2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Evaluation of progression free survival as surrogate endpoint for overall survival for prostate cancer treatment

Promotor :
Prof. dr. Ziv SHKEDY

Promotor :
Dr. SUZY VAN SANDEN

Theophile Bigirumurame

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

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

De transnationale Universiteit Limburg is een uniek samenwerkingsverband van twee universiteiten in twee landen: de Universiteit Hasselt en Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

Maastricht University

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

2011
2012

FACULTY OF SCIENCES

Master of Statistics: Biostatistics

Masterproef

Evaluation of progression free survival as surrogate endpoint for overall survival for prostate cancer treatment

Promotor :
Prof. dr. Ziv SHKEDY

Promotor :
Dr. SUZY VAN SANDEN

Theophile Bigirumurame

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

Maastricht University

universiteit hasselt
KNOWLEDGE IN ACTION

ABSTRACT

In some clinical research, the endpoint of greatest relevance to inferences concerning therapeutic efficacy is not available or cannot be measured easily. Sometimes the determination of the true endpoint is difficult, requiring an expensive, invasive or uncomfortable procedure. In some trials, however, the main endpoint of interest, for example death, is rare and/or takes a long period of time to reach. In such trials, there would be benefit in finding a more proximate endpoint to determine more quickly the effect of an intervention.

In this report, a meta-analytic approach was used to validate progression free survival as a surrogate for overall survival. Firstly the logarithm of both endpoints were assumed to be normally distributed and fixed effects models were applied to them. Secondly both endpoints were considered as failure time and appropriate methods were applied. In all cases, the individual and trial level surrogacy estimates were too low to be useful. It was concluded that, for this study, progression free survival can't be used as a surrogate for the overall survival.

TABLE OF CONTENTS

ABSTRACT i

ACKNOWLEDGEMENTS iii

CHAPTER 1. INTRODUCTION 1

1.1 Objectives 4

CHAPTER 2. DATA DESCRIPTION 5

CHAPTER 3. METHODOLOGY 9

3.1 Single trial validation 10

3.2 Meta-Analytic approach 12

3.2.1 Normal/Normal endpoints case 13

3.2.2 Failure time endpoints case 17

CHAPTER 4. RESULTS 21

4.1 Normal/Normal endpoints case 21

4.2 Failure time endpoints case 22

4.2.1 Country specific Cox-ph model 22

4.2.2 Joint modeling of two failure time endpoints 23

CHAPTER 5. DISCUSSION AND CONCLUSION 25

BIBLIOGRAPHY 27

APPENDIX A. 30

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my thanks to those who helped me with various aspects in my studies at Hasselt University.

First and foremost, my supervisors Prof. dr. Ziv Shkedy and Dr. Suzy Van Sanden for their guidance, patience and support throughout this thesis.

I would additionally like to thank Prof. dr. Burzykowski Tomasz for his guidance throughout the initial stages of my thesis.

I would also like to thank all my classmates for their constant help and encouragement, but more especially my group members: Tadele, Aissami, Rashida, Hu, Aneth, and Thien, from whom I learnt a lot through many discussions we had.

My gratitude is expressed to all the Hasselt University Staff. There are people who live for helping and sharing what they have, among them Mrs Martine and Mrs Karin Daniels. Many thanks.

My deepest appreciation go to all my family members for the love they have shown me all through the two years I have been away from them.

Finally, I thank the Almighty God for making all this possible.

CHAPTER 1. INTRODUCTION

Prostate cancer is a form of cancer that develops in the prostate, a gland in the male reproductive system. Most prostate cancers are slow growing; however, there are cases of aggressive prostate cancers. Rates of detection of prostate cancers vary widely across the world, with South and East Asia detecting less frequently than in Europe, and especially the United States. Prostate cancer tends to develop in men over the age of fifty (Siegel et al. 2011). Globally it is the sixth leading cause of cancer-related death in men (in the United States it is the second) (Baade et al. 2009; Siegel et al. 2011). Prostate cancer is most common in the developed world with increasing rates in the developing world (Baade et al. 2009).

Different types of treatment are available for patients with prostate cancer. Some treatments are standard (the currently used treatment), and some are being tested in clinical trials. The dataset analyzed in this report is from a clinical trial where a new treatment is compared to standard treatment (Due to confidentiality, the names of the two treatment are not given), and two endpoints which are overall survival and progression free survival are considered. Overall survival was defined as the time interval from randomization to death from any cause. One secondary endpoint was progression-free survival, which was defined as the time from randomization to disease progression or death.

The conventional approach to evaluate the efficacy of therapeutic agents is to conduct clinical trials with clinical endpoints that reflect tangible benefits to patients. Such

endpoints include disease occurrence (for example, infection, cancer recurrence and heart attack) and death. Unfortunately, conventional clinical trials require hundreds of patients and take years to complete. Researchers and patients wish to assess the effectiveness of promising new agents as quickly as possible, which has led investigators to explore laboratory markers that may serve as surrogate endpoints in clinical trials. Replacement of a rare or late-occurring clinical endpoint with a frequent or short-term outcome variable can lead to substantial reduction in sample size and trial duration (Lin et al. 1997).

In some clinical research, the endpoint of greatest relevance to inferences concerning therapeutic efficacy is not available or cannot be measured easily. Sometimes the determination of the true endpoint is difficult, requiring an expensive, invasive or uncomfortable procedure. Sometimes we find it unobservable for an impractically long interval. Occasionally the true endpoint is not directly measurable at all, at least with current technology. In these cases we must rely on alternative, or surrogate, endpoints (Ellenberg and Hamilton 1989).

Surrogate endpoints are referred to as endpoints that can be used in lieu of other endpoints in the evaluation of experimental treatments or other interventions. They are useful when they can be measured earlier, more conveniently or more frequently than the endpoints of interest, which are referred to as the ' true' or 'final' endpoints (Ellenberg and Hamilton, 1989). In this project two endpoints were considered, progression free survival (surrogate endpoint), and overall survival (true endpoint). Progression free survival, which is the time from randomization to the progression of the disease or death was chosen here, because it can be observed earlier compared to overall survival. The prostate cancer grows slowly.

Prentice (1989) define a surrogate endpoint to be a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint. In

this project this definition is equivalent to say that a test of the null hypothesis of no relationship to the treatment groups under comparison on progression free survival is also valid test of the corresponding null hypothesis based on overall survival.

In this paragraph we give some advantages of using surrogate endpoint. First, the length of time required for follow-up in a trial that uses a surrogate endpoint is often much shorter than a trial that uses a true endpoint. Second, in some cases a surrogate may be easier to measure than the true endpoint. For example, for cardiovascular diseases if the true endpoint is the size of the infarction as measured by a myocardial scintigraphy, a measure of enzymes is an easily obtainable surrogate (Wittes et al. 1989). Third, the prevalence of certain rare diseases may be so low that a study of the 'true' endpoint is impossible. Perhaps the most important practical advantage of a surrogate is that the sample size in a trial with a surrogate may be considerably lower than in a trial of the true endpoint. For rare diseases for example, the use of survival time as the primary endpoint presents problems. First, the comparison requires a very large number of patients followed for an extended period to observe the necessary number of deaths.

