2012•2013

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Statistical modeling for correlate of protection using accelerated failure time models and piecewise methods.

Promotor :
Prof. dr. Cristina SOTTO

Promotor :
Mrs. MARTINE DOUHA
Dr. FABIAN TIBALDI

Polycarp Mogeni

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

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

2012•2013
FACULTY OF SCIENCES
Master of Statistics: Biostatistics

Masterproef

Statistical modeling for correlate of protection using accelerated failure time models and piecewise methods.

Promotor :
Prof. dr. Cristina SOTTO

Promotor :
Mrs. MARTINE DOUHA
Dr. FABIAN TIBALDI

Polycarp Mogeni

Master Thesis nominated to obtain the degree of Master of Statistics , specialization Biostatistics

universiteit hasselt | Maastricht University

Acknowledgment

First is to thank the almighty God for giving me the opportunity to pursue the Master of Science in Biostatistics. Special thanks to my sponsor (VLIR-UOS) who made it possible to pursue the dream of becoming a Biostatistician and without which my dream of attaining a Master degree in Biostatatistics would just have remained a dream.

I am heartily thankful to my internal supervisor Professor Cristina Sotto of Hasselt University and the external supervisors Dr. Fabian Tibaldi and Martine Douha of GlaxoSmithKline vaccines, whose constant support and guidance towards understanding the subject matter and the write-up has been invaluable. I would also like to thank the GSK administration for granting me an opportunity to carry out my research work with them. I extend my appreciation to Hasselt University professors through whom I gained a lot of knowledge and expertise in the various areas of Biostatistics. My profound gratitude to Mrs. Martine Machiels, whom despite her busy schedule was always willing to assist whenever I sought assistance from her.

To my friends, colleagues and especially team members in my various projects; your close collaboration, assistance and contribution in all the difficult times that we went through together will forever remain valuable in my entire life. Similarly I wish to express my gratitude to my parents, my brothers, and sister for the support, prayers, encouragement and belief in my potential. Special thanks to Emilly Kerubo for your motivation, prayers, support and above all patience during this period.

Polycarp Mogeni

September 11, 2013

Table of Contents

List of Abbreviations

AFT: Accelerated Failure Time

AIC: Akaike's Information Cliterion

ATP: According-to-protocol

BIC: Bayesian Information Criterion

GSK: GlaxoSmithKline

IDMC: Independent Data Monitoring Committee

MMVR: Priorix-Tetra-(two-dose live varicella vaccine)

MMR +V: Varilrix – (one-dose live varicella vaccine)

MMR: Measles-Mumps-Rubella

PCR: Polymerase Chain Reactions

VZV: Varicella-Zoster Virus

WHO: World Health Organisation

Abstract

Chicken pox is an important childhood illness affecting mostly school-going children. The disease can be spread through contacts between infected and susceptible individuals. It is a very contagious disease caused by the varicella-zoster virus. Its main symptoms are: blister-like rash, tiredness, itching, and fever. Chicken pox can be serious, especially in adults, babies, and people with weakened immune systems. The objective of this study was to apply parametric survival models to determine whether there is a relationship between the risk of developing a varicella case and the titer obtained after vaccination and also to apply accelerated failure time models and the piecewise exponential model to determine the threshold that corresponds to the titer which best reflects the change in the risk of breakthrough varicella disease. In this analysis we applied the accelerated failure time models (Weibull model, the Log-logistic, the lognormal model) and the piecewise exponential model. The Akaike's Information Criterion and Bayesian Information Criterion values for the different thresholds were obtained and compared within each model to determine the most plausible range of values that can be regarded as cutoff values. In all the four models applied, the log antibody titer post vaccination was significant. These results show that varicella antibody titer 42 days after vaccination strongly correlate with long-term disease breakthrough. Based on the methodology applied and the set of AIC values obtained from the different thresholds, the data does not seem to support the existence of a cutoff value.

Key Words: *AFT; Weibull; Log-logistic; Piecewise exponential; Log-normal*

1. Introduction

Measles, mumps and rubella (MMR) are common viral childhood illnesses. These highly infectious diseases and their complications are responsible for considerable morbidity and mortality throughout the world [WHO, 2004; WHO, 2001; WHO, 2000]. In countries where immunization against MMR is routinely practiced, significant reduction in disease incidence is reported.

Varicella or chicken pox is a highly contagious disease that most children contract. It is commonly regarded as a mild childhood illness. However, serious complications such as secondary bacterial infection and pneumonia from varicella infections can occur, leading to hospitalizations and, in rare cases, even to death [WHO,1998]**.** Although the varicella vaccine is licensed in many countries, it is not routinely used, especially in healthy subjects. The universal varicella vaccination is only implemented in a limited number of countries (e.g. United States, Germany, Sicily, Qatar and Uruguay). The incidence of varicella disease and the rate of related hospitalization in the US have declined by 80% since the introduction of routine vaccination in 1995 [Galil, 2002]. Though a vaccinated person can get chicken pox, the symptoms are usually very mild and only last for a few days.

The human adaptive immune response relies on a complex combination of cellular and humoral immunity, mediated by T- and B-lymphocytes. Although vaccination aims to activate cellular and humoral immunity, vaccine-induced immunity is typically evaluated by means of the antibody titer secreted by B-cells and plasma cells (PC). Memory B-cells permit a faster and more effective immune response upon further exposure to the antigen, whereas PC are the main antibody secreting cells [Mathieu et al, 2012].

The main objective of vaccination is to stimulate an individual's immune system to develop adaptive immunity to a target pathogen. Vaccine administration seeks to protect an individual against an infection or minimize the unpleasant clinical features associated with the target disease in case of breakthrough after vaccination. The development of vaccines has been made possible through the use of weakened live viruses, inactivated viruses, purified bacterial proteins and glycoproteins, and recombinant, pathogen-derived proteins. Typically the most convenient

measure of vaccine-induced immunity is the antibody titre after vaccination. The relationship between antibody titre and protective efficacy has led to the concept of a correlate of protection, which is often described as a protective level of antibody titre that can be used to determine whether a person is not protected and still susceptible to the disease or completely protected [Wilson et al, 2004; Lachenbruch et al, 2000]. Although establishing a protective antibody level is desirable, identifying a clear-cut value is very challenging and sometimes impossible with more complex pathogens. For example, clinical trials with VARIVAX^{TM} (a live attenuated varicella [Oka/Merck] vaccine) have shown the vaccine to have high efficacy against varicella [Weibel et al, 1984; Kuter et al, 1991]. Long-term follow-up of vaccine recipients has suggested that the frequency of disease breakthroughs after vaccination is inversely related to the post-vaccination antibody level [White et al, 1992]. However, as far as these studies are concerned, no particular antibody level can be considered an absolute protective level.

In this study, GlaxoSmithKline (GSK) vaccines conducted a randomised clinical trial to evaluate the protection against varicella afforded by administration of two doses of MMRV or one dose of Varilrix. To achieve more robust efficacy estimates than in previous multi-year live varicella vaccine trials, the first phase of this trial (reported here) was observer-blinded and used a concurrent control group, active surveillance for varicella, and rigorous case confirmation and severity-grading procedures. This trial was initiated in countries with endemic varicella among toddlers, permitting a stringent test of vaccine efficacy.

1.1 Objectives of this work

The objective of this work is firstly to apply the parametric survival model to determine whether there is a relationship between the risk of developing a varicella case and the titer obtained after vaccination. Secondly, to apply accelerated failure time (AFT) models and the piecewise exponential model to determine the threshold that corresponds to the titre which best reflects the change in the risk of breakthrough varicella disease.

2. Study design

This study is the first part (September 2005 to June 2009) of an observer-blinded, randomised, controlled trial conducted at multiple sites in ten European countries.

Eligible subjects were randomised according to a 3:3:1 ratio into one of three treatment groups, respectively (Figure 1): (i) Group MMRV, where MMRV(*Priorix-Tetra*™, GlaxoSmithKline Vaccines [GSK]) was administered at Dose 1 and 2; (ii) Group MMR+V, where MMR was administered at Dose 1 (*Priorix*™, GSK) and V (*Varilrix*™, GSK) at Dose 2; or (iii) Group MMR, where MMR (*Priorix*™) was administered at Dose 1 and 2. Doses were administered 42 days apart (Day 0 and Day 42). After the completion of this first phase of the clinical trial, Group MMR+V subjects were offered the second dose of MMR in accordance with the immunisation schedule of their respective country. The eligible subjects were healthy children aged 12–22 months at the time of the first vaccination; had a negative history of varicella, mumps, measles and rubella diseases/vaccinations, and were either (i) at home with at least one sibling (with negative history of varicella/vaccination), or (ii) attending a childminder (where at least one child was without a known positive history of varicella/vaccination), or (iii) playing for more than five minutes weekly with children without a known positive history of varicella/vaccination, or (iv) registered to attend a daycare center from 24 months of age. The study was conducted in 111 study centers.

2.1 Study endpoints

The primary efficacy endpoint was the occurrence of confirmed varicella from 42 days after the second vaccine dose to the end of the first phase of the trial.

2.2 Data and Sample Size

The dataset was obtained according to the study design described in the previous section. The number of subjects to be enrolled was estimated to be 5,754 children in order to obtain the required number of evaluable confirmed varicella cases within two years of follow-up assuming an annual attack rate of 5% for children in the control group and a vaccine efficacy of 80%. Details of sample size calculation are not shown here. In total 5,803 subjects were enrolled.

5,285 subjects were included in the according to protocol (ATP) analysis, but in this analysis 143 subjects were eliminated from the main analysis due to missing immunological results.

