2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Evaluation of the change-point piecewise exponential model

Promotor :
Prof. dr. Geert MOLENBERGHS

Promotor :
Mr. EDOUARD LEDENT

Sofia Kanavou

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

2014•2015
FACULTY OF SCIENCES
Master of Statistics

Master's thesis

Evaluation of the change-point piecewise exponential model

Promotor :
Prof. dr. Geert MOLENBERGHS

Promotor :
Mr. EDOUARD LEDENT

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

universiteit hasselt | Maastricht University

Abstract

Herpes zoster is a reactivated form of the latent varicella-zoster virus, commonly known as chickenpox, causing damage to the nervous system, followed by persistent neuropathetic pain. This characteristic pain can be divided into three phases: acute, following the exanthema phase, sub-acute and chronic or Post-Herpetic Neuralgia (PHN) when the pain persists for 4 weeks or more regardless the healing state. In change-point analysis of Desmond (2002) antiviral treatments, Acyclovir and Valacyclovir have been used and piecewise exponential model with two change-points was applied to obtain treatment effect and corresponding hazard rates, for both cases when the change-point is known or has to be estimated. Main objective of this study was to evaluate the use of a piecewise exponential model with two change-points in a paradigm data set, similar to the one used in Desmond analysis. Applying first the Cox model, Valacyclovir has been proven more effective than Acyclovir, regardless the dose group. Then, proportional hazards assumption was tested both graphically and computationally. Results led to the use of a more complex model; piecewise exponential model was applied for two change-points, in two cases. First for known change-points (30,120) as initially defined by Dworkin and Portenoy in 1994. Hazard rate estimates showed a decline as patient moves from acute to chronic phase, while there was statistically significant effect evident only during sub-acute phase. Moreover, the pair of change-points was estimated via exhaustive grid search in R and in SAS via the macro code of Mahdi Sadat-Hashemi et al. Corresponding hazard rate estimates appeared to be close in both models. Graphical representation of the cumulative hazard was used, where the first cut-off points were visible in the area around the 1^{st} trimester, yet it was not that distinct where the second one lies on. Based on graphical representation again, it was shown that above the 140^{th} day the likelihood becomes "flat" and the detection of a change-point there is rather difficult. Even though method applied in SAS was an ad-hoc procedure while in R cutoff estimates were jointly found, the results were close. Furthermore, it occurs that the length of the chosen grid matters significantly, since the same one was used in both methods. In the end, it is questioned whether a direct comparison of treatments across the different phases of pain is possible within the context of a randomized clinical trial, based on discussions of Arani [4], Goodman [5] and Kay [14].

Acknowledgments

First of all, I would like to express my gratitude and appreciation to both my supervisors Edouard Ledent and Professor dr. Geert Molenberghs for their support and valuable guidance throughout this thesis.
I would also like to thank Associate professor Dimitris Karlis, for his encouragement and inspiration led me to the field of Biostatistics.
I could not leave out my most valuable team: my parents and my brother and best friend, Fotis, for the limitless support in every path I have chosen in my life so far.
Last but not least, I want to express my sincere apologies and thank cordially all of my friends who tolerated me and continued to support me throughout this master.

Contents

Chapter 1

Introduction

1.1 Varicella - Zoster Virus

Herpes Zoster occurs when latent varicella-zoster virus (VZV) reactivates and multiplies within a sensory ganglion and travels along the sensory nerve to the skin. Varicella virus is most commonly known as *chickenpox*; *zoster* is the evolution of varicella infection. Patients above 50 years of age are usually tormented by this neurological disease, whose main symptom is pain. This neuropathetic pain is a result of the damage caused to patient's nervous system. Pain is divided into three different phases, depending on its volume and duration. **Acute** pain is a symptom appearing during the homonymous acute exanthem phase, following a dermatomal rash. Especially in the areas of the affected sensory nerves, the pain discomfort can be prolonged for days, weeks, or months and in some cases even years. A pain that persistent is termed **chronic** or Post-Herpetic Neuralgia (PHN) and usually appears after the healing of skin lesions or 4 weeks after the onset of lesions regardless the stage of healing. Although accute pain stimulates one's immune system to avoid further damage of the tissue area, PHN leads only to discomfort and distress. Main cause of PHN is the activity appearing in the spinohypothalamic and pontine - hypothalamic pathways. Next, the stress response is initiated and once this stress response is extended it becomes dysregulated and maladaptive. This leads to fatigue, impaired functioning and adaptive behavior, experienced as sickness. The further psychological effect renders one to be incapable of performing a daily routine at home, community or workplace, making PHN a burden not only to the patient but also to the healthcare system and society at large [1].

Main target of the treatment is to minimize the risk of further pain discomfort. The medication used in order to reduce complications and limit PHN are: *steroids, nerve blocks, analgesics, antidepressants* and *antivirals*. The most suitable way to minimize pain is to prevent any further nerve damage; antivirals achieve that by canceling the replication of VZV. *Acyclovir* is an orally administrated drug and the first one proved to be effective. Wood states based on clinical trials conducted in the past, that Acyclovir reduces significantly zoster-associated pain (ZAP) duration, without having the same effect on PHN though. *Valacyclovir*, a bioavailable prodrug of Acyclovir, also showed significant advantages regarding pain resolution.

Nonetheless, assessing the effect of antiviral therapies on the different phases of pain is the most challenging part. The main difficulty that needs to be overcome is to detect the time point where acute pain ends and chronic pain begins respectively. As a solution, piecewise regression modeling of the hazard function of ZAP has been proposed. Detection of the areas where curve's exponential decline changes, combined with the estimation of corresponding change - points, will explain how zoster pain evolves from acute to chronic.

Transition points. Dworkin and Portenoy [2] suggested as transition points τ_1 and τ_2, the time points of 30 and 120 days respectively. These suggestions were based on the fact that acute infection usually heals after 4 weeks. However it is mentioned that criteria used by International Association for the Study of Pain (IASP) for the definition of PHN were rather arbitrary. Three months time was considered to be the most convenient way to describe persisting pain symptoms after the end of acute period. Finally, for patients experiencing pain in between, *subacute* phase was defined.

1.2 Objectives

Main objective of this dissertation is to evaluate the behavior of a piecewise exponential model (PEM) with multiple change-points, applied in real data. The data set used was similar to the one applied in the change-point analysis made by Desmond in 2002 [3]. Though a direct comparison cannot be made, the results of the paper will be used as as a benchmark. More importantly, the ways of estimating these change-points will be explored based on the methodology described by Arani et al (2001) [4]. The analysis is done step by step, applying first Cox proportional hazards model and computing the Kaplan-Meier product limit estimate, followed by corresponding graphical representation. Then, the proportionality of the hazard function is tested, given the nature of PEM (constant within defined time intervals). Finally, PEM is applied in both cases: (i) given the pair of change - points is known (30 and 120 days) and (ii) in the case of the change points being unknown. In the latter case, an exhaustive grid search is applied in order to detect and estimate the suitable pair of transition points, using a maximum likelihood optimization technique. For these purposes, two different types of statistical software and techniques were used and compared (SAS 9.4 and R 3.1.2).

1.3 Data Description

The analyses done by Desmond [3] was based upon two studies, thus two different data sets. Both studies focus on the computation of hazard rates for treatment effect and baseline pain resolution across each of the phases of pain and the comparison of parameters related to herpes zoster. It must be underlined that only the first study will be used as a benchmark, while only treatment effect was considered for the model.
On the data set used as paradigm, 1141 patients are enrolled in a controlled clinical trial of *Acyclovir (ACV) versus Valacyclovir (VACV)*. Treatment was administrated orally and doses among treatment

groups differ as follows: VACV at 1000 mg for 7 days, VACV at 1000 mg for 14 days and ACV at 800 mg for 7 days. *Time to complete cessation of ZAP* was used as survival time, were patients declared the number of days completed feeling pain, starting from 0 days and reaching maximum of almost 6 months (174 days). Moreover, average age of patients is approximately 69 years, the majority of which consists of caucasian (94.65%), confirming baseline pain (90.23%), while the percentages of male and female population are kept almost in balance.