In some cases, to measure the true endpoint may be more costly compared to the surrogate endpoint. For diseases where no effective therapies exist, surrogate endpoints are used to accelerate the approval mechanism by regulatory agencies. Before a surrogate endpoint can replace a final endpoint in the evaluation of an experimental treatment, it must be formally 'validated', a process that has caused much controversy and has not been fully elucidated so far (Burzykowski et al., 2005). A surrogate to endpoint should be clinically relevant and biologically plausible. It should be simple and cheap to measure and should be sensitive and specific to treatment effect. It should share a causal mechanism with the clinical endpoint and must be as close as possible to the clinical endpoint in the chain of events leading from drug-receptor interaction to the therapeutic response (Weir and Walley 2006).

The key motivation for validating a surrogate endpoint is to be able to predict the effect of treatment on the true endpoint, based on the observed effect of the treatment on the surrogate endpoint, with sufficient precision to distinguish safely between effects that are clinically worthwhile and effect that are not (Buyse et al. 2000).

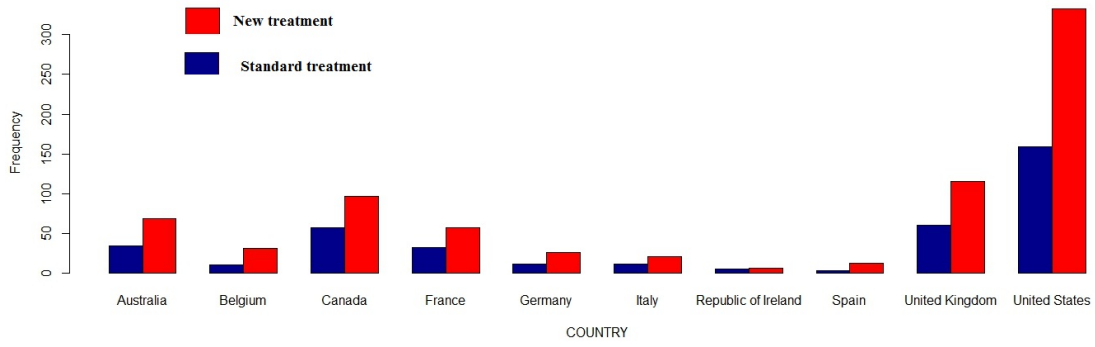
1.1 Objectives

The main goal of this report is to validate progression free survival as a surrogate for overall survival for prostate cancer for a new treatment. The structure of this report is as follows. The second chapter describes the dataset used for the analysis. In chapter three, different approaches used to validate the surrogate were introduced and described. In chapter four we present results obtained, and in the last chapter we discuss and conclude from the obtained results.

CHAPTER 2. DATA DESCRIPTION

Data analyzed in this report come from a clinical trial conducted in the following countries: Austria (AT), Australia (AU), Belgium (BE), Canada (CA), France (FR), Germany (GE), Hungary (HU), Italy (IT), Netherlands (NL), Republic of Ireland (IR), Spain (SP), United kingdom (UK) and United State of America (USA). In these countries, patients were enrolled in investigator sites. In total there was 147 sites. The number of site per country range from 1 (Netherlands) to 60 (USA). In the site, patients were assigned either to the new treatment or to the standard treatment (coded as 1 for the new treatment and 0 otherwise). The number of patients per site range from 1 to 48. Among those sites, 23 of them were not having any patient under the standards treatment arm, and 13 without any patient under the new treatment arm. 69 out of 147 (47%) sites have less than 5 patients in both treatment arms. Figure 2.1 shows the distribution of patients per country and per treatment arm. As it can be seen from it, in all countries, most patients were randomized to the new treatment arm.

Figure 2.1 Patients distribution per country and per treatment arm



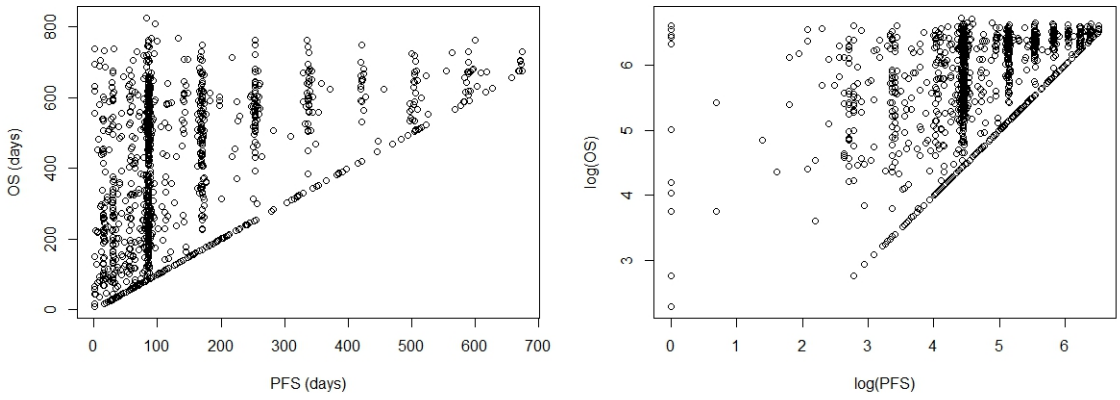
In total, there was 1183 patients, 789 of them were assigned to the new treatment, and 394 assigned to the standard treatment. Table 2.1 presents the distribution of the patients and sites per country.

Table 2.1 Patients and sites distribution per country

	AT	AU	BE	CA	FR	GE	HU	IT	NL	IR	SP	UK	USA
Sites	3	17	10	12	9	5	2	5	1	6	5	12	60
Patients	12	104	43	154	90	38	7	33	6	13	16	176	491

In what follows the country from which the patients come from was treated as the unit of analysis. Thus the term "trial-specific" should be understood as meaning country specific. The analysis was restricted to countries with at least 3 patients on each treatment arm. This constraint was adopted to ensure estimability. As a result, data from 10 countries were used, with a total sample size of 1158 patients. Data from Austria, Hungary, and Netherlands were not used in analysis. The overall survival time (true endpoint) and the progression free survival time (surrogate endpoint to be validated) raw data and log transformed data were shown in Figure 2.2.