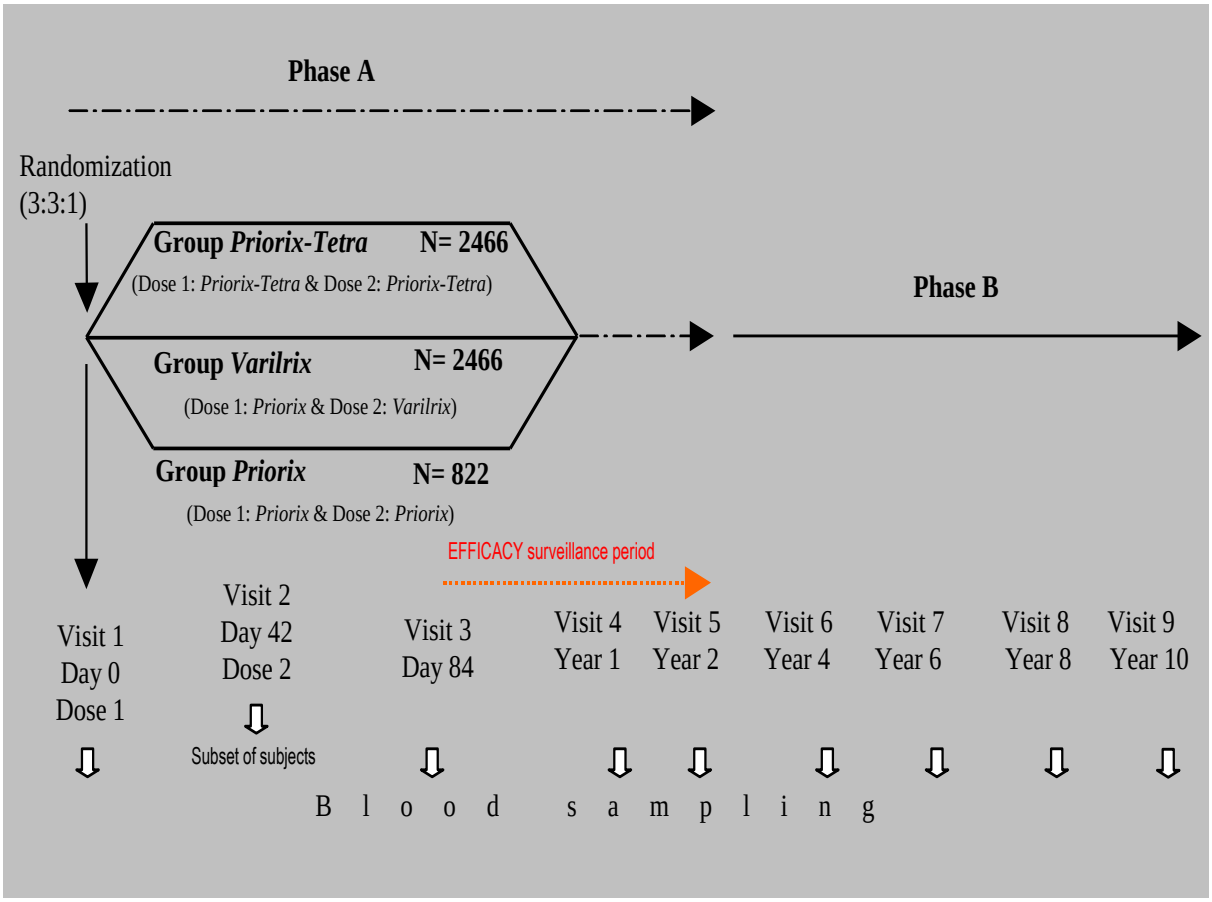


Figure 1: *Trial Design*

2.3 Case definition

Confirmed varicella cases are reached during the time frame between the subjects' 42 days post vaccination and the appearance of breakthrough varicella case. The disease was initially identified through parents examining their children for possible development of skin reactions that might be indicative of varicella/zoster after vaccination. The rash onset that is, the date when the first lesion appears and the date the rash ended, that is, date of the first day when no new lesion appears were recorded by the parents, while the investigators were required to collect biological samples (that is dermal lesions) for viral identification. An Independent Data Monitoring Committee (IDMC) consisting of experts in varicella disease and microbiology and

other appropriate disciplines reviewed and classify suspected varicella cases according to the case definition illustrated in Figure 2.Varicella severity grading was based on the scoring system of Vázquez *et al.* with modifications: mild disease, ≤7 points; moderately severe disease, 8–15 points; severe disease: ≥16 points.

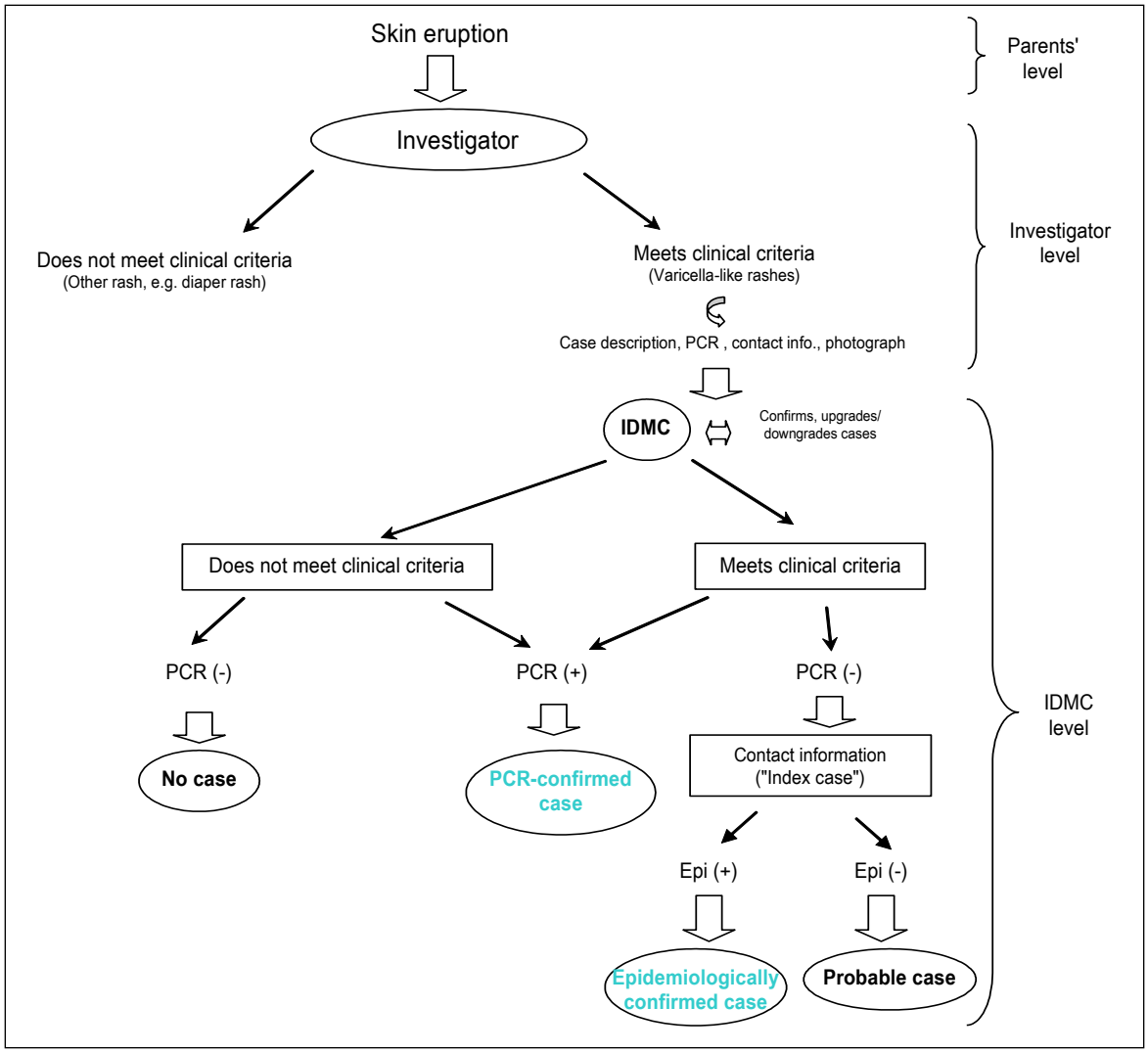


Figure 2: *Varicella case definition*

2.4 Data Description

In this section we will briefly describe the variables used in the statistical modeling exercise. The time at risk was measured as the number of days from the 42^{nd} day after the date of vaccination to the date of onset of a confirmed case of breakthrough disease or the duration in days of follow-up starting the 42^{nd} day post vaccination to the end of last contact with the subject in case of no breakthrough disease. Other subject characteristics were measured. A binary covariate which took the value 1 if there was a breakthrough disease and 0 if there was no varicella case, the continuous covariate representing the antibody concentration, the vaccination group as categorical variable denoting the type of vaccine that was administered to the subject and the categorical variable indicating the severity of the breakthrough disease.

3. Statistical Methods

The methodology discussed in this section is contained in the paper by Ivan et al, 2002, looking at the use of statistical models for evaluating antibody response as a correlate of protection against varicella. For a detailed description refer to the mentioned paper.

3.1 Exploratory Data Analysis

To gain insight into the dataset, summary statistics, tables of frequencies and graphics were considered.

The vaccine efficacy analyses were performed with a method that can handle uncensored as well as censored data. The Cox proportional hazards regression model on time to first event fulfills this requirement. By including the vaccine group effect as the only regressor in a Cox proportional hazards regression model, the relative risk (RR) of being a varicella case in a vaccine group compared to being a varicella case in the control group was estimated through the hazard ratio (i.e. RR=exp(B), where B is the estimated regression coefficient for the dummy regressor). The vaccine efficacy estimate was then obtained as (1–RR)*100%. A 97.5% confidence interval for the vaccine efficacy was calculated using the same regression analysis for each comparison between a varicella vaccine and the control group [Dipika et al, 2011].

3.2 Accelerated Failure Time (AFT) Models

The main interest of this study was also to assess the effect of the antibody titer 42 days after vaccination on the survival function, and thus, AFT models were applied.

Let T_i be the time to a varicella event after vaccination for the *i*th person (i=1,…, n), and (X_{i1},…, X_{ik}) be the set of baseline covariates for that person. The general AFT model specifies that

$$\ln(T_i) = \beta_0 + \beta_1 X_{i1} + \cdots \beta_k X_{ik} + \sigma\varepsilon_i \,, \qquad (1)$$

where ε_i is a random disturbance term, usually assumed to be independent and identically distributed with some density function $f(\varepsilon)$. $\beta_0, \beta_1, \ldots, \beta_k$ are regression parameters of interest, and σ is a scale parameter. Different distributional assumptions can be assumed for the random disturbance term in model (1) and hence the type of AFT model.

To assess the relationship between the probability of having a varicella event after vaccination and the 42 days varicella antibody titre, three popular AFT models were considered; the Weibull, log-logistic and log-normal hazards [Lawless, 1982; Kalbfleisch et al, 1980] and the piecewise exponential model. The main reason for allowing these different distributions in our analysis is that they encompass a variety of hazard functions that are widely used to analyze time-to-event data [Friedman, 1982; Allison, 1995].

Motivation for the choice of AFT models

The AFT models are effective methods for regression analysis of censored survival data. Though these models are less robust than the more widely used Cox regression analysis, in most cases, the results produced by the two approaches are very similar [Allison, 1995]. Moreover, unlike the cox regression, the AFT methods make it possible to test hypothesis about the shape of the hazard function. All the AFT models to be considered in the next section assume that the hazard is a smooth, relatively simple function of time. Cox model is much less restrictive in this regard, but it lacks the facility to test hypotheses about the shape of the hazard function. One way to get some of the flexibility of the Cox model without losing the hypothesis testing capability is to employ the piecewise exponential model. The piecewise exponential model tries to gain most of the strength of both the AFT models and the cox regression model while minimizing their weaknesses. The major challenge with the piecewise exponential model is the difficulty in determining the cut points or the size of the pieces.

