Prostate Cancer: Prognostic Factors, Markers of Outcome and Design of Clinical Trials

Prostaatkanker: prognostische factoren, merkers voor behandelingsresultaten en opzet van klinische studies

Laurence Collette

The publication of protocol 30892 in chapter 2 was supported by Schering AG, Berlin, Germany, and by Schering Plough Inc., Bloomfield, NJ, USA. This publication was also supported by grant numbers 5U10-CA11488-30 and 2U10-CA11488-31 from the National Cancer Institute (US NCI, Bethesda, MD, USA)
The publication of EORTC trial 22863 in chapter 4 was supported by grants numbers 2U10 CA488-16 through 5U10 CA488-32 from the US NCI
The publications of Trial 22911 in chapter 5 were supported by grants number 5U10 CA488-21 through 5U10-CA488-34 from the US NCI and by a grant from Ligue Nationale contre le Cancer (Comité de l'Isère, Grenoble, France).
The publication of the joint analysis of EORTC trials 30903, 30944 and 30921 in chapter 8 was supported by grants number 5U10-CA11488-30 through 5U10-CA11488-33 from the National Cancer Institute (Bethesda, MD, USA).
The publication on the statistical aspects of trials of MAB in chapter 7 was supported by grant number 5U10-CA11488-29 from the US NCI.

Their contents are solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or of the Ligue Nationale contre le Cancer

The printing of this thesis was supported by grants from the Stichting Urologisch Wetenschappelijk Onderzoek (SUWO) Rotterdam, the EORTC Genito-urinary Tract Cancer Group, the EORTC Radiotherapy Group, Schering AG Berlin, Sanofi-Aventis and AstraZeneca.

Prostate Cancer: Prognostic Factors, Markers of Outcome and Design of Clinical Trials

Prostaatkanker: prognostische factoren, merkers voor behandelingsresultaten en opzet van klinische studies

Thesis

to obtain the degree of doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. S.W.J. Lamberts

and in accordance with the decision of the Doctorate Board
The public defence shall be held on

Wednesday 25 January 2006 at 15.45 hrs
by

Laurence Alice Jeanne Marie Collette
born at Namur, Belgium

Doctoral Committee

Promotors:	Prof.dr. F.H. Schröder
	Prof.dr. C.H. Bangma
Other members:	Prof.dr. TH. Stijnen
	Prof.dr. U.E. Studer
	Prof.dr. D.W.W. Newling

Que sais-je?
Montaigne

Au Professeur Paul Mainguet

This thesis is based on the following publications:

Chapter 2
L. Collette, Th.M. de Reijke, F.H. Schröder and members of the EORTC GU group Prostate Specific Antigen – A prognostic marker of overall survival in good prognosis M1 prostate cancer? *European Urology* 2003; 44(2):182-9. © 2003 European Association of Urology. Reprinted with permission from European Association of Urology.

Chapter 3
L. Collette, T. Burzykowski, K.J. Caroll, D. Newling, T. Morris, F.H. Schröder. PSA is not a valid surrogate endpoint for overall survival in patients with metastatic prostate cancer. A joint research of the European Organisation for Research and Treatment of Cancer, the Limburgs Universitair Centrum and AstraZeneca Pharmaceuticals. *Journal of Clinical Oncology* 2005; 23(25): 6139-48. © 2005 American Society of Clinical Oncology. Reprinted with permission from American Society of Clinical Oncology.

Chapter 4
M. Bolla, L. Collette, L. Blank, P. Warde, J.B. Dubois, R.O. Mirimanoff, G. Storme, J. Bernier, A. Kuten, C. Sternberg, J. Mattelaer, J. Lopez Torecilla, J.R. Pfeffer, C.L Cutajar and M Pierart. Long term results with immediate androgen suppression and external irradiation in patients with locally advanced prostate cancer (an EORTC study): a phase III randomized trial. *The Lancet* 2002; 360: 103-8. © 2002 Elsevier Ltd. Reprinted with permission from Elsevier

Chapter 5
M. Bolla, H. van Poppel, L. Collette, P. Van Cangh, K. Vekemans, L. Da Pozzo, Th. De Reijke, A. Verbaeys, J.F. Bosset, R. Van Velthoven, J.M. Maréchal, P. Scalliet, K. Haustermans, M. Pierart. Post-operative radiotherapy after radical prostatectomy: a randomized controlled trial (EORTC trial 22911). *The Lancet* 2005; 366: 572-578. © 2005 Elsevier Ltd. Reprinted with permission from Elsevier
L. Collette, H. van Poppel, M. Bolla, P. van Cangh, K. Vekemans, L. Da Pozzo, Th.M. de Reijke, A. Verbaeys, J.F. Bosset9 for the European Organisation for Research and Treatment of Cancer (EORTC) Radiotherapy and Genito-urinary Groups. Patients at high risk of progression after radical prostatectomy: do they all benefit from immediate post-operative irradiation? (EORTC trial 22911). *European Journal of Cancer* 2005; 41: 2662-2672. © 2005 Elsevier Ltd. Reprinted with permission from Elsevier.

Chapter 6
L. Collette, G. Van Andel, A. Bottomley, G.O.N. Oosterhof, W. Albrecht, Th.M. de Reijke, S.D. Fossa and members of the EORTC Genitourinary group. Is baseline quality of life useful for predicting overall survival with hormone refractory prostate cancer? *Journal of Clinical Oncology* 2004; 22(19): 3877-85. © 2004 American Society of Clinical Oncology. Reprinted with permission from American Society of Clinical Oncology.

Chapter 7
L. Collette, U.E. Studer, F.H. Schröder, L. Denis and R. Sylvester. Why trials of maximal androgen blockade versus castration in M1 prostate cancer rarely show statistically significant differences. *The Prostate* 2001; 48(1): 29-39, Copyright © 2001 Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc. Reprinted with permission of John Wiley & Sons, Inc.

Contents

Index of Tables

Index of Figures

Chapter 1

Introduction

1.1 The need for surrogate endpoints

Phase III clinical trials that evaluate the clinical benefit of new treatment options often require large patient numbers and long follow-up. In diseases with a long natural history, such as prostate cancer (PCa), the final result of comparative trials with survival endpoints is often not known before five to ten years after the study start[1-3]. This time lag reduces the relevance of the study results and delays the availability of level I evidence to support effective changes in clinical practice. It is indeed not rare that new treatment strategies or new markers of prognosis are implemented into clinical practice long before randomized clinical trial results become available.

With increasing knowledge of the genetic background of cancer and cancer progression and of the mechanisms of signal transduction, a large number of new compounds are under development[4-6]. In addition, there is increasing public pressure for promising new drugs to receive marketing approval as rapidly as possible, in particular for life threatening diseases such as cancer. For these reasons, there is an urgent need to find ways of speeding up the new drug development process by optimizing the design of (especially phase III) clinical trials.

The long duration of phase III clinical trials mostly results from the use of a long term clinical endpoint (clinical progression, survival) especially when studying a slowly developing disease like prostate cancer. Therefore, to replace that long term endpoint (the "true" endpoint) by another (a "surrogate" endpoint), that could be measured earlier, more conveniently or more frequently, and that would adequately reflect the benefit of new treatments on the true endpoint would seem attractive[7]. Such replacement endpoints are called "surrogate" endpoints.

The need to accelerate drug development in human cancer is reflected by changes in the regulation of cancer drug approval by the Food and Drug Administration (FDA) in the US[8]. In 1992, the regulation was amended to allow accelerated approval for serious or life-threatening diseases, when adequate and well controlled clinical trials establish that the new drug product has an effect on a surrogate endpoint that is "reasonably likely", based on epidemiologic, therapeutic, pathophysiologic, or other evidence, to predict clinical benefit on the basis of an effect on a endpoint other than survival[9]. In 2004, the FDA held a Public Workshop entirely devoted to discussing "Clinical Trial Endpoints in Prostate Cancer" during which the endpoints that would be regarded as acceptable for the regulatory

(accelerated) approval of new drugs in prostate cancer according to the above definition were discussed at length.

From a methodological perspective as well, the conditions for the use of surrogate markers as endpoints of clinical trials have to be carefully defined and it is useful to clarify some of the terminology.

1.2 Definition of terms

1.2.1 Prognostic factor

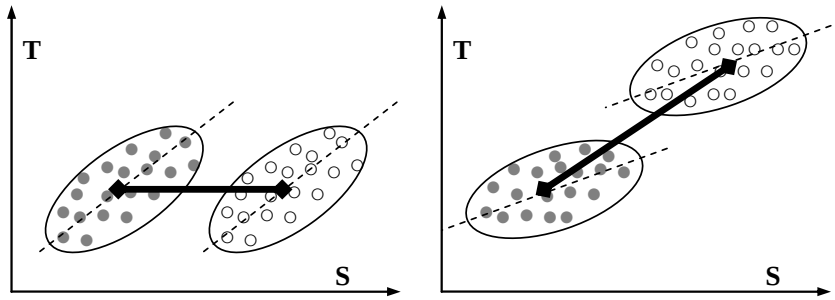
Biomarkers or prognostic factors are physical signs or laboratory measurements that occur in association with a pathological process or that have putative diagnostic and/or prognostic utility[10]**.** They are generally regarded as best candidate surrogate endpoint. A biomarker (or prognostic factor) is an intermediate outcome (S) that is correlated with the true clinical outcome (T) for an *individual* patient. The knowledge of the biomarker S may be useful for diagnostic or prognostic assessment of an individual patient. It is especially useful whenever S is observed earlier or less invasively than the true outcome T. (This notion is sometimes called "individual-level surrogacy" of S for T but to avoid confusion, we shall use the term "prognostic factor" in the rest of the text).

1.2.2 Surrogate endpoint

It is a common misconception that established prognostic factors *necessarily* make valid surrogate endpoints. For a prognostic factor to be a surrogate endpoint (S), it is further required that "the effect of treatment on a surrogate endpoint must be "reasonably likely" to predict clinical benefit"[11]. In other words, a biomarker S will be a good surrogate for the true endpoint T if the results of a trial using outcome S can be used to infer the results of the trial if T had been observed and used as endpoint and this with sufficient precision. *Surrogacy* is thus a concept that relates to *groups* of patients. To demonstrate surrogacy, a high association between the treatment effects on the surrogate and on the true endpoint thus needs to be established *across groups of patients* treated with the new and standard interventions. (This notion is also called "trial-level" surrogacy of S for T).

Figure 1.1 shows a schematic of two situations were
(a) S is a good prognostic factor for the true endpoint T in both treatment groups (grey and white dots) but is a not a surrogate for T and where
(b) S is a prognostic indicator of the endpoint T at the individual level and is also a good surrogate for T and thus can be used for predicting the effect of a treatment intervention (white versus grey dots) onto the true endpoint T. In theory, the factor may even not be prognostic but could be a surrogate of the true endpoint T (in which instance, the ovals would be horizontal)

Figure 1.1
Prognostic factor versus Surrogate endpoint



(a) S prognostic for T
But not surrogate

(b) S surrogate for T
(and weakly prognostic)

1.2.3 An example

To further illustrate the notions of prognostic factor and a surrogate endpoint, we shall refer to the EORTC trial 22911 that is presented in details in chapter 5 of the present thesis.

This trial considered with pT2-3pN0M0 prostate cancer patients who presented post-prostatectomy with at least one out of three pathological risk factors of: positive surgical margins (SM+, including at the level of the prostate apex where the capsule is non existent), capsule perforation (pT3a), or invasion of seminal vesicles (pT3b). The patients were randomized between no adjuvant treatment (standard arm) and immediate adjuvant irradiation (experimental arm). The trial endpoints were biochemical progression-free survival and clinical progression-free survival. Biochemical progression was defined as the first PSA increase over the lowest postoperative value to a value >0.2 ng/ml, confirmed twice at min 2-week intervals. Clinical progression was assessed by imaging. The trial showed a significant benefit for biochemical progression-free survival (P<0.0001, HR=0.48, 98% CI: 0.37-0.62) and for clinical progression-free survival (P=0.0009, HR=0.61, 98%CI: 0.43-0.87).

An example of a prognostic factor in this trial is baseline PSA. To simplify, we will consider only the patients in the reference group, in whom the natural disease history can be best assessed. In that group, patients with an elevated PSA prior to surgery (>20 ng/ml) appeared to have a worse outcome than the patients with a low initial PSA. Indeed, the patients with pre-operative PSA>20 ng/ml have about double the risk of clinical progression than the patients with pre-operative PSA<20 ng/ml (HR=1.93, 95%CI: 1.32, 2.83, P=0.0006).

Another example of a (time-dependent) prognostic factor is PSA relapse post-surgery. Irrespective of the treatment group, patients with a biochemical failure are at increased risk of clinical failure. Once they have had a PSA relapse, the risk of clinical relapse is increased 11-fold in comparison to patients without a PSA relapse (HR=11.3, 95%CI: 8.2-15.5, P<0.0001).

For post-treatment biochemical failure to be a surrogate for clinical progression-free survival, we would further need that there is a strong association between the treatment effects on PSA relapse and on the clinical endpoint. This association between the treatment effects is needed to enable to predict the treatment effect onto clinical progression-free survival from the treatment effect observed on biochemical progression-free survival. If there would be such an association and if we would have observed a significant treatment hazard ratio for the endpoint biochemical progression-free survival, say HR=0.48, we would be in a position to predict that there would certainly be a significant treatment effect on clinical progression-free survival, were it of smaller magnitude, say 0.61. Conversely, surrogacy would require that observing no no significant impact of post-operative irradiation on biochemical failure would with certainty be indicate the absence of a significant impact on clinical progression-free survival.

1.3 The Prostate Specific Antigen (PSA)

In the field of PCa, prostate-specific antigen (PSA) has probably been the most studied biomarker[12,13].

PSA is a glycoprotein that is produced by the prostate and is usually only detectable at very low levels in the blood. PSA is produced by the glandular tissue of the prostate and is normally secreted into semen or lost in urine. PSA does not however originate solely from the prostatic epithelium. PSA and/or PSA gene expression has been detected in the endometrium, normal breast tissue, breast tumor, adrenal neoplasms and renal cell carcinomas at extremely low levels[12]. An increased leakage of PSA into the circulation appears to be what causes most PSA elevations in prostatic diseases. The PSA level in prostatic fluid is approximately one million-fold higher than that found in the serum. An epithelial layer, a basal cell layer and a basement membrane separate the intraductal PSA from the capillary and lymphatic drainage of the prostate. It is believed that the leakage of PSA into the serum increases, causing elevated PSA levels in the blood, when a prostatic disease such as prostatitis, benign prostatic hyperplasia or prostate cancer interferes with this natural barrier. PSA levels in the serum also increase after inflammation or urologic manipulations (e.g. after a biopsy)[14]. Although hyperplastic, neoplastic and non cancerous prostatic epithelial cells make PSA, prostate cancer is associated with PSA levels at least 10-fold higher per gram tissue than BPH[15,16]. Thus for clinical utility, PSA is organ-specific but not cancer-specific. As a prostate cancer grows, PSA levels in the blood tend to increase as well. This finding suggests that response to prostate cancer treatment or failure after treatment may be followed with serial PSA levels.

The production of PSA by a normal prostate as well as by prostate cancer cells is dependent on male hormones (androgens) being present in the body at normal levels. Thus the mainstay of treatment of prostate cancer is based on blockage of the androgen production by endocrine treatment. Caution is however required when assessing PSA after treatment since PSA in itself is an endocrine dependent enzyme and its expression is regulated by a promoter that contains androgen responsive elements[17,18]. The effects of endocrine treatments seen on PSA may thus at least in part result from a direct, non tumor mass related effect whereas treatments targeting other pathways might act on the tumor without affecting PSA to a great extent.

PSA has been studied as a diagnostic indicator of prostate cancer to be used in screening studies (for instance in European Randomized Study of Screening for Prostate Cancer or ERSPC[19]**)** as well as an indicator of disease stage and extent but is not alone sufficiently sensitive or specific for staging[e.g.20]. Baseline PSA was assessed as a potential prognostic factor for long term outcome in all disease stages but did not consistently demonstrate significant association[12]. PSA progression was also studied as an early indicator of clinical disease progression after irradiation[e.g.**21**], after radical prostatectomy[e.g.**3**] as well as in advanced disease[e.g.**22**] and PSA doubling time and PSA velocity were studied as predictors of relapse in early diseas[**e.g.23,24**] as well as in more advanced disease[7,25].

PSA was also studied as a potential surrogate for long term clinical outcome across disease stages[12,13,26,27], and in hormone refractory patients in particular[28-32]. In 1999, a Consensus Statement provided guidelines to the use PSA as endpoint in phase II studies in hormone refractory disease[33]. Many studies have demonstrated an association between PSA levels and long term outcome, thus showing the value of PSA as a prognostic factor at the individual patient level. However, many of the surrogate endpoint validation studies carried out thus far have failed to provide statistically definitive evidence of the value of PSA as a surrogate endpoint for long term clinical outcome such as (disease-free) survival[12,13,26-32] or were based on the Prentice criteria, the validity of which we shall discuss later[34-36]. Thus the use of PSA endpoints as surrogate endpoint in phase III trials is until now not warranted.