Figure 2.2 original overall survival time vs progression free survival data (left plot). Log transformation of the original data (right plot)

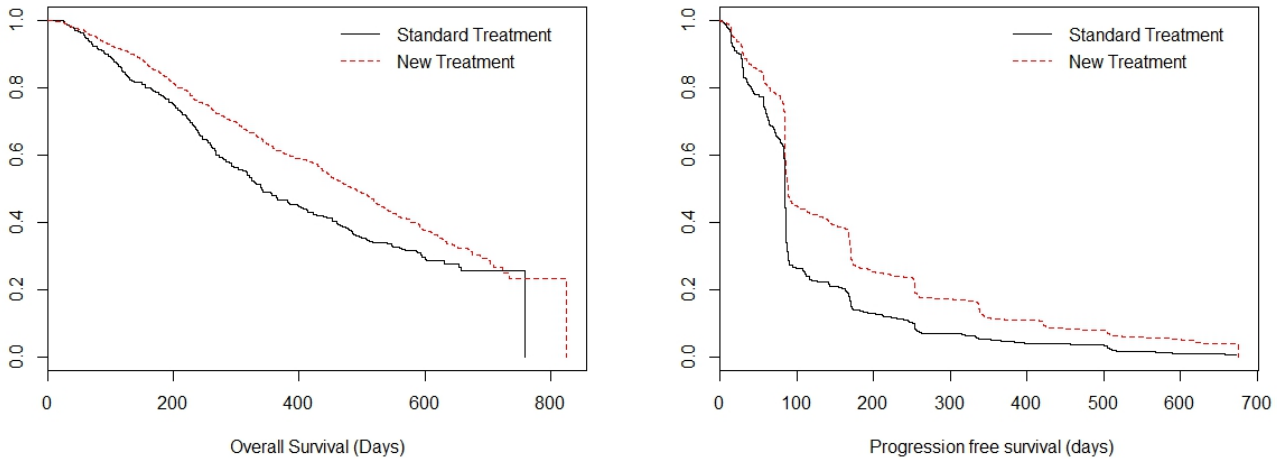


From the above figure, it can be seen that on the original scale most of the people had an event (death or progression) before 300 days, and at the end (about 700 days)

they were few people still surviving. The diagonal shows those patients for whom the overall survival was equal to the progression free survival. For patients who were above the diagonal, they had an event on their progression free survival (disease progression) before having an event on their overall survival (either died very late or still alive at the end of the study). In most patients, longtime was needed to have an event on overall survival endpoint compared to progression free survival endpoint.

Figure 2.3 shows that the medium survival time is longer on overall survival endpoint, as expected, compared to progression free survival endpoint. No significant results were found on both endpoints.

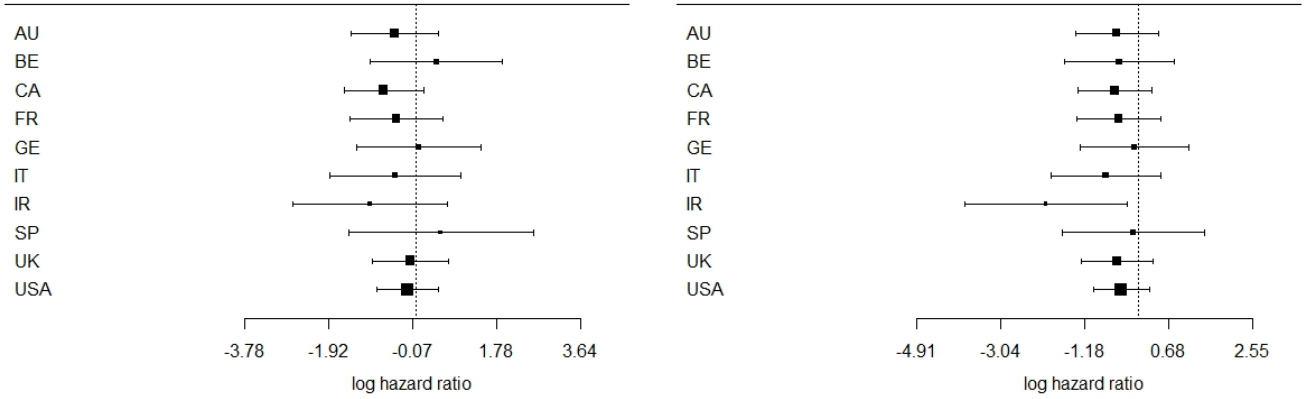
Figure 2.3 Kaplan Meier plot per endpoint



754 out 1158 (65%) patient had an event on overall survival time, while 1093 out of 1158 (94%) had an event on progression free survival time. 754 patients had an event on both endpoints.

Figure 2.4 presents the forest plots for log hazard obtained for both endpoints in different countries. It can be seen that all intervals contain zero, this suggests that there was no treatment effect on each endpoints, except for one country (Ireland), on the progression free survival endpoint. The vertical dashed line represents reference value zero.

Figure 2.4 Forest plots:log hazard ratio on overall survival (left) and log hazard ratio on progression free survival (right)



In this report SAS macro using IMLPlus programming language, developed by Burzykowski, were used to estimate the Copula. Other analysis were done using SAS version 9.2 and R version 2.15.1.

CHAPTER 3. METHODOLOGY

In validation of possible surrogate, two major branches have been developed, methods in single-trial settings and meta-analytical evaluation methods. However, there are many drawbacks of using single-trial data to validate or evaluate a possible surrogate endpoint (for example require of extremely large sample size, and/or require an unrealistic highly significant treatment effect on clinical endpoints to obtain estimates with sufficient precision) (Shi and Sargent 2009).

In order to reach the standard of validity of a surrogate endpoint, a trial-level assessment is required (Buyse and Molenberghs 1998, Buyse et al. 2000). The nature of the single trial restricts the generalizable information one can gain to evaluate surrogate endpoints within any single trial. Without multitrial data, it is almost impossible to make any direct inference about the association between the treatment effects on the surrogate and clinical endpoints, because one pair of data cannot provide sufficient evidence of any association (Shi and Sargent 2009). Systematically gathering data across various previous randomized trials can provide a more complete understanding of how well the treatment effect on the surrogate endpoint can predict the effect on the clinical endpoint (Hughes 2002).

As pointed out previously, the validity of a surrogate endpoint needs to be assessed at both the patient and the trial-level (Buyse and Molenberghs 1998, Buyse et al. 2000). The individual-level surrogacy, measures the association between the potential surrogate endpoint and the clinical endpoint, adjusting for the effect of treatment across all the trials included. On the other hand, the trial-level surrogacy, describes how well one

can predict the treatment effect on the clinical endpoint in a future trial based on the observed association between the treatment effects on the surrogate and clinical endpoints observed in previous trials (Shi and Sargent 2009).

In this chapter we describe methodology used in this report. First we discuss briefly single trial validation, based on Buyse and Molenberghs (1998) and then after the meta-analytic approach is discussed. In the first part of the analysis, the endpoints are assumed to be normal distributed (logarithm of the endpoints), ignoring the censoring in the data, but the main focus is put on them when they are failure time endpoints.