3.2.1 Weibull model

The Weibull model assumes that ε_i has a standard extreme value distribution, and the log-hazard function for the *i*th person is given by

$$ln\,(h_i(t)) = \alpha \log t + \beta_0{}^* + \beta_1{}^* X_{i1} + \cdots + \beta_k{}^* X_{ik}\,, \qquad (2)$$

where $\alpha = (1/\sigma - 1)$, and $\beta_j{}^* = \frac{-\beta_j}{\sigma}$ for j=1,…k .

The hazard function is monotonic in the Weibull model. When $\sigma = 1$, the model simplifies to the exponential model in which the hazard is constant over time. The hazard decreases with time

if $\sigma > 1$ and increases with time if $\sigma < 1$. The Weibull model is also a proportional hazards model.

The hazard function is given as $h(t) = \sigma\lambda t^{\sigma-1}$, and the survival time is given as $S(t) = \exp(-\lambda t^{\sigma})$. Having defined the hazard and survival functions, the Weibull model maximizes the likelihood:

$$l = \prod_{i=1}^{N} \left\{\sigma\lambda(\lambda t)^{\sigma-1}\, e^{(-\lambda t)^{\sigma}}\right\}^{d_i} \left\{e^{(-\lambda t)^{\sigma}}\right\}^{1-d_i} , \quad (3)$$

where λ is reparameterized in terms of predictor variables and regression parameters.

The cumulative probability distribution function of the Weibull model for the *i*th person is

$$C_i(t) = \Pr(T_i \le t) = 1 - \exp\left\{-\left[te^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_k X_{ik})}\right]^{1/\sigma}\right\}. \quad (4)$$

3.2.2 Log-normal model

The log-normal model assumes that ε_i has a standard normal distribution. Its log-hazard function has no closed form expression, but can be expressed by the following relationship:

$$ln\,(h_i(t)) = \log h_0(t\, e^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_k X_{ik})}) - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}) , \quad (5)$$

where $h_0(.)$ is the common hazard function when all covariates ($X_{i1},\ldots, X_{ik}$) are 0. Unlike the Weibull model, the log-normal model has a non-monotonic hazard function that starts at 0 when t=0 and increases to a peak and then declines toward 0 as time goes to infinity. The density function is given as $f(t) = \frac{1}{t\sqrt{2\pi\gamma}} \exp\left[-\frac{1}{2\gamma}(log(t) - \mu)^2\right]$ and the survival time is given as $S(t) = 1 - F_N\left(\frac{log(t)-\mu}{\sqrt{\gamma}}\right)$. The cumulative probability distribution function of the log-normal model for the *i*th person is

$$C_i(t) = \Phi\left\{\frac{1}{\sigma}[\log t - (\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})]\right\} , \quad (6)$$

where $\Phi(.)$ is the standard normal distribution function and γ is the variance. In addition, the log-normal and the Weibull models are special cases of the generalized gamma model [Allison, 1995].

3.2.3 Log-logistic model

The log-logistic model assumes that ε_i has a logistic distribution. Its log-hazard function has a relationship similar to (5) except that $h_0(.)$ takes a different form. The hazard function behaves similarly to log-normal hazard with a longer right tail when $\sigma < 1$. When $\sigma > 1$, the hazard is decreasing and is similar to the Weibull hazard [Allison, 1995]. The density function is given as $f(t) = \frac{\sigma t^{\sigma-1}\lambda^{\sigma}}{[1+(t\lambda)^{\sigma}]^2}$ and the survival time is given as $\mathrm{S(t)} = \frac{1}{1+(t\lambda)^{\sigma}}$.

The cumulative probability distribution function of the log-logistic model for the ith person is

$$C_i(t) = \frac{[te^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_k X_{ik})}]^{1/\sigma}}{1-[te^{-(\beta_0+\beta_1 X_{i1}+\cdots+\beta_k X_{ik})}]^{1/\sigma}} . \quad (7)$$

The log-logistic model is also a proportional odds model since the log-odds of the survival function, $1 - C_i(t)$, is linear:

$$\ln[\frac{1-C_i(t)}{C_i(t)}] = \frac{1}{\sigma}(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} - \log t) . \quad (8)$$

3.2.4 Piecewise exponential model

Piecewise exponential model [Friedman, 1982; Allison, 1995] allow more flexible estimation of underlying hazards. Suppose the follow-up time is divided into intervals with cut points $0 = a_0 < a_1 < a_2 < \cdots < a_J = \infty$; the piecewise exponential model assumes a constant hazard in each time interval. Given the covariates, the hazard function for the *i*th person is

$$h_i(t) = \lambda_j \exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik}), \quad (9)$$
$$for\ a_{j-1} \le t < a_j, j = 1, \ldots, J ,$$

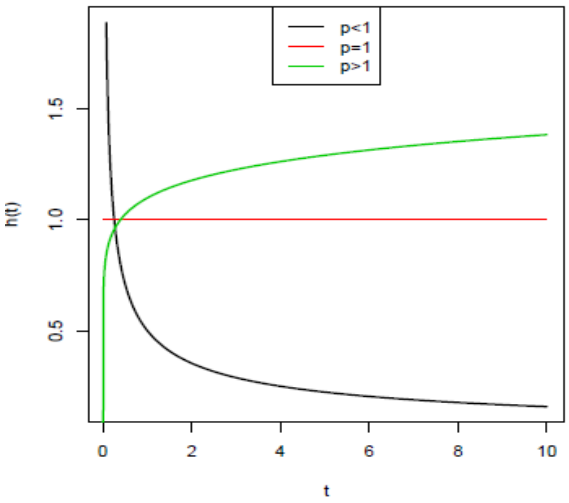
and the cumulative event rate can be calculated as, for $a_{j-1} \le t < a_j$,

$$C_i(t) = 1 - \exp\left\{-\left[\sum_{k=1}^{j-1} \lambda_k(a_k - a_{k-1}) + \lambda_j\left(t - a_{j-1}\right)\right] exp(\beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik})\right\}. \quad (10)$$

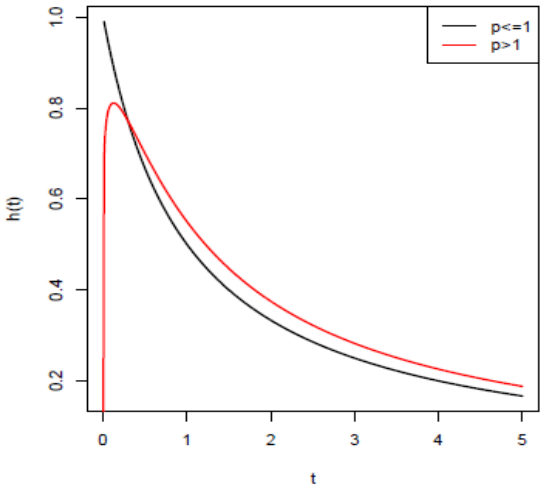
The piecewise exponential model is a proportional hazards model, and it provides a simple way of approximating a variable hazard function that may change abruptly at some threshold time points. This model is similar to the Cox proportional hazards model for estimating effects of covariates, but it gives a simple functional form of hazard that can be used for estimation and prediction purposes. Clearly, sensible choice of the cut points should allow one to approximate reasonably well almost any baseline hazard, using closely-spaced boundaries where the hazard varies rapidly and wider intervals where the hazard changes more gradually. The regression parameter estimates from the piecewise exponential model are generally close to those obtained from the Cox model, especially if small time intervals are used [Allison, 1995].

The possible hazard function for the Weibull, log-logistic and the piecewise exponential model are presented in Figure 3.

Weibull hazard function



Log-logistic hazard function



Piecewise exponential hazard function

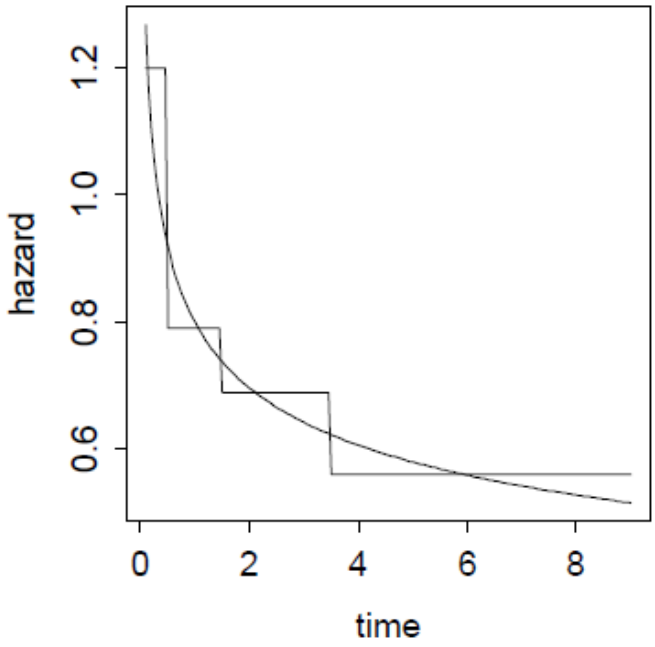


Figure 3: *Hazard functions for different models*

3.3 Software

3.3.1 Procedures for determining the cutoff values

The AFT model for right censored data can be implemented using the PROC LIFEREG or PROC NLMIXED in SAS. However, due to its flexibility in terms of programing, PROC NLMIXED was preferred. This procedure allows iterations to be specified within its body and hence facilitates the programming. To guarantee the correct specification in PROC NLMIXED the parameter estimates were compared with those obtained from PROC LIFEREG. To determine the set of values that best reflects the change in risk of breakthrough disease, all the four models (Weibull, the log-normal, the Log-logistic and the piecewise exponential models) were programmed using PROC NLMIXED. Equation 11 presents the AFT model specification for the different thresholds.

$$ln(T_i) = \beta_0 + \beta_1 X_i + \beta_2 * (X_i - log10(threshold)) * Y + \sigma\varepsilon_i\,, \qquad (11)$$

where X is the log concentration of antibody titer,

The threshold takes values from 50-300 in intervals of 5 units and

$$Y = \begin{cases} 1 & if\ \ \mathrm{X} > \mathrm{Log10(threshold)} \\ 0 & otherwise \end{cases}.$$

An iterative procedure running from a threshold of 50 through 300 in intervals of 5 was implemented through a macro. The Akaike's Information Criterion (AIC) and Bayesian Information Criterion (BIC) values for fitted models at different thresholds were obtained and compared within each model to determine the most plausible range of values that can be regarded as cutoff values. These are the set of thresholds that yield the minimum BIC and AIC values. Statistical analysis was performed using SAS version 9.2 and R version 2.15.2.