1.4 How to demonstrate surrogacy?

1.4.1 Validation methods based on data from a single trial: the Prentice Criteria

The validation of a candidate surrogate endpoint is not straightforward. Until recently, the statistical methods developed for this purpose used the data from a single trial[37-39]. These methods suffer from numerous drawbacks: some of them are too stringent to be of practical value, while others are based on non testable assumptions[40,41]. The most widely used single-trial validation method is known as the "Prentice Criteria"[37]. Prentice proposed defining a surrogate endpoint, as "a response variable for which a statistical 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". Prentice proposed four criteria to verify that the definition holds true. These criteria require that

(a) The treatment must have a statistically significant effect upon the surrogate endpoint
(b) The treatment must have a statistically significant effect upon the true endpoint
(c) The putative surrogate must have statistically significant effect upon the true endpoint
(d) The full effect of the treatment onto the true endpoint must be mediated through the surrogate

Condition (c) requires that the putative surrogate endpoint be a prognostic factor of the true endpoint, i.e. it requires strong association between the endpoints within the individual patient. Condition (b) limits the applicability of the criteria to trials that showed a statistically significant treatment effect on the true endpoint, a condition which is rarely fulfilled by clinical trials in prostate cancer. Condition (c) raises a conceptual difficulty as it requires proving the absence of remaining treatment effect onto the true endpoint in a model that is adjusted for the surrogate endpoint and thus requires that a statistical tests shows the treatment effect to be *non* statistically significant in such a model. Statistical tests however are designed to reject a null hypothesis (of no treatment difference) in favor of an alternative hypothesis (that there is a difference between treatments). The non-rejection of the null hypothesis never stands as definitive proof that the null hypothesis is true: it is statistically impossible to prove the absence of a treatment effect. Finally, it was shown that the set of four criteria is equivalent to the definition only in the case of binary (response) endpoints. In other settings, like in phase III trials with time-to-event endpoints, the operational criteria are necessary but not sufficient to demonstrate that surrogacy holds true[39]. Thus, failure to demonstrate that the four criteria hold is useful to disregard a biomarker as being a surrogate endpoint, but successful demonstration that the four criteria hold true is not sufficient to actually demonstrate that a biomarker is a surrogate for a time-to-event endpoint.

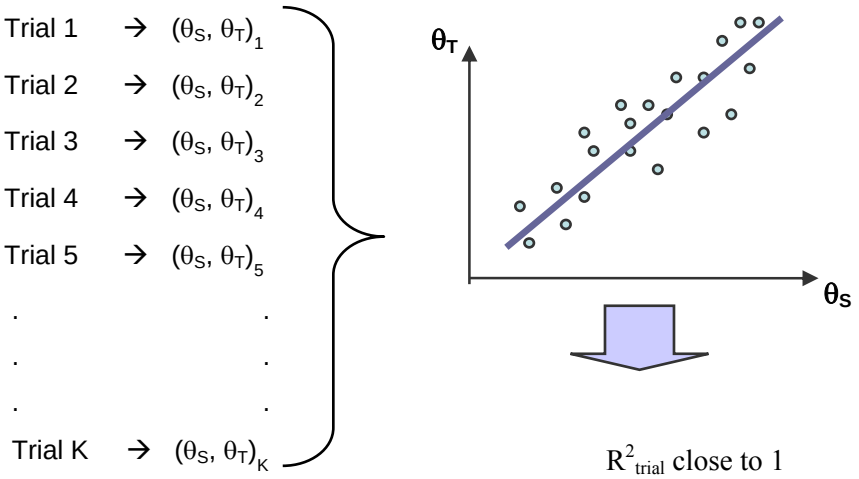
1.4.2 Validation method based on data from a several trials: the meta-analytic approach

To overcome the limitations detailed above, a new methodology known as the "meta-analytic" validation approach was recently developed[42-44]. Unlike the Prentice criteria that are deeply rooted into the concept of statistical testing, the meta-analytic method is rooted in the concept of estimation. With this new method, a surrogate endpoint is considered statistically valid if it enables us to predict precisely enough the treatment effect on the true clinical endpoint from the treatment effect observed on the surrogate endpoint. This new methodology does not require that any of the treatment effects be statistically significant. However, the method necessitates large databases from multiple randomized clinical trials with similar design and treatments.

Indeed, when data from several trials are available, the method consists of simultaneously estimating the relative treatment effects on the true (survival) endpoint and on the surrogate (PSA) endpoint in each trial. A bivariate model is then adjusted to estimate the association between the treatment effects on the true endpoint and the corresponding effects on the surrogate endpoint in a way similar to standard linear regression **(**Figure 1.2), although mathematically more sophisticated. The model also takes into account the fact that the treatment effects in the various studies were themselves estimated and thus are subject to measurement and estimation error. Alike in linear regression, the strength of the association is measured by the squared correlation coefficient (R^2_{trial}). This coefficient also indicates the precision with which the treatment effect on the true (survival) endpoint can be predicted from the observed treatment effect on the surrogate (PSA). The maximal possible value of R^2_{trial} is 1 which indicates a perfect prediction. In practice, observing R^2_{trial}=1 is not possible and one rather seeks a value close to one which indicates a strong association between the treatment effects and thus a relatively precise prediction[43,45]. Additionally, this model also provides measures of the prognostic association between the surrogate endpoint and the true endpoint at the individual patient level (prognostic factor association).

This new methodology is most appropriate for studying the value of PSA as a surrogate endpoint as in the field of prostate cancer, several studies of similar drugs are often available whereas between the advent of hormonal treatment in the 1940's and that of Taxotere in 2004, few trials demonstrated statistically significant treatment benefit.

Figure 1.2
The Meta-Analytic validation method



R^2_{trial} close to 1
To allow a precise prediction of the treatment effect on the true endpoint (θ_T) from the treatment effect observed on the surrogate endpoint (θ_S)

1.5 PSA as an endpoint: definitions

Before embarking in a validation exercise of a PSA as surrogate for long term clinical endpoints in any given stage of prostate cancer, one ought to carefully define the actual PSA endpoint that will be assessed. Many variants of PSA endpoints exist from PSA response using specified threshold of response (undetectability, normalization, decline by 80%, by 50% with or without a Landmark); to PSA doubling time, PSA velocity and a variety of definitions of PSA progression (ASTRO definition, time to PSA relapse after being undetectable, time to PSA rise above a threshold and/or by a given amount above previously observed nadir, with or without confirmation…).

The endpoint is defined in function not only of the disease group being studied but also in relation to the treatments that are being applied. For example in early disease, the ASTRO definition of PSA progression may be appropriate for trials investigating radiotherapy, whereas it is not after radical prostatectomy[46]. Some definitions are well accepted, for example, PSA response in hormone refractory disease[33] whereas other definitions are subject to debate such as the most appropriate cut-point to define PSA relapse after radical prostatectomy[46], the definition to use for defining biochemical failure after radical prostatectomy[47] or in advanced disease stages treated with hormonal therapy[31]. In addition, it is known that the studied treatments may affect PSA differently[48,49] and that this must be taken into account when defining any PSA-based endpoint.

Misspecification of the target endpoint may lead to uninformative clinical trials. For example assume a new drug that does not affect PSA in hormone refractory disease and this fact is not known in advance. Specification of the trial primary endpoint in terms of PSA decline or PSA progression will induce negative trial results. Furthermore guidance to stop therapy in case of PSA progression will lead to early treatment termination in most patients so that no other endpoint than the PSA endpoint will be assessable in the study. Scher[48] also demonstrated that the definition of the PSA progression itself may impact on the trial results.

1.6 Heterogeneity of treatment results

1.6.1 Patient selection, impact of prognostic factors

Although a careful definition of the trial endpoints is crucial to the design and success of clinical trials, patient selection to trials is of importance as well. Indeed, heterogeneity of the patient groups entered into clinical trials and of their responsiveness to the tested treatment modalities can result in increased variability in the estimation treatment effect. In addition, when comparing two or more treatment modalities in the frame of a randomized clinical trial, lack of balance of important prognostic factors between treatment groups may result in severe bias since difference in patient prognosis between groups will be

confounded with the treatment effect and may either artificially dilute or inflate it, thus leading to inflation of the type II or type I trial error rates, respectively.

1.6.2 Randomization and stratification

Randomization into clinical trials is used to correct for such imbalances. Randomization per se can however only achieve balance on average and for large patient numbers. In practice, there always remains a small risk that with randomization alone, some imbalance in important prognostic factors may remain by the play of chance alone. The technique known as stratification that can be implemented into trials either as stratified (block) randomization[50] or by the minimization technique[51] forces the randomization to balance specified characteristics of the patient groups across the randomized treatments. The identification of prognostic factors for the endpoint being studied is thus extremely important because stratification of the randomized trials for these factors will reduce the risk of bias in the results and increase the statistical power of comparisons by reduction of the model variability[52]. Besides, the identification of prognostic factors of long term outcome is useful to the clinician for patient counseling and treatment decisions as well as to the patient himself [53].

1.6.3 Heterogeneity due to predictive factors

Besides heterogeneity of the patient population, the treatment results themselves may also be heterogeneous across groups of patients characterized by *predictive* factors.

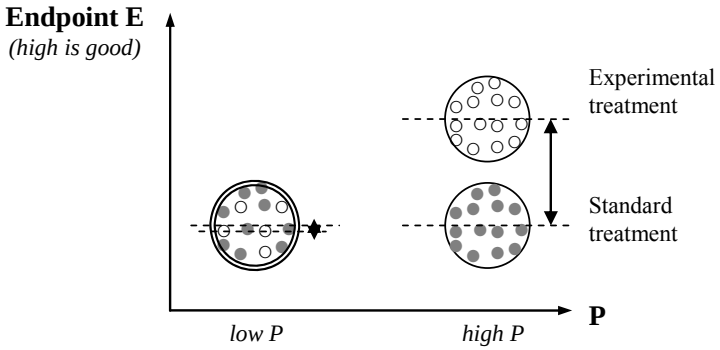
Predictive factors, unlike prognostic factors, may or may not be relevant to the assessment of a patient fate in absence of treatment but are modulators of the patient sensitivity or responsiveness to a given treatment (Figure 1.3). Predictive factors are factors that modulate the magnitude of the expected treatment benefit. Typically, in a clinical trial, a predictive factor will be characterized by a significant interaction between the predictive factor and the treatment effect.

The identification of predictive factors of treatment benefit is important because the applicability of the overall treatment benefit observed in a clinical trial may be limited whenever a significant interaction between a predictive classification of the patients and the treatment is present. Indeed, in such instances, the treatment benefit may be less or even not clinically relevant in some subgroups of patients whereas it may be more striking in others, so that the average trial results are not applicable to the overall patient group but should be described separately in the groups defined by the predictive factor.

For planning future studies, the knowledge of predictive factors may be helpful to defining appropriate selection criteria. By selecting the group of patients that is most sensitive to the experimental treatment, one increases the likelihood of finding significant treatment

benefits whilst sparing unnecessary treatment to patients that are less or not responsive to that treatment.

Figure 1.3
A predictive factor



The factor P is predictive for a benefit from the new treatment:
Of the patients randomized to the new treatment (white dots), only those with a high value of P demonstrate an increased response in terms the endpoint E in comparison to the standard treatment (grey dots)

1.6.4 An example

To illustrate the difference between the notions of prognostic and of predictive factor, we shall refer again to the EORTC trial 22911. The trial showed a significant benefit for clinical progression-free survival (P=0.0009, HR=0.61, 98%CI: 0.43-0.87).

We identified earlier baseline PSA as a significant prognostic factor for clinical progression-free survival in the standard treatment group. Without post-operative irradiation, patients with pre-operative PSA>20 ng/ml have about double the risk of clinical progression than the patients with pre-operative PSA<20 ng/ml (HR=1.93, 95%CI: 1.32, 2.83, P=0.0006).

For pre-operative PSA to be a predictive factor, we would further require that the magnitude of the treatment benefit be different in the group with low pre-operative PSA than in the group with high pre-operative PSA. In this study, there was borderline statistically significant evidence of an interaction between the pre-operative PSA level and the magnitude of the treatment benefit (interaction test P-value=0.08). The magnitude of the treatment benefit seemed to be less in the group with pre-operative PSA<20 ng/ml than in the group with elevated pre-operative PSA. In the group with PSA<20 ng/ml, the hazard ratio for post-operative irradiation is 0.69 (95% CI: 0.48-0.99) with P=0.042 indicating a moderate benefit from irradiation. In the group with PSA>20 ng/ml, the hazard ratio is 0.39 (95%CI: 0.24 – 0.48) with P=0.00023, indicating a larger benefit in that subgroup. Thus in this study, there is some evidence

that pre-operative PSA level be also predictive of treatment benefit in terms of clinical progression-free survival.

1.7 Rationale and outline of this thesis

Prostate cancer is a significant health problem in most industrialized Western countries where it is the most commonly diagnosed cancer in men after middle age. It is also a leading cause of death from cancer among men diagnosed with cancer in the US and in Europe.

Prostate cancer is a relatively slow growing disease and patients diagnosed early often benefit from long survival time. Phase III clinical trials in this disease that use long term clinical endpoints such as progression/disease-free survival or cancer-specific or overall survival often require a very long follow-up. This induces a long time-lag between the formulation of a scientific therapeutic hypothesis and its resolution by level I evidence from randomized trials. Additionally, prostate cancer is a rather heterogeneous disease and many prognostic factors (such as Gleason score, number of positive biopsies ...) severely affect the natural course of the disease. Some factors that relate to the patient prognosis (e.g. for instance pathologic stage after prostatectomy) may also act as predictive factors for the benefit of new treatments, thus inducing extra heterogeneity into clinical trials.

In this thesis we argue that in order to improve the efficiency of clinical trials in prostate cancer, a thorough understanding of *prognostic factors* of outcome is needed, as well as an exploration of potential *predictive factors* that might affect treatment benefit and cause heterogeneity of results. In addition, the development of *surrogate endpoints* of the long term clinical outcome is needed in order to speed up the drug development process by reducing the observation time required to assess the final endpoint.

Prostate-Specific Antigen (PSA) is the most widely studied marker for prostate cancer diagnosis and prognosis. A wide variety of definitions of PSA endpoints have been studied as prognostic factors but few were assessed for their true value as surrogate endpoints of long term clinical benefit. We argue that PSA should only be used as the primary endpoint in clinical trials after successful statistical validation of its quality of surrogate endpoint.

We maintain that irrespective of the endpoint chosen in a clinical trial, good statistical practice is needed in the analysis and presentation of results, in order not to inflate the risk of publishing spuriously significant or non significant results. We also recommend that lessons from past experiences be taken by studying in detail the reasons why, apart from genuine treatment ineffectiveness, some clinical trials failed to demonstrate significant benefit.

We utilize existing clinical trial data from past studies, most of which involve endocrine treatments against prostate cancer, to challenge a number of pathways of improvement of the efficiency of phase III clinical trial designs in prostate cancer:

In **Chapter 2**, we study various definitions of time to PSA progression as prognostic factors for overall survival in a phase III clinical trial that compared the anti-androgens Flutamide and Cyproterone Acetate in a population of good prognosis metastatic prostate cancer patients (EORTC 30892). We assess the sensitivity and specificity of five alternative definitions of PSA-endpoints for predicting survival at five years.

In **Chapter 3**, we use the meta-analytic method to study the value of several PSA-based endpoints as potential surrogates for overall survival in a database of over 2000 metastatic prostate cancer patients treated in AstraZeneca's Casodex Development Program. We studied PSA response defined as a 50% decline of PSA from baseline level, PSA normalization, two definitions of PSA including the one that was shown to be the most powerful prognostic factor in chapter 2, as well as the complete series of PSA measurements which is thought to contain maximal information from the marker.

In **Chapter 4**, we consider a phase III clinical trial in locally advanced prostate cancer that compares a group that is treated exclusively with external beam irradiation and does not receive endocrine treatment and a group that received immediate adjuvant hormonal treatment (EORTC 22863). We assess prognostic factors for survival and we explore the heterogeneity of the treatment results across defined subgroups of patients (predictive groups).

In **Chapter 5**, we consider a more recent phase III trial in patients with known pathological risk factors after radical prostatectomy (pT3) that compared immediate adjuvant irradiation to no immediate adjuvant treatment (EORTC 22911) This trial is especially interesting because the use of PSA as a marker of disease progression became widespread during the course of the trial. The primary trial endpoint was thus amended to incorporate PSA changes ("biochemical failure") as an event into the primary endpoint. We study prognostic factors of the disease history after radical prostatectomy alone and we assess the value of this classification as a predictive classification by assessing the heterogeneity of the treatment benefit across groups of patients defined by the risk factors of disease progression determined from the control arm of the study. We also investigate the heterogeneity of the results according to the pathological risk factors that were stratified for at the time of entry on study here regarded as putative predictive factors, in order to evaluate if all subgroups benefited from adjuvant irradiation.

In **Chapter 6**, we turn to the (bone) metastatic hormone refractory patient population. This population represents the most advanced category of prostate cancer patients. Until recently, palliation of symptoms was the primary goal of the treatment in this patient group. We pooled the data from three EORTC studies that were carried out in the 1990's and that assessed health related quality of life (HRQOL) in addition to clinical endpoints. We use the baseline HRQOL data together with previously identified clinical and biochemical parameters to identify independent prognostic factors for overall survival. The precision of the predictions from the resulting prognostic model is compared to that of models based on clinical and biochemical factors only in order to assess the value of HRQOL parameters at improving the quality of the predictions.