3.1 Single trial validation

In the case where both the surrogate and the true endpoint are continuously and jointly normally distributed Buyse and Molenberghs (1998) assumed the following model (adapted to our dataset)

$$log(PFS)_i = \mu_S + \alpha Z_i + \varepsilon_{Si} \tag{3.1}$$

and

$$log(OS)_i = \mu_T + \beta Z_i + \varepsilon_{Ti} \tag{3.2}$$

Where μ_S and μ_T are fixed intercepts and α and β are the fixed effect of the treatment Z on the progression free survival and overall survival endpoints, respectively. $Z = 1$ if the patient is under new treatment arm, and $Z = 0$ if the patient is under the standard treatment arm. Further, ε_{Si} and ε_{Ti} are error terms having a joint zero-mean normal distribution with the variance-covariance matrix :

$$\Sigma = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ & \sigma_{TT} \end{pmatrix} \tag{3.3}$$

Buyse and Molenberghs (1998) outlined some conceptual difficulties with the PE (proportion of the effect of Z on T (log(OS)) that can be explained by the surrogate

(log(PFS)), defined as :

$$PE(T,S,Z) = \frac{\beta - \beta_S}{\beta} = 1 - \frac{\beta_S}{\beta} \tag{3.4}$$

where β and β_S are the estimates of the effects of Z on T without and with adjustment for S. In some cases PE is not a proportion: it can be estimated to be anywhere on the real line, which complicates its interpretation. Buyse and Molenberghs (1998) argued that PE can advantageously be replaced by two related quantities. The first, defined at the population level and termed ' relative effect' (RE), is the ratio of the overall treatment effect on the true endpoint over that on the surrogate endpoint. Using (3.1)-(3.2), RE is formally defined as follows:

$$RE = \frac{\beta}{\alpha} \tag{3.5}$$

The second is the individual-level association between both endpoints, after accounting for the effect of treatment, and referred to as 'adjusted association' ρ_Z. For normally distributed endpoints, the adjusted association is defined as follows:

$$\rho_Z = \frac{\sigma_{ST}}{\sqrt{\sigma_{SS}\sigma_{TT}}} \tag{3.6}$$

Additionally, one would expect a good surrogate to have a strong association with the true endpoint *within* individuals, hopefully reflecting some biological pathway from the surrogate endpoint to the true endpoint (Buyse and Molenberghs 1998). Such an association could be captured by ρ_Z.

However, there are also problem with the RE as remarked by Buyse and Molenberghs (1998); the confidence interval for RE can be wide. This difficulty can be overcome by sufficiently large sample sizes, though. More importantly, in order to use the estimate of RE for predicting the treatment effect on T for a new trial (given the effect on S), it is necessary to assume that the relationship between the treatment effects on the

surrogate and the true endpoints is multiplicative (Buyse and Molenberghs 1998; Buyse et al. 2000). This assumption may be too stringent and, if the estimate of RE is based on a single trial, is unverifiable. A verification is possible if data from multiple trials are available (Buyse and Molenberghs 1998).

In our case, we have a meta analytic setting , and more emphasize are given to this setting. The meta-analytic framework is now a well-accepted one. It allows to increase the accuracy of the validation process (for example, increase the precision of the estimation of PE or RE), and to cast the evaluation in terms of two important concepts and ultimately quantities: trial-level and individual-level surrogacy. A surrogate being valid if it is both trial-level and individual-level valid. If the data are available on a single trial (or, more generally, on single experimental unit), the above concept (trial and individual surrogacy) is only partially possible. While the individual-level surrogacy (producing ρ_Z) carries over by virtue of the within-trial replication, the trial-level reasoning breaks down and one cannot go beyond the relative effect as suggested by Buyse and Molenberghs (1998).

3.2 Meta-Analytic approach

In meta-analytic approach, parameters estimates are more accurate compared to single trial validation. Here, the logarithm transformation of progression free survival and overall survival are assumed to be jointly normally distributed. Two distinct modeling strategies are followed, based on two-stage fixed effects representation on the one hand and random effects on the other hand (Buyse et al. 2000).

3.2.1 Normal/Normal endpoints case

3.2.1.1 Reduced fixed effects model

The first stage (from the two-stage model representation) is based upon a fixed effects model, fitted on progression free survival (log(PFS)) and overall survival (log(OS)), ignoring censoring. We start with the following reduced fixed effect model:

$$log(PFS)_{ij} = \mu_S + \alpha_i Z_{ij} + \varepsilon_{Sij} \tag{3.7}$$

and

$$log(OS)_{ij} = \mu_T + \beta_i Z_{ij} + \varepsilon_{Tij} \tag{3.8}$$

Where μ_S and μ_T are common intercepts, α_i and β_i are trial specific effects of treatment Z on progression free survival and overall survival , respectively, in trial i and ε_{Sij} and ε_{Tij} are correlated normally distributed error terms, assumed to be mean zero with covariance matrix:

$$\Sigma = \begin{pmatrix} \sigma_{SS} & \sigma_{ST} \\ & \sigma_{TT} \end{pmatrix} \tag{3.9}$$

With this reduced fixed effect model, trial level surrogacy is assessed by the determination coefficient obtained by fitting the following linear regression model:

$$\widehat{\beta}_i = \lambda_0 + \lambda_1 \widehat{\alpha}_i + \varepsilon_i \tag{3.10}$$

Where $\widehat{\beta}_i$ is the estimate of treatment effect on $log(OS)_{ij}$ in the i^{th} trial and $\widehat{\alpha}_i$ is treatment effect on $log(PFS)_{ij}$ in the i^{th} trial, obtained from models (3.7)-(3.8). ε_i are error terms, normally distributed with mean zero and a constant variance.

The individual level is assessed by the squared correlation between S and T after adjusting for both the trial effects and treatment effect, and is given by

$$R^2_{indiv} = \frac{\sigma^2_{ST}}{\sigma_{SS}\sigma_{TT}} \tag{3.11}$$

3.2.1.2 Reduced mixed effects model

At the second stage, it is assumed

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} a_i \\ b_i \end{pmatrix} \tag{3.12}$$

Where the second term on the right hand side of (3.12) is assumed to follow a zero mean normal distribution with dispersion matrix:

$$D_r = \begin{pmatrix} d_{aa} & d_{ab} \\ & d_{bb} \end{pmatrix} \tag{3.13}$$

The reduced mixed effects model is obtained by combining the first stage in (3.7)-(3.8) and second stage in (3.12) to give

$$log(PFS)_{ij} = \mu_S + (\alpha + a_i)Z_{ij} + \varepsilon_{Sij} \tag{3.14}$$

and

$$log(OS)_{ij} = \mu_T + (\beta + b_i)Z_{ij} + \varepsilon_{Tij} \tag{3.15}$$

where now μ_S and μ_T are fixed intercepts, α and β are the fixed effects of treatment Z on the endpoints, and a_i and b_i are the random effects of treatment Z on the endpoints in trial i . The vector of random effects (a_i, b_i) is assumed to be mean-zero normally distributed with covariance matrix (3.13). The error term ε_{Sij} and ε_{Tij} have a covariance matrix (3.9).