3.3.2 Data preparation for piecewise exponential model

The piecewise model was constructed by splitting the surveillance period of 4 years into 4 pieces of 1 year. A record is created for each year during which an individual was at risk of experiencing an event. Four records were created for persons who experienced an event in the fourth year or did not experience an event during the four year follow up. Three, two and one records were created for those who experienced an event in the third, second and first year,

respectively. Each record is treated as a distinct observation, with the time reset to 0 at the beginning of the year. If an event occurred in the quarter, a censoring indicator variable for that person-year is set to 1; otherwise, it is set to 0. If neither an event nor censoring occurred in the year, the time variable is assigned the full 365 days. If a breakthrough or censoring occurred for any given reason, the time variable is coded as the length of time from the start of that year until the event/ censoring. For the antibody titer concentration, the piecewise model was constructed by first splitting the antibody concentration into pieces of 50, and a similar procedure as that applied for time is used to create the piece.

3.4 Model selection and determination of cutoff

The descriptive assessment of the model fit was done by comparing the predicted probability of developing clinically–diagnosed varicella during year 1 to 4 after vaccination from the four models with the observed Kaplan-Meier probabilities [Lawless, 1982; Kalbfleisch et al, 1980]. AFT models can be compared using the likelihood ratio test if they are nested [Allison, 1995]. However, this is not the case when one wants to make a choice among the log-normal, Weibull, log-logistic and piecewise exponential methods. Hence they can be compared using the AIC or BIC. The Akaike's method penalizes each model's log likelihood by the number of parameters that are being estimated. The lower the AIC the better is the model.

The AIC is calculated as:

$$AIC = -2lnl + 2k.$$

The BIC is calculated as:

$$BIC = -2lnl + kln(n),$$

where k is the number of parameters in the model and n is the sample size.

The BIC and AIC were used to determine the set of values that best reflect the change in risk of breakthrough disease. In this case, these are the set of values that gives the lowest BIC and AIC in a given range of antibody titer.

4. Results

4.1 Exploratory data analysis

Between September 2005 and May 2006, 5803 subjects were enrolled (Figure 2). The majority (56·4%) of the 5285 subjects in the efficacy cohort were from the Czech Republic, Poland, and the Russian Federation. The mean duration of follow-up in the per protocol efficacy cohort was 35 months and was similar in each of the treatment groups.

In the MMR (control) group, there were 201 varicella cases, giving an incidence of 10·4 per 100 person years, of these, more than half of the cases were graded as moderately severe or severe. In the MMR+V group (one-dose live varicella vaccine), there were 243 varicella cases, giving an incidence of 3·8 per 100 person years. In the MMRV group (two-dose live varicella vaccine), there were 37 varicella cases, giving an incidence of 0·6 per 100 person years. The efficacy of two-dose vaccination against all varicella was 94·9% and against moderate to severe varicella, was 99·5%. For one-dose, these efficacy rates were 65·4% and 90·7%, respectively. The summary of the incidence and their respective confidence intervals are presented in Table 1.

Table 1**:** *Summary of the incidence of varicella cases in each treatment group per 100 person years*

Group	Varicella severity	n / N	Total Time in year	Incidence rate (97·5%CI) 100 person year^{-1}	Vaccine Efficacy (97·5%CI)
MMRV	all	37 / 2279	6690	0·6 (0·4–0·8)	94·9 (92·4–96·6)
	mod./sev.	2 / 2279	6740	0·0 (0·0–0·1)	99·5 (97·5–99·9)
MMR+V	all	243 / 2263	6455	3·8 (3·3–4·3)	65·4 (57·2–72·1)
	mod./sev.	37 / 2263	6698	0·6 (0·4–0·8)	90·7 (85·9–93·9)
MMR	all	201 / 743	1934	10·4 (9·1–11·9)	-
	mod./sev.	117 / 743	2047	5·7 (4·8–6·9)	-

The distribution of the antibody titer was found to be right skewed and was therefore log transformed. The box plot (Figure 4) hints on presence of number of outlying observations. The median antibody titer for the MMVR group is higher than that MMR+V and MMR group.

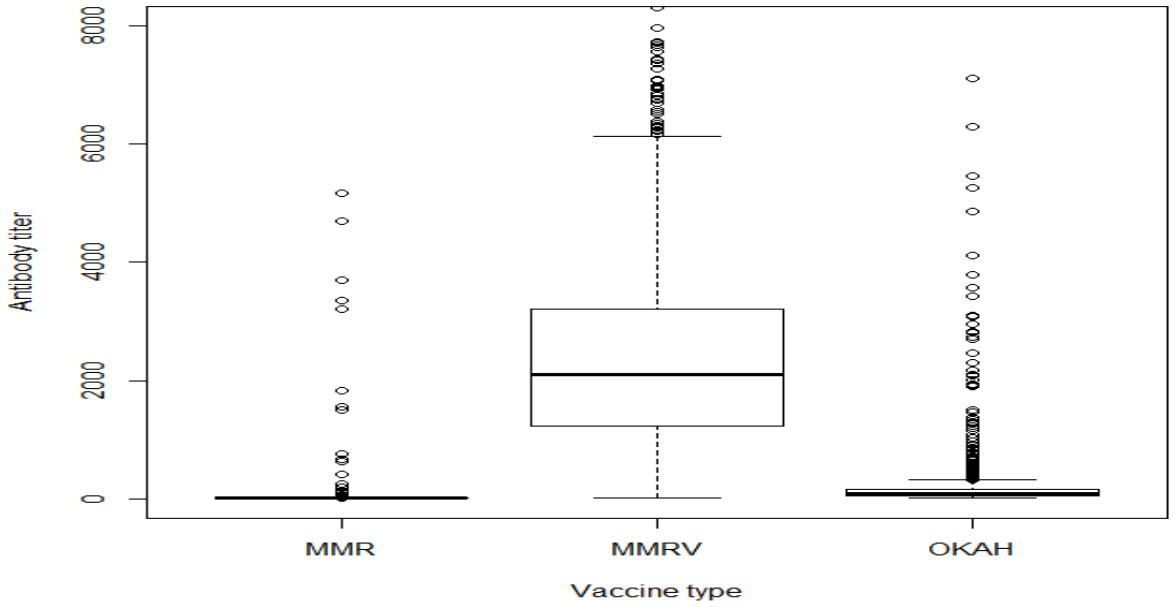


Figure 4: *Box plot displaying the distribution of antibody titer in each treatment group*

To assess the possible correlation between the antibody titer concentration and breakthrough disease among the vaccinated patients only, different cutoff values were assessed graphically as displayed in the Figure 5. From the graph we observe that while breakthrough cases occurred across a range of VZV titers, higher titers were associated with less breakthrough disease.

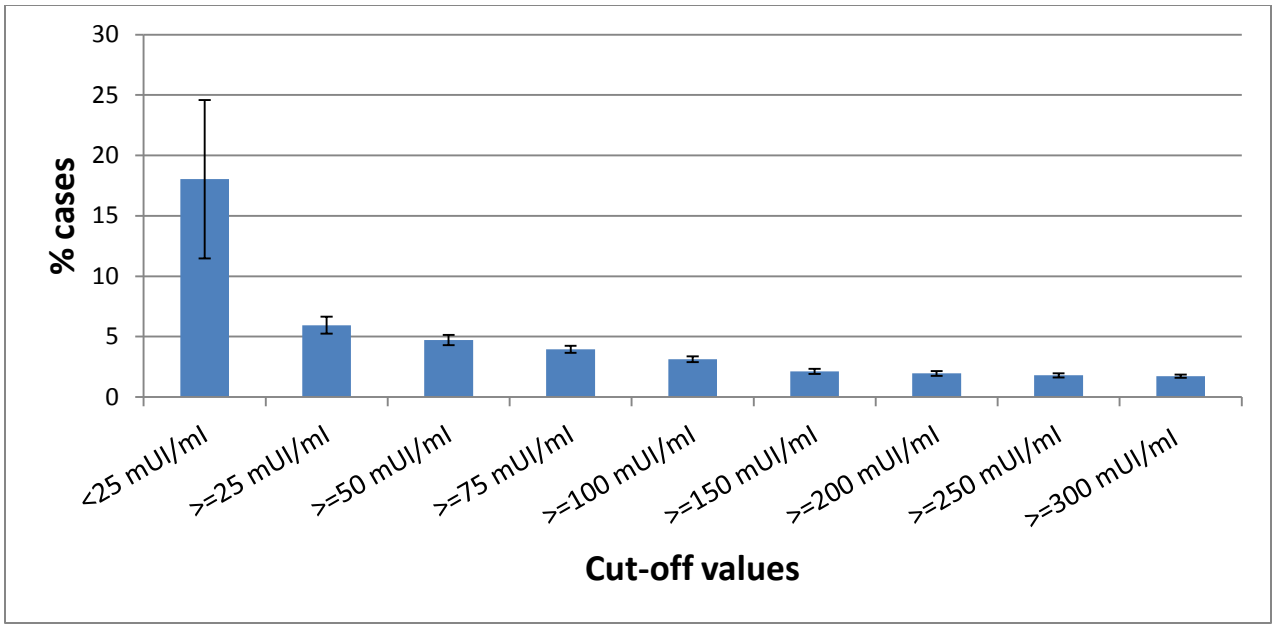


Figure 5: *Breakthrough rate of confirmed varicella as a function of anti-VZV concentration 42 days after vaccination*

4.2 Parametric model estimates

The parameter estimates obtained from fitting the AFT models and the piecewise exponential model are shown in Table 3 below. In all the four failure time models, the log antibody titer 42 days post vaccination was significant. These results consistently show that varicella antibody titer 42 days after vaccination strongly correlate with long-term varicella breakthrough.

Table 3: *Statistical model parameter estimates of the time to varicella event after vaccination.*

Method	Covariate	Estimate [95% CI]	P-value	AIC
Weibull	Log antibody titer	0.81 [0.70, 0.91]	<0.001	9202.516
Lognormal	Log antibody titer	0.90 [0.79, 1.01]	<0.001	9283.54
Loglogistic	Log antibody titer	0.82 [0.71, 0.92]	<0.001	9211.026
Piecewise (piece on time)	Log antibody titer	1.34 [1.26, 1.43]	<0.001	21645.29
Piecewise (piece on both time and titer concentration)	Log antibody titer	1.28 [1.20, 1.37]	<0.001	54343.16

The piecewise levels 1, 2, 3, 4, corresponding to the yearly pieces covered by each record in the piecewise exponential model, show a significant effect, implying that the hazard is not constant over time (Table A3, Appendix). The Wald chi-square value is 18.17 on 3 degrees of freedom and a p-value 0.0004. From this result the Weibull model (AIC=9202.516) fits the data better than the log-normal, log-logistic and the piecewise model. It can be observed from the model fitted that subjects with higher antibody titer 42 days after vaccination have better disease free survival. For the Weibull model, for one unit increase in log-antibody titer, the subjects' average disease free survival is twofold (2.24 times) (Table A1, appendix).