In **Chapter 7**, we revisit the trials that compared maximal androgen blockade to castration in metastatic prostate cancer. We investigate the reasons why the trials failed to meet their goals. We review several aspects of the design and of the statistical analysis of those trials that may have limited their ability to demonstrate a survival benefit.

In **Chapter 8**, the general discussion, we critically review the results presented in chapters 2-7 in terms of statistical requirements and of clinical relevance and in the light of evidence from other existing publications in this area. In particular, we discuss the methodology of surrogate endpoint validation and its practical limitations. We also address the issue of the transportability of surrogacy of an endpoint across disease stages and across treatment modalities. Finally, we derive some recommendations for the use of PSA as endpoint and for the design of future trials in prostate cancer. We also delineate some avenues for future research.

1.8 References

1. F.H. Schroeder, K.H. Kurth, S.D. Fossa, W. Hoekstra, P.P.M. Karthaus, M. Debois, L. Collette and members of the EORTC GU Group. Early versus delayed endocrine treatment in pN1-3M0 prostate cancer without local treatment – Results of EORTC 30846 – a phase III study. J Urol 2004: 172, 923–927.
2. Prostate Cancer Trialists' Collaborative Group. Maximum androgen blockade in advanced prostate cancer: an overview of the randomized trials. Lancet 2000; 3555:1491-8.
3. Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC. Natural history of progression after PSA elevation following radical prostatectomy. JAMA 1999; 282(17):1591.
4. Morris MJ, Scher HI. Novel therapies for the treatment of prostate cancer: current clinical trials and development strategies. Surg Oncol 2002; 11: 13-23.
5. Oh WK. The evolving role of chemotherapy and other systemic therapies for managing localized prostate cancer. J Urol 2003; 170: S28-32; discussion S33-4.
6. van der Poel HG: Smart drugs in prostate cancer. Eur Urol 2004; 45: 1-17.
7. Kelloff GJ, Coffey DS, Chabner BA, Dicker AP, Guyton KZ, Nisen PD, Soule HR and D'Amico AV. Prostate-specific antigen doubling time as a surrogate marker for evaluation of oncologic drugs to treat prostate cancer. Clin Cancer Res. 2004;10(11):3927-3
8. Clinton B, Gore A. Reinventing the regulation of cancer drugs. Accelerating approval and expanding access. National performance review. Washington, DC: Food and Drug Administration, 1996; 1-12.
9. Johnson JR, Williams G, Pazdur R. Endpoints and United States Food and Drugs Administration Approval of Oncology Drugs. J Clin Oncol 2003; 21(7): 1401-1411.
10. Lesko LJ, Atkinson AJ. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. Annu Rev Pharmacol Toxicol 2001;41:347-66.
11. Biomarkers Definitions Working Group: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework Clin Pharmacol Ther 2001;69:89-95

12. Polascik TJ, Oesterling JE, Partin AW. Prostate specific antigen: a decade of discovery – what we have leraned and were we are going. J Urol 1999; 162: 293-306.
13. Small EJ, Mack Roach III. Prostate-specific antigen in prostate cancer: a case study in the development of a tumor marker to monitor recurrence and assess response. Seminars in Oncology 2002; 19(3): 264-273.
14. Brawer M.K. Tumor markers in International Handbood of Prostate Cancer (eg. Kirk D). Euromed Communications Ltd., Haslemere, Surrey, England, 1999.
15. Stamey TA, Yang N, Hay A, McNeal JE, Freiha FS and Redwine E. Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate. New Engl J Med 1987; 317: 909-916.
16. Stamey TA, Kabalin JN, McNeal JE, Jonhnstone IM, Freiha F, Rdwine EA and Yang N. Prostate-specific antigen in the diagnosis and treatment of adenocarcinoma of the prostate II: radical prostatectomy treated patients. J Urol 1989;141: 1076-1083.
17. Cleutjens KB, van Eekelen CC, van der Kroput HA, Brinkmann AO, Trapman J. Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter. J Biol Chem 1996; 271(11): 6379-88.
18. Riegman PH, Vlietstra RJ, van der Korput JA, Brinkmann AO, Trapman J. The promoter of the prostate-specific antigen gene contains a functional androgen responsive element. Mol Endocrinol 1991; 5(12):1921-30.
19. Robool MJ and Schröder FH. European Randomized Study of Screening for Prostate Cancer: achievements and presentation. BJU Int 2003, suppl. 92: 88.
20. Partin AW, Kattan MW, Subong ENP, Walsh PC, Wojno KJ, Oesterling JE, Scardino PT and Pearson JD. Combination of prostate-specific antigen, clinical stage and Gleason score to predict pathological stage of localized prostate cancer: a multi-institutional update. JAMA 1997; 277: 1445-1451.
21. American Society for Therapeutic Radiology and Oncology Consensus Panel. Consensus statement: guidelines for PSA following radiation therapy. Int J Rad Oncol Biol Phys 1997; 37: 1035-1041.
22. Newling D, Denis L and Vermeylen K, Orchiectomy versus Goserelin and Flutamide and the treatment of newly diagnosed metastatic prostate cancer. Cancer 1993; 72 (12): 3793-98
23. Cannon GM, Walsh PC, Partin AW, Pound CR. Prostate-specific antigen doubling time in the identification of patients at risk for progression after treatment and biochemical recurrence for prostate cancer. Urology 2003; 62(suppl. 6B): 2-8.
24. D'Amico AV, Moul JW, Carroll PR, Sun L, Lubeck D, Chen MH. Surrogate endpoint for prostate cancer specific mortality after radical prostatectomy or radiation therapy. J Natl Cancer Inst. 2003; 95(18): 1376-83
25. D'Amico AV, Moul JW, Carroll PR, Cote K, Sun L, Lubeck D, Renshaw AA, Loffredo M, Chen MH. Intermediate end point for prostate cancer-specific mortality following salvage hormonal therapy for prostate-specific antigen failure. J Natl Cancer Inst. 2004;96(7):509-15.
26. Ornstein DK, Pruthi RS. Prostate-specific antigen. Expert Opin Pharmacother 2000;1: 1399-1411.
27. Schroeder FH. Evidence in advanced disease: the challenge of new therapies. Eur Urol 2002;S1: 11-16.

28. Sridhara R, Eisenberger MA, Sinibaldi VJ, Reyno LM, Egorin MJ. Evaluation of prostate-specific antigen as a surrogate marker for response of hormone-refractory prostate cancer to suramin therapy. J Clin Oncol 1995;13:2944-2953.
29. Smith DC, Dunn RL, Strawderman MS, Pienta KJ. Change in serum prostate-specific antigen as a marker of response to cytotoxic therapy for hormone-refractory prostate cancer. J Clin Oncol 1998;16:1835-1843.
30. Kelly WK, Scher HI, Mazumdar M, Vlamis V, Schwartz M, Fossa SD. Prostate-specific antigen as a measure of disease outcome in metastatic hormone-refractory prostate cancer. J Clin Oncol 1993;11:607-615.
31. Scher HI, Kelly WK, Zhang ZF, Ouyang P, Sun M, Schwartz M, Ding C, Wang W, Horak ID, Kremer AB. Post-therapy serum prostate-specific antigen level and survival in patients with androgen-independent prostate cancer. J Natl Cancer Inst 1999; 91:244-251.
32. Buyse M, Vangeneugden T, Bijnens L, Renard, D., Burzykowski, T., Geys, H., and Molenberghs, G. Validation of Biomarkers ad Surrogates for Clinical Endpoints. In Biomarkers in Clinical Drug Development (Bloom JC, ed) New-York: Springer-Verlag, 2003. p. 149-168.
33. Bubley GJ, Carducci M, Dahut W, Dawson N, Daliani D, Eisenberger M, Figg WD, Freidlin B, Halabi S, Hudes G, Hussain M, Kaplan R, Myers C, Oh W, Petrylak DP, Reed E, Roth B, Sartor O, Scher H, Simons J, Sinibaldi V, Small EJ, Smith MR, Trump DL, Wilding G: Eligibility and response guidelines for phase II clinical trials in androgen-independent prostate cancer: recommendations from the Prostate-Specific Antigen Working Group. J Clin Oncol 1999;17:3461-3467.
34. Valicenti R, Desilvio M, Hanks G, Porter A, Brereton H, Shipley W, Sandler H. Surrogate Endpoint for Prostate Cancer-specific Survival: Validation from an Analysis of Radiation Therapy Oncology Group Protocol 92-02. Proc Am Soc Clin Oncol 2005. J Clin Oncol 2005; 23: 390s,abstract 4549.
35. D'Amico AV, Moul J, Carroll PR, Sun L, Lubeck D, Chen MH. Surrogate endpoint for prostate cancer specific mortality in patients with nonmetastatic hormone refractory disease. J Urol 2005; 173: 1572-6.
36. Crawford ED, Pauler DK, Tangen CM, Hussain MHA, Small EJ, Taplin EJ, Burch PA, Greene GF, Lara P, Petrylak DP. Three-month change in PSA as a surrogate endpoint for mortality in advanced hormone-refractory prostate cancer (HRPC): Data from Southwest Oncology Group Study S9916. Proc Am Soc Clin Oncol 2004; 22: 14S, abstract 4505.
37. Prentice RL. Surrogate endpoints in clinical trials: definitions and operational criteria. Stat Med. 1989;8:431-440.
38. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med. 1992; 11:167-178.
39. Buyse M, Molenberghs G. Criteria for the validation of surrogate end-points in randomized experiments. Biometrics. 1998;54: 1014-29.
40. Buyse M, Molenberghs G, Burzykowski T, Renard D and Geys H. Statistical validation of surrogate endpoints: problems and proposals. Drug Inf J. 2000; 34: 447-454.
41. Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials.. Control Clin Trials. 2002; 23: 607-625.
42. Daniels MJ, Hugues MD. Meta-analysis for the evaluation of potential surrogate markers. Stat Med. 1997; 16: 1515-27.

43. Buyse M, Molenberghs G, Burzykowski T, Renard D and Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000;1: 49-68.
44. Gail MH, Pfeiffer R, Van Houwelingen HC, Carrol RJ. On meta-analytic assessment of surrogate outcomes. Biostatistics. 2000 ; 1(3):231-46.
45. Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. J Roy Stat Soci Series C (Appl Stat). 2001;50: 405-22.
46. Amling CL, Bergstralh EJ, Blute ML, Slezak JM, Zincke H. Defining prostate specific antigen progression after radical prostatectomy: what is the most appropriate cut point? J Urol 2001; 165: 1146-1151.
47. Thames H, Kuban D, Levy L, Horwitz EM, Kupelian P, Martinez A, Michalski J, Pisansky T, Sandler H, Shipley W, Zelefsky M and Zietman A. Comparison of alternative biochemical failure definitions based on clinical outcome in 4839 prostate cancer patients treated by external beam radiotherapy between 1986 and 1995. Int J Radiation Oncol Biol Phys 2003; 57(4): 929-943.
48. Scher HI, Eisenberger M, D'Amico AV, Halabi S, Small EJ, Morris M, Kattan MW, Roach M, Kantoff P, Pienta KJ, Carducci MA, Agus D, Slovin SF, Heller G, Kelly WK, Lange PH, Petrylak D, Berg W, Higano C, Wilding G, Moul JW, Partin AN, Logothetis C, Soule HR. Eligibility and outcomes reporting guidelines for clinical trials for patients in the state of a rising prostate-specific antigen: recommendations from the Prostate-Specific Antigen Working Group. J Clin Oncol. 2004;22(3):537-56 (with erratum in J Clin Oncol. 2004;22(15):3205)
49. Dixon SC, Knopf KB, Figg WD. The control of prostate-specific antigen expression and gene regulation by pharmacological agents. Pharmacol Rev 2001;53: 73-91.
50. Palta M. Investigating maximum power losses in survival studies with non stratified randomization. Biometrics 1985; 41: 497-504.
51. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. Biometrics 1975;31:103-15.
52. Korn EL, Simon R. Measures of explained variation for survival data. Stat Med 9: 487-503, 1990.
53. Weeks JC, Cook EF, O'Day SJ, Peterson LM, Wenger N, Reding D, Harrell FE, Kussin P, Dawson NV, Connors AF Jr, Lynn J, Phillips RS. Relationship between cancer patients' predictions of prognosis and their treatment preferences. JAMA 279:1709-1714, 1998.

Chapter 2:

Is Prostate specific antigen a prognostic factor for overall survival in good prognosis metastatic prostate cancer?

2.1 Abstract

Background: We study the value of PSA response and PSA progression as prognostic factors for survival in good prognosis metastatic prostate cancer.
Methods: Data from 257 patients treated with Flutamide or Cyproterone acetate within the EORTC GU Group protocol 30892 have been used and analysis by Cox models.
Results: A PSA response defined as a decrease to $\underline{\leq}$1 ng/ml and to between 1 and 10 ng/ml were associated with a hazard ratio of 0.30 and 0.61 for overall survival, respectively, as compared to the non responders (PSA>10 ng/ml). Five definitions of PSA progression were considered: i) a confirmed or ii) a repeated doubling of the PSA over nadir and unconfirmed iii) 100%, iv) 50% and v) 20% increase of the PSA over nadir, each to a value > 4 ng/ml. Definition 5 was the most sensitive with sensitivity 76.20% and specificity 32.08%. With this definition, 70.0% of the patients had a PSA progression, which occurred in median 1.98 years before death.
Conclusion: For good prognosis metastatic prostate cancer patients under anti-androgen treatment, PSA response at 6 months with cut-off levels of ≤1 ng/ml and ≤10 ng/ml is prognostic for survival. A 20% increase over nadir to a value > 4 ng/ml is prognostic for a poor survival with a 76.20% sensitivity. In this study, confirmation of the increase by a second observation did not seem necessary. Genuine surrogacy is not established in this study.

2.2 Introduction

Clinical trials in prostate cancer are difficult undertakings for several reasons: even metastatic prostate cancer is relatively slow growing, requiring long observation times. Survival as an endpoint is diluted by relatively high proportions of men dying of intercurrent disease, depending on the extent of the disease at presentation. Phase II studies are even more troublesome because of the lack of measurable lesions in most patients. For less advanced disease, e.g. rising PSA after radical treatment, exceedingly long observation times are necessary, reaching as much of 13 years after radical prostatectomy[1,2].
With increasing knowledge of the genetic background of cancer and cancer progression and of the mechanisms of signal transduction, a large number of new compounds are under development[3]. To fasten new treatment development, surrogate markers are needed to shorten the duration of phase II and III trials. The conditions for the use of such markers will have to be carefully defined as was recently attempted for hormone unresponsive PSA progression[4-6].

In the present report we use data of the EORTC GU Group protocol 30892, a phase III study in patients with previously untreated metastatic prostate cancer and favorable prognostic factors that compared Flutamide 250 mg t.i.d. to Cyproterone acetate100 mg t.i.d. This study was subject to 3 previous reports[7-9]. In the present report, we attempt to establish the value PSA response in the first 6 months after entry on study and PSA progression under treatment as predictors for overall survival. Disease-specific survival is used as secondary endpoint, for description mainly since this endpoint is affected by competing risks of death.

2.3 Material and Methods

Trial design

This randomized trial comparing endocrine treatment with Flutamide 250 mg t.i.d. or Cyproterone acetate 100 mg t.i.d. in patients with painless metastatic prostate cancer with favorable prognostic factors recruited 310 patients between 1990 and 1996[9]. 310 patients were randomized. "Good prognosis" was defined according to the EORTC classification[10] as the presence of two out of three criteria: WHO performance status 0, alkaline phosphatase $\leq$ Normal, T category <T4. Patients with painful metastases were also excluded. The allocated treatment was given until disease progression.
Progression was either subjective or objective progression (new soft tissue or new symptomatic bone metastases). The former required the observation of two of three parameters: increase of acid or alkaline phosphatase or PSA to $\geq$ 2.5×Normal; pain increase by 2 scores; or worsening of performance status by 2 scores. Treatment after progression was left at the investigator's discretion. Most patients had some form of castration as second line treatment.

The follow-up schedule is presented in Table 2.1. According to this schedule, PSA was measured until the patient stopped protocol treatment, at each participating institution. For the purpose of this analysis, the PSA measurements have been standardized to a reference upper normal limit of 4 ng/ml:
($PSA_{ana}=4\times PSA_{institution}/Upper\_normal\_limit\_PSA_{institution}$).

2.3.1 Statistical methods

The main endpoint is overall survival. Cox proportional hazard regression models stratified for randomized treatment were used in all analyses. Statistical significance was claimed at the 0.05 level. PSA progression under hormonal treatment was modeled as a time dependent variable. A landmark period of 6 months was applied to the analysis of PSA response to prevent lead time bias[11]. The duration of the Landmark was decided upfront on the basis of the frequency of follow-up.

Specificity and sensitivity of PSA progression under hormonal treatment as predictor for survival outcome were estimated on the basis of the first 5-year follow-up from entry on study.