Trial-level surrogacy is given by following coefficient:

$$R^2_{trial(r)} = R^2_{b_i|a_i} = \frac{d^2_{ab}}{d_{aa}d_{bb}} \tag{3.16}$$

The individual level surrogacy is assessed as in the previous case, and is given by the quantity in (3.11).

3.2.1.3 Full mixed effects model

The full fixed effects model is obtained by adding random intercepts on the right side of equations (3.14)-(3.15)

$$log(PFS)_{ij} = (\mu_S + m_{Si}) + (\alpha + a_i)Z_{ij} + \varepsilon_{Sij} \tag{3.17}$$

and

$$log(OS)_{ij} = (\mu_T + m_{Ti}) + (\beta + b_i)Z_{ij} + \varepsilon_{Tij} \tag{3.18}$$

where now μ_S and μ_T are fixed intercepts, α and β are the fixed effects of treatment Z on the endpoints, m_{Si} and m_{Ti} are random intercepts, and a_i and b_i are the random effects of treatment Z on the endpoints in trial i. The vector of random effects $(m_{Si}, m_{Ti}, a_i, b_i)$ is assumed to be mean-zero normally distributed with covariance matrix:

$$D = \begin{pmatrix} d_{SS} & d_{ST} & d_{Sa} & d_{Sb} \\ & d_{TT} & d_{Ta} & d_{Tb} \\ & & d_{aa} & d_{ab} \\ & & & d_{bb} \end{pmatrix} \tag{3.19}$$

The error terms ε_{Sij} and ε_{Tij} follow the same assumptions as in fixed-effects model (3.7)-(3.8), with covariance matrix (3.9).

The quality of the surrogate at the trial level may then be calculated as the coefficient of determination for predicting the effect of Z on T, given the effect of Z on S:

$$R^2_{trial(f)} = R^2_{b_i|m_{Si},a_i} = \frac{\begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}^T \begin{pmatrix} d_{SS} & d_{Sa} \\ d_{Sa} & d_{aa} \end{pmatrix}^{-1} \begin{pmatrix} d_{Sb} \\ d_{ab} \end{pmatrix}}{d_{bb}} \tag{3.20}$$

Coefficient (3.20) is unitless and ranges in the unit interval if the corresponding variance-covariance matrix is positive definite, two desirable features for its interpretation (Buyse et al. 2000).

At the individual level, the association between endpoints is the squared correlation coefficient between S and T after adjusting for the trial and treatment effects:

$$R^2_{indiv} = R^2_{\epsilon_{Ti}|\epsilon_{Si}} = \frac{\sigma^2_{ST}}{\sigma_{SS}\sigma_{TT}} \tag{3.21}$$

Buyse et al. (2000) suggests to term a surrogate 'trial-level valid' if $R^2_{trial(f)}$ (or $R^2_{trial(r)}$) is sufficiently close to one, and 'individual-level valid' if R^2_{indiv} is sufficiently close to one. Finally, a surrogate is termed 'valid' if it is both trial-level and individual-level valid. In order to replace the words 'valid' with 'perfect', the corresponding R^2 values are required to equal one. To be useful in practice, a valid surrogate must be able to predict the effect of treatment upon the true endpoint with sufficient precision to distinguish safely between effects that are clinically worthwhile and effects that are not.

3.2.1.4 simplified models

Buyse et al. (2000) showed that fitting mixed-effects model (3.17)-(3.18) can be surprisingly difficult task in a number of situations. This is particularly true when the number of trials or the number of patients per trial is small. Also situations with extreme correlations pose problems. These authors studied one alternative approach in the sense that they replaced the mixed-effects model by their fixed-effect counterparts. Tibaldi et al. (2003) considered three dimensions along which simplifications can be made.

Trial dimension: whether the trial-specific effects are treated as either random or fixed. A full random-effects is then distinguished from a two-stage approach.

Endpoint dimension: whether the surrogate and true endpoints are modeled as bivariate outcome or two univariate ones. In the latter case the correlation between both endpoints is not incorporated into the modeling strategy, rendering the study of the individual-level surrogacy more involved.

Measurement error dimension: whenever the full random-effects model is abandoned, one is confronted with measurement error since the treatment effects in the various trial are estimated with error. The magnitude of this error is likely to depend on several characteristics, such as trial size, which will vary across trials. Three ways to account for measurement error are considered: unadjusted (i.e., no correction at all), adjusted by trial size, and an approach suggested by Van Houwelingen, Arends, and Stijnen (2002). Recall that the measurement error dimension is irrelevant when the full random effects model is assumed, but is crucial when a fixed effects approach is selected on the trial dimension and/or when a univariate model is chosen on the endpoint dimension.

3.2.2 Failure time endpoints case

In this setting, the surrogate and the true endpoint are failure time variable. The validation of surrogate in this setting is complicated by several factors, like presence of censoring and competing risks, or absence of a unifying framework such as multivariate normal distribution (Burzykowski et al. 2005).

In this setting two approaches were applied to assess the validity of progression free survival as surrogate for the overall survival. In the first case, Cox proportional hazard model were fitted univariately on both endpoints and per country. the following model was used to this end.

$$\lambda_{ij}(t) = \lambda_{i0}(t)exp(Z_{ij}\beta) \tag{3.22}$$

where:

- $\lambda_{ij}(t)$ is the hazard function for the j^{th} individual in the i^{th} country,
- $\lambda_{i0}(t)$ is the baseline hazard function for the i^{th} country and Z_{ij} is the treatment at which the j^{th} individual from the i^{th} country, is assigned to,
- β is the regression coefficient assumed to be common for all individuals across country.

After obtaining these coefficients, the one corresponding to the overall survival were

regressed on those corresponding to the progression free survival. The country-level (trial-level) surrogacy was assessed by the coefficient of determination of this regression line.

In the second case, we considered method developed by Burzykowski et al. (2001). Burzykowski et al. (2001) developed a method based on an extension of the meta-analytic proposed by Buyse et al. (2000). To extend the latter approach to the case where both surrogate and the true endpoint are failure time, Burzykowski et al. (2001) proposed to use a copula model. More specifically they assumed that the joint survival function of (S_{ij}, T_{ij}) can be written as:

$$F(s,t) = P(S_{ij} \geq s, T_{ij} \geq t) = C_\theta\{F_{Sij}(s), F_{Tij}(t)\}, \quad s,t \geq 0, \tag{3.23}$$

where F_{Sij} and F_{Tij} denote marginal survival functions for both endpoints (overall survival and progression free survival) and C_θ is a copula, i.e., a bivariate distribution function on $[0,1]^2$ with uniform margins. An attractive feature of model (3.23) is that the margins do not depend on the choice of the copula function.