Chapter 2

Methodology

2.1 Survival Analysis

Cox semi-parametric proportional hazards model is applied first, introducing only treatment effect to the model. Plot of Kaplan - Meier (KM) estimator will help to compare pain duration among different treatment groups, examine the scenario of switching to a binary treatment variable and search for possible hints of change-point existence. Then, proportionality hypothesis is tested by introducing time-varying covariates to the model and plotting the Schoenfeld residuals. That will eventually show whether the proportional hazards assumption holds or the use of a more complex model is really necessary.

Graphical representation holds a main part in the analysis; it cannot replace hypothesis testing though. Its role is to reveal the areas in which transition point lies upon, if there is one. Furthermore, what is expected from one to see is a ”jump” of the hazard function at the transition point of one (time) interval to another, while the hazard function must look (approximately) constant within those intervals defined. It must be mentioned that no hypothesis testing was done when proceeding from Cox model to PEM. Normally a Wald test is applied first as to test the need of a more simple model (1 change-point) [5]. Primary goal is to duplicate the analysis procedure carried out by Desmond and explore the complexities of this model. Thus, the need of a two-change point model was considered known from the start.

In paradigm data set, PEM was applied for both known transition times, where (τ_1, τ_2) are set to 30 and 120 days respectively, and unknown. However, since the estimation method applied by the author is different, in the second case the scenario of having slightly different estimates should be mentioned. Parameter estimates obtained from single change - point piecewise model were used as initial values for the final one. Before exploring the use of PEM in real data though, the model was applied on simulated data and the sample size produced was close to the one of herpes zoster study. Several cases were examined (e.g. data with and without censoring) and various number of replications were tried.

2.2 Exhaustive Grid Search

Grid search was used in order to estimate the pair of transition times. In simulated data, known transition times of 50 and 120 units were used, whereas the *grid* expected to detect the transition times defined by intervals (21,70) and (91,140). After taking all possible pairs (2500 in total), the program targets the pair that achieves the global maximum likelihood estimate (MLE) from a vector of estimates for all possible pairs. Consequently, since it is both a computationally intensive and time consuming method:

(i) the smaller the grid is the closer the estimates will be to its boundaries.

(ii) the less the repetitions are, the faster the estimation will complete.

Grid search was applied in both statistical packages. In R programming, cases of both one and two change-point models were considered on simulated data first while 2 change-point model was also used on real data. The choice of the intervals, (21,70) and (91,140) respectively, was based on graphical representation (e.g. Kaplan - Meier plot). An effort was made to obtain similar results in SAS. For that matter, macro code created by researchers Mahdi Sadat-Hashemi, Emmanouil Rampakakis, John S. Sampalis and Behrooz Kavehie was used `http://www.runmycode.org/companion/view/675`. This macro aims to find a transition point in the hazard curve by optimizing the likelihood function over an interval (l,u) set by the programmer, where $l = p_l$ and $u = p_u$ for $p_l < p_u$, are respectively the lower and upper bound of the grid. In the end, a data set is produced containing all candidate break points in descending order MLE outcome, along with their corresponding hazard rates estimated before and after cut-off point, hazard ratios and a KM plot automatically produced with the cut-off point of the global MLE. However, this code replies to a single change - point estimation; thus, for a two-point estimation in this case, the program was applied twice each time to each of the chosen grids.

2.3 Theoretical Background

2.3.1 Cox Semi-Parametric Model

Cox Semi-parametric proportional hazards modeling lies upon two basic assumptions: (i) the hazard ratio to be constant over time across groups and (ii) the hazard functions of two compared groups (e.g. treatment groups) must run in parallel over time. If the latter condition holds, then covariate - adjusted hazard ratios can be produced. Its semi - parametric nature is due to the constant baseline hazard which remains unspecified. Therefore, a partial likelihood maximization technique is used for estimation of model parameters. The model is defined as follows:

$$h_i(t) = \lambda_0(t) \exp(\beta_1 x_{i1}, \ldots, \beta_k x_{ik}) \tag{2.1}$$

where $\lambda_0(t)$ indicates the baseline hazard, constant for all i units, $(x_1, \ldots, x_k)$ are the k covariates included to the model and $\beta = (\beta_1, \ldots, \beta_k)'$ is the vector of coefficients. In case of ties, there is a

variety of methods that can be applied (depend on the statistical program). Efron's approximation is preferred, a method similar to the exact conditional probability based-method [7]. Given the fact that proportionality assumption holds, the expression for the Hazard ratio of two categories i and j respectively is:

$$HR_{i,j} = \frac{h_i(t)}{h_j(t)} = \frac{\lambda_0(t)\exp(\beta_1 x_{i1},\ldots,\beta_k x_{ik})}{\lambda_0(t)\exp(\beta_1 x_{j1},\ldots,\beta_k x_{jk})} = \exp(\beta_1(x_{i1}-x_{j1}),\ldots,\beta_k(x_{ik}-x_{jk})) \tag{2.2}$$

The proportional hazards assumption can be however very restricting. It can be examined by applying a formal test or graphically. Moreover, in the first case, an interaction term of the covariate of interest with the time variable (or its logarithmic form) is added in the model and a Wald test is performed. If the interaction is not statistically significant, the assumption holds. In the latter, (weighted) Schoenfeld residuals are plotted against time; if the hypothesis holds then no time-dependent pattern is expected to be seen.

2.3.2 Piecewise Exponential Model

Piecewise exponential model appears to be more flexible than Cox regression in matters of hypothesis testing. Given j intervals, baseline hazard λ_j will be constant along the interval, but not necessarily across the different intervals defined by the change-points [8]. A discussion on how likelihood function is defined follows above [9].

General Case. Given a sample of n units in total, where unit i is observed at time point t_i. For a given unit that fails at time t_i its contribution to the likelihood function will be as follows:

$$L_i = f(t_i) = \lambda(t_i)S(t_i) \tag{2.3}$$

In addition, for a given unit i that is still alive at time t_i (censored observation), its contribution to the likelihood will be S(t_i). Hence, likelihood function is defined as follows:

$$L = \prod_{i=1}^{n} f(t_i) = \prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S(t_i)^{1-\delta_i} \tag{2.4}$$

Thus, the logarithm of the likelihood function will be:

$$log(\prod_{i=1}^{n} \lambda(t_i)^{\delta_i} S(t_i)) = \sum_{i=1}^{n} log(\lambda(t_i)^{\delta_i} S(t_i)) = \delta_i log(\lambda(t_i)) + log(S(t_i)) \tag{2.5}$$

where δ_i is the binary event indicator (e.g. 1: failure), while S(t_i) and $\lambda(t_i)$ indicate the survival and the hazard function respectively.

PEM case. Let $\tau_0, \tau_1, \tau_2, \ldots, \tau_\kappa$,where $\kappa = 1, \ldots, j-1$ be specified values that define j time intervals. For these intervals it is known that $\tau_0 = 0$ and $\tau_\kappa = \infty$. Consequently, hazard rates within each interval will be written as follows:

$$h_0(t; \lambda, \tau_1, \tau_2) = \begin{cases} \lambda_1, & 0 < t \leq \tau_1 \\ \lambda_2, & \tau_1 < t \leq \tau_2 \\ \ldots & \\ \lambda_\kappa, & \tau_{j-1} < t \leq \tau_{j=\infty} \end{cases} \tag{2.6}$$

In Herpes Zoster virus study the phases of pain studied are 3 thus, the related hazard rates will be formed as follows:

$$h_0(t; \lambda, \tau_\kappa) = \begin{cases} \lambda_1, & 0 < t \leq \tau_1 \\ \lambda_2, & \tau_1 < t \leq \tau_2 \\ \lambda_3, & t > \tau_2 \end{cases} \tag{2.7}$$

where the general expression λ_{ij} indicates the hazard rate of the jth patient observed in the ith interval, here $j = 1, 2, 3$; or else in each phase of ZAP. The hazard function can also be rewritten as follows: $h(t|x; \theta) = h_0(t; \lambda) \exp(\beta' x)$, where $\theta = (\lambda, \beta)$ are the unknown parameters to be included into the model, representing baseline hazard and vector of coefficients respectively.