Table 4 below presents the various model-based estimates of cumulative varicella event rates through years 1 to 4. All the four yielded cumulative event rates estimates that were quite similar to those obtained from the life-table estimates. In addition, the piecewise exponential model gave year-by-year estimates that are almost identical to the life-table estimates except for the 4th year. This could be attributed to the low number of subjects in the fourth year compared to the previous years. By comparison the year-by-year estimates from the three AFT models and the piecewise exponential model were similar to those obtained from the life-table estimates.

Table 4: *Estimated cumulative rate of varicella event rates from year 1 through year 4 after vaccination calculated as a life table estimate using the Kaplan Meier method and predicted cumulative rates as estimated by different statistical models.*

Time Interval	Number of Reported Varicella cases	life-table Estimate Per 100 person year	Model-based Cumulative estimate per 100 person year			
			Weibull	Log-normal	Log-logistic	Piecewise exponential
Day 42-Year1	48	1.07%	0.91%	1.07%	0.91%	1.06%
Year1-Year2	69	2.73%	3.07%	3.41%	3.10%	2.71%
Year1-Year3	111	5.61%	6.12%	6.13%	6.10%	5.62%
Year1-Year4	52	8.92%	9.85%	8.84%	9.59%	10.90%

The cumulative probability of events over 3 year follow-up period as a function of titer was predicted from the models at different levels of the antibody titer concentration (Table 5). It can be seen that the cumulative probability of breakthrough varicella decreases with increasing thresholds. This observation is consistent across all the AFT models and the piecewise exponential model.

Table 5: *Cumulative probability of breakthrough varicella over 3 year follow-up period as a function of anti-VZV concentration among vaccinated subjects.*

Threshold	Weibull	Log-normal	Log-logistic	Piecewise Exponential
25	19.0%	18.9%	19.3%	17.5%
50	13.0%	13.8%	13.3%	11.8%
75	10.3%	11.3%	10.6%	9.5%
100	8.8%	9.7%	9.0%	7.9%
125	7.7%	8.6%	7.9%	7.1%
150	7.0%	7.8%	7.1%	6.4%
200	5.9%	6.6%	6.0%	5.4%
300	4.6%	5.2%	4.7%	4.2%

The estimated predictions for the three years follow-up at different thresholds are very similar, especially among the three AFT models (Weibull model, lognormal model and the loglogistic model). However, the piecewise exponential model estimates are slightly different from the rest of the models.

Figure 6 below presents the plot of the cumulative probability estimates against the log varicella antibody titer for the four models. For all the AFT and piecewise models presented in Figure 6, it can be observed that the cumulative varicella event rates decreases monotonically with increasing antibody titer.

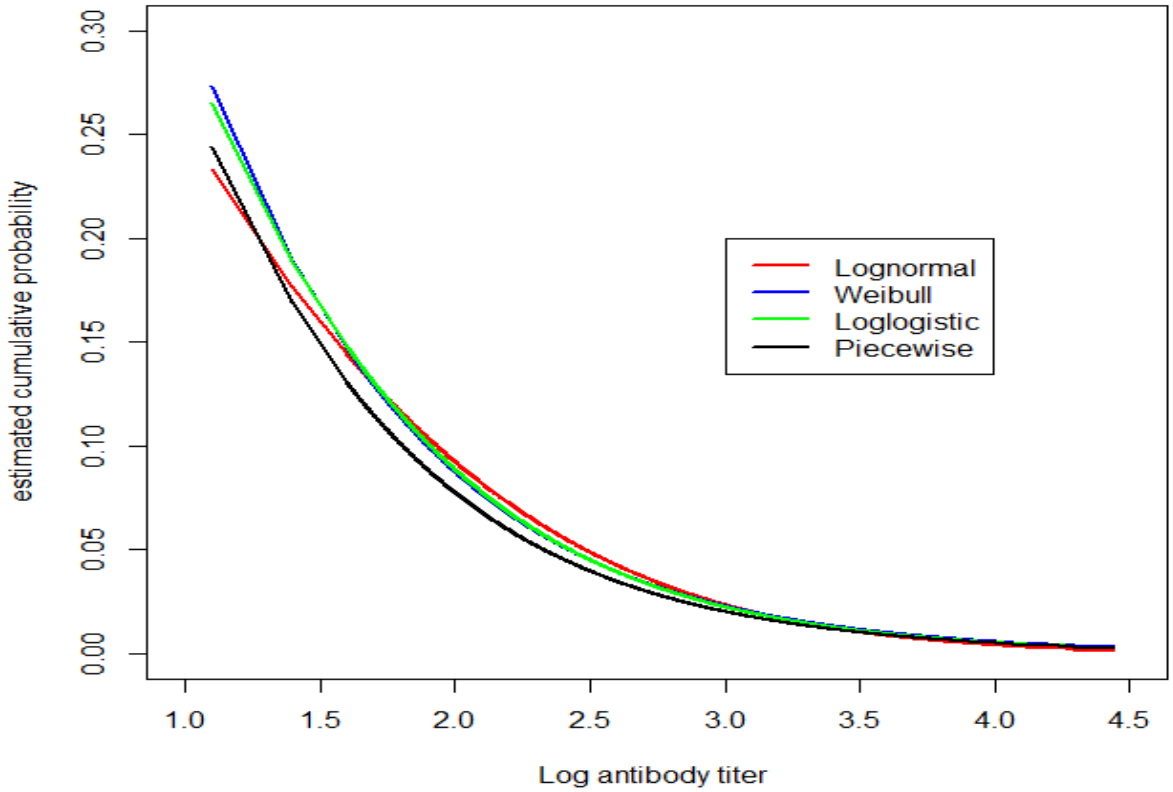
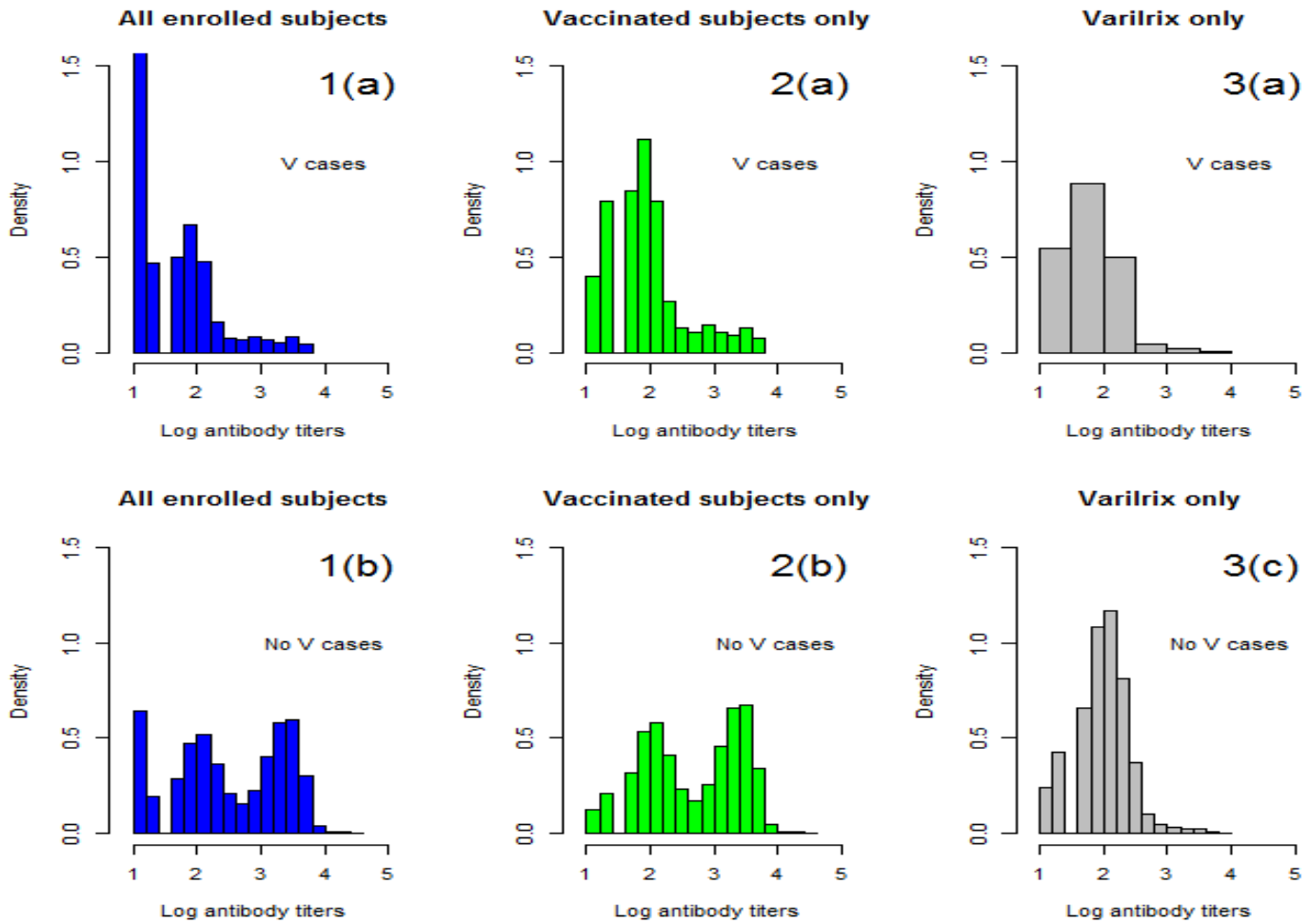


Figure 6: *Estimated cumulative probability of a varicella event through 3 years follow-up period*

4.3 Distribution of log varicella antibody titers by event

The distribution of varicella titer by varicella cases was assessed among all enrolled subjects, vaccinated subjects (that is, subjects who received MMRV or MMR+V) and among subjects vaccinated with MMR+V. The objective was to gain insight to the possibility of obtaining the threshold that correspond to the titer which best reflects the change in the risk of breakthrough of varicella disease. Figure 7; 1(a) compared to 1(b), 2(a) compared to 2(b) and 3(a) compared to 3(b) indicates that the distribution of varicella titer among varicella cases (V cases) overlaps with that of no varicella cases (No V cases). This implies that if there exists a single cutoff value, then one has to contend with some amount of misclassification.