In our setting, the following three copula functions were considered:

- The Clayton copula
- The Hougaard copula
- The Placket copula.

To model the effect of the treatment on the marginal distribution of S_{ij} and T_{ij} in (3.23) we use proportional hazards model proposed by Burzykowski et al. (2001):

$$F_{Sij}(s) = exp\{-\int_0^s \lambda_{Si}(x)exp(\alpha_i Z_{ij})dx\}, \tag{3.24}$$

$$F_{Tij}(t) = exp\{-\int_0^t \lambda_{Ti}(x)exp(\beta_i Z_{ij})dx\}, \tag{3.25}$$

where λ_{Si} and λ_{Ti} are trial-specific marginal baseline hazard functions and α_i and β_i are trial-specific effects of treatment Z on the endpoints in the trial i . The hazard

functions can be specified parametrically or can be left unspecified. In our setting the Weibull distribution for the marginal baseline hazards have been used. At the second stage, Burzykowski et al. (2001) proposed to use the model :

$$\begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + \begin{pmatrix} a_i \\ b_i \end{pmatrix} \tag{3.26}$$

where the second term on the right hand side of (3.26) is assumed to follow a zero mean normal distribution with dispersion matrix

$$D_r = \begin{pmatrix} d_{aa} & d_{ab} \\ & d_{bb} \end{pmatrix} \tag{3.27}$$

In view of using model (3.26) at the second stage of the two-stage approach, the quality of surrogate S at the trial level is assessed based on the coefficient of determination

$$R^2_{trial(r)} = \frac{d^2_{ab}}{d_{aa} d_{bb}} \tag{3.28}$$

To assess the quality of the surrogate at the individual level, a measure of association between S_{ij} and T_{ij} , calculated while adjusting the marginal distribution of the two endpoints for both the trial and treatment effects, is needed. Burzykowski et al. (2001) proposed to used Kendall's τ , as it depends only on the copula function C_θ and is independent of the marginal distribution of S_{ij} and T_{ij} . It describes the strength of the association between the two endpoints remaining after adjustment, through the marginal model (3.24)-(3.25), for trial and treatment effects. Kendall's τ has a straightforward interpretation as a measure of individual-level association: its range is $[-1, 1]$ and 0 means independence (Weir and Walley 2006).

CHAPTER 4. RESULTS

4.1 Normal/Normal endpoints case

As described in the previous chapter, different technics were applied to our dataset. The method developed by Buyse et al. (2000) was first tested on the data. In some cases there were convergence problems with the random effects models approach. For the full and reduced random effects model, there was no convergence for our data.

Some alternative methods described by Tibaldi et al (2003) were applied. Joint fixed effects (full and reduced) models were fitted. In this approach, there is heterogeneity because we have people from different countries with different sizes. One way to account for this heterogeneity is to use weights proportional to country size. Another way is to use method developed by Van Houweling et al. (2002).

To this end, weighted regression and unweighted regression were fitted in order to account for heterogeneity. There was convergence problem when the method developed by Van Houweling at al (2002) was applied to our data. In table 4.1, we give trial level surrogacy results.

Table 4.1 Trial level surrogacy

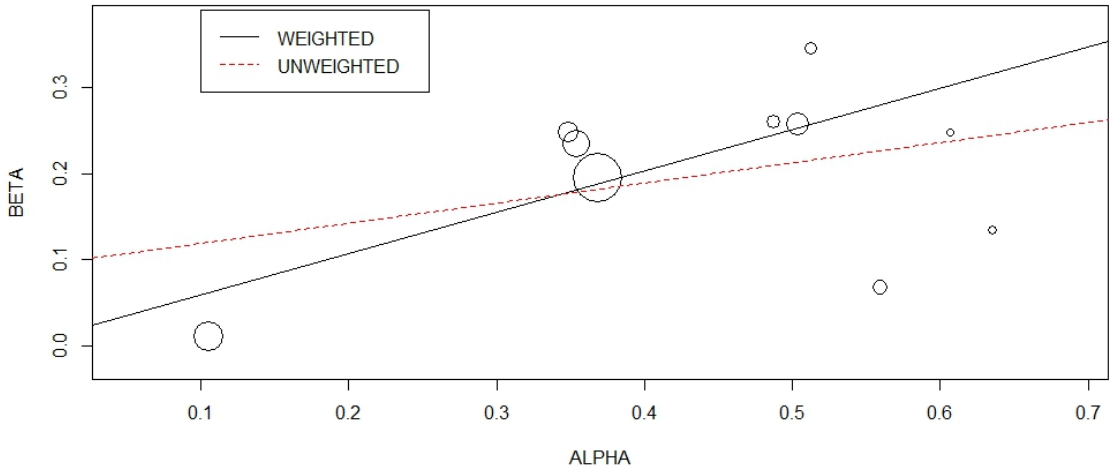
	Full Fixed effects model			**Reduced effects model**		
Level	**Unweighted**	**weighted**	**Van H.**	**Unweighted**	**weighted**	**Van H.**
Trial	0.42	0.29	-	0.14	0.51	-

The reduced fixed effects model gave the highest trial level surrogacy , $R^2_{trial(r)} = 0.51$, but this value is too low to be useful. The individual level surrogacy gave the same value

from both cases (full and reduced effects), $R^2_{indiv} = 0.44$.

The next plot shows the fitted regression line. The size of the circle is proportional to the number of patients from each trial (country).

Figure 4.1 Treatment effect on the logarithm of overall survival time vs treatment effect on logarithm of progression free survival

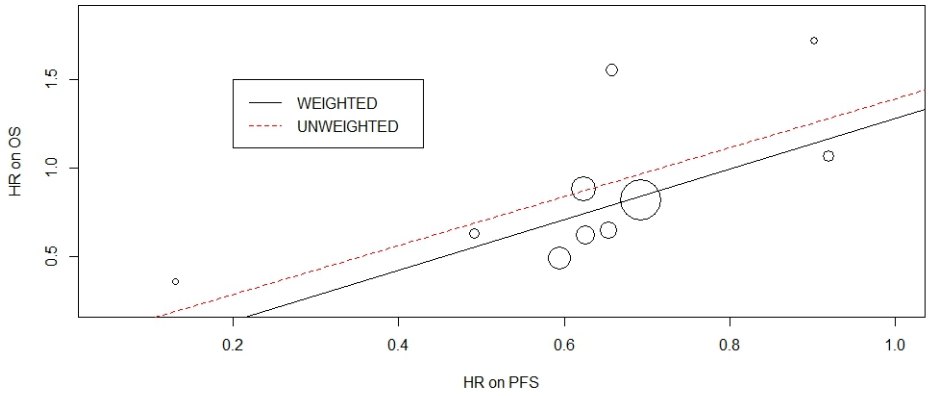


4.2 Failure time endpoints case

4.2.1 Country specific Cox-ph model

Beside fixed effects models (full and reduced), failure time endpoints were considered, and Cox's regression model fitted on both endpoints (per country). The hazard ratio from the overall survival endpoint were regressed on the one obtained from the progression free survival endpoint. The coefficient of determination from that linear regression model was used to assess trial level surrogacy. A value of 0.30 was obtained with a weighted regression line . A value of 0.46 was obtained with unweighted regression. The next figure shows both fitted lines. The circles sizes are proportion to the trial size.

