patrick Royston, and Roger L Holder. The design of simulation studies in medical statistics. *Statistics in medicine*, 25(24):4279–92, December 2006. ISSN 0277-6715. doi: 10.1002/sim.2673.

[18] Christopher Baethge and Peter Schlattmann. A survival analysis for recurrent events in psychiatric research. *Bipolar disorders*, 6(2):115–21, April 2004. ISSN 1398-5647. URL `http://www.ncbi.nlm.nih.gov/pubmed/15005749`.

[19] Ingeborg Van Der Tweel and Paulus A H Van Noord. Early stopping in clinical trials and epidemiologic studies for " futility ": Conditional power versus sequential analysis. *Journal of Clinical Epidemiology*, 56:610–617, 2003. doi: 10.1016/S0895-4356(03)00117-3.

[20] Scott S Emerson, John M Kittelson, and Daniel L Gillen. UW Biostatistics Working Paper Series Group Sequential Clinical Trials Group Sequential Clinical Trials. 2005.

Appendix A

Appendices

SAS Simulation and Analysis code

```
/*data _null_;
a='10OCT2013'd;
put 'SAS date='a;
put 'formatted date='a date9.;
run;

SAS date=19175
formatted date=01JUL2012
SAS date=19632
formatted date=01OCT2013
SAS date=19641
formatted date=10OCT2013

*/

proc iml;
accr=do(19175,19632,30);
f=quantile('normal',0.926);
print f;quit;

Proc iml;
beta1=-0.1863;
```

```
/***create id matrix***/
id=1:1540;id=id‘;
do sample=1 to 1000;
/**randomly sample 770 subjects***/;
call randseed(sample);
t_id=sample(id,770,"NoReplace"); t_id=t_id‘;
p_id=setdif(id,t_id);p_id=p_id‘;
/**treatment allocation***/
trt_1=j(nrow(t_id),1,1);
trt_0=j(nrow(p_id),1,0);
/**combine id with its allocated treatment**/
treat=t_id||trt_1;placebo=p_id||trt_0;
/**append treat and placebo***/
dataset=treat//placebo;
/*Allocate accrual dates*/
accr=do(19175,19632,30);accr=accr‘;
accr=repeat(accr,96,1);
accr1={19175,19565,19805,19355};
accr=accr//accr1;
dataset=dataset||accr;
/*Generate number of events in treatment group*/
num_evts_trt=j(nrow(treat),1,.);
call randseed(482 + sample);
call randgen(num_evts_trt,’poisson’,0.68);
/*Generate number of events in placebo group*/
num_evts_pcb=j(nrow(placebo),1,.);
call randseed(482 + sample);
call randgen(num_evts_pcb,’poisson’, 1);
num_evts=num_evts_trt//num_evts_pcb;
samp=repeat(sample,nrow(dataset),1);
dataset=dataset||num_evts||samp;
/**time to events**/
do k=1 to nrow(dataset);
num=dataset[k,4];
linpred=exp(beta1*dataset[k,2]);
if num=0 then do;
```

```
x=j(1,1,.);
call randseed(k);
call randgen(x,'uniform');
time=-log(x)/(0.6*linpred);
if time>1.8 then time=1.8;else time=time;
status=0;
idd=dataset[k,];cum_time=time;samp=sample;
row_sub=idd||time||status||cum_time;
end;
else if num>0 then do;
xk=0;
do until (xk=1);
x=j(num,1,.);
call randseed(k);
call randgen(x,'uniform');
time=-log(x)# 1/(0.6*linpred);
sum_time=sum(time);if sum_time>1.8 then xk=0;else xk=1;
idd=repeat(dataset[k,],num,1); status=repeat(1,num,1);
samp=repeat(sample,num,1);
if num=1 then cum_time=time;else cum_time=cusum(time);
row_sub=idd||time||status||cum_time;
end;
end;
/***combine all subjects***/
allsub=allsub//row_sub;
end;
Aom_data=sub_data//dataset;
end;

create thesis from allsub[colname={"id" "treat" "date" "num_events" "sample" "ti
append from allsub;
create Aom from sub_data[colname={'id' 'trt' 'date' 'num_events' 'sample'}];
append from Aom_data;
quit;
proc sort data=Thesis;
by sample id time cum_time;
```

```
run;
/***A-G Model Data Manipulation***/
Data Thesis2;
Set Thesis;
By sample id;
Var=days;
lag_Var = lag(days);
lag_time=lag(cum_in_days);
Retain Tstart;
If first.id then Tstart=0 and (tstop=days);
else tstart=lag_var;
Tstop=days;
if Last.id then Tstop=cum_in_days;
output;
drop lag_var var lag_time ;
run;
/****Andersen and Gill Model 2 sided alpha***/
ods output ParameterEstimates = ParmEst;
proc phreg data= thesis2 alpha=0.148 covs(aggregate) covm;;
by sample;
class status;
model (tstart tstop)*status(0) = treat / rl;
id id;
run;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Computation of probability of study success based on data accumulated until an Interim Analysis

Richting: **Master of Statistics-Biostatistics**
Jaar: **2014**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Milanzi, Edith

Datum: **11/09/2014**