*V cases --→ Varicella events ; No V cases-→ No varicella events

Figure 7: *Distribution of the varicella titers as measured by BELISA in MMRV and Varilrix groups*

4.4 Determination of the cutoff value

To determine the set of antibody titer values that best reflect the change in the risk of event, we applied the three AFT models and the piecewise exponential model. This was done on all subjects, only the vaccinated groups (MMRV and the MMR+V) and on the Varilrix group only (MMR+V). The MMRV group only was not investigated due to the low number of events. The results in terms of the AIC are presented in the appendix (Tables A4 and A5).

The choice of the range of antibody titer used in the investigation of the cutoff value was motivated by the knowledge that the antibody titer of 50 is the limit of quantification while the upper limit was motivated by the knowledge that not many Subjects get an antibody titer greater than 300 post vaccination.

Plotting the graph of AIC against the thresholds, we expect to a U-shape relationship. That is, we expect the model to fit poorly with thresholds far away from the cutoff values. The obtained graphs (Figure A1, A2 and A3, Appendix) do not seem to capture the expected U-shape. From the given dataset and the AIC values obtained, there seems to be no set of values that can be said to best reflect the change in risk of breakthrough varicella disease. All the three AFT models and the piecewise exponential model seem to yield similar conclusion regarding to the possibility of obtaining a set of values that could be regarded as possible cutoff values.

5. Discussion

Varicella is a highly contagious disease that is caused by the varicella-zoster virus (VZV) and predominantly affects preschool and school aged children and therefore its importance as a health problem cannot be over emphasized. Vaccination against varicella aims to educate the immune system by introducing varicella memory specific pathogens in its absence. To measure the effect of varicella vaccine, the antibody titer is normally used as a marker for the vaccine's efficacy. After vaccinating the children, the antibody titer concentration in the blood is expected to rise to a maximum after approximately 42 days which explains why the follow up time variable and the antibody titer were measured 42 days post vaccination.

From data exploration we observed that breakthrough events occurred across the whole range of antibody titers, though higher titers were associated with less breakthrough events. The distribution of antibody titer was highly skewed which motivated the log transformation of the titer concentration. For the 4 years follow-up period, higher proportions of events were observed among patients in the control group when compared to the vaccinated groups. From the box plot and the table of means for the log antibody titer, it was observed that the vaccinated children had a higher mean log antibody titer compared to those in the MMR group.

The efficacy of the two-dose and one-dose vaccination against moderate to severe varicella, was high, that is 99·5% and 90·7% respectively which is similar to some clinical trials with VARIVAX^{TM} where the vaccine was shown to have high efficacy against varicella [Weibel et al, 1984; Kuter et al, 1991]. Moderate to severe breakthrough varicella was highest among the control group (MMR), with an incidence of between 4.8 and 6.9 per 100 person years and lowest among subjects in the MMRV group.

The relationship between the varicella antibody response and the risk of breakthrough disease has been previously assessed by using AFT models and the piecewise exponential model [Ivan et al, 2002]. This study aimed at applying this methodology to assess the association between the varicella antibody titer after vaccination and the rate of breakthrough disease. Using these statistical methods we established an association between the distribution of varicella antibody response after vaccination and the long term protection against varicella in all the 4 fitted

models. These findings were consistent to those reported in the paper by Ivan et al, 2002, though the two studies were quite different in terms of study design and especially the age group; While this study focused on children between the age of 12 and 22 months in Europe, the study by Ivan et al was carried out among children aged between 1 year and 12 years in the United States of America.

Based on the AIC values, the Weibull model seems to perform better than the rest of the models fitted; this contradicts the finding of the paper by Ivan et al where the exponential model was shown to perform better. This is an indication that the choice of the best model among this group of models highly depends on the data at hand. The descriptive assessment of the model based prediction comparison with the life table estimates indicated that the four models performed quite similar. That is all the four models yielded cumulative event rate estimates that were very close to those obtained from the life-table estimates. In addition, the piecewise exponential model gave year-by-year estimates that were almost identical to the life-table estimates except for the 4^{th} year. The prediction of the probability of varicella events through 3 years follow-up period are quite similar for the Weibull, log-normal and the log-logistic models and slightly different for the piecewise exponential model with pieces based on the yearly time interval.

To estimate the set of point that best reflects the change in risk of breakthrough disease, a set of models were fitted at different thresholds specifically between the limit of quantification 50 and 300 titer in an interval of 5. The aim was to obtain a set of antibody titer that yields the lowest AIC values and hence the cutoff. Based on the methodology applied and the set of AIC values obtained the four year follow-up data does not seem to support the existence of a cutoff value.

We therefore recommend that the same models be fitted once again after the second phase of follow-up to check whether the result will change.

References

Allison PD. *Survival Analysis using the SAS System - a Practical Guide*. SAS Institute: Cary, North Carolina, 1995; chapter 4, 61–110.

Chen RT, Markowitz LE, Albrecht P, Stewart JA, Mofenson LM, Preblud SR, Orenstein WA. Measles antibody: reevaluation of protective titres. *Journal of Infectious Disease* 1990; **162**:1036 –1042.

Cox DR. Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society*, Series B 1972, **34**:187–220.

David C. Modelling Survival Data in Medical Research. Chapman & Hall CRC Second Edition, 2003

Dipika S, Suman K, Binod S, Byomkesh M, Mohammad A, Allison M, Swapan K et al. Efficacy of a Low-Cost, Inactivated Whole-Cell Oral Cholera Vaccine: Results from 3 Years of Follow-Up of a Randomized, Controlled Trial. PLoS Neglected Tropical Diseases 2011.

Friedman M. Piecewise exponential models for survival data with covariates. *Annals of Statistics* 1982; **10**: 101–113.

Galil K, Lee B, Strine T, Carraher C, Baughman A, Eaton M, Montero J, Seward J. Outbreak of varicella at a day-care center despite vaccination. *New England Journal of Medicine* 2002; **347**(24):1909-15.

Ivan C, Shu L, Holly M, Christina C, Rupert V, Jerald S, J Heyse. Use of statistical models for evaluating antibody response as a correlate of protection against varicella. *Statistics in Medicine* 2002; **21**:3411–3430

Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. Wiley: New York, 1980.

Kuter BJ, Weibel RE, Guess HA, Mathews H, Morton DH, Ne_ BJ, Provost PH, Watson BA, Starr SE, Plotkin SA. Oka=Merck varicella vaccine in healthy children: final report of a 2-year efficacy study and 7-year follow-up studies. *Vaccine* 1991; **9**:643–647.

Lachenbruch PA, Horne AD, Lynch CJ, Tiwari J, Ellenberg SS. Biologics. *In Encyclopedia of Biopharmaceutical Statistics*. Marcel Dekker: New York, 2000; 47–54.

Lawless JF. *Survival Models and Methods for Lifetime Data*. Wiley: New York, 1982.

Mathieu A, Olivier L, Jammbe Z.M, Benson O, Philippe B, Niel H. Living on Three Time Scales: The Dynamics of Plasma Cell and Antibody Populations Illustrated for Hepatitis A Virus. PLoS *Computational Biology*, 2012.

Orenstein WA, Bernier RH, Dondero TJ, Hinman AR, Marks JS, Bart KJ. Field evaluation of vaccine efficacy. *Bulletin of the World Health Organization* 1985; **63**(6):1055-68

Vázquez M, LaRussa PS, Gershon AA, Steinberg SP, Freudigman K, Shapiro ED. The Effectiveness of the Varicella Vaccine in Clinical Practice. *New England Journal of Medicine* 2001; **344**(13):955-60.

Weibel RE, Neff BJ, Kuter BJ, Guess HA, Rothenberger CA, Fitzgerald AJ, Connor KA, McLean AA, Hilleman MR, Buynak EB, Scolnick EM. Live attenuated varicella virus vaccine: efficacy trial in healthy children. *New England Journal of Medicine* 1984; **310**:1409 –1415.

WHO (World Health Organization). WHO Position Paper on Measles Vaccines. *Weekly Epidemiological Record.* 2004; **79**:129-44.

WHO (World Health Organization). WHO Position Paper on Mumps Vaccines. *Weekly Epidemiological Record.* 2001; **76**:345-56.

WHO (World Health Organization). WHO Position Paper on Rubella Vaccines. *Weekly Epidemiological Record.* 2000; **75**:161-72.

WHO (World Health Organization). WHO Position Paper on Varicella Vaccines. *Weekly Epidemiological Record.* 1998; **73**:241-48.

Appendix

Table A1*: Ratio of survival times per unit increase in the log antibody titer among the vaccinated subjects*

Method	Covariate	Ratio of survival times [95% CI]
Weibull	Log antibody titer	2.24[2.01,2.48]
Lognormal	Log antibody titer	2.46 [2.20, 2.75]
Loglogistic	Log antibody titer	2.26 [2.03, 2.51]
Piecewise (piece on time)	Log antibody titer	3.81 [3.53, 4.18]
Piecewise (piece on both time and titer concentration)	Log antibody titer	3.60 [3.32, 3.94]

Table A2: *Summary Statistics of the log antibody titer per treatment group*

Vaccine group	N	Mean	95% CI	Minimum	Maximum
MMVR (Log titer)	2,216	3.26	[2.496, 4.024]	1.10	4.44
MMR+V (Log titer)	2,202	1.98	[1.157, 2.803]	1.10	3.85
Placebo (Log titer)	724	1.15	[0.562,1.738]	1.10	3.71

Table A3: *Parameter estimates from the piecewise exponential model*

Parameter	Estimate	95%CI	p-value
intercept	4.7530	[4.626, 4.880]	< .0001
Log antibody titer	1.3446	[1.257, 1.433]	< .0001
Lamba1	0.8885	[0.796, 0.981]	< .0001
Lamba2	0.9559	[0.858, 1.054]	< .0001
Lamba3	1.1912	[1.079, 1.304]	< .0001
Lamba4	0.8174	[0.642, 0.992]	< .0001