Figure 4.2 Hazard ratio from overall survival endpoint vs hazard ratio from progression free survival endpoint



4.2.2 Joint modeling of two failure time endpoints

Lastly, results obtained with the copula models are presented. From these models, the trial surrogacy are assessed by the correlation coefficient (3.28). As stated before, three different copula were used. The individual level surrogacy was assessed by the Kendall's tau coefficient obtained from each copula. The next table summarizes the obtained values.

Table 4.2 Trial and individual level surrogacy results

Model	Individual level	Trial level (unadjusted)
Clayton	0.3039 [0.2635;0.3442]	0.3175 [-0.1592;0.7942]
Hougaard	0.3565 [0.3189;0.3941]	0.4262 [-0.0382;0.8905]
Plackett	0.3461 [0.3371;0.3552]	0.4754 [0.0270;0.9237]

The results shown in the above table, show that the highest trial-level surrogacy was given by the Plackett copula (0.47), and it is somehow closed to the one obtained using reduced fixed model 0.51 (by ignoring censoring and assuming normal distribution for the logarithm of both endpoints). The highest individual level surrogacy value was given by Hougaard copula (0.356). it is a little bit small compared to the one obtained using fixed effects models (0.44). we tried to find the adjusted R^2 using the van Houweling et

al. (2002), but there was convergence problem as in the fixed effects models cases.

Figures in the appendix, show treatment effect on overall survival vs treatment effect on the progression free survival from the three copula. The straight lines are the predictions from a (weighted by country size) simple linear regression model. As it can be seen, the relation seems to be not strong, which is in accordance with the above results.

CHAPTER 5. DISCUSSION AND CONCLUSION

The main objective of this study was the evaluation of progression free survival as a surrogate for the overall survival for prostate cancer . The dataset used in this study was obtained from a clinical trial conducted in 13 countries . The data was first explored at site level and at the level of the country. The exploratory analysis gave some insight into the dataset. Due to estimability problem, the country was chosen as analysis unit, and countries with at least three patients per treatment arm were keep for analysis.

In the first part, the censoring was ignored and the logarithm of the endpoints were assumed to be normally distributed. Fixed effects models (full and reduced) were fitted. Due to convergence problems, the random effects could not be fitted. In order to account for heterogeneity in the countries, the method described by Tibaldi et al. (2003) was used. It was not possible to obtain results from the approach developed by Van Houweling et al. (2002). There were convergence problems.

The estimates for R^2_{trial} and R^2_{indiv} were too small to be useful. For this dataset, progression free survival can't be used as a surrogate for overall survival.

When the endpoints were considered as failure time, two approaches were applied. Univariate Cox's propotional hazard model and the copula models. Results obtained from the Cox regression showed that progression free survival can't be used as surrogate for overall survival.

Results obtain from the copulas, were quite close to those obtained using fixed effects models. Progression free survival couldn't be validated as surrogate for the overall

survival in this study.

To conclude, in all approaches, it was possible to assess trial-level surrogacy. The obtained values point in the same direction. They are all too small to conclude that progression free survival can be used as a surrogate for the overall survival. Values obtained for individual surrogacy assessment were also small to conclude. In all models, where it was possible, the Van Houweling et al (2002) approach's didn't produce results, there were convergence problems.

BIBLIOGRAPHY

Baade, P.D., Youlden, D.R., Krnjacki L.J., (2009) International epidemiology of prostate cancer: geographical distribution and secular trends. *Molecular Nutrition Food Research*, **53**,171-184

Burzykowski, T., Molenberghs, G., Buyse, M., Renard, D., and Geys, H. (2001) Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. *Applied Statistics*, **50**, 405-422.

Burzykowski,T., Molenberghs, G., and Buyse, M. *The Evaluation of Surrogate Endpoints*. NewYork: Springer-Verlag (2005).

Buyse , M. and Molenberghs, G. (1998). The validation of surrogate endpoints in randomized experiments. *Biometrics*, **54**, 1014-1029.

Buyse, M., Molenberghs, G., Burzykowski, T., Renard, D., and Geys, H. (2000) The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics*, **1**, 49-67.

Daniels, M. J. and Hughes, M. D. (1997) Meta-analysis for the evaluation of potential Surrogate Markers. *Statistics in Medicine*, **16**, 1965-1982.

Hughes, M. D, (2002) Evaluating surrogate endpoints. *Control Clin Trials*, **23**, 703-707.

Lin, D., Fleming, T. R. and De Gruttola ,V. (1997) Estimating the proportion of treatment effect explained by a surrogate marker. *Statistics in Medicine*, **16**, 1515-1527

Ellenberg, S.S. and Hamilton, J.M (1989) Surrogate endpoints in clinical trials: cancer. *Statistics in Medecine* , **8**, 405-413.

Prentice, R.L. (1989) Surrogate endpoints in clinical trials: definitions and operational criteria. *Statistics in Medecine*, **8**, 431-440.

Shi, Q., Sargent, D.J. (2009) Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials. *Int J Clin Oncol*, **14**, 102-111.

Siegel, R., Ward, E., Brawley, O. and Jemal, A. (2011), Cancer statistics, 2011. CA: *A Cancer Journal for Clinicians*, **61**: 212-236. doi: 10.3322/caac.20121

Tibaldi, F.S., Cortias Abrahantes, J., Molenberghs, G., Renard, D., Burzykowski, T., Buyse,. M., Parmar, M., Stijnen, T., and Wolfinger, R. (2003) Simplified hierarchical linear models for the evaluation of surrogate endpoints. *Journal of Statistical Computing and Simulation*, **73**, 643-658.

Van Houwelingen, J.C., Arends, L.A., and Stijnen, T. (2002) Advanced methods in meta-analysis: multivariate approach and meta-regression. *Statistics in Medecine*, **21**, 589-624.

Weir, C.J. and Walley, R.J. (2006) Statistical evaluation of the biomarkers as surrogate endpoints: a literature review. *Statistics in Medecine*, **25**, 183-203.

Wittes, J., Lakatos, E. and Probstfield, J. (1989) Surrogate endpoints in Clinical Trials:Cardiovascular diseases. *Statistics in Medicine*, **8**,415-425.