Thus, Survival function is formed accordingly:

$$S(t, \lambda, \tau_1, \tau_2) = \begin{cases} \exp[-\lambda_1 t], & 0 < t \leq \tau_1 \\ \exp[-\lambda_1 \tau_1 - \lambda_2 (t - \tau_1)], & \tau_1 < t \leq \tau_2 \\ \exp[-\lambda_1 \tau_1 - \lambda_2 (\tau_2 - \tau_1) - \lambda_3 (t - \tau_2)], & t > \tau_2 \end{cases} \tag{2.8}$$

Finally, combined with equation (3), the log-likelihood will be written as follows:

$$logL(\lambda, \tau_1, \tau_2) = \sum_{i=1}^{n} \delta_i \log h(y_i, \lambda, \tau_1, \tau_2) + \log S(y_i, \lambda, \tau_1, \tau_2) \tag{2.9}$$

where $(y_1, \delta_1), (y_2, \delta_2), \ldots, (y_n, \delta_n)$ are values of the vector (Y, Δ) with δ_i being the event indicator and y_i being the observed survival times though *rescaled*, in a way that indicates in which interval they belong to.
Moreover, there is an equivalent expression of likelihood by Friedman [9] that can be applied. Instead of rescaling data points, an indicator I_j for j intervals can be used:

$$I_{ij} = \begin{cases} 1, & \text{if the } j \text{ th individual fails during } i\text{th interval} \\ 0, & otherwise \end{cases} \tag{2.10}$$

with t_{ij} = max[0, min(T_i - T_{i-1}, t_j - T_{i-1})] and l being the vector with components $l_{ij} = log(\lambda_{ij})$, $l \in \Omega(n)$ for a given linear subspace $\Omega(n)$ of $R^{nI(n)}$. Whereas where λ_{ij} indicates the hazard rate of the jth patients observed in the ith interval. Finally the expression of log-Likelihood function will be:

$$L(l) = \sum_{i,j} l_{ij} - \sum_{i,j} t_{ij} \exp l_{ij} \tag{2.11}$$

Let (i,j) be partitioned in $A = (i,j) : t_{i,j} > 0$ and $B = (i,j) : t_{i,j} = 0$; the value of L(l) does not change when the summation is only over A. If the likelihood function has a maximum, the following condition reassures that MLE exists.

Condition 1: For any pair of vectors x,y $\in \Omega(n)$, if $x_{ij} = y_{ij}$ for every (i,j) $\in A$, then $x = y$. Let L' and L" have the first and the second differentials respectively of the likelihood at l (likelihood estimate). Then for any x,y $\in \Omega(n)$, it is known:

$$\begin{aligned} L_l'(x) &= \sum x_{ij} I_{ij} - \sum x_{ij} t_{ij} \exp(l_{ij}) \\ L_l''(x) &= \sum x_{ij} y_{ij} t_{ij} \exp(l_{ij}) \end{aligned} \tag{2.12}$$

From previous Condition it is evident that L(l) is a concave function.

Note. As far as change-point estimation is concerned, a problem occurs regarding the second derivative, when the change - point is introduced as a parameter to the model (τ_j). Consequently, the calculated Hessian matrix (square matrix of second order partial derivatives) appears not to be positive definite. This particular type of modeling can be rather complex and sometimes implausible even, as the hazard function will be discontinuous at the time point of transition [10]. However, if boundaries are properly defined and the correct number of cut-off points is used (avoid overfitting), this model can approximate arbitrary shapes of the hazard and the survival function and eventually, provide a better insight to the data [4].

2.4 Software

The analysis was carried out partially with SAS 9.4 and R 3.1.2. Flexibility and potential of NLMIXED procedure makes it a suitable for many non-standard applications (e.g. non-linear models), even when random effects are not applied [8]. Piecewise exponential modeling was applied in SAS via NLMIXED procedure. Other ways suggested are via PHREG procedure with *bayes* statement or MCMC procedure for bayesian analysis. For a classical frequentist approach, NLMIXED procedure in SAS allows the user to set the Likelihood via *general* statement. PHREG on the other hand, uses a different calculation method based on the partial likelihood [11]. A partial Likelihood estimation approach in PEM -for one change-point- has been described by Liang et al. (1990) [12]. Nevertheless, the sensitivity of this procedure to initial values must be underlined. Moreover, when change-points were added as individual

parameters to the model lead to problems regarding Hessian matrix calculation and optimization procedure. Normally problems as such can be overheld by rescaling or choosing more appropriate initial values [13]. Here, however, it is a matter of how likelihood occurs to be ”flat” at some time points.

Finally, for time efficiency, R programming was used in the part of simulation and grid search. Library *msm* provided the environment to generate data points directly from a piecewise exponential model. Grid search was also applied ad-hoc in SAS, based on a macro code written by Mahdi Sadat-Hashemi, Emmanouil Rampakakis, John S. Sampalis and Behrooz Kavehiel.

Note : Library *msm* was created under version 3.1.2 of R. In case of an earlier version, one can install library(*installr*) as to apply the latest version and updates of R program.

Chapter 3

Results

3.1 Exploratory Data Analysis

The mean duration of time until complete pain cessation is approximately 2 months (60.67 days) with median time of 36 days. The minimum amount of observed time till complete pain cessation is 0 days. Thus, patient stopped experiencing any pain symptoms right after taking the assigned treatment. Also, the maximum number of days with pain symptoms was almost 6 months (174 days). According to histogram of *time till pain cessation* variable in figure B.1 it is evident that most patients have declared pain continuation before the 90^{th} day of observation. In addition, almost 32% of observations are censored. Finally, there seems to be patients declaring 0 days of pain who are censored at the same time. Apart from that, there were only 3 (censored) observations missing and therefore have been omitted. No other intervention has been made to the data.

Note: Hypothesis testing was carried out in a 5% level of significance.

3.2 Cox Semi-Parametric Proportional Hazards Model

Kaplan - Meier plots were considered for both cases: including both distributions of Valacyclovir and the case were the dose categories are merged. According to figure B.2 survival curves for different distribution of Valacyclovir are really close, following a common trend. In the case of ACV versus VACV in figure 3.1, there is an instant drop close to the 30^{th} day of observation, while both curves seem to develop in time.

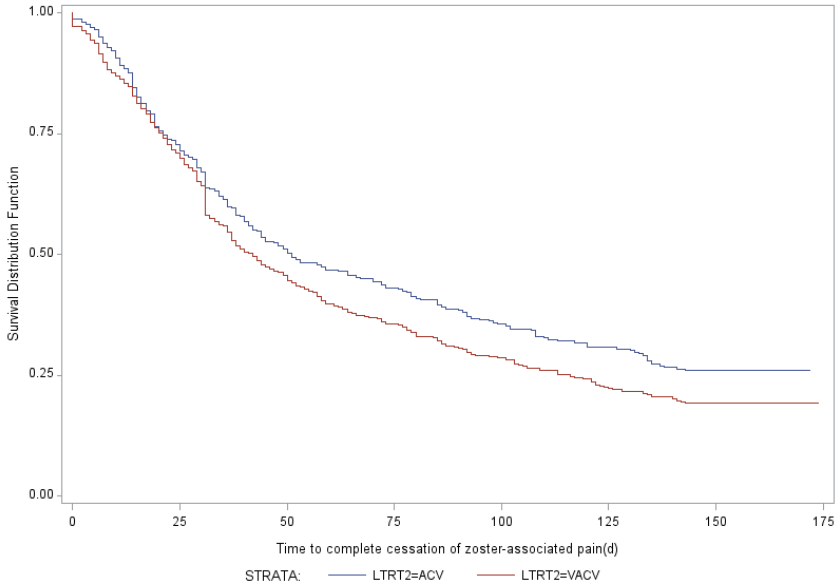


Figure 3.1: Plot of Kaplan - Meier product limit estimator for Acyclovir (ACV) versus Valacyclovir (VACV) with merged dose categories in one.