Table 2.1: Follow-up schedule

Baseline examinations	Histological or cytological proof of prostate cancer Physical examination and CT-scan Blood tests including PSA Chest X-ray, Bone scan and X-rays of areas with hot spots Optional prostate, kidney and liver ultrasound
Follow-up examinations	
At 6 weeks, 3 months and then every 3 months in the first 2 years, then every 6 months until death	Physical examination Blood tests including PSA (until the end of the treatment) Chest X-ray (once yearly)
In case of elevated markers or symptoms	Bone scan
In case of symptoms or suspicious signs of recurrence	Other imaging (Ultrasounds, X-rays, etc…)

2.3.2 Patient characteristics

310 patients were randomized in this study, 156 to Flutamide and 154 to Cyproterone acetate. Their characteristics are displayed in table 2.2 and were well balanced between the two treatment groups (data not shown), except for age and presence of visceral metastases. At the time of the most recent publication[9], 205 patients (66%) had died after a median follow-up of 5.1 years, 135 (66%) due to prostate cancer. The median survival time was 3.2 years. Time to progression, disease-specific and overall survival were similar in the two groups, irrespective of adjustment for imbalances in baseline characteristics. On average, 8 PSA measurements were available per patient (range: 1-26).

Fifty-three of 310 patients had to be excluded from all present analyses (for the reasons of exclusion, see footnote to table 2). The landmark excluded another 5 patients from the PSA response analysis.. As seen in table 2, the 58 excluded patients have similar characteristics as the other patients, except for the initial PSA which tends to be higher in the study group (median: 64.5 ng/ml vs 39.7 ng/ml, P=0.175).

Table 2.2: Baseline characteristics and reasons for exclusion from the analysis

	Excluded from The analysis* (N=58)	**Included In the analysis (N=252)**
	N (%)	N (%)
Treatment		
Flutamide	26 (44.8)	128 (50.8)
CPA	32 (55.2)	124 (49.2)
Age		
Median (range)	74.8 (59.7-84.3)	71.3 (48.9-85.7)
Performance status		
WHO 0	51 (87.9)	224 (88.9)
WHO 1	7 (12.1)	28 (11.1)
T category		
T0	3 (5.2)	4 (1.6)
T1	1 (1.7)	9 (3.6)
T2	9 (15.5)	58 (23.0)
T3	39 (67.2)	158 (62.7)
T4	3 (5.2)	23 (9.1)
TX	3 (5.1)	0 (0.0)
N category (Nodal Status assessed by physical examinination and CT-scan)		
N0	18 (31.0)	110 (43.7)
N+	7 (10.3)	27 (10.7)
NX	31 (56.9)	115 (45.6)
M category		
M0	0 (0.0)	2 (0.7)
M1	58 (100)	250 (99.3)
WHO histopathological grade (G)		
G0	0 (0.0)	1 (0.4)
G1	6 (10.3)	46 (18.3)
G2	25 (43.1)	121 (48.0)
G3	24 (41.4)	82 (32.5)
GX	3 (5.1)	2 (0.8)
Visceral metastases present	3 (5.2)	3 (1.2)
Number of hot spots: median (range)	4 (0-11)	4 (0-11)
Serum creatinine		
Missing	3 (5.2)	12 (4.8)
<=Upper Normal Limit	41 (70.7)	173 (68.7)
> Upper Normal Limit	14 (24.1)	67 (26.6)

	Excluded from The analysis* (N=58) N (%)	**Included In the analysis (N=252) N (%)**
Alkaline phosphatase		
Missing	4 (6.9)	11 (4.4)
<=2.5x Upper Normal Limit	48 (82.8)	226 (89.7)
>2.5x Upper Normal Limit	6 (10.3)	15 (6.0)
Prostate Acid Phosphatase		
Missing	27 (46.6)	98 (38.9)
<=5x Upper Normal Limit	21 (36.2)	83 (32.5)
>5x Upper Normal Limit	10 (17.2)	72 (28.5)
PSA at entry (ng/ml)		
Missing	10 (17.2)	0 (0.0)
Median	39.7	64.5
(Range)	(2-2104)	(1.1-3350)

* The reasons for exclusion from the analysis were: (1) no baseline PSA available (10 patients); (2) Less than 2 follow-up PSA measurements (43 patients of whom 6 died due to prostate cancer and 5 of other cause during the first 6 months on study, 4 stopped treatment due to early progression of the disease and for 28 the reason for missingness of the PSA levels is unrelated to the evolution of the disease); (3) excluded by the landmark used in the analysis of PSA response(5 patients)

2.4 Results

PSA response as a prognostic factor for survival.

PSA response was determined by the lowest PSA value measured during the 6-month landmark period. Three cut-off values of response were initially considered: ≤ 1 ng/ml, ≤ 4 ng/ml and ≤ 10 ng/ml. For each cut-off the date of PSA response was the first date when the PSA was below the cut-off level. The response had to be confirmed by a subsequent observation below or equal to the cut-off value. The patients were then classified according to the best response achieved: ≤ 1 ng/ml, above 1 but ≤ 4 ng/ml, above 4 but ≤10 ng/ml, non reponse (>10 ng/ml)

Fifty-one of 252 patients (20.2%) met the criteria of PSA response at the ≤1 ng/ml cut-off, another 60 (23.8%) had a confirmed PSA decline to a value between 1 and 4 ng/ml, and 53 (21.0%) met the ≤10 ng/ml cut-off criterion. The remaining 88 (34.9%) patients did not match any of the PSA response criteria within the Landmark period.

The Cox model (Table 2.3) indicates that a PSA response at the ≤1 ng/ml level is associated with a death hazard ratio of 0.30 (95% CI: 0.18 – 0.49; P<0.0001) relative to the group without PSA response (PSA>10 ng/ml). There was no difference between the groups having a PSA nadir between 1 and 4 ng/ml or between 4 and 10 ng/ml. A response at any level between 1 and 10 ng/ml was associated with a hazard ratio of 0.61

(95% CI: 0.44 – 0.86, P=0.004) relative to those group without PSA response. These conclusions remain unchanged in a multivariate model accounting for baseline factors such as age, number of hot spots, initial PSA, histological grade or alkaline phosphatase level.

Table 2.3: PSA decrease within 6 months of entry (PSA response) as prognostic factor for overall survival

PSA response	N patients / total (%)	HR[a] (95% CI[b])	Median (95% CI[b])	P-value	P-value multivariate
≤ 1 ng/ml	51 / 252 (20.2)	0.30 (0.18 – 0.49)	6.4 years (5.6 – 8.4)	<0.0001	<0.0001
1-10 ng/ml	113 / 252 (44.8)	0.61 (0.44 – 0.86)	3.8 years (3.0 – 4.7)	0.004	0.012
> 10 ng/ml	88 / 252 (34.9)	1.00	2.3 years (1.9 – 3.3)	-	-

[a]HR stands for hazard ratio, a value > 1 favours the reference category "no response: PSA>10 ng/ml";

[b]CI stands for confidence interval

Figure 2.1 displays overall survival by PSA response. The median survival was 6.4 years for patients who reached a nadir PSA ≤ 1 ng/ml within 6 months after entry on study , 3.8 years for those who achieved a nadir PSA between 1 and 10 ng/ml and 2.3 years for those whose PSA remained above 10 ng/Ml.

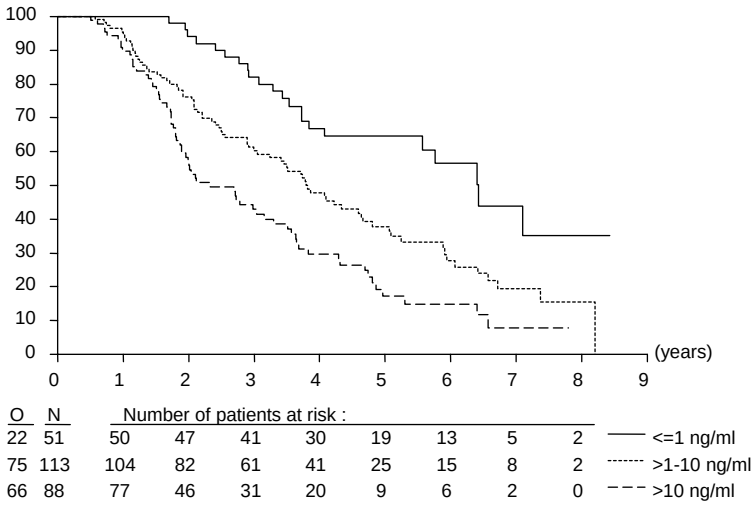
PSA progression under hormonal treatment as a prognostic factor for survival.

Five definitions of PSA progression were considered.

1) A confirmed one hundred percent increase over nadir value to > 4 ng/ml
2) Two successive doublings over nadir value to > 4 ng/ml
3) A first 100% increase over nadir value to > 4 ng/ml
4) A first 50% increase over nadir value to > 4 ng/ml
5) A first 20% increase over nadir value to > 4 ng/ml

Definitions 3, 4 and 5 do not require confirmation of the PSA elevation at a subsequent measurement Definition 3 is the unconfirmed variant of definition 1.

Figure 2.1
Overall survival by PSA response using a Landmark of 6 months.



O represents the number of events (deaths), N represents the total number of patients. The survival curves are estimated using the Kaplan-Meier technique. Patients are classified according to the lowest observed PSA value under treatment.

The results of the time-dependent Cox models are presented in Table 2.4. The hazard ratio represents the increase in the risk of dying at any time t after PSA progression for those patients with PSA progression in comparison to the risk of dying at the same time t for the patients without PSA progression. The median time to PSA progression and median survival after PSA progression have been estimated by the Kaplan-Meier technique.

The sensitivity and specificity for predicting survival at 5 years from entry on study are reported in table 2.5 for overall survival (a) and for disease-specific survival (b). As seen in Table 4, all five definitions result in a hazard ratio of 3 or more. Definition 2 is the most stringent of all, as it requires a second doubling to confirm PSA progression and only 44.0% of the patients match that criterion of PSA progression. Definition 2 is therefore the most specific, but its ability to predict death due to prostate cancer is the lowest (66%). The other definition requiring confirmation (definition 1) is somewhat more sensitive but the sensitivity for overall survival remains 10% less than the unconfirmed variant (definition 3).

Table 2.4: PSA progression as prognostic factor for overall survival

Definition of PSA progression	N with PSA progression / N total (%)	HR (95% CI) On overall survival	P-value	Median time to PSA progression (in years with 95% CI)	Median time to death after PSA progression (in years with 95% CI)
(1) Confirmed 100% increase over nadir to > 4 ng / ml	132/257 (51.4)	2.94 (2.12-4.08)	<0.0001	3.04 (2.51-3.93)	1.84 (1.46-2.15)
(2) Two successive 100% increase over nadir to > 4 ng/ml	113/257 (44.0)	3.50 (2.54-4.82)	<0.0001	4.10 (3.52-5.14)	1.51 (1.23-1.90)
(3) A first 100% increase over nadir to > 4 ng/ml	161/257 (62.7)	3.32 (2.37-4.67)	<0.0001	2.07 (1.76-2.82)	1.86 (1.49-2.31)
(4) A first 50% increase over nadir to >4 ng/ml	170/257 (66.2)	3.30 (2.32-4.69)	<0.0001	1.81 (1.45-2.17)	1.88 (1.49-2.39)
(5) A first 20% increase over nadir to >4 ng/ml	180/257 (70.0)	3.71 (2.56-5.37)	<0.0001	1.42 (1.21-1.99)	1.98 (1.51-2.46)

Definitions 3, 4 and 5 are all associated with hazard ratios above 3.30, corresponding to a 3.3-fold or more increase in the risk of dying once the event of PSA progression has occurred in comparison to the risk in absence of a PSA progression. The hazard ratio is largest for definition 5 and the median time to PSA progression is shortest (1.48 years). The median survival time after PSA progression is about 1 month longer than with definitions 2 and 3. This because this definition is the least stringent of all (70% of the patient satisfied this definition) and therefore the event tends to occur earlier on in the course of the follow-up. Definition 5 is the most sensitive definition of PSA progression with sensitivity 76.19% for overall survival and 85.71% for specific survival. It is also the least specific of all definitions for both overall and specific survival.

Table 2.5: Specificity and Sensitivity of PSA progression as predictor of overall survival (a) and specific survival (b) at 5 years from entry on study

PSA progression	(a) 5-year overall survival		(b) 5-year Specific Survival	
	Specificity	**Sensitivity**	**Specificity**	**Sensitivity**
(1) Confirmed 100% increase over nadir to > 4 ng / ml	47.17	55.78	60.78	71.43
(2) Two successive 100% increase over nadir to > 4 ng/ml	60.38	51.02	69.61	66.33
(3) A first 100% increase over nadir to > 4 ng/ml	39.62	65.99	49.02	78.57
(4) A first 50% increase over nadir to >4 ng/ml	37.74	71.43	43.14	81.68
(5) A first 20% increase over nadir to >4 ng/ml	32.08	76.19	37.25	85.71

2.4.1 Discussion

The establishment of surrogate markers for survival endpoints would allow their use as primary endpoint in phase II and III clinical trials. For patients with metastatic disease and good prognostic factors, as selected to this protocol, the overall median time to death is 3.3 years. PSA progression (20%, 50% or100% increase over nadir value to > 4 ng/ml) occurred at a median close to 2 years before death and in median 2.07, 1.81 and 1.2 years, respectively after entry on study. It is evident that the trial duration could be shortened if using PSA progression as a surrogate endpoint. With a landmark of 6-months trials having

PSA response as primary endpoint would require a follow-up of only six months, while progression could be evaluated within 2 years.

The present study reveals a strong prognostic value of both PSA response and PSA progression for predicting overall survival in metastatic prostate cancer patients of good prognosis receiving anti-androgen monotherapy. Since 66% of the deaths were due to prostate cancer, the impact of intercurrent death is relatively moderate in this trial. Unfortunately, the study does not enable to establish the surrogacy of the intermediate endpoints because of the absence of a difference in survival between the treatment arms which prevents the establishment of a correlation between differences in the PSA endpoints with *differences* in survival between the treatment groups.

The information in current literature on the use of PSA as a longitudinal marker for response and progression is scarce and no attempts have been made to systematically explore the value of different definitions of response and progression. In this report, the choice for criteria of PSA response and progression follows those used in routine clinical practice and observations described in the current literature. A continuous model for the evaluation of PSA nadirs and definitions for PSA progression would be desirable, but is not achievable at this time. On this background the choices made will remain arbitrary to a certain degree.

PSA response as prognostic factor was studied by a large number of authors. Almost unanimously, significant correlation with survival and/or progression were found[12-16]. The definitions of PSA response varied between $\leq$ 1, $\leq$ 4, $\leq$ 10, 19% decrease and <10% of baseline. Landmark periods were either 3 or 6 months.

In our study a confirmed decrease to a value of $\leq$ 1 ng/ml within the first 6 months on study is associated with the best prognosis for survival and a confirmed decrease to a value between 1 and 10 ng/ml with a slightly less favorable prognosis. No difference in prognosis was seen between a decrease of PSA to $\leq$4 ng/ml and a decrease to values between 4 and 10 ng/ml. The risk of death for those who achieve a PSA response is 0.30 (PSA $\leq$ 1 ng/ml) to 0.61 (PSA $\leq$ 10 ng/ml) compared to those whose PSA remains > 10 ng/ml. The median survival time differs by about 2 years between those whose PSA decreases to $\leq$10 ng/ml and those whose PSA remains>10 ng/ml.

In another two arm randomized study[12] showing a statistically significant difference in time to progression and cancer specific survival between orchiectomy plus placebo and orchiectomy plus Anandron, a PSA nadir of < 2.5 ng/ml correlated with a longer progression free time (median 24 versus 17 months), and also with a longer time to cancer death (median 49 versus 28 months). Eisenberger et al[17] in protocol INT-105, comparing castration plus placebo to castration plus Flutamide did not find a difference in progression free nor in overall survival between the treatment arms despite a greater PSA response rate on maximal androgen blockade.

Five definitions for PSA progression were evaluated in our study and unconfirmed and confirmed PSA rises were compared (Data not shown). Of the five definitions that are shown in Table 5, a first increase of 20% over nadir value to a PSA higher than 4.0 ng/ml was the most sensitive. This definition was associated with a 3.71-fold increase in risk of death after PSA progression relative to the risk of men without PSA progression. Its specificity in predicting death due to prostate cancer was however below 40%. Being the less stringent definition, this is also the one according to which the PSA progression occurred earliest in the course of the disease. When considering the confirmed variants of definitions 3 to 5, the conclusions remained the same. A 20% increase over nadir (definition 5) remained associated the largest increase in the risk of death (3.0-fold), and with greatest sensitivity (61.9% for overall survival and 76.5% for specific survival). Its specificity was 55% for both endpoints. Reports in the current literature are confirmatory for these findings. Miller et al[18] in a study of 48 patients defined PSA progression as a rise above 4 ng/ml after normalization or a first PSA increase to any level if no normalization occurred. The median time to PSA progression was 7.3 months, a statistically predictive of clinical progression. Newling et al[19] found that in M1 disease an increase in PSA to a value above 10 ng/ml or a 100% increase compared to baseline if the initial PSA was > 10 ng/ml correlated significantly with survival. The relative risk for dying was 1.93, 50 of 57 patients had PSA progression as their first sign of progression, and the median survival after PSA progression was 52 weeks. Others[20] reported similar results.

The choice of the best criteria of progression depends in our view of the future use of the prognostic factor. If one is willing to consider that sensitivity is the most relevant characteristic of a prognostic marker with possible consequence to unnecessarily change treatment earlier in a number of patients, then, among our definitions, definition 5 is best. If however, one is willing to balance between sensitivity and specificity, so as to avoid unnecessary procedure, then confirmed definitions of PSA progression need to be considered at the cost of decreased sensitivity (60% or less for overall survival). It is also be reminded, that the value of a given definition of PSA response or progression as prognostic factor for overall survival, as tested in the present trial, remains valid for patient groups similar to the one considered in our study (metastatic patients of relatively good prognosis) and for classes of hormonal treatment that affect PSA levels in a similar way as Flutamide or CPA, the two anti-androgen monotherapies used in this trial. A generalization of the results to the group of metastatic prostate cancer patients with less favorable prognostic factors requires further validation on an independent dataset in an unselected metastatic population.