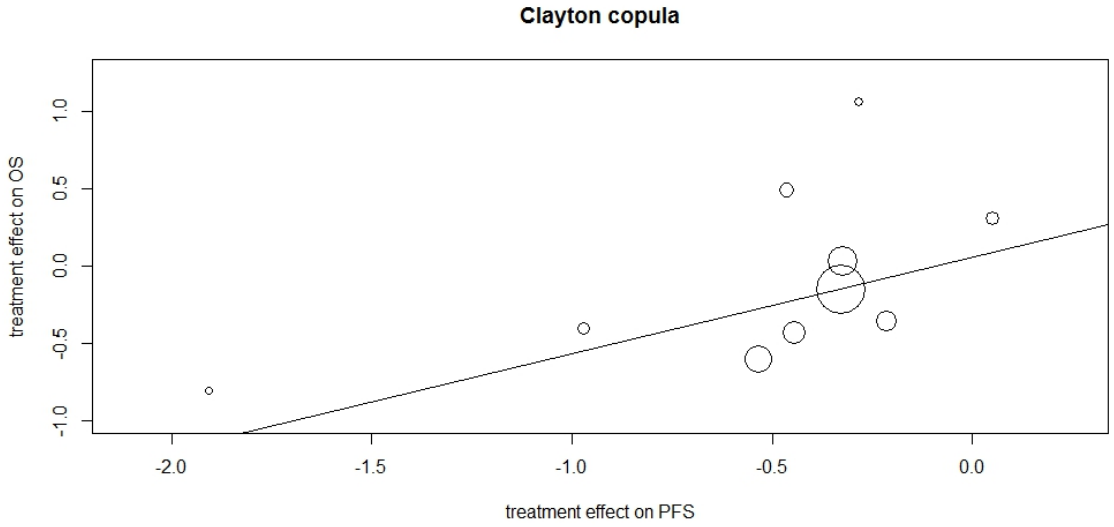
Prostate cancer overview http://www.cancer.org/Cancer/ProstateCancer/OverviewGuide/prostate-cancer-overview-what-is-prostate-cancer (accessed on 23/08/2012).

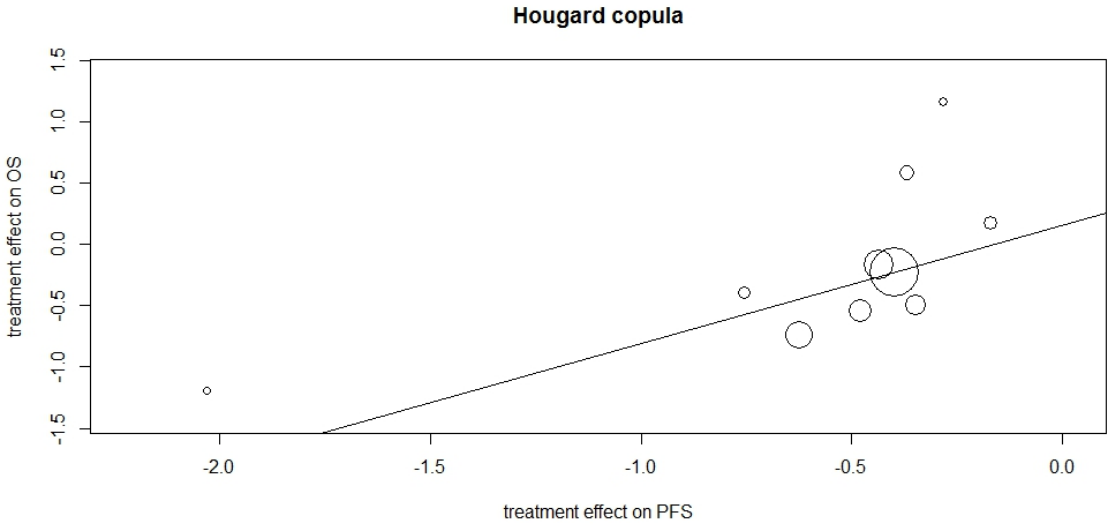
Prostate cancer http://en.wikipedia.org/wiki/Prostate_cancer#cite_note-1 (accessed on 23/08/2012).

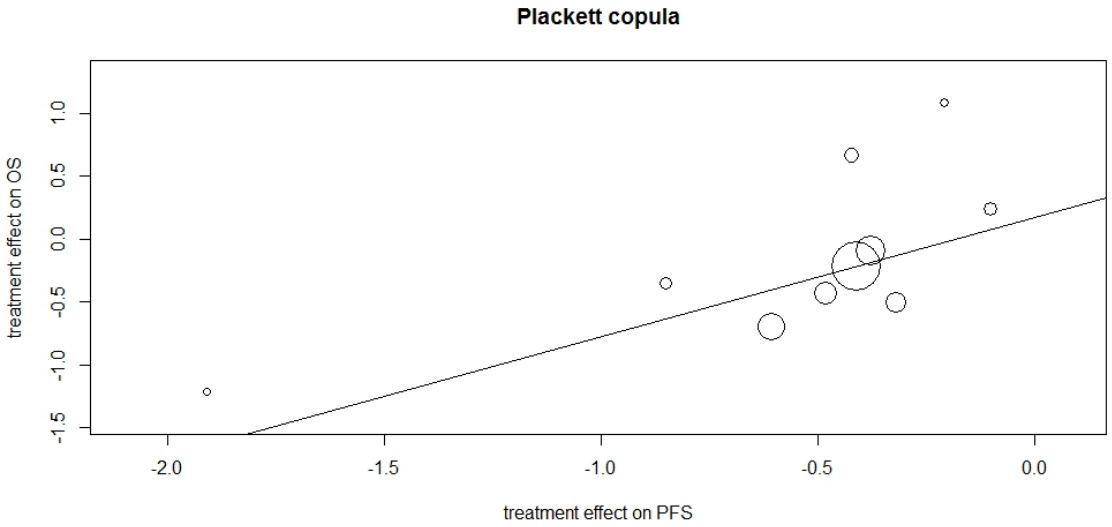
National Cancer Institute: Prostate cancer treatment http://www.cancer.gov/cancertopics/pdq/treatment/prostate/Patient/page4.(accessed on 23/08)

APPENDIX A.

Treatment effect on overall survival vs treatment effect on progression free survival




Hougard copula
treatment effect on OS
1.5
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-1.5
-1.0
-0.5
0.0
treatment effect on PFS


Plackett copula
treatment effect on OS
1.0
0.5
0.0
-0.5
-1.0
-1.5
-2.0
-1.5
-1.0
-0.5
0.0
treatment effect on PFS

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
Evaluation of progression free survival as surrogate endpoint for overall survival for prostate cancer treatment

Richting: **Master of Statistics-Biostatistics**
Jaar: **2012**

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,

Bigirumurame, Theophile

Datum: **14/09/2012**