Moreover, according to results of Cox regression in table A.6, both doses of VACV seem to have statistically significant effect on zoster-associated pain resolution. Using ACV level of treatment as reference category, it is evident that patients treated with VACV have greater hazard of a ZAP - cessation event. Consequently, patients in either VACV group are expected to experience pain for a smaller amount of time compared to those treated with ACV. Based on the same table, there seems to be no difference in survival for the two dose groups of Valacyclovir (HR approximately equal to 1). Thus, from this moment on, the two categories of VACV distribution to patients, will be merged in one.
In addition, in order to test proportional hazards assumption a time-varying covariate of treatment effect is added to the model. Based on Wald test results, it is found to be statistically not significant (p-value: 0.8410) and therefore, the assumption holds. Nevertheless, the conclusion made when plotting Schoenfeld residuals against time B.3 is contradictory. There seems to be a time pattern regarding treatment effect, thus, the use of a more complex model (i.e. PEM) is explored below.

3.3 Piecewise Exponential Proportional Hazards Model

For the exploration of the piecewise exponential proportional hazards model, two approaches were made: one using known transition times of τ_1=30 and τ_2=120 days and one for unknown transition times. In both cases, phase - specific hazard rates were obtained, defining the level of hazard for all units based on the state of pain. On the latter case, exhaustive grid search was applied as to obtain jointly the transition times based on their maximum likelihood estimate (MLE). Finally, standard errors and 95% confidence intervals of the estimates were obtained with bootstrap re-sampling method.

3.3.1 Known Change-Points

According to table 3.1, there both treatments seem to affect significantly pain relief in patients experiencing sub-acute pain ($p-value_{VACV}$=0.0082 and $p-value_{ACV}$ <0.001 respectively), while the hazard rate appears to be slightly smaller for patients treated with Acyclovir. Hazard rate for patients treated with Valacyclovir equals: $\exp(-(\beta_0+\beta_1)) = 0.0124$.

Moreover, hazard rates for both treatments in acute and chronic phase are relatively close. There is an obvious decrease in baseline hazard as moving to PHN phase of pain, with the larger drop appearing as we move from the sub-acute to patients feeling chronic pain. Consequently, it will take longer for patients in PHN state to be in a painless state.

Table 3.1: Parameter estimates of the piecewise exponential model, using change-points of 30 and 120 days. Last column contains baseline hazards (λ_i) as caluclated for Acyclovir (ACV) and Valacyclovir (VACV) respectively.

Phases	**Parameters**	**Estimate (s.e.)**	**p-value**	λ_{ACV}	λ_{VACV}
I	α_0	4.3675 (0.0945)	<0.001		
	α_1	-0.1161 (0.1144)	0.3103	0.0127	0.0142
II	β_0	4.6849 (0.0921)	<0.001		
	β_1	-0.2941 (0.1110)	0.0082	0.0092	0.0124
III	γ_0	5.3916 (0.2357)	<0.001		
	γ_1	-0.1655 (0.2981)	0.579	0.0046	0.0054
	-2Loglik: 8482		*AIC: 8494*		

3.3.2 Unknown Change-Points

First, 2000 data points were simulated from a piecewise exponential model. The purpose of this was to apply grid search method and evaluate its findings. Moreover, half of the observations were generated from a PEM with known change - points set at 50 and 120 respectively and vector of baseline hazards λ_j for j=1,2,3 intervals set at λ_j =(0.02,0.01,0.005), while censored observations were generated from a simple exponential model with rate equal to 0.002. The grid was initially set for a small number of pairs (e.g. 100) close to the region of change-points (τ_1, τ_2). The program executed in R 3.1.2 was set for 2500 candidate pairs of points, taken from intervals (21,70) for τ_1 and (91,140) for τ_2 respectively. However, this approach is considered computationally intensive and involves quite an amount of time to complete its replications. Therefore, the procedure was only repeated up to maximum 50 times. Estimates of change - points and baseline hazards were really close to their initial values as seen in table A.7.
Grid - search on zoster paradigm data set was applied next. The first and second cut-off point are estimated τ_1=52 and τ_2=136 respectively. Standard errors of the estimates appear to be really small, based on results of table 3.2 after reproducing 100 bootstrap samples. According to the results, there is a decreasing trend in hazard rate with respect to different phases of pain and different treatment groups.

According to the table of change-point estimates, a bigger decline is visible as one goes from sub-acute to chronic phase of pain.

Table 3.2: For i=1,0 treatments Acyclovir (ACV) and Valacyclovir (VACV) respectively and j=1,2,3 intervals, standard errors and 95% CI for baseline hazards (λ_{ij}) are produced from Bootstrap for *B= 100* replications.

	Parameters	**Estimate (s.e.)**	**95% CI**
change-point	τ_1	52 (7.104)	(39, 60)
	τ_2	136 (3.7346)	(131, 139)
Hazard	λ_{01}	0.0134 (0.0011)	(0.0117, 0.0155)
rate	λ_{02}	0.0069 (0.0008)	(0.0057, 0.0087)
for ACV	λ_{03}	0.0016 (0.0011)	(0.0, 0.0040)
Hazard	λ_{11}	0.0156 (0.0009)	(0.0146, 0.0177)
rate	λ_{12}	0.0092 (0.0009)	(0.0076, 0.0109)
for VACV	λ_{13}	0.0023 (0.0010)	(0.0010, 0.0040)
	MLE: 4204.902		

Note. It would be legitimate to test the significance of hazard rate differences among different treatment groups, across different phases of pain. Desmond (2002) [3] applied this hypothesis testing which appears to be the same used in Arani (2001) [4]. However, there is an interesting review by Kay [14], according to which, even though it is clinically meaningful to consider different phases of pain, the different treatment effects cannot be evaluated across those phases within the context of randomized clinical trial.

Furthermore, it is interesting to see the findings of the graphical representation -likelihood estimates for each candidate pair of change-points- in figure 3.2. A black region appears around all pairs for which, the 1^{st} "jump" is from the 40^{th} day and above and it takes up to the 145^{th} until the second one appears. Around the 145^{th} day of observation, there is almost a black line "drawn", proving that likelihood function is "flat" around that area. As a result, detecting an instant change, either increase or drop, around that day is rather difficult. In addition, "flatness" is related to the calculation of the 2^{nd} derivative and thus, the Hessian matrix. Hence, that leads back to the main drawback of this complex model: MLE cannot be calculated at all time points observed.

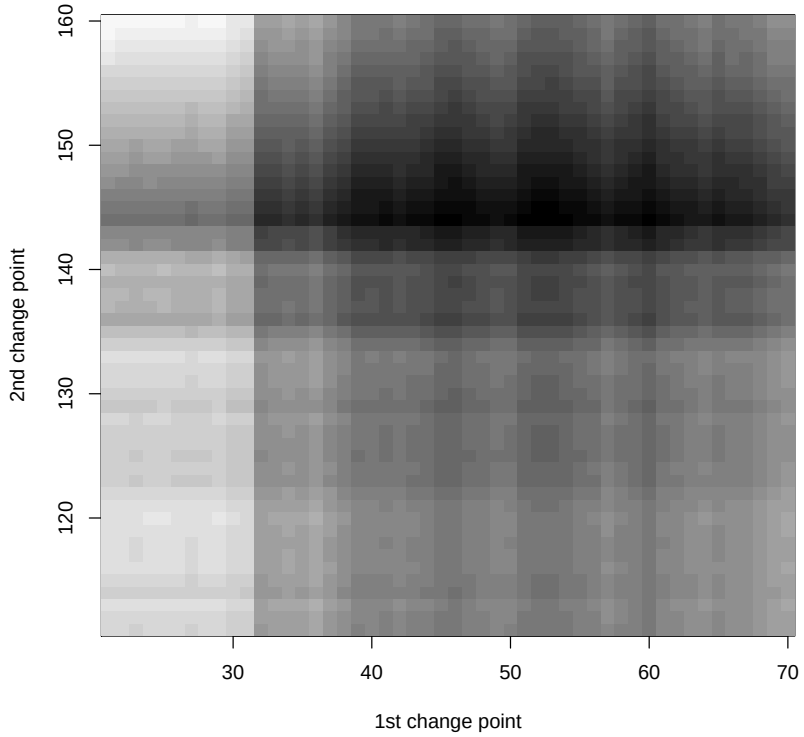


Figure 3.2: Graphical representation of the log likelihood estimates regarding the pair of candidate change-point