In interpreting PSA related data under endocrine treatment one has to be aware, that the expression of PSA in itself is an endocrine dependent process. However, as long as a sufficiently high correlation with major endpoints such as survival can be shown this aspect can probably be neglected. The possibility of PSA response without a clinical response, however, cannot be excluded. Also, some of the patients are likely to suffer from clinical progression, without the PSA ever matching any criteria of biological progression. This is a clear limitation to the use of PSA as surrogate endpoint. Further studies are needed to assess the true value of PSA as a surrogate and in particular the proportion of a treatment effect on

survival that PSA would be able to capture. New methods for validating surrogate endpoints, aside the Prentice Criteria, may be useful in this respect, but require multi-studies assessments[21].

2.5 References

1. Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC : Natural history of progression after PSA elevation following radical prostatectomy. JAMA 1999; 282(17):1591.
2. Schröder FH: Endocrine treatment of prostate cancer - recent developments and the future. Part 1: maximal androgen blockade, early versus delayed endocrine treatment and side-effects. Br J Urol 1999;83: 161-170.
3. Belldegrun A: What's hot in the prostate? Prostate Cancer and Prostatic Diseases 1999; 2:1-2.
4. Thalmann GN, Sikes RA , Chang S-M, Johnston DA, von Eschenbach AC, Chung LW: Suramin-induced decrease in prostate-specific antigen expression with no effect on tumor growth in the LNCaP model of human prostate cancer. J Nat Cancer Inst 1996 88(12): 794-801.
5. Bubley GJ, Carducci M, Dahut W, Dawson N, Daliani D, Eisenberger M, Figg WD, Freidlin B, Halabi S, Hudes G, Hussain M, Kaplan R, Myers C, Oh W, Petrylak DP, Reed E, Roth B, Sartor O, Scher H, Simons J, Sinibaldi V, Small EJ, Smith MR, Trump DL, Wilding G: Eligibility and response guidelines for phase II clinical trials in androgen-independent prostate cancer: recommendations from the Prostate-Specific Antigen Working Group. J Clin Oncol 1999;17:3461-3467.
6. Schröder FH, Kranse R, Barbet N, Hop WC, Kandra A, Lassus M: Prostate-Specific Antigen: a surrogate endpoint for screening new agents against prostate cancer? The Prostate 2000;42:107-115.
7. Schröder FH: Endocrine treatment of prostate cancer: The future after maximum androgen blockade and the role of PSA monitoring (Chapter 20). In: Renal, Bladder and Prostate Cancer - An Update. Proceedings of the V congress on "Progress and Controversies in Oncological Urology" (PACIOU V), Rotterdam, October 1998. Kurth KH, Mickisch GH, Schröder FH (eds). Parthenon Publishing Group. London, New York, 1999, pp. 173-181.
8. Schröder FH, Collette L, de Reijke TM**,** Whelan P: Prostate cancer treated by anti-androgens; is sexual function preserved? Brit J Cancer 2000;82(2):283-290.
9. Schröder FH, Whelan P, de Reijke ThM, Pavone-Macaluso M, Mattelaer J, Van Velthoven RFP, Collette L and members of the EORTC Genito-Urinary Group: Metastatic prostate cancer treated by Flutamide versus Cyproterone acetate. (EORTC 30892, Final analysis). Eur Urol 1999; 37 (suppl 2):127.
10. de Voogt HJ, Suciu S, Sylvester R, Pavone-Macaluso M, Smith PH, de Pauw M and members of the EORTC GU group: Multivariate analysis of prognostic factors in patients with advanced prostate cancer: results from two EORTC trials. J Urol 1989; 141: 883-8.
11. Anderson RJ, Cain KC and Gelber RD: Analysis of survival by tumor response, J Clin Oncol 1983;1(11): 710-19.
12. Dijkman GA, Janknegt RA, de Reijke ThM, Debruyne FM: Long-term efficacy and safety of nilutamide plus castration in advanced prostate cancer, and the significance of early prostate specific antigen normalization. International Anandron Study Group. J Urol 1997;158:160-163.
13. Oosterlinck W, Mattelaer J, Casselman J, Van Velthoven R, Derde MP, Kaufman L.: PSA evolution: a prognostic factor during treatment of advanced prostatic carcinoma with total androgen blockade. Data from a Belgian multicentric study of 546 patients. Acta Urol Belg 1997;65:63-71.

14. Reynard JM, Peters TJ, Gillatt D. Prostate-specific antigen and prognosis in patients with metastatic prostate cancer - a multivariate analysis of prostate cancer mortality. Brit J Urol 1995;75: 507-515.
15. Cooper EH, Armitage TG, Robinson MRG, Newling DW, Richards BR, Smith PH, Denis L, Sylvester R: Prostatic specific antigen and the prediction of prognosis in metastatic prostatic cancer. Cancer 1990;66:1025-1028.
16. Matzkin H, Soloway MS, Schellhammer PF , Chodak G, Smith JA, Caplan R, Kennealey GT: Prognostic factors in stage D2 prostate cancer treated with a pure non-steroidal anti-androgen. Cancer 1993;72:1286-1290,.
17. Eisenberger MA, Blumenstein BA, Crawford ED, Miller G, McLeod DG, Loehrer PJ, Wilding G, Sears K, Culkin DJ, Thompson IM Jr, Bueschen AJ, Lowe BA: Bilateral orchiectomy with or without Flutamide for metastatic prostate cancer. N Engl J Med 1998; 339:1036-1042.
18. Miller JI, Ahmann FR, Drach GW, Emerson SS, Bottaccini MR: Serum PSA levels predict duration of remission and survival post hormone therapy of metastatic prostate cancer. J Urol 1991; 145: 384A.
19. Newling DW, Denis L, Vermeylen K. Orchiectomy versus goserelin and flutamide in the treatment of newly diagnosed metastatic prostate cancer. Analysis of the criteria of evaluation used in the European Organization for Research and Treatment of Cancer - Genitourinary Group Study 30853. Cancer 1993;72(suppl 2): 3793-3798.
20. Strohmaier WL, Keller T, Bichler KH. Follow-up in prostate cancer patients: which parameters are necessary? Eur Urol 1999;35: 21-25.
21. Burzykowski T, Molenberghs G, Buyse M, Geys H, and Renard D. Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. *Applied Statist* 2001; **50**: 405-422.

Chapter 3:

Is Prostate Specific Antigen a surrogate endpoint for overall survival in metastatic prostate cancer?

3.1 Abstract

Background: The long duration of phase III clinical trials of overall survival (OS) slows down the treatment development process. It could be shortened by using surrogate endpoints. Prostate-specific antigen (PSA) is the most studied biomarker in prostate cancer (PCa). This study attempts to validate PSA endpoints as surrogates for OS in advanced PCa.

Methods: Individual data from 2161 advanced PCa patients treated in studies comparing bicalutamide to castration were used in a meta-analytic approach to surrogate end-point validation. PSA response, PSA normalization, time to PSA progression, and longitudinal PSA measurements were considered.

Results: The known association between PSA and OS at the individual patient level was confirmed. The association between the effect of intervention on any PSA endpoint and on OS was generally low (determination coefficient < 0.69).

Conclusion: It is a common misconception that high correlation between biomarkers and true endpoint justify the use of the former as surrogates. In order to statistically validate surrogate endpoints, a high correlation between the treatment effects on the surrogate and true endpoint needs to be established across groups of patients treated with two alternative interventions. The levels of association observed in this study indicate that the effect of hormonal treatment on OS cannot be predicted with a high degree of precision from observed treatment effects on PSA endpoints and thus statistical validity is unproven. In practice, non null treatment effects on overall survival can be predicted only from precisely estimated, very large effects on time to PSA progression (TTPP; hazard ratio <0.50).

3.2 Introduction

Phase III cancer clinical trials that evaluate the clinical benefit of new treatment options often require large patient numbers and long follow-up. Recent advances in the understanding of the biological mechanisms of disease development have resulted in the emergence of a large number of potentially effective new agents. There is also increasing public pressure for promising new drugs to receive marketing approval as rapidly as possible, in particular for life threatening diseases such as cancer. For these reasons, there is an urgent need to find ways of shortening the duration of cancer clinical trials. The duration of phase III trials results from the use of long term clinical endpoint (clinical progression, survival). Therefore, to replace this endpoint (the "true" endpoint) by another ("surrogate") endpoint that could be measured earlier, more conveniently or more frequently, and would

adequately reflect the benefit of new treatments on the clinical endpoint(s) seems to be an attractive solution.

"Biomarkers" (i.e., physical signs or laboratory measurements that occur in association with a pathological process or that have putative diagnostic and/or prognostic utility[1]) are generally regarded as best candidate surrogate end points. A biomarker is an intermediate outcome that is correlated with the true clinical outcome for an *individual* patient. It may be useful for diagnostic or prognostic information on a particular patient. It is a common misconception that established biomarkers necessarily make valid surrogate end points. To this aim, it is required that "the effect of treatment on a surrogate endpoint must be reasonably likely to predict clinical benefit"[2]. Thus a "*surrogacy*" is a concept that relates to *groups* of patients. To demonstrate surrogacy, a strong association between the treatment effects on the surrogate and on the true endpoint needs to be established *across groups of patients* treated with the new and standard interventions.

The validation of a candidate surrogate end point is not straightforward. Until recently, the statistical methods developed for this purpose used the data from a single trial[3-5]. These methods suffer from numerous drawbacks: some of them are too stringent to be of practical value, while others are based on nontestable assumptions[6,7]. To overcome these limitations, a new methodology, known as the "meta-analytic" validation approach, was developed recently[8-10]. This method uses large databases from multiple randomized clinical trials and aims at measuring directly the association between the treatment effects on the surrogate and the true endpoint.

In the field of prostate cancer (PCa), prostate-specific antigen (PSA) has probably been the most studied biomarker. It has been investigated as a potential surrogate endpoint across disease stages[11-14], and in hormone refractory patients in particular[15-18]. In a recent paper, Buyse et al[19] have considered several PSA-based end points in androgen-independent patients treated with liarozole (an imidazole-like compound that causes elevation of retinoic acid, postulated to have anti tumour activity), cyproterone acetate or flutamide. They showed that despite a strong association at the individual patient level, none of the end points qualified as a surrogate for overall survival (OS). In early prostate cancer, Newling et al.[20] found a modest correlation between the effect of Casodex on time to PSA progression (TTPP) and on objectively confirmed progression. In primary metastatic PCa, several studies demonstrated some level of association between a post-therapy fall in PSA or a PSA relapse on treatment and long term survival prognosis[21-25]. However this merely qualifies PSA as a biomarker. In trial NCI-INT-105, treatment differences in post-therapy PSA levels did not translate into survival differences[26]. Thus, whether PSA is a valid surrogate for survival in hormonally treated PCa remains an open question. This question is of importance, because the use of PSA could shorten the time to the endpoint from between several months in advanced disease[27] to several years in early disease[28].

The objective of the present research is to assess PSA-based end points as surrogates for OS in hormone naïve metastatic PCa using the meta-analytic approach. The data from > 2000

patients treated with bicalutamide (Casodex) that were made available by AstraZeneca Pharmaceuticals were used for this purpose.

Table 3.1: Trials used in the analysis

Trial 301/302[28,29]	Patients	Stage D2, fit for orchidectomy; ECOG performance status 0-2 No prior systemic therapy for prostate cancer, no previous radiotherapy to the prostate within 3 months of entry
	Treatments	Bicalutamide (50 mg daily) versus castration (orchidectomy in trial 301, orchidectomy or Goserelin 3.6 mg monthly injection in trial 302)
	Design	Open 2-arm randomization
	Objective	To compare Bicalutamide to castration , in a pooled analysis
	Efficacy endpoints	Time to treatment failure (objective progression, change of treatment, death due to any cause)*; Overall survival
	Results	Bicalutamide 50 mg daily demonstrated significantly worse time to progression and survival in trial 301. The trend was not significant in trial 302. By pooled analysis, both endpoints were significantly worse with bicalutamide than with castration.
Trial 306/307[30]	Patients	Metastatic (M1) or locally advanced with PSA fivefold in excess of the upper normal limit (T3-4 M0) – Only the M1 patients were included in the presently reported analyses Fit for orchidectomy; ECOG performance status 0-2; No prior systemic therapy for prostate cancer, no previous radiotherapy to the prostate within 3 months of entry
	Treatments	Bicalutamide (100 mg/day) or Bicalutamide (150 mg/day) or castration (medical or surgical at the patient's discretion)
	Design	Initially 2 (Casodex 100 mg):2 (Casodex 150 mg):1 (castration) then changed to 2:1 randomization between Casodex 150 mg and castration
	Objective	To demonstrate non-inferiority of Casodex 150 mg in comparison to castration by excluding a risk increase of 25%
	Efficacy endpoints	Time to treatment failure (addition of systemic therapy or withdrawal from therapy, objective progression or death)*; Overall survival; Objective response
	Results (in M1)	Significant difference in favour of castration were found for time to treatment failure (HR=1.43, 95% CI: 1.20-1.71 in favour of castration) and for overall survival (HR=1.30, 95% CI: 1.04-1.64)
US trial[31,32]	Patients	Stage D2 only; ECOG performance status 0-2; no prior systemic therapy
	Treatments	Bicalutamide (50 mg daily) versus Flutamide (250 mg three times daily) in combination with Goserelin acetate (3.6 mg monthly injection) or leuprolide acetate (7.5 mg monthly injection)
	Design	2x2 factorial design, blinding for the LHRH-A randomization
	Objective	To demonstrate non inferiority of Bicalutamide + LHRH-A relative Flutamide+ LHRH-A by excluding a relative risk increase of 25%
	Efficacy endpoints	Time to treatment failure (addition of systemic therapy or withdrawal from therapy, objective progression or death)*; Overall Survival
	Results	Non-inferior time to treatment failure (HR=0.93 in favour of bicalutamide, 95% CI: 0.79 – 1.10); Non-inferior overall survival (HR=0.87 in favour of bicalutamide, 95% CI: 0.72-1.05)

* A rising PSA was not considered a sign of progression in any of the studies

3.3 Patients

Individual data from three large international randomized trials of AstraZeneca's Casodex Development Program were used (301/302[29,30], 306/307[31], US trial 0001[32,33], Table 3.1). In studies 301/302 and 306/307, Casodex monotherapy (50 and 150 mg/day, respectively) was compared to medical or surgical castration. In the US trial, Casodex (50 mg/day) in combination with goserelin or leuprolide acetate was compared to the combination of Flutamide (750 mg/day) and castration in a 2x2 factorial design. All patients were newly diagnosed with metastatic PCa. Four hundred eighty patients with T3-4 M0 disease and elevated PSA from trial 306/307 were excluded. Survival was an endpoint in all studies although time to treatment failure (Table 3.1) was the primary endpoint in most. PSA was monitored at months 1, 2 (except US trial) and 3, then every 3 months until month 18 (trial 301/302) or death (other trials). For the analysis the PSA test date was assumed to be the visit date.

3.4 Endpoints

We considered OS calculated from randomization to the date of death or last visit as true end point. PCa-specific survival was defined similarly but with deaths unrelated to PCa or treatment censored at the last visit. PSA response, PSA normalization, TTPP and the complete series of PSA measurements ("PSA profile") were successively assessed as potential surrogate end points for OS.

Patients who had a baseline PSA level at least five times above the normal range (>20 ng/mL) were included in the analyses of PSA response and PSA normalization. Patients qualified for PSA response if their PSA declined by at least 50% from baseline level at 2 subsequent observations at least 4 weeks apart. Patients in whom the decline reached a value below or equal to normal (4 ng/mL) qualified for PSA normalization[25].

Two definitions of TTPP were assessed:

(1) For TTPP-1, PSA progression was defined as a PSA value above normal (4 ng/mL) representing a first increase ≥ 20% above the nadir[25] (eg, with a PSA nadir of 2 ng/mL a minimum increase to 4 ng/mL [100% increase] is required, whereas with a PSA nadir of 3.5 ng/mL a 20% increase to 4.2 ng/ml is enough).

(2) For TTPP-2, PSA progression was defined as a PSA value > 2.5 time the normal range (10 ng/mL) representing a first increase ≥ 50% above the moving average (based on three consecutive measurements) nadir. This increase had to be either the last observed value or be sustained for at least 4 weeks[19] (eg, with a nadir of 2 ng/mL at three consecutive occasions, a 500% increase to 10 ng/ml is needed to reach the end point, whereas after a nadir of 7 ng/mL, a 50% increase to 10.5 ng/mL is enough).

Patients who died or are alive without PSA progression were censored at the time of death or last visit, respectively.