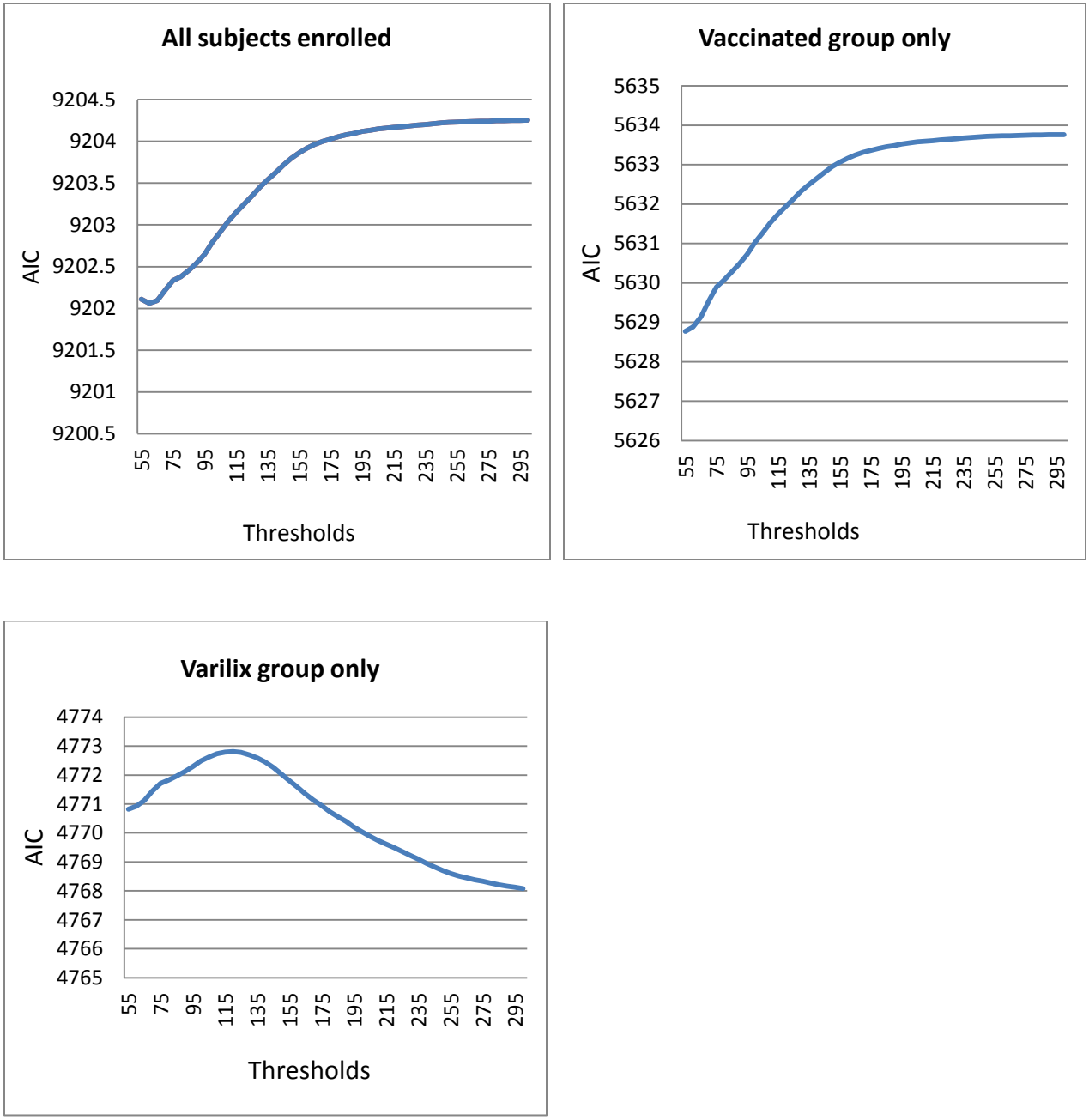


Figure A1: *Graph of AIC against thresholds for the different groups of subjects for the Weibull model*

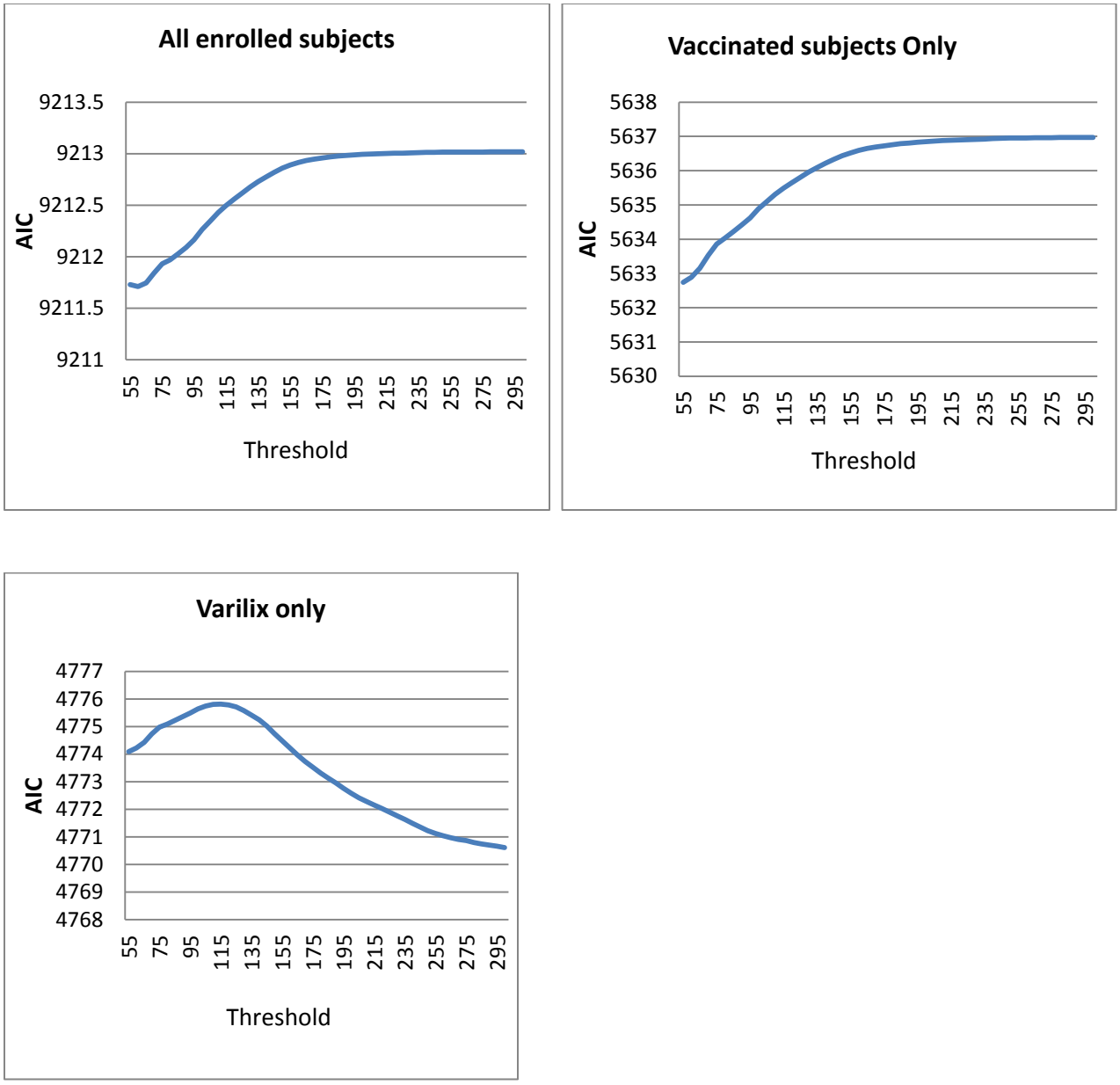
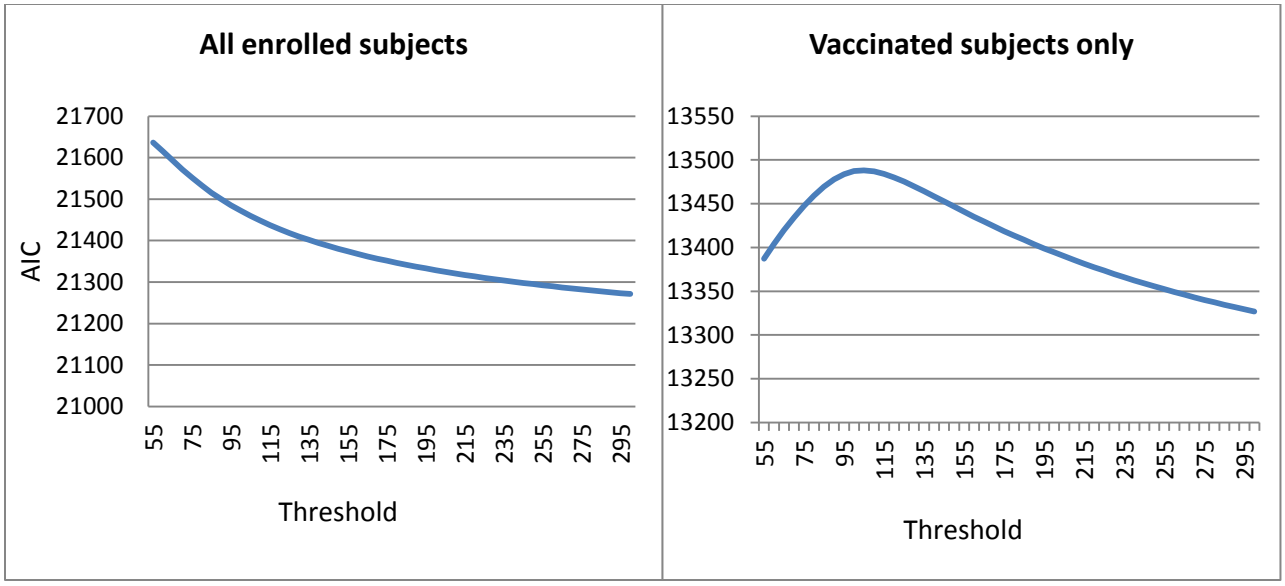


Figure A2: *Graph of AIC against thresholds for the different groups of subjects for the Log-logistic model*



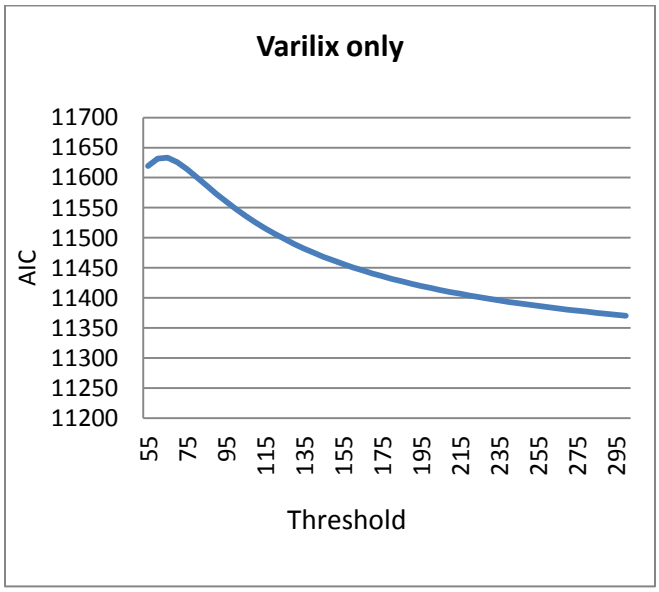


Figure A3: *Graph of AIC against thresholds for the different groups of subjects for the piecewise exponential model*