Even though estimates differ yet they appear to be in the ”neighborhood” of the initial values. In addition, the small standard errors produced vanish any doubts regarding the validity of the estimation method. It should be underlined though, that the choice of initial values applied as well as the length of the chosen grid play both a crucial role in this procedure. If the grid is forced to be minimized around 30 and 120 minimizing the amount of candidate pairs to a 100, the estimates will occur closer to boundaries and therefore, closer to the initial change-point values. In addition, according to graph B.1 it is evident that there is more information gathered until the 3^{rd} trimester (90 days) of the observational period. Combining those two facts, a strong argument can be made about the variation of the current results compared to the estimations made by Desmond [3]. In addition, according to figure B.4 estimated curve of cumulative hazard rates for each treatment (i.e. negative logarithm of survival plotted against time till pain cessation), there seems to be a ”break” close to the 30^{th} and 130^{th} day. Nonetheless, in figure B.5 describing the fitted cumulative hazards curves for each pair of change-points, seems that the ”break” is more visible in the right plot, where estimated points of 52 and 136 days were applied.

Grid search in SAS. Moreover, a macro code programmed by Mahdi Sadat-Hashemi and his fellow researchers `http://www.runmycode.org/companion/view/675` that applies grid search for 1 change-point in SAS was used. This program not only detects the change-point but also produces a

Kaplan-Meier plot indicating the estimated "jump". This time, the pair of change-points was detected not jointly yet separately. Thus, the code is applied twice, using each time each one of the intervals previously defined ((21,70) and (91,140) respectively). The procedure is based on likelihood estimation according to theory discussed by Friedman [9]. Results are quite close to the ones produced in R, with the two change-points estimates being 59 and 135 respectively. However this procedure can be considered ad-hoc; maximum likelihood function is better to take both cut-off points into consideration when calculated. Finally, regardless the differences of these two methods, it is evident from table A.9 that the fit of these models does not differ much. That is something to be expected from the models applying similar change-points, with reference to the 2^{nd} and 3^{rd} model of the table, still it is not necessarily expected compared to the 1^{st} model.

Chapter 4

Discussion

Herpes Zoster occurs when latent varicella-zoster virus (or zoster) is reactivated, damaging patient's nervous system. Most common symptom of zoster is neuropathetic pain. This characteristic pain can be divided into three phases: acute, which follows the appearance of exanthemas, chronic or Post-Herpetic Neuralgia (PHN) which usually appears 4 weeks after the onset of lesions regardless the healing process and the sub-acute phase which lies in between. Antiviral treatment, like Acyclovir and Valacyclovir, aims at cancelling the replication of zoster [1].

Main objective of this study was to evaluate the use of a piecewise exponential model with multiple change-points. For that purpose, the model was applied within data similar to the one analyzed by Desmond in 2002. Hence, this analysis [3] was used as a benchmark, while it was considered known from the beginning that a two change-point model is needed. Moreover, treatment groups are defined as follows: Acyclovir, 800mg 5 times per day for 7 days, Valacyclovir 1000mg, 3 times per day for 14 days and last, Valacyclovir, 1000mg for 7 days.
Data analysis begins with the Cox semi-parametric proportional hazards model, followed by a proportionality assumption test, done both graphically and computationally. Then, the piecewise exponential model is applied for 2 change-points in both cases: (i) known change-points (30,120) as defined by Dworkin and Portenoy [2] and (ii) unknown, where they have to be estimated. In the latter case, grid search was applied both in R and SAS, first by taking all possible pairs among two intervals subjectively chosen (exhaustive grid search) in R. Next, by applying one change-point macro in SAS using the same intervals as before but with an ad-hoc way and comparing the two methods.
From the start, the Cox model has proven that Valacyclovir is more effective than Acyclovir, regardless the way it was administrated. Hence, the two doses of Valacyclovir were merged in one category and treatment was used in binary form. Passing forward to piecewise regression, hazard rates show a decline as patient moves from acute to chronic phase. However, both treatments had a statistically significant effect regarding sub-acute phase only . Moreover, patients in Valacyclovir group appear to have a higher hazard rate, leading as a consequence to a lower amount of time experiencing pain.

Then, change-points were estimated and corresponding hazard rates for each treatment were obtained regarding each phase of zoster associated pain. Estimates appear to be close, regardless the fact that both

change - points are estimated to be approximately 20 days later in time (τ_1 =52 and τ_2= 136 respectively). Nonetheless, in the graphical representation of the fitted cumulative hazard of the two models, when cut-off points are known and estimated respectively, it seems that the expected ”breaks” are more distinct the second model rather than the first one. Another matter that puzzles, is the fact that results from SAS macro and exhaustive grid search applied in R seem to be close. In a way, it can be justified from the fact that above the 140^{th} day the likelihood gets ”flatter” and the observations left to estimate are much less after the 90^{th} day of observation, where only PHN patients are more likely to be found. In the first method an ad-hoc procedure is followed, while in the last one the estimation is done jointly and thus is preferred. It occurs that the length of the chosen grid matters significantly, as well as the initial values used.

In the end, it is practically proven that a shorter observational period can be defined, for change-points being 30 and 120 days, and have the same results as if the change-point were set approximately 15 - 20 days later. Thus, a series of questions arise; first it is questionable whether a direct comparison of treatments across different states of pain can be made. Although Goodman [5] has thoroughly discussed the use of Wald test statistic for that matter and Arani [4] applied a test as such, Kay [14] finds it statistically implausible to do so, under the context of randomized clinical trials. Second, it would be interesting to test the effect of other zoster-associated covariates and evaluate the risk factors. From that moment on, it is a matter of sample size requirements and appropriate level of significance in order for a researcher to detect a treatment effect, in suitable level of power.

Bibliography

[1] Wood M.J. (2002). *Understanding Pain in Herpes Zoster: An Essential for Optimizing Treatment*, The Journal of Infectious Diseases, 186(Suppl 1):S78-82.

[2] Dworkin R.H., Portenoy R.K. (1994). Proposed Classification of herpes zoster pain, Lancet, 343:1648.

[3] Desmond R.A., Weiss H.L., Arani R.B., Soong S.J., Wood M.J., Fiddian P.A., Gnann J.W., Whitley R.J. (2002). *Clinical Applications for Change-Point Analysis of Herpes Zoster Pain*, Journal of Pain and Symptom Management, Vol.23;No.6, Elsevier, New York, USA.

[4] Arani R.B., Soong S.J., Weiss H.L., Wood M.J., Fiddian P.A., Gnann J.W., Whitley R. (2001). *Phase specific analysis of herpes zoster associated in pain data*, Statistics in Medicine; 20:2429-2439, John Wiley & Sons,Ltd.

[5] Goodman M.S., Li Y., Tiwari R.C. (2011). *Detecting multiple change-points in piecewise constant hazard functions*, Vol. 38, No. 11, 2523-2532, Journal of Applied Statistics.

[6] Lahiri S., Schick A., SenGupta A., Sriram T.N. (2014). *Contemporary Developments in Statistical Theory: A Festschrift for Hira Lal Koul*, Springer, Cham,Heidelberg,New York,Dordrecht,London.

[7] Lu J., Shen D. (2014). *Survival Analysis Approaches and New Developments using SAS*, PharmaSUG conference 2014, Paper PO02.

[8] Allison P.D. (2010). *Survival Analysis Using SAS: A Practical Guide, Second Edition*, SAS Institution Inc., Cary, NC, USA.