3.5 Statistical methods

The meta-analytic approach to surrogate endpoint validation has been extensively detailed elsewhere[6,9,34-36]. We shall thus only summarize the key features. The method is rooted in the concept that a valid surrogate end point must enable one to predict with sufficient precision the treatment effect on the true clinical end point (OS) from the observed treatment effect on the surrogate (PSA-based) end point. Unlike traditional validation methods such as the Prentice Criteria[3], this new methodology does not require that any of those be statistically significant. Indeed, when data from several trials are available, the method consists of simultaneously estimating the relative treatment effects on the survival end point and on the PSA end point (log odds ratio of PSA response or normalization, log hazard ratio [HR] of PSA progression, treatment effect on the longitudinal PSA measurements) in each trial. A model that estimates the association between the treatment effects on the true end point and the corresponding effects on the PSA end points (PSA response[34], time-to-PSA progression[35] or longitudinal PSA measurements[36]) in a way similar to standard linear regression (although mathematically more sophisticated) is then adjusted. As in linear regression, the strength of the association is measured by the squared correlation coefficient that we shall denote R^2_{trial}. This coefficient also indicates the precision with which the treatment effect on the survival endpoint can be prediced from the observed treatment effect on the surrogate. The maximal possible value of R^2_{trial} is 1, which indicates a perfect prediction. In practice, observing $R^2_{trial}=1$ is not possible and one rather seeks a value close to one, which indicates a strong association between the treatment effects and thus a relatively precise prediction[9,35]. Additionally, the model quantifies the association between the PSA-based end point and the survival end point at the individual patient level. Parameters quantifying the strength of the association at this level will be denoted by the subscript "patient". They can be regarded as measures of validity of the PSA endpoint as a biomarker for predicting duration of survival.

Only three trials were available. This is too few to allow a precise estimation of R^2_{trial}. Therefore, the patients were grouped by the trial they entered and by their country of residence, as done by Buyse et al[19]. These groups will be henceforth referred to as "trial-units".

3.5.1 Results

After excluding non-metastatic patients and those with no baseline or follow-up PSA measurements, the individual data from 2161 patients classified into 21 trial-units were available for the analysis (Table 3.2). Their baseline and treatment characteristics are listed in Table 3.3. More than half of the patients presented with six or more bone metastases. After a median follow-up of 3.25 years, 1018 patients (52.9%) had died, 815 (71.3%) as a result of PCa (Table 3.4). The median OS was 2.2 years (95% CI, 2.1-2.5) for the Casodex treated patients and 2.3 years (95% CI, 2.1-2.6) in the pooled control groups (Figure 3.1). The average number of PSA assessments per patient was 6.9 (range, 1-23)

Table 3.2: Trial-units available for the analysis

Trial	Country	N	Trial	Country	N
301	Denmark	158	307	Australia	35
301	Norway	75	307	Austria	14
301	Sweden	63	307	Belgium	95
302	Austria	46	307	Germany	47
302	The Netherlands	29	307	The Netherlands	35
302	UK	159	307	Italy	11
306	Denmark	83	307	Republic of South Africa	48
306	Finland	69	307	Spain	22
306	Norway	83	307	UK	242
306	Sweden	86	US	Canada	114
			US	USA	647
Total 2161 patients					

Table 3.3: Patient characteristics

	Age mean (SE); *median (Q1, Q3)*	**Performance status 0/1/2/3/4 (%)**	**Baseline PSA: mean (SE); median (Q1, Q3)**
301/302			
Total (N=530)	data not available	not available	839.1 (1551.3) *267.9 (98.6, 784.7)*
Casodex 50mg (N=262)	data not available	data not available	811.2 (1477.8); *273.2 (98.3, 840.0)*
Castration (N=268)	data not available	data not available	866.3 (1622.2); *266.7 (99.4, 713.3)*
306/307 (UICC M1 pts.)			
Total (N=870)	71.6 (8.2) *72 (66, 78)*	53.8/32.8/13.3/0/0.1	747.3 (1657.2); *179.1 (65.7, 634.7)*
Casodex 100/150mg (N=617)	71.2 (8.2) *72 (66, 77)*	54.0/31.9/14.1/0/0	772.6 (1772.5); *189.8 (64.5, 658.4)*
Castration (N=253)	72.7 (8.1) *73 (67, 78)*	53.4/34.8/11.5/0/0.4	685.6 (1336.0); *156.0 (67.0, 587.3)*
US (D2 pts.)			
Total (761)	70.2 (8.7) *70 (65, 76)*	51.4/37.2/11.4/0/0	694.2 (1444.2); *174.3 (45.6, 580.6)*
Casodex + castration (N=377)	69.8 (8.2) *70 (65, 75)*	53.8/36.1/10.1/0/0	650.4 (1382.8); *170.0 (53.8, 588.1)*
Flutamide + castration (N=384)	70.5 (9.2) *71 (65, 77)*	49.0/38.3/12.8/0/0	737.3 (1502.6); *178.3 (38.7, 576.5)*

Figure 3.1
Overall survival by randomized treatment

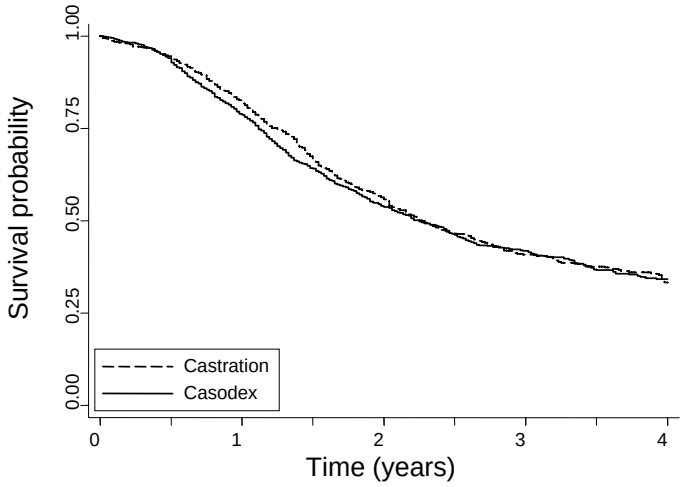


Table 3.4: Survival and PSA outcome

	Casodex (N=1256)	**Control (N=905)**	**Total (2161)**
Alive	571	447	1018
Dead	685	458	1143 (52.9%)
due to prostate cancer	496	319	815 (71.3%)
due to another cause	189	139	328 (28.7%)
PSA response			
evaluable	1090	763	1853
Decline to ≤ 4 ng/ml	399 (36.6%)	380 (49.8%)	779 (42.0%)
Decline by ≥ 50% of baseline	575 (52.8%)	307 (40.2%)	882 (47.6%)
No response	116 (10.6%)	76 (10.0)	192 (10.4%)
not evaluable	142	166	308
PSA progression (TTPP-1)	415	729	1144
PSA progression (TTPP-2)	432	233	665
not evaluable for PSA progression	35	32	67

PSA response (≥50% decline from baseline) and PSA normalization

PSA response could be assessed for 1853 patients. A total of 974 (89.4%) and 687 (90.0%) assessable patients on the Casodex and control groups, respectively, achieved a PSA response (Table 3.4). Only thirteen trial-units representing 1606 patients were used in the analysis: two trial-units were removed because no deaths were observed in the castration group, six because all patients responded in one or both treatment arms. At the individual level, PSA response was a strong predictor of prolonged survival with a survival odds ratio

$\theta_{patient}$ of 1.94 (SE, 0.33), representing a two-fold increase in the odds of surviving beyond any specified time *t* for the PSA responders compared to the non responders. At the trial level, the effects of hormonal intervention on PSA response and on OS were poorly correlated with R^2_{trial}=0.08 (SE,0.14; 95% CI, 0.0-0.49). Figure 3.2A presents the estimated treatment effects on the response (log odds ratio) and OS (log HR). One should be careful in interpreting these results, because eight trial-units with extreme results were excluded from the analysis.

In 399 (36.6%) and 380 (49.8%) of the assessable patients, the PSA declined to a value ≤ 4 ng/mL. Seventeen trial units representing 1778 patients could be used for this analysis: four were excluded for same reasons as above. At the individual level, the survival odds ratio $\theta_{patient}$ for patients with PSA normalization compared to those without was 4.90 (SE, 0.52) indicating a 4.9-fold greater odds of surviving any specified time *t* for the patients whose PSA normalized. At the trial level, the treatment effects on PSA and on OS were moderately correlated with R^2_{trial} 0.41 (SE, 0.18; 95%CI, 0.05-0.72, Table 3.5). Figure 3.2B presents the estimated treatment effects on PSA normalization and OS.

PSA progression

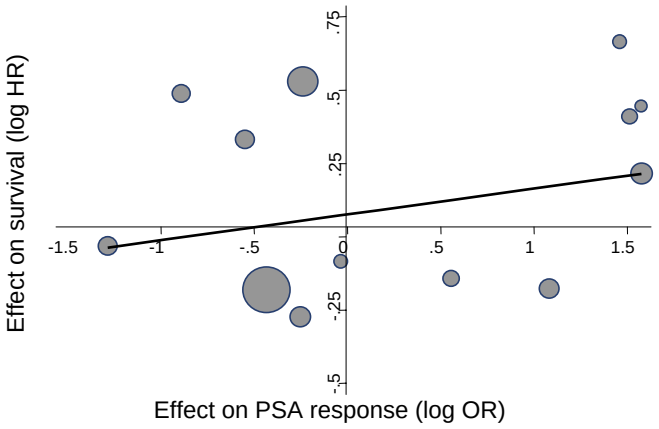
Nineteen trial-units (2070 patients) and eighteen trial-units (2043 patients) could be used for the analysis of TTPP-1 and TTPP-2, respectively (two trial-units were excluded from both analyses due to absence of deaths in the castration arm and one from the TTPP-2 analysis due to the absence of PSA progressions in both treatment arms).

The TTPP-1 is presented in Figure 3.3A: 54.6% of the patients progressed according to this definition (Table 3.4) within a median time of 11.1 months after being randomly assigned. TTPP-1 was somewhat shorter for the pooled Casodex group than for the control group. TTPP-1 was moderately associated with OS at the individual patient level: the concordance coefficient $\tau_{patient}$=0.52 (SE, 0.004) indicates that for each individual patient there is an approximately 50% chance to observe a long (short) OS given a long (short) TTPP. At the trial-unit level, the association between the effects of Casodex on TTPP-1 and on OS was low with R^2_{trial}=0.21 (SE, 0.17; 95%CI, 0.0-0.56; Table 3.5). This analysis is depicted in Figure 3.4A where the treatment effect on survival is regressed against the treatment effect on TTPP-1: the size of the circles represents the trial-unit size. The low trial-level association may be partly because of the outlying data from one trial unit. Excluding this unit from the analysis leaves the individual-level association unchanged ($\tau_{patient}$ = 0.52; SE, 0.004), but increases R^2_{trial} to 0.58 (SE, 0.15; 95%CI, 0.20-0.81).

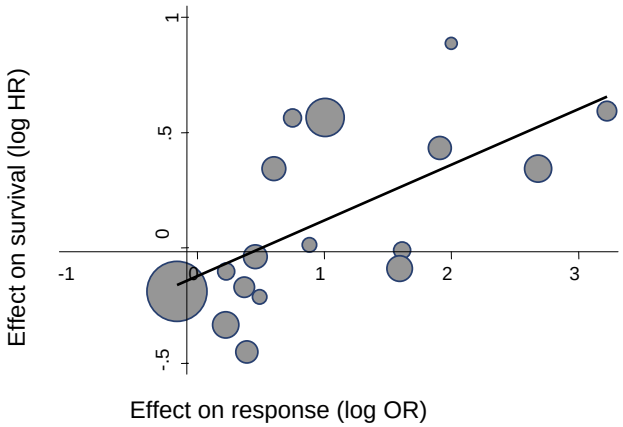
Only 31.8% of the patients met the more stringent criterion TTPP-2 (Table 3.4) at a median time of 24.9 months (Figure 3.3B). At the patient level, the association of TTPP-2 and OS was somewhat stronger than for TTPP-1, with a concordance coefficient $\tau_{patient}$=0.61 (SE, 0.02). The association between the treatment effects on TTPP-2 and OS was somewhat higher than for TTPP-1, with R^2_{trial} =0.66 (SE, 0.13; 95%CI, 0.30-0.85; Figure 3.4B and Table 3.5).

Figure 3.2
The treatment effects on survival and on PSA response

A. ≥ 50% decline from baseline level: R^2_{trial}=0.08.



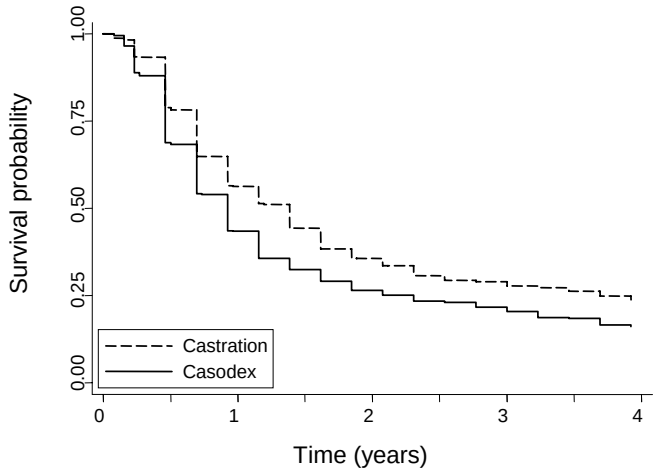
B. PSA normalization (PSA≤4 ng/ml): R^2_{trial}=0.41



The circles represent the observations in the trial-units, their size is proportionate to the trial-unit sample size. The line represents the prediction from an estimated (weighted) regression line.

Figure 3.3
Time to PSA progression by randomized treatment

A. TTPP-1
(Time to the first 20% increase of PSA over previously observed nadir to a value above the upper limit of the normal PSA range (4 ng/ml))



B. TTPP-2
(Time to the first 50% increase of PSA over previously observed moving average nadir to a value above 2.5x the upper limit of the normal PSA range (10 ng/ml), sustained for at least 4 weeks)

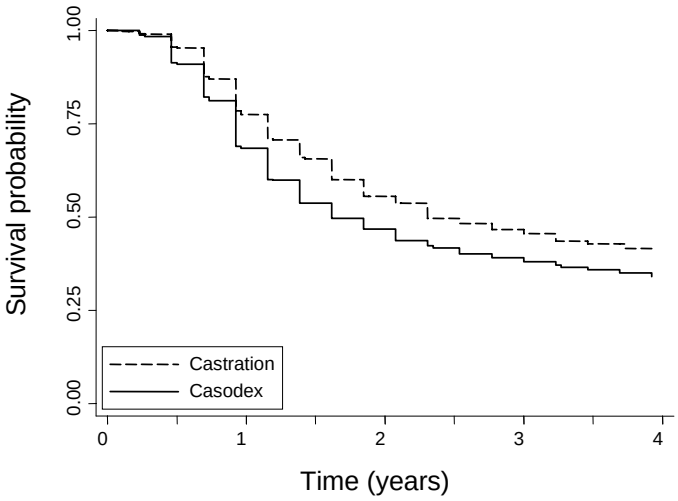
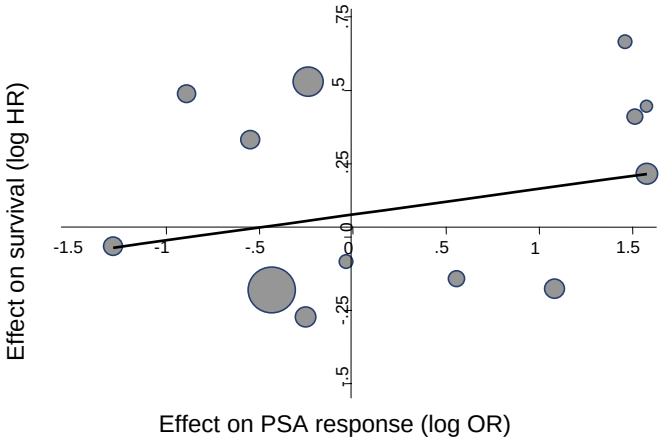
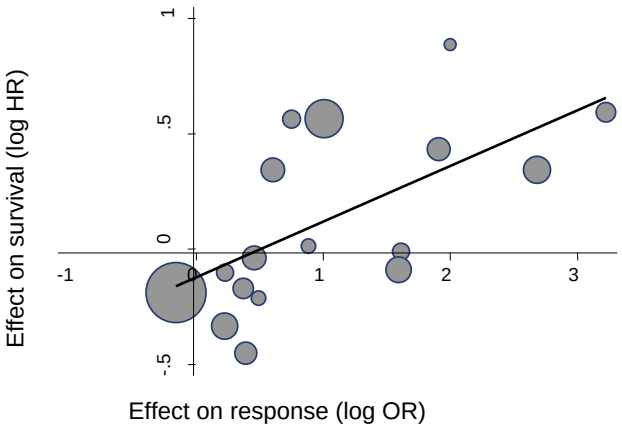


Figure 3.4
The treatment effects on time to PSA progression and on overall survival

A. TTPP-1 (R^2_{trial}=0.21)



B. TTPP-2 (R^2_{trial}=0.66)



The circles represent the observations in the trial-units, their size is proportionate to the trial-unit sample size. The line represents the prediction from an estimated (weighted) regression line.

Longitudinal measurements of PSA

All previously considered PSA-based endpoints are summary measures derived of the longitudinal PSA measurements and use only a limited amount of the available information. It thus seemed logical to investigate if the longitudinal series of PSA measurements would not be a better surrogate endpoint for OS. Figure 3.5A presents the mean profiles of log-transformed PSA measurements for groups of patients with similar observation time: all profiles eventually end with a PSA increase (progression) and patients with an early progression tend to have a higher initial PSA that does not decrease as much early on.