Table A4: *Model comparison at different thresholds*

Weibull model							Log-logistic Model						
	All subjects		Vaccinated subjects only		Varilix only			All subjects		Vaccinated subjects only		Varilix only	
Threshold	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**	**Threshold**	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
55	9202.10	9228.29	5628.77	5654.34	4770.82	4793.60	55	9211.73	9237.91	5632.74	5658.31	4774.09	4796.87
65	9202.09	9228.27	5629.14	5654.71	4771.12	4793.91	65	9211.74	9237.92	5633.15	5658.72	4774.42	4797.21
75	9202.33	9228.51	5629.89	5655.46	4771.70	4794.49	75	9211.93	9238.11	5633.85	5659.43	4774.97	4797.76
85	9202.45	9228.63	5630.27	5655.85	4771.96	4794.75	85	9212.02	9238.20	5634.22	5659.79	4775.21	4798.00
100	9202.79	9228.97	5631.02	5656.60	4772.48	4795.27	100	9212.26	9238.44	5634.88	5660.45	4775.64	4798.43
115	9203.19	9229.32	5631.74	5657.31	4772.79	4795.58	115	9212.49	9238.67	5635.48	5661.06	4775.81	4798.60
125	9203.34	9229.52	5632.13	5657.70	4772.78	4795.57	125	9212.61	9238.79	5635.80	5661.37	4775.70	4798.49
140	9203.62	9229.80	5632.65	5658.22	4772.45	4795.23	140	9212.77	9238.95	5636.21	5661.78	4775.24	4798.03
150	9203.79	9229.97	5632.95	5658.52	4772.03	4794.82	150	9212.86	9239.04	5636.43	5662.00	4774.74	4797.53
165	9203.96	9230.14	5633.24	5658.82	4771.34	4794.13	165	9212.93	9239.11	5636.64	5662.22	4773.96	4796.75
175	9204.02	9230.20	5633.36	5658.93	4770.93	4793.72	175	9212.95	9239.13	5636.72	5662.30	4773.53	4796.31
190	9204.09	9230.27	5633.48	5659.05	4770.39	4793.18	190	9212.98	9239.16	5636.80	5662.38	4772.96	4795.74
200	9204.13	9230.31	5633.55	5659.12	4770.03	4792.82	200	9212.99	9239.17	5636.85	5662.42	4772.58	4795.37
215	9204.16	9230.34	5633.61	5659.18	4769.61	4792.40	215	9213.00	9239.18	5636.88	5662.46	4772.15	4794.94
225	9204.18	9230.36	5633.64	5659.21	4769.35	4792.14	225	9213.00	9239.18	5636.90	5662.47	4771.89	4794.68
240	9204.21	9230.39	5633.69	5659.26	4768.95	4791.74	240	9213.01	9239.19	5636.93	5662.50	4771.48	4794.27
250	9204.22	9230.40	5633.71	5659.29	4768.70	4791.49	250	9213.01	9239.19	5636.95	5662.52	4771.22	4794.01
265	9204.23	9230.41	5633.73	5659.30	4768.44	4791.23	265	9213.01	9239.19	5636.95	5662.53	4770.96	4793.75
275	9204.242	9230.42	5633.74	5659.31	4768.33	4791.12	275	9213.01	9239.19	5636.96	5662.53	4770.86	4793.64
290	9204.25	9230.43	5633.75	5659.33	4768.16	4790.95	290	9213.01	9239.19	5636.97	5662.54	4770.69	4793.48
300	9204.252	9230.43	5633.76	5659.33	4768.07	4790.86	300	9213.01	9239.19	5636.97	5662.54	4770.61	4793.40

Table A5: *Model comparison at different thresholds*

Piecewise exponential model							Log-normal Model						
	All subjects		Vaccinated subjects only		Varilix only			All subjects		Vaccinated subjects only		Varilix only	
Threshold	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**	**Threshold**	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**
55	21629.95	21684.45	13387.27	13440.78	11619.64	11668.18	55	9285.47	9311.65	5668.07	5693.65	4797.19	4819.98
65	21588.75	21643.25	13419.07	13472.58	11633.46	11682.00	65	9285.45	9311.63	5668.47	5694.04	4797.58	4820.37
75	21547.93	21602.43	13445.78	13499.29	11616.00	11664.54	75	9285.38	9311.56	5669.00	5694.57	4798.14	4820.93
85	21512.09	21566.59	13466.75	13520.27	11576.24	11624.79	85	9285.31	9311.49	5669.30	5694.88	4798.44	4821.23
100	21467.72	21522.22	13484.00	13537.51	11551.04	11599.58	100	9285.13	9311.31	5669.73	5695.30	4798.82	4821.61
115	21432.34	21486.84	13482.37	13535.89	11518.23	11566.78	115	9284.86	9311.04	5670.02	5695.60	4798.84	4821.63
125	21412.56	21467.06	13474.63	13528.15	11499.59	11548.14	125	9284.67	9310.85	5670.126	5695.7	4798.60	4821.39
140	21387.25	21441.75	13459.13	13512.65	11475.73	11524.27	140	9284.35	9310.53	5670.181	5695.75	4797.95	4820.74
150	21372.71	21427.20	13448.00	13501.52	11462.1	11510.64	150	9284.11	9310.29	5670.164	5695.73	4797.34	4820.13
165	21353.64	21408.14	13431.48	13485.00	11444.41	11492.96	165	9283.82	9310.00	5670.1	5695.67	4796.44	4819.23
175	21342.44	21396.93	13420.93	13474.44	11434.15	11482.7	175	9283.68	9309.86	5670.05	5695.63	4795.93	4818.72
190	21327.47	21381.97	13406.01	13459.53	11420.64	11469.18	190	9283.52	9309.70	5669.99	5695.57	4795.29	4818.08
200	21318.53	21373.03	13396.71	13450.23	11412.68	11461.23	200	9283.42	9309.60	5669.96	5695.53	4794.88	4817.67
215	21306.41	21360.90	13383.71	13437.23	11402.06	11450.6	215	9283.33	9309.51	5669.92	5695.49	4794.42	4817.21
225	21292.24	21346.73	13375.64	13429.16	11395.72	11444.26	225	9283.28	9309.46	5669.90	5695.47	4794.12	4816.91
240	21289.00	21343.49	13364.36	13417.88	11387.16	11435.7	240	9283.20	9309.38	5669.86	5695.43	4793.66	4816.45
250	21282.83	21337.33	13357.35	13410.87	11382.00	11430.54	250	9283.15	9309.33	5669.84	5695.41	4793.38	4816.17
265	21274.31	21328.80	13347.53	13401.05	11374.96	11423.51	265	9283.12	9309.30	5669.82	5695.40	4793.09	4815.87
275	21269.04	21323.54	13341.41	13394.92	11370.68	11419.23	275	9283.10	9309.28	5669.82	5695.39	4792.96	4815.75
290	21261.70	21316.19	13332.79	13386.31	11364.79	11413.34	290	9283.08	9309.26	5669.81	5695.38	4792.76	4815.55
300	21257.13	21311.63	13327.40	13380.92	11361.18	11409.73	300	9283.08	9309.26	5669.80	5695.38	4792.66	4815.45

SAS Codes

```
/**Weibull Model*/
proc nlmixed data=efficacy11;
      bounds gamma > 0;
parms b0=8 b1=1 gamma=1;
      linp = b0+b1*LOGN;
      alpha = exp(-linp);
      G_t= exp(-(alpha*tte)**gamma);
      g=gamma*alpha*((alpha*tte)**(gamma-1))*G_t;
      ll=(vevent=1)*log(g) + (vevent=0)*log(G_t);
      model tte~general(ll);

/*Weibull model predictions*/
      predict 1-exp(-((exp(-(b0)))*1460)**gamma) out=cdf2;
  run;
```

```
/**Log-Normal Model*/

PROC NLMIXED data=efficacy1;
   parms  b0=-1 b1=0 sigma=1;
          bounds sigma>0;

   mu=(b0+b1*LOGN);
  *survival function;
         st=(1-cdf('LOGN',tte,mu,sqrt(sigma)));
   *Density fuction;
   g = exp(-0.5*((log(tte)-
mu)**2/sigma))/((y*(2*sigma*CONSTANT('PI'))**0.5));
   ll=(vevent=1)*log(g)+ (vevent=0)*log(st);
   model tte~general(ll);

         /*Lognormal model predictions*/
predict 1-(1-cdf('LOGN',1095,(b0+b1*logN),sqrt(sigma))) out=cdf1;
  run;
```

```
/**Log-logistic Model*/

proc nlmixed data=efficacy1;
 bounds gamma > 0;
 parms b0=8 b1=1 gamma=1;
 linp  = b0 + b1*LOGN;
 alpha = exp(-linp);
 G_t   = 1/(1+(alpha*tte)**gamma);
 g = (gamma*((alpha)**gamma)*(tte)**(gamma-1) /(1+(alpha*tte)**gamma))*G_t;
 ll = (vevent=1)*log(g) + (vevent=0)*log(G_t);
 model tte ~ general(ll);

 /*Loglogistic model predictions*/
 predict 1-1/(1+(exp(-(b0 + b1*LOGN))*1095)**gamma) out=cdf3;
  run;
```

```
/*Data preparation for the piecewise model*/
data effc1;
  set efficacy1;
    year=ceil(tte/365) ;
      do j=1 to year;
        time=365;
          event =0 ;
          if j=year and vevent=1 then do;
         event=1;
       time=tte-365*(year-1) ;
      end;
     output;
    end; run;

data effc;
set effc1;
if event=0 and tte>0 and tte<365 and j=1 then time=tte;
if event=0 and tte>365 and tte<730 and j=2 then time=tte-365;
if event=0 and tte>730 and tte<1095 and j=3 then time=tte-365*2;
if event=0 and tte>1095 and tte<1460 and j=4 then time=tte-365*3;
y1=log10(time);
run;
```

```
/***********Piecewise Exponential Model****************/
proc nlmixed data=effc;
     parms b0=8 b1=1 a1=1.1 a2=2 a3=2 a4=2;
     linp  = b0 + b1*LOGN +a1*(J=1)+a2*(J=2)+a3*(J=3)+a4*(J=4);
     alpha = exp(-linp);
     G_t   = exp(-(alpha*time));
     g     =alpha*G_t;
     ll    = (vevent=1)*log(g) + (vevent=0)*log(G_t);
     model time ~ general(ll);
     /*piecewise exponential  model predictions*/
     predict 1-exp(-(exp(-(b0 + b1*LOGN +a1*(J=1)+a2*(J=2)+ a3*(J=3)
+a4*(J=4)))*365)) out=cdf4;
  run;
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Statistical modeling for correlate of protection using accelerated failure time models and piecewise methods.

Richting: **Master of Statistics-Biostatistics**
Jaar: **2013**

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,

Mogeni, Polycarp

Datum: **11/09/2013**