[9] Friedman M. (1982). *Piecewise Exponential Models for Survival Data with Covariates*, The Annals of Statistics, Vol. 10, No. 1, pp.101-113, Institute of Mathematical Statistics.

[10] Lawless J.F. (2002). *Statistical Models and Methods in Lifetime Data, Second Edition*, John Wiley & Sons,Ltd.

[11] Liu D. (2010). *PROC LIFEREG or PROC PHREG*, Northwestern University, Chicago, IL.

[12] Liang K.Y., Self S.G., Liu X. (1990). *The Cox Proportional Hazards Model with Change Point: An Epidemiologic Application*, Biometrics 46, 783-793.

[13] Kiernan K., Tao J., Gibbs P. (2012). *Tips and Strategies for Mixed Modeling with SAS/STAT procedures*, SAS Institute Inc., Cary, NC, USA.

[14] Kay R. (2006). *LETTER TO THE EDITOR: Phase specific analysis of herpes zoster associated pain data: a new statistical approach*, 25:359-360, Statistics in Medicine.

Appendix A

Tables

Table A.1: Quantitave data description.

Variable	Description	mean (median)	std dev	(min , max)	missing
ZTIMETO	Time till pain cessation	60.67 (36.0)	58.13	(0, 174)	3
Age		68.48 (69.0)	9.96	(49, 99)	
Total: 1141 observations					

Table A.2: Qualitative data description.

Variable	N (%)
Sex	
Male	493 (43.21)
Female	648 (56.79)
Race	
White	1080 (94.65)
Black	35 (3.07)
Other	26 (2.28)
Baseline Pain	*5 obs. missing*
Yes	1025 (90.23)
No	111 (9.77)
Total	1141 observations

Table A.3: Treatment and corresponding dose in herpes zoster study.

Treatment	Dose	N(%)
Acyclovir	800mg,5/day for 7 days	376 (32.95)
Valacyclovir	1000mg,3/day for 14 days	381 (33.39)
Valacyclovir	1000mg for 7 days	384 (33.65)

Table A.4: Censoring in herpes zoster study.

Censored	**N (%)**
Yes (0)	359 (31.55)
No (1)	779 (68.45)
Missing	3 (0.26)

Table A.5: Summary of events for Acyclovir at 800 mg 5 times daily for 7 days (ACV-7d), Valacyclovir at 1000 mg for 7 days (VACV-7d) and Valacyclovir at 1000 mg 3 times daily for 14 days.

Treatment	**Events**	**Censored (%)**
ACV-7d	248	127 (33.87)
VACV-14d	268	112 (29.47)
VACV-7d	263	120 (31.33)

Table A.6: Hazard ratios between treatment groups: Acyclovir at 800 mg 5 times daily for 7 days (ACV-7d), Valacyclovir at 1000 mg for 7 days (VACV-7d) and Valacyclovir at 1000 mg 3 times daily for 14 days with corresponding p-values and 95% Wald Confidence Limits.

Parameters	**HR**	**p-value**	**95 % Wald CL**
ACV vs VACV - 7d	0.824	0.0290	(0.693, 0.980)
ACV vs VACV - 14d	0.833	0.0387	(0.701, 0.991)
VACV - 7d vs VACV - 14d	0.989	0.8993	(0.834, 1.173)

Table A.7: Standard errors and 95% CI produced after *B= 50* replications, on censored data simulated from piecewise exponential model with τ_1=50 τ_1=120.

	Parameters	**Estimates (s.e.)**	**95% CI**
change-point	τ_1	50 (2.9829)	(46,52)
	τ_2	117 (7.1407)	(100.675 126.0)
baseline hazard	λ_1	0.02 (0.0009)	(0.0188,0.0214)
	λ_2	0.0102 (0.0009)	(0.0089,0.0214)
	λ_3	0.005 (0.0004)	(0.0039,0.0061)

Table A.8: Estimates of change-points (τ_{est}), corresponding hazard rates (h_1, h_2) and (absolute) value of maximum likelihood, after applying separately SAS Macro.

τ_{est}	**MLE**	h_1	h_2
59	4221.65	0.0146	0.0066
135	4234.31	0.0122	0.0020

Table A.9: Fit Statistics of 3 different piecewise exponential model: (i) defined by Desmond, (ii) estimated change - points with Grid search from R and (iii) estimated change-points ad-hoc, with macro in SAS

2 change-point PEM	**-2Loglik**	**AIC**
(30,120)	8482.0	8494.0
(52,136)	8409.8	8421.8
(59,135)	8422.4	8434.4

Appendix B

Figures

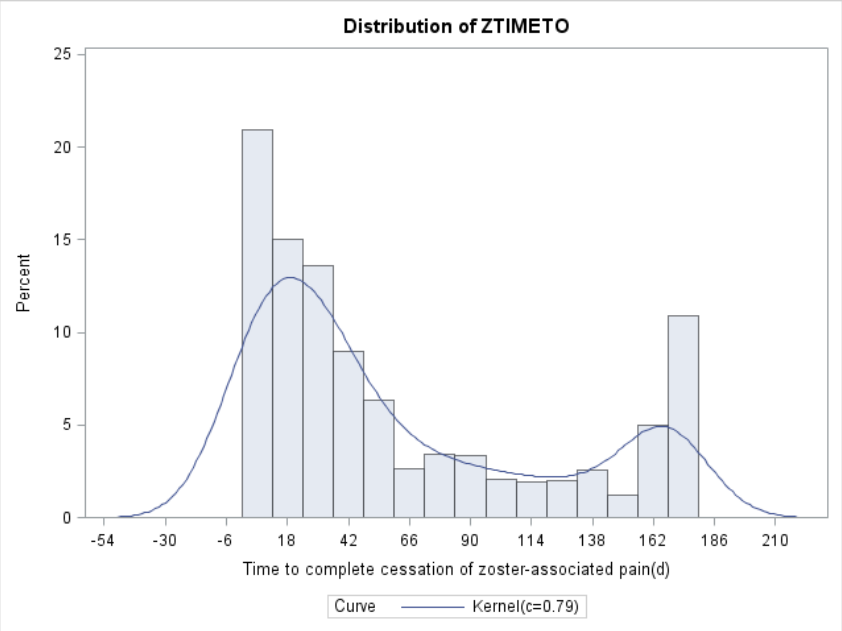


Figure B.1: Histogram of variable describing time in days until termination of zoster-associated pain.

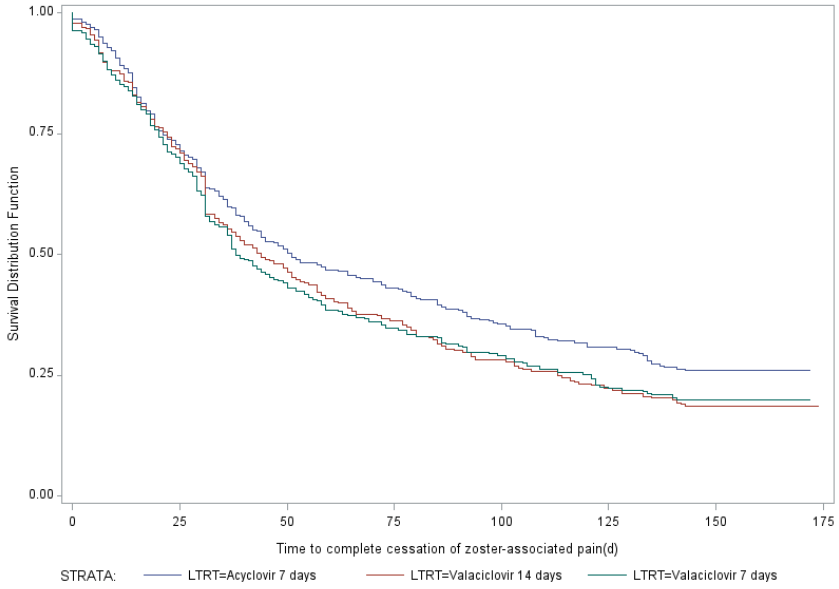


Figure B.2: Plot of Kaplan - Meier product limit estimator for Acyclovir versus Valacyclovir for the 3 different doses, whith merged dose categories.