Figure 3.5B displays the mean PSA profiles per treatment group: starting from week 52 the curves show a relatively stable linear decrease rather than the increasing curvature observed in Figure 3.5A. This distortion results from attrition: progressive patients, in whom PSA increases, tend to leave the study, and thus the curve in Figure 3.5B reflects only those with stable PSA.

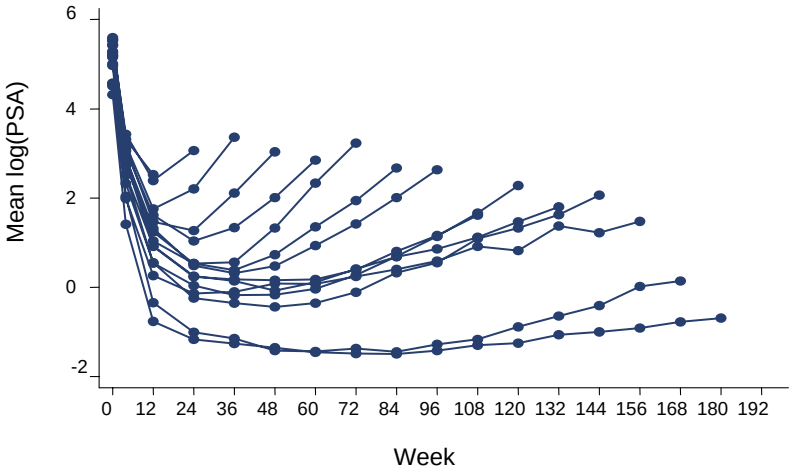
In view of Figure 3.5A, the treatment effect on the log-transformed PSA levels was expressed as a function of time and its square root in a joint model of PSA measurements and survival times. In that model, the individual patient-level association between the PSA process and the hazard of dying is a function of time and cannot be easily summarized into a single measure[35]. The results indicated that the correlation between the individual PSA and mortality hazard processes was > 0.90 at any time > 7 months, which suggests a strong association between the PSA profile and the hazard of dying for individual patients. At the trial-unit level, the association between the effect of Casodex on the longitudinal PSA and OS was slightly higher than that for TTPP-2 (R^2_{trial} = 0.68; SE, 0.12; Table 3.5).

Table 3.5: Summary of the results

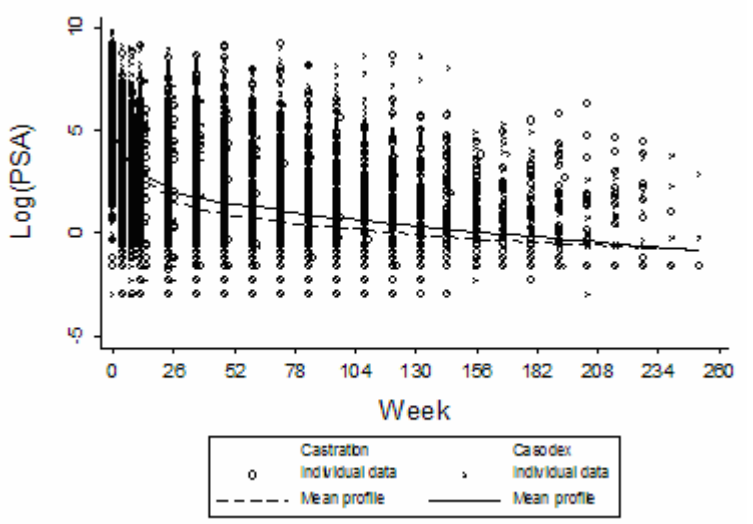
PSA endpoint	Patient-level association between PSA and survival (standard error)	Trial-level association between PSA and survival (standard error)
PSA response (decline by ≥ 50% from baseline)	Survival odds ratio $\theta_{patient}$ = 1.94 (SE=0.33)	R^2_{trial} = 0.08 (SE=0.14) 95% CI: 0 – 0.49
PSA normalization (≤4 ng/ml)	Survival odds ratio $\theta_{patient}$ = 4.90 (SE=0.52)	R^2_{trial} = 0.41 (SE=0.18) 95% CI: 0.05 – 0.72
Time to PSA progression (TTPP-1)	Concordance coefficient between tir to PSA progression and duration of survival $\tau_{patient}$ = 0.52 (SE=0.004)	R^2_{trial} = 0.21 (SE=0.17) 95% CI: 0 – 0.56
Time to PSA progression (TTPP-2)	Concordance coefficient between tir to PSA progression and duration of survival $\tau_{patient}$ = 0.61 (SE=0.02)	R^2_{trial}= 0.66 (SE=0.13) 95% CI: 0.30 – 0.85
Longitudinal PSA measurements	$R^2_{patient}$>0.9 at all times > 7 months	R^2_{trial}=0.68 (SE=0.12) 95% CI: undetermined

Figure 3.5
Log(PSA) measurements over time

A: Average log(PSA) profiles by drop-out time



B: Average log(PSA) profile by treatment arm



3.6 Discussion

Using data from the Casodex Development Program, we investigated whether the biomarker PSA could be used to define a valid surrogate for OS in patients with metastatic PCa. The analyses confirm the known value of PSA as a biomarker of prognosis and disease activity (individual-level association). When comparing groups of patients treated with Casodex-based or control treatment, however, the association between the treatment effect on any PSA-based endpoint and the treatment effect on OS was low in general ($R^2_{trial} < 0.69$ with wide confidence intervals).

The choice of the threshold for R^2_{trial} required for a valid surrogate is still a matter of debate[6]. Nevertheless, one can argue that the precision of the prediction of the treatment effect on OS from the effect on the PSA-based endpoints, indicated by the R^2_{trial} values observed in the present study, is insufficient to claim any of the assessed PSA-based endpoints a statistically valid surrogate end point for overall survival in phase III clinical trials of hormonal treatment in metastatic PCa.

To illustrate the problem, let us consider a new trial with TTPP as primary end point (defined as TTPP-2) where data analysis occurs after 400 events and yields an HR of 0.75 for PSA progression (with 400 events, SE [log(HR)] would be of the order of 0.10, resulting in P<.01). Without adjusting for the estimation error in the parameters of the prediction model, one could predict with approximately 95% confidence that the corresponding survival hazard ratio would lie within the interval 0.48 to 1.12. Adjustment for the estimation error would widen the confidence interval even further; thus, non null treatment effects on survival would potentially be identifiable only in large new trials showing a large effect on the PSA end point (eg, HR approximately 0.50 with SE=0.10).

Buyse et al[19] assessed similar PSA-based end points as candidate surrogates for OS in androgen-independent PCa patients treated with Liarozole versus anti-androgen monotherapy. In their study, the association between treatment effects at the trial level were generally low with $R^2_{trial} < 0.45$ for all tested PSA end points. They concluded that PSA end points could not be regarded as valid surrogates for OS. The reasons for the lack of association in their study may be different than ours; the disease was more advanced and treatment mode of action differed. In early disease, for which time savings of using PSA could be greater than in advanced disease, Newling at al.[20] also found only moderate correlation between the effect of Casodex on PSA progression and objective clinical progression.

Unfortunately, in cancer and other diseases, biomarkers that are strong predictors of the clinical end point for the individual patient often proved to be poor surrogate endpoints[37-43]. Several authors have discussed biological and medical reasons why biomarkers often fail to validate as surrogate end points[2,37,39,44]. The principal explanation is that only part of the treatment effect on the true clinical endpoint will be reflected in the biomarker, which may lead to over- or under-estimation of the treatment effect on the true end point from the observed effect on the biomarker. Baker and Kramer[45] mention that perfect predictors of the true end point at the patient level do not necessarily make good surrogate end points,

because the prediction function could differ between randomized treatments and thus would induce incorrect inference on the true end point.

The inability thus far to demonstrate surrogacy for PSA can be explained by several biological mechanisms. PSA is also produced by normal prostatic tissue, and the amount present may vary between patients. Poorly differentiated tumours may produce proportionally less PSA for the level of tumour burden compared with better differentiated tumors. In addition, PSA is studied in the serum while the source is prostatic tissue. Conceptually, serum levels can be related to unknown factors promoting or inhibiting leakage from prostate cancer cells into to the blood, to cellular levels of PSA and their interindividual variation and obviously to the total tumor mass present in a given patient[46,47]. PSA in itself is an endocrine dependent enzyme and its expression is regulated by a promoter that contains androgen-responsive elements[48,49]. The treatment effects seen on PSA in trials of endocrine treatment of PCa may thus at least in part result from a direct, non tumor mas- related effect. Such considerations led Scher et al[18] to conclude that PSA may not be an appropriate end point for clinical trials of first-line hormonal treatment.

Part of the imprecision in the prediction achieved in our study may be due to the limited number of observations available in each trial unit. The database we used, however, is the largest available. There is also some heterogeneity in trial design between the two monotherapy trials and the combined androgen blockade trial. However re-analysis excluding the latter did not change the results.

One could also argue about the use of overall rather than disease specific survival as true end point in our study. Analyses using disease specific survival as the true end point, however, led to essentially similar conclusions (R^2_{trial} for TTPP-2 was then 0.49; SE, 0.17; 95%CI, 0.11-0.76).

Finally, it was not possible to assess dynamic measures of PSA such as PSA doubling time or PSA velocity in these analyses; as suggested by Kelloff et al[50] and D'Amico et al[51], such measures of PSA may carry more information than the ones we could assess.

Our study indicates that PSA surrogacy could not be statistically validated in trials of hormonal treatments against metastatic prostate cancer. However, if very large effects on time to PSA end point (HR, <0.50) could be demonstrated with high precision in a new trial, the results of the present study would still provide evidence of a likely non null effect (upper bound of the 95% prediction interval for HR, <1) on OS. This suggests that, in such an instance, TTPP could potentially serve as a basis for accelerated drug approval, together with other trial data documenting safety and other measures of patient benefit, until firm evidence on the basis of the true endpoint becomes available. Nevertheless, additional research for more powerful surrogate end points in PCa is still needed. Such research should probably focus on dynamic PSA measurements, on new, hopefully more specific markers, or combinations of markers.

3.7 References

1. Lesko LJ, Atkinson AJ. Use of biomarkers and surrogate endpoints in drug development and regulatory decision making: criteria, validation, strategies. Annu Rev Pharmacol Toxicol. 2001;41:347-66.
2. Biomarkers Definitions Working Group: Biomarkers and surrogate endpoints: preferred definitions and conceptual framework Clin Pharmacol Ther. 2001;69:89-95.
3. Prentice RL. Surrogate endpoints in clinical trials: definitions and operational criteria. Stat Med. 1989;8:431-440.
4. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med. 1992; 11:167-178.
5. Buyse M, Molenberghs G. Criteria for the validation of surrogate end-points in randomized experiments. Biometrics. 1998;54: 1014-29.
6. Buyse M, Molenberghs G, Burzykowski T, *et al*. Statistical validation of surrogate endpoints: problems and proposals. Drug Inf J. 2000; 34: 447-454.
7. Molenberghs G, Buyse M, Geys H, *et al*. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. Control Clin Trials. 2002; 23: 607-625.
8. Daniels MJ, Hugues MD. Meta-analysis for the evaluation of potential surrogate markers. Stat Med. 1997; 16: 1515-27.
9. Buyse M, Molenberghs G, Burzykowski T, *et al*. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000;1: 49-68.
10. Gail MH, Pfeiffer R, Van Houwelingen HC, *et al*. On meta-analytic assessment of surrogate outcomes. Biostatistics. 2000 ; 1(3):231-46
11. Polascik TJ, Oesterling JE, Partin AW. Prostate specific antigen: a decade of discovery – what we have leraned and were we are going. J Urol. 1999; 162: 293-306.
12. Ornstein DK, Pruthi RS. Prostate-specific antigen. Expert Opin Pharmacother. 2000;1: 1399-1411.
13. Schroeder FH. Evidence in advanced disease: the challenge of new therapies. Eur Urol. 2002; Suppl 1: 11-16.
14. Small EJ, Mack Roach III. Prostate-specific antigen in prostate cancer: a case study in the development of a tumor marker to monitor recurrence and assess response. Sem Oncol. 2002; 3: 264-273.
15. Sridhara R, Eisenberger MA, Sinibaldi VJ, *et al*. Evaluation of prostate-specific antigen as a surrogate marker for response of hormone-refractory prostate cancer to suramin therapy. J Clin Oncol. 1995;13:2944-2953.
16. Smith DC, Dunn RL, Stawderman MS, *et al*. Change in serum prostate-specific antigen as a marker of response to cytotoxic therapy for hormone-refractory prostate cancer. J Clin Oncol. 1998;16:1835-1843.
17. Kelly WK, Scher HI, Mazumdar M, *et al.* Prostate-specific antigen as a measure of disease outcome in metastatic hormone-refractory prostate cancer. J Clin Oncol. 1993;11:607-615.
18. Scher HI, Kelly WK, Zhang ZF, et al. Post-therapy serum prostate-specific antigen level and survival in patients with androgen-independent prostate cancer. J Natl Cancer Inst. 1999;91:244-251.

19. Buyse M, Vangeneugden T, Bijnens L, *et al*. Validation of Biomarkers ad Surrogates for Clinical Endpoints. In *Biomarkers in Clinical Drug Development* (Bloom JC, ed) New-York: Springer-Verlag, 2003. p. 149-168.
20. Newling D, Carroll K, Morris T. Is prostate-specific antigen progression a surrogate for objective clinical progression in early prostate cancer? *J Clin Onco.l*; 22 (14S): 4652.
21. Matzkin H, Soloway MS, Schellhammer PF, *et al*. Prognostic factors in stage D2 prostate cancer treated with a pure non-steroidal anti-androgen. Cancer. 1993;72:1286-1290.
22. Miller JI, Ahmann FR, Drach GW, *et al*. Serum PSA levels predict duration of remission and survival post hormone therapy of metastatic prostate cancer. J Urol. 1991; 145: 384A.
23. Newling DW, Denis L, Vermeylen K. Orchiectomy versus goserelin and flutamide in the treatment of newly diagnosed metastatic prostate cancer. Analysis of the criteria of evaluation used in the European Organization for Research and Treatment of Cancer - Genitourinary Group Study 30853. Cancer. 1993;72 Suppl 2: 3793-3798.
24. Cooper EH, Armitage TG, Robinson MRG, *et al*. Prostatic specific antigen and the prediction of prognosis in metastatic prostatic cancer. Cancer. 1990;66:1025-1028.
25. Collette L, de Reijke TM, Schroeder FH et al. Prostate Specific Antigen – A prognostic marker of overall survival in good prognosis M1 prostate cancer? Eur Urol. 2003; 44(2):182-9.
26. Eisenberger MA, Blumenstein BA, Crawford ED, *et al*. Bilateral orchiectomy with or without Flutamide for metastatic prostate cancer. N Engl J Med. 1998; 339:1036-1042.
27. Newling D, Denis L and Vermeylen K, Orchiectomy versus Goserelin and Flutamide and the treatment of newly diagnosed metastatic prostate cancer. Cancer 1993; 72 (12): 3793-98
28. Pound CR, Partin AW, Eisenberger MA, Chan DW, Pearson JD, Walsh PC. Natural history of progression after PSA elevation following radical prostatectomy. JAMA. 1999; 281(17): 1591-7.
29. Iversen P, Tveter K, Varenhorst E. Randomised study of Casodex 50 MG monotherapy vs orchidectomy in the treatment of metastatic prostate cancer. The Scandinavian Casodex Cooperative Group. Scand J Urol Nephrol. 1996;30(2):93-8.
30. Kaisary AV, Tyrrell CJ, Beacock C, *et al*. A randomised comparison of monotherapy with Casodex 50 mg daily and castration in the treatment of metastatic prostate carcinoma. Casodex Study Group. Eur Urol. 1995;28(3):215-22.
31. Tyrrell CJ, Kaisary AV, Iversen P, *et al*. A randomised comparison of 'Casodex' (bicalutamide) 150 mg monotherapy versus castration in the treatment of metastatic and locally advanced prostate cancer. Eur Urol. 1998;33(5):447-56.
32. Schellhammer P, Sharifi R, Block N, *et al*. A controlled trial of bicalutamide versus flutamide, each in combination with luteinizing hormone-releasing hormone analogue therapy, in patients with advanced prostate cancer. Casodex Combination Study Group. Urology. 1995;45(5):745-52.
33. Schellhammer PF, Sharifi R, Block NL, *et al*. Clinical benefits of bicalutamide compared with flutamide in combined androgen blockade for patients with advanced prostatic carcinoma: final report of a double-blind, randomized, multicenter trial. Casodex Combination Study Group. Urology. 1997;50(3):330-6.
34. Burzykowski, T., Molenberghs, G., Buyse, M. The validation of surrogate endpoints using data from randomized clinical trials: a case study in advanced colorectal cancer. *J R Stat Soc A*. 2004; 167 (1): 103-124.
35. Burzykowski T, Molenberghs G, Buyse M, *et al*. Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. Appl Stat. 2001;50: 405-22.
36. Renard D, Geys H, Molenberghs G, *et al*. Validation of a longitudinally measured surrogate marker for time-to-event endpoint. *J Applied Stat.* 2003; 30: 235-47

37. Fleming TR. Evaluating therapeutic interventions: Some issues and experiences. Stat Sciences. 1992; 7: 428-456.
38. Temple RJ. A regulatory authority's opinion about surrogate endpoints. In: Clinical measurement in drug evaluation (W. Nimmo, G. Ticker, Eds.) John Wiley, Chichester, 1995.
39. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? Ann Intern Med. 1996;125:605-613.
40. Buyse M, Thirion P, Carlson RW, *et al*. Relation between tumour response to first-line chemotherapy and survival in advanced colorectal cancer: a meta-analysis. Lancet 2000;356: 373-378.
41. The Cardiac Arrhytmia Suppression Trial (CAST) Investigators. Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. N Engl J Med. 1989; 321: 406-412.
42. The Cardiac Arrhytmia Suppression Trial (CAST) II Investigators. Effect of the antiarrythmic agent moricizine on survival after myocardial infarction. N Engl J Med 1992; 327: 227-33.
43. Burzykowski T, Molenberghs G, Buyse M, *et al*. The validation of surrogate endpoints using data from randomized clinical trials: a case study in advanced colorectal cancer. J R Stat Soc A. 2004. 167 (part 1): 103-124
44. Frank R and Hargreaves R. Clinical Biomarkers in Drug Discovery and Development. Nat Rev Drug Discov 2003;2(7):566-80.
45. Baker SG, Kramer BS. A perfect correlate does not a surrogate make. BMC Med Res Methodol 2003; 3(1):16
46. Stege RH, Tribukait B, Calrlstron KA, *et al*. Tissue PSA from fine-needle biopsies of prostatic carcinoma as related to serum PSA, clinical stage, cytological grade and DNA ploidy. Prostate. 1999; 38(3): 183-8
47. Schmid HP, McNeal JE, Stamey TA. Observations on the doubling time of prostate cancer. The use of serial prostate-cancer antigen in patients with untreated disease as a measure of increasing cancer volume. Cancer. 1993; 71(6): 2031-40
48. Cleutjens KB, van Eekelen CC, van der Kroput HA, *et al*. Two androgen response regions cooperate in steroid hormone regulated activity of the prostate-specific antigen promoter. J Biol Chem 1996; 271(11): 6379-88
49. Riegman PH, Vlietstra RJ, van der Korput JA, *et al*. The promoter of the prostate-specific antigen gene contains a functional androgen responsive element. Mol Endocrinol. 1991; 5(12):1921-30
50. Kelloff GJ, Coffey DS, Chabner BA, *et al*. Prostate-specific antigen doubling time as a surrogate marker for evaluation of oncologic drugs to treat prostate cancer. Clin Cancer Res. 2004;10(11):3927-3
51. D'Amico AV, Moul JW, Carroll PR, *et al*. Surrogate endpoint for prostate cancer specific mortality after radical prostatectomy or radiation therapy. J Natl Cancer Inst. 2003; 95(18): 1376-83.

Chapter 4:

Prognostic factors of clinical progression-free survival in a phase III trial of adjuvant hormone treatment versus nil in a population of high risk T3-4 or G3 prostate cancer patients treated with external beam irradiation: Does adjuvant hormone therapy benefit to all patients?

4.1 Abstract

Background: We did a randomised phase III trial comparing external irradiation alone and external irradiation combined with an analogue of luteinising-hormone releasing hormone (LHRH) to investigate the added value of long-term androgen suppression in locally advanced prostate cancer.
Methods: Between 1987 and 1995, 415 patients were randomly assigned radiotherapy alone or radiotherapy plus immediate androgen suppression. Eligible patients had T1–2
tumours of WHO grade 3 or T3–4 N0–1 M0 tumours; the median age of participants was 71 years (range 51–80). In both treatment groups, 50 Gy radiation was delivered to the pelvis over 5 weeks, and 20 Gy over 2 weeks as a prostatic boost. Goserelin (3·6 mg subcutaneously every 4 weeks) was started on the first day of irradiation and continued for 3 years; cyproterone acetate (150 mg orally) was given for 1 month starting 1 week before the first goserelin injection. The primary endpoint was clinical disease-free survival. Analyses were by intention to treat.
Results: 412 patients had evaluable data, with median follow-up of 66 months (range 1–126). 5-year clinical disease-free survival was 40% (95% CI 32–48) in the radiotherapy-alone group and 74% (67–81) in the combined-treatment group (p=0·0001). 5-year overall survival was 62% (52–72) and 78% (72–84), respectively (p=0·0002) and 5-year specific survival 79% (72–86) and 94% (90–98).
Conclusions: Immediate androgen suppression with an LHRH analogue given during and for 3 years after external irradiation improves disease-free and overall survival of patients with locally advanced prostate cancer.

4.2 Introduction

The long-term outcome after external irradiation alone in locally advanced prostate cancer, staged T3–4N0M0 according to the classification of the International Union against Cancer[1] is poor[2], especially for biochemically defined disease-free survival.[3] From the mid-1980s, two approaches to improve these results were investigated: first, the combination of androgen suppression and external irradiation in an attempt to

decrease the local failure rate and to destroy hormone-dependent micrometastases outside the planning target volume; and second, three-dimensional conformal radiotherapy to improve local control,[4] by increasing the dose delivered to the prostate. Three phase III randomised trials have shown an improvement of overall survival with the combination of radiotherapy and androgen suppression. For two of these trials,[5,6] this improvement was greatest for a subset of patients with Gleason 8–10 T2c–T4 tumours, whereas an EORTC trial[7] found a significant difference in 5-year overall survival, irrespective of the histological grade. Here, we present the long-term results of this trial, with a multivariate prognostic-factor analysis, and assessment of the serum testosterone profile after the end of the long-term androgen suppression.

4.3 Methods

Patients

Eligible patients were younger than 80 years, with histologically proven T1–2 prostatic adenocarcinoma of WHO histological grade 3, or T3–4 prostatic adenocarcinoma of any histological grade. The clinical investigation was based on bone scan, chest radiograph, and ultrasonography or CT of the abdomen. Lymph nodes were assessed by CT scan, bipedal lymphangiography, or extraperitoneal lymphadenectomy. The laboratory studies included complete blood count and measurements of creatinine, serum testosterone, and prostate specific antigen (PSA) assessed by radioimmunoassaay or enzyme immunoassay. Patients had to have newly diagnosed prostate cancer and to have given informed consent in writing. Approval by local ethics committees was obtained in all centres. We excluded patients with a history of previous malignant disease, except for adequately treated basal-cell carcinoma of the skin, or evidence of distant metastases, including involvement of common iliac or para-aortic lymph nodes. Pathological samples were centrally reviewed.

Design and procedures

Randomisation was centralised at the EORTC Data Centre. Patients were stratified by institution, clinical stage of disease (T1–2 grade 3 vs T3–4 grade 1–3), results of pelvic-lymph-node dissection (N0 vs N1), and irradiation technique (extended fields vs limited fields). Randomisation was done by two independent secretaries with the minimisation technique.[7] Patients were assigned external irradiation alone or external irradiation combined with an analogue of luteinising-hormone releasing hormone (LHRH) by a computer-generated system.

The methods of treatment have been described previously[8] Briefly, irradiation with photons of 10 MV and above was recommended. The first planning target volume was the whole pelvis and the second encompassed the prostate and the seminal vesicles. A four-field technique was used to irradiate the whole pelvis; at some centres limited fields were preferred. The second planning target volume was irradiated with the same technique or with three fields. The specification of the dose was given at the intersection of the beam axes according to report 29 of the International Commission on Radiation Units and Measurements.[9] Patients were treated once a day, 5 days a week, for 7 weeks:

the first planning target volume was irradiated up to 50 Gy and the second received an additional 20 Gy. Goserelin acetate, the LHRH analogue, was administered subcutaneously every 4 weeks, starting on the first day of pelvic irradiation, and continued for 3 years; this method of administration is now obsolete, owing to the availability of a 3-monthly depot preparation. A steroidal antiandrogen (cyproterone acetate) was given orally for 1 month, 50 mg three times daily, starting a week before the start of goserelin, to prevent the flare resulting from the surge in testosterone that arises after LHRH.

We measured PSA 2 months after the end of external irradiation, then every 3 months for 3 years, and every 6 months thereafter; in practice, the mean number of PSA measurements available per patient during follow-up in addition to the baseline assessment was seven (range none to 23). Follow-up PSA measurements were available for 388 patients (94%), 197 in the radiotherapy-alone group and 191 in the combined-treatment group. Measurement of testosterone concentration was asked for as a check of the validity of adjuvant hormonal treatment by LHRH analogues; these measurements had to be done with the same frequency as PSA measurements. However, owing to poor compliance, follow-up testosterone data were available for only 110 patients.

Acute side-effects of radiotherapy were scored according to the WHO scale.[1] Late toxic effects were scored according to the Radiotherapy Oncology Group. Local failure was defined as an increase of more than 50% in the product of the two maximum perpendicular diameters of the primary lesion as measured digitally, by CT or transabdominal ultrasonography; in case of doubt, biopsy was strongly recommended. Local progression was defined as recurrence of a palpable tumour after initial regression. Regional failure, in the area of the pelvis or the para-aortic lymph nodes, was detected by ultrasonography or CT and confirmed by biopsy. Distant metastases in bones, parenchymal organs, or soft tissues were identified radiologically and then by biopsy if deemed necessary.

Quality assurance in calibration of linear accelerators and treatment technique have been described elsewhere.[10] Quality-of-life data were not gathered in this study.

Statistical analysis

The primary endpoint was disease-free survival at 5 years. With the assumption of 5-year disease-free survival of 40% in the group assigned radiotherapy alone, we estimated that 75 patients had to be followed up until relapse or death in each treatment group for us to detect a 15% difference in the 5-year disease-free survival with power of 80% and a type-I error probability of 0·05.[11] This difference corresponds to an increase from 3.8 to 5.8 years in the median disease-free survival (on the assumption of exponentiality).

Clinical disease-free interval (time to first clinical evidence of progression), clinical disease-free survival (time to first clinical progression or death from any cause), and overall survival were calculated from the date of randomisation to the date of event and

have already been defined.[8] Survival curves were estimated by the Kaplan- Meier technique. We compared duration of overall and disease-free survival and the disease-free interval by a two-sided log-rank test. Cumulative incidence curves were used to assess the rates of loco-regional failure and distant failure. The time until the first treatment failure after a biological response was measured from the date of randomisation to the date of clinically determined progression, PSA-defined progression, or the latest follow-up. PSA-defined progression was a concentration higher than 1·5 µg/L and increasing on two consecutive measurements.[3]

The multivariate prognostic-factor analysis used Cox's proportional-hazards regression model. The prognostic index was obtained as the sum of the log (hazard ratio) associated with each patient's profile; the value was 0 plus 0.36 (if WHO performance status was above 0), plus 0.56 (if grade G3), plus 0.55 (if PSA concentration was above 10 µg/L). The prognostic index could therefore be between 0 and 1.52. Three risk groups were formed by classing together patients with similar values of the prognostic index to obtain the maximum separation between survival of the groups. All analyses were by intention to treat.

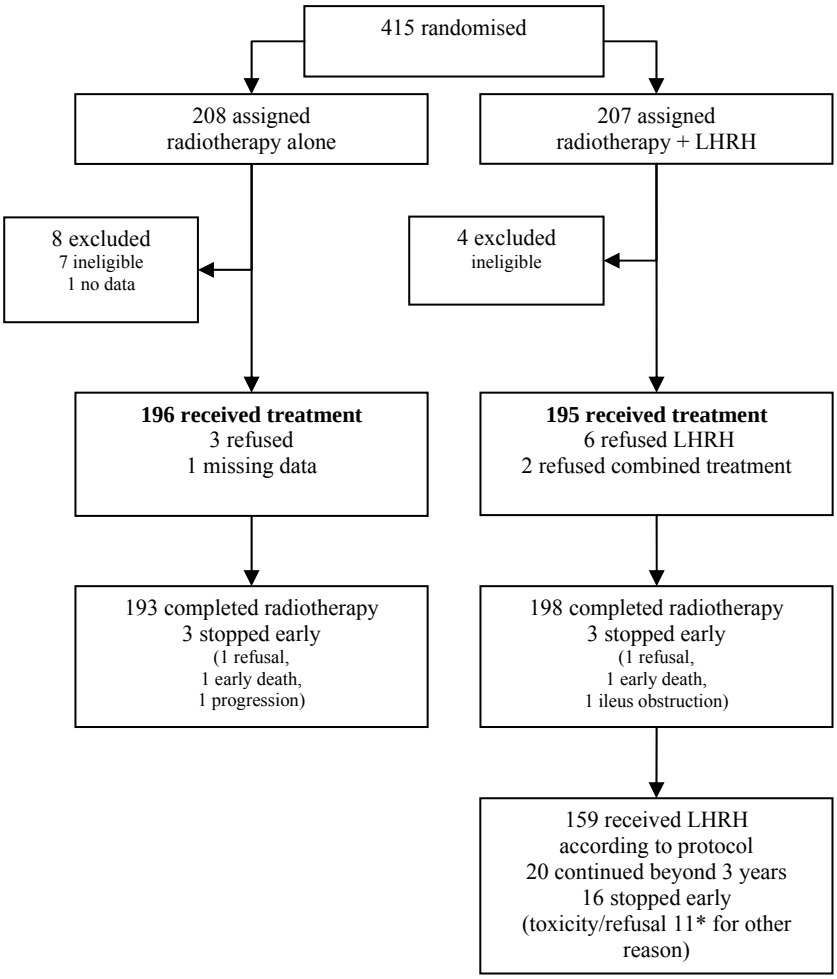
Role of the funding source
Astra Zeneca provided the goserelin, but had no role in study design, data collection, data analysis, data interpretation, or in the writing of the report.

4.4 Results

Between May, 1987, and September, 1995, 415 patients entered the study: 208 were randomly assigned to the radiotherapy-alone group and 207 to the combined-treatment group (Figure 4.1). At the time of the analysis reported here, median duration of follow-up was 66 months (range 1–126). 412 patients could be centrally evaluated and 11 were ineligible (four assigned combined-treatment and seven assigned radiotherapy alone). The reasons for ineligibility were incomplete examination before randomisation (one), inadequate disease staging or histopathology (nine), or poor physical condition (one).

The characteristics of eligible patients showed good balance between the treatment groups in age, WHO performance status, clinical stage, pelvic-lymph-node status, WHO histological grade, Gleason grade, and baseline PSA (Table 4.1). Baseline PSA concentration was defined in relation to the normal value of each PSA measurement. Cardiovascular disease was present in 58 (29%) patients in the radiotherapy group and 49 (24%) in the combined-treatment group; no chronic disease was mentioned for 95 (48%) patients in the radiotherapy group compared with 108 (53%) in the combined-treatment group.

Figure 4.1
Trial Profile



All randomised patients were included in the intention-to-treat analyses. *Five refused further treatment owing to hot flushes and three refused further treatment at the end of radiotherapy; two stopped because of depression and one because of mastodynia and galactorrhoea. †Two underwent orchiectomy because of poor compliance with goserelin; one needed urethrotomy and reconstruction owing to recurrent strictures; one was lost to follow-up after 15 months; and one had poor compliance owing to the cost of treatment so underwent orchiectomy.

Table 4.1: Baseline characteristics by treatment group

	Radiotherapy alone (N = 198)		Combined-treatment (N = 203)	
Median (range) age, years	70	(51-80)	71	(54-80)
Performance status				
0	157	(79)	158	(78)
1	37	(19)	38	(19)
2	4	(2)	7	(4)
WHO grade				
G1	37	(19)	44	(22)
G2	90	(46)	96	(47)
G3	67	(34)	62	(30)
Gx	4	(2)	1	(1)
Gleason grade				
2-4	16	(8)	10	(5)
5-6	38	(19)	49	(24)
7-10	71	(36)	66	(33)
unknown	73	(37)	78	(38)
Classification				
T1	0	(0)	2	(1)
T2	17	(7)	15	(7)
T3	163	(82)	167	(82)
T4	18	(9)	19	(9)
T according to grade				
T1-T2 G3	17	(9)	17	(8)
T3-T4 all grades	181	(91)	186	(92)
N classification				
N0	176	(89)	181	(89)
N1	5	(3)	4	(2)
N2	1	(1)	4	(2)
Unknown	16	(8)	14	(7)
PSA concentration (µg/L)				
≤ 4.0	10	(5)	15	(7)
4·1 – 10	21	(11)	24	(12)
10·1 - 20	36	(18)	28	(14)
20·1-40	48	(24)	47	(23)
> 40	62	(31)	72	(35)
unknown	21	(11)	17	(8)

Data are number of patients otherwise indicated

Treatment information was available for 403 patients (Figure 4.1; 200 in the radiotherapy group and 203 in the combined-treatment group). 196 patients in the radiotherapy-alone group received external irradiation; three refused treatment. In the combined-treatment group, 195 of the 203 patients received the treatment; two patients refused all treatment, and six refused hormonal therapy. 21 (11%) patients received the hormone injections with irregular intervals. 159 (87%) patients received the hormonal treatment planned by the protocol; 127 (65%) received the full 3 years of treatment, 23

(12%) stopped earlier than 3 years owing to progression or death, and 20 (10%) received the 3 years of treatment but did not stop at that point. For another nine patients, total follow-up is presently less than 3 years and they remain on treatment. 11 (6%) stopped earlier than 3 years because of toxic effects or the patient's wish, and five stopped earlier than 3 years for other reasons.

Table 4.2: Sites of disease progression