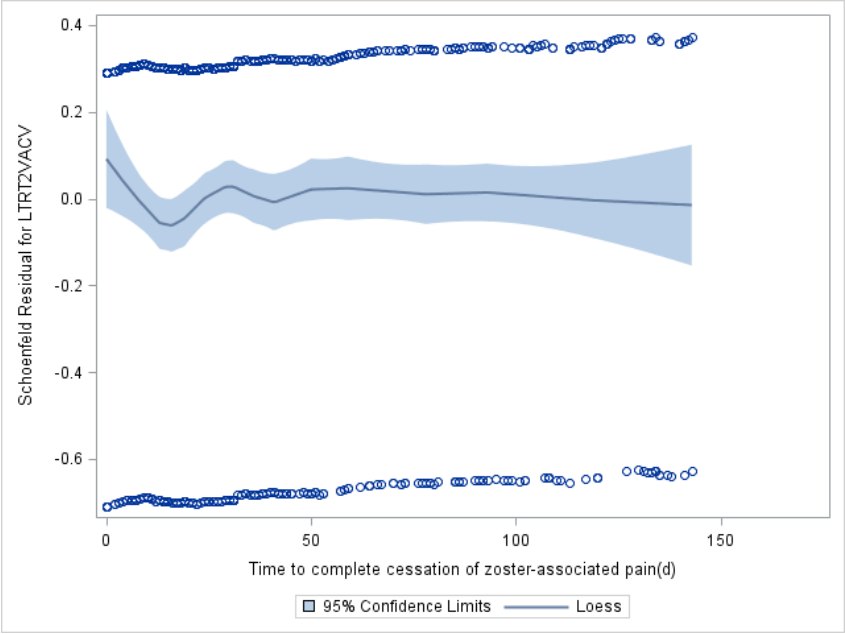


Figure B.3: Plot of Schoenfeld residuals, for the case of Acyclovir versus Valacyclovir of merged dose categories.

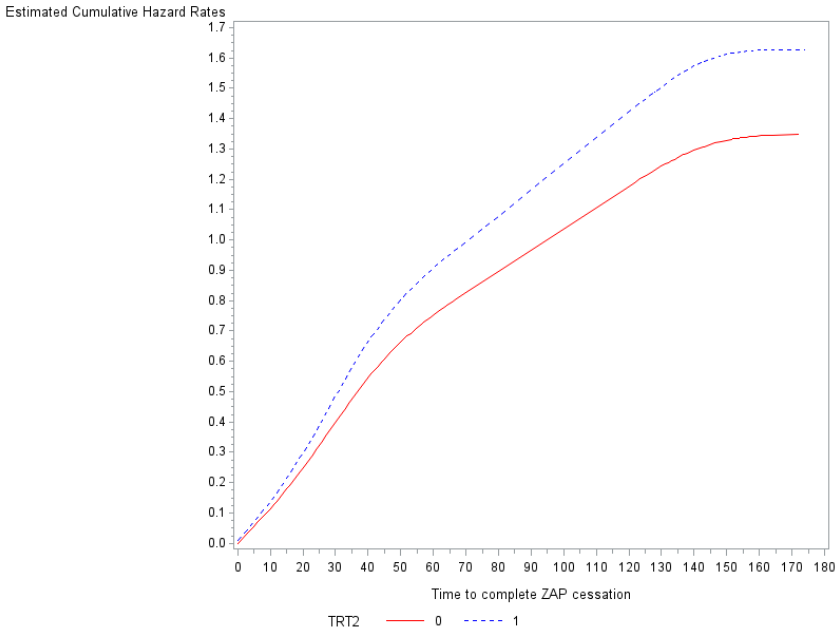
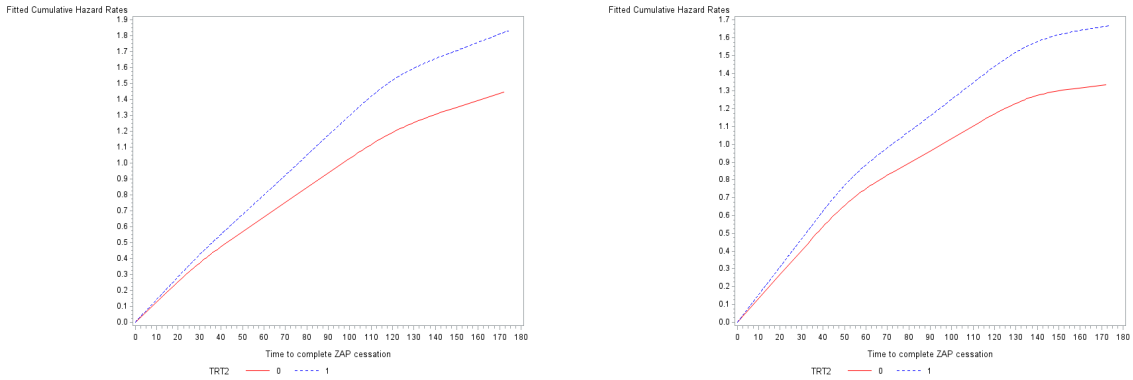


Figure B.4: Plot of Estimated Cumulative Hazard rated for Acyclovir (red) and Valacyclovir (blue).



(a) PEM with change points: 30 & 120 days

(b) PEM with change points: 52 & 136 days

Figure B.5: Fitted Cumulative Hazards plot of piecewise exponential models for given and estimated change-points (a) (30,120) and (b) (52,136) respectively. Each curve belongs to a treatment (Acyclovir: red, Valacyclovir: blue).

Appendix C

Codes

C.1 SAS Code

```
/* EDA */
DATA trial1; SET aa.trial1;
KEEP ZTIMETO LTRT TRT
AGE RACE SEX BASEPAIN
CENS_Z RASHON;
RUN;
/* QUALITATIVES */
PROC FREQ DATA=trial1;
TABLE CENS_Z TRT LTRT
RACE SEX BASEPAIN;
RUN;
/* QUANTITATIVES */
PROC MEANS N MEAN MIN MAX
MEDIAN STD DATA=aa.trial1;
VAR ZTIMETO AGE RASHON;
run;
/* CENSORING INDICATOR */
DATA trial1; SET trial1;
IF CENS_Z="Y" THEN DO;
CENSOR=0; END;
IF CENS_Z="N" THEN DO;
CENSOR=1; END;
RUN;
/* MISSING VALUES */
PROC FREQ DATA= trial1;
TABLE ZTIMETO AGE RASHON;
RUN;/*check freq of 0 obs.*/

/* Binary trt */
/* 1138 obs */
DATA trial; SET trial1;
LENGTH LTRT2 $10.;
IF LTRT="Acyclovir 7 days"
THEN DO;
TRT2=0;LTRT2="ACV";END;
IF LTRT="Valaciclovir 7 days"
THEN DO;
TRT2=1;LTRT2="VACV";END;
IF LTRT="Valaciclovir 14 days"
THEN DO;
TRT2=1;LTRT2="VACV";END;
/* Variables of interest */
KEEP ZTIMETO TRT LTRT TRT2
LTRT2 CENSOR;
IF CENSOR=. THEN DELETE;
RUN;
/* 1138 obs to analyze */


/* COX MODEL. PART */
```

```
/*LOG-RANK test for TRT*/
ODS GRAPHICS OFF;
PROC LIFETEST DATA=TRIAL
PLOTS=(s,ls,h) CS=NONE;
TIME ZTIMETO*CENSOR(0);
STRATA LTRT;
/*TRT can be used as well*/
/*LTRT has trt with dose*/
RUN;

/*COX REGRESSION*/
/* initial trt variable */
/* VACV14 -->baseline  */
ODS GRAPHICS ON;
PROC PHREG DATA=TRIAL;
CLASS LTRT(REF="VACV14");
MODEL ZTIMETO*CENSOR(0)=LTRT
/TIES=EFRON RL;
HAZARDRATIO LTRT2;
RUN;

/* binary trt*/
PROC PHREG DATA=TRIAL;
CLASS TRT2;
MODEL ZTIMETO*CENSOR(0)= TRT2
/TIES= EFRON RL;
run;

/* PROPORTIONALITY TEST */
/*Add time-dep covariate*/
PROC PHREG DATA=TRIAL;
CLASS TRT2;
MODEL ZTIMETO*CENSOR(0)=TRT2
TRT_TIME;
TRT_TIME=TRT2*ZTIMETO;
RUN;

/*Schoenfeld resid plot*/
PROC PHREG DATA=TRIAL;
CLASS LTRT2(ref="ACV");
MODEL ZTIMETO*CENSOR(0)=LTRT2;
OUTPUT OUT= PROPOTEST
RESSCH = SCHOENRES;
RUN;
PROC SGPLOT DATA=PROPOTEST;
LOESS X=ZTIMETO
Y=SCHOENRES /CLM;
RUN;

/*Empirical estim of Cum Haz*/
/*method=ch= NA estim*/
PROC PHREG DATA= trial;
MODEL ZTIMETO*censor(0)=TRT2;
output out=figure
LOGSURV=ls
/method = ch;
run;
DATA fig;
set figure;
haz=-ls;
run;
PROC SORT
DATA= fig;
by TRT2 ZTIMETO;
run;
proc gplot data = fig;
plot haz*ZTIMETO=TRT2 ;
symbol1 i=sm50
line=1 c =red;
symbol2 i=sm50
line=2 c =blue;
label haz="Estimated CH";
run;


/*** PEM ***/

/*for known change-points*/
PROC LIFEREG DATA=TRIAL;
```

```
MODEL ZTIMETO*CENSOR(0)=TRT2
/DIST=EXPONENTIAL;
RUN;
PROC NLMIXED DATA=TRIAL;
PARMS a0=4.6558 a1=-0.2155;
*Interval Indicators;
D1= (ZTIMETO <30);
D2= (ZTIMETO>=30);
linp1 = a0 + a1*TRT2;
linp2 = b0 + b1*TRT2;
* Baseline hazard ;
l1= exp(-linp1);
l2= exp(-linp2);
* Hazard function ;
h = l1*D1 + l2*D2;
* Survival function ;
S = exp(-l1*ZTIMETO)*D1 +
exp(-l1*30 -l2*(ZTIMETO-30))*D2;
loglik = CENSOR*log(h) + log(S);
MODEL ZTIMETO ~ GENERAL(loglik);
RUN;
/* 2 CP PEM*/
PROC NLMIXED DATA=TRIAL;
PARMS a0=4.3591 a1=-0.182
b0=5.1296 b1=-0.2359;
*automatically set c0=c1=1;
D1= (ZTIMETO<30);
D2= ( (ZTIMETO>=30)
and (ZTIMETO<120) );
D3= (ZTIMETO>=120);
linp1= a0 + a1*TRT2;
linp2= b0 + b1*TRT2;
linp3= c0 + c1*TRT2;
l1 = exp(-linp1);
l2 = exp(-linp2);
l3 = exp(-linp3);
h = l1*D1 + l2*D2 + l3*D3;
S = exp(-l1*ZTIMETO)*D1 +
exp(-l1*30 -l2*(ZTIMETO-30))*D2 +
exp(-l1*30 - l2*(120-30)
-l3*(ZTIMETO-120))*D3;
loglik = censor*log(h) + log(S);
MODEL ZTIMETO ~ GENERAL(loglik);
PREDICT 1-S out=cdfnew;
RUN;
```

C.2 R Code

```
install.packages("msm")
library(msm)

### SIMULATION ###
loglik<-function(lambda,censor,
data,points=c(50,120)){
n<-length(data)
fx<-dpexp(data,rate=lambda,
t=c(0,points))*(censor==0)+
(1-ppexp(data,rate=lambda,
t=c(0,points)))*(censor==1)
-sum(log(fx))
}
candidates<-cbind(rep(21:70,
each=50),rep(91:140,50))

## 50 repetitions ##
myestimates<-NULL
for (j in 1:50) {
print(j)
data<-rpexp(1000,
rate=c(0.02,0.01,0.005),
t=c(0,50,120))
censd<-rexp(1000,0.002)
censor<-as.numeric(censd<data)
data<-apply(cbind(data,censd),
```

```
1,min)

## for each grid point
## calculate maxim. likel.
logl<-estim<-NULL
for (i in 1:dim(candidates)[1]){
points<-candidates[i,]
myml<-optim(c(0.05,0.02,0.01),
loglik,censor=censor,
data=data, points=points)
logl<-c(logl,myml$value)
estim<-rbind(estim,myml$par)
}
maxML<-candidates[which.min(logl),]
maxMLlambda<-estim[which.min(logl),]
myestimates<-rbind(myestimates,
c(maxML,maxMLlambda))
}
## for s.e. and 95% CI
## Check at the end of Bootstrap

## HERPES ZOSTER DATA
## named zoster data set
## 3 columns:
## ZTIMETO, TRT2, censor

## Set the likelihood function
loglik<-function(lambda,censor,
data,trt,points=c(30,120)){

lambda1<-lambda[1:3] #ACV
lambda2<-lambda[4:6] #VACV
n<-length(data)
## censor==1 for events
## censor==0 for censored
t<- trt==0
fx1<-dpexp(data[t],rate=lambda1,
t=c(0,points))*(censor[t]==1)+
(1-ppexp(data[t],rate=lambda1,
t=c(0,points)))*(censor[t]==0)
t<- trt==1
fx2<-dpexp(data[t],rate=lambda2,
t=c(0,points))*(censor[t]==1)+
(1-ppexp(data[t],rate=lambda2,
t=c(0,points)))*(censor[t]==0)
-sum(log(fx1))-sum(log(fx2))
}

## choose grid
## form all possible pairs
candidates<-cbind(rep(20:60,
each=40),
rep(90:130,40))
## Grid search
logl<-estim<-NULL
for(i in 1:dim(candidates)[1]){
print(i)
points<-candidates[i,]
## initialvalues= vector of
## initial values for lambda
myml<-optim(initialvalues,
loglik,
censor=zoster$censor,
data=zoster$ZTIMETO,
points=points,trt=zoster$TRT2,
control=list(maxit=5000))
logl<-c(logl,myml$value)
estim<-rbind(estim,myml$par)}
maxML<-candidates[which.min(logl),]
maxMLlambda<-estim[which.min(logl),]

## BOOTSTRAP ##
```

```
keepmaxML<-logl
keepmaxMLlambda<-estim
myestim<-NULL
n<-dim(zoster)[1]

## for j=100 reps
for (j in 1:100) {
print(j)
nn<-sample(1:n,n,replace=TRUE)
new<-zoster[nn,]
logl<-estim<-conv<-NULL

for(i in 1:dim(candidates)[1]){
points<-candidates[i,]

myml<-optim(keepmaxMLlambda,
loglik,
censor=new$censor,
data=new$ZTIMETO,
points=points,trt=new$TRT2,
control=list(maxit=5000))

logl<-c(logl,myml$value)
conv<-c(conv,myml$conergence)
estim<-rbind(estim,myml$par)
}maxML<-candidates[which.min(logl),]
maxMLlambda<-estim[which.min(logl),]
myestim<-rbind(myestim,
c(maxML,maxMLlambda))}

## s.e. and 95% CI
se<-apply(estim,2,sd);se
CI<-matrix(NA,ncol(bootzost),2)
for (i in 1:nrow(CI)){
CI[i,]<-round(quantile(bootzost[,i],
prob=c(0.025, 0.975)),4)
};CI
## image plot (Likelihood)
image(seq(21,70,by=1),
seq(91,140,by=1),matrix(logl,50,50),
col=gray((0:32)/32),
xlab="1st change point",
ylab="2nd change point")
```

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Evaluation of the change-point piecewise exponential model

Richting: **Master of Statistics-Biostatistics**
Jaar: **2015**

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,

Kanavou, Sofia

Datum: **10/02/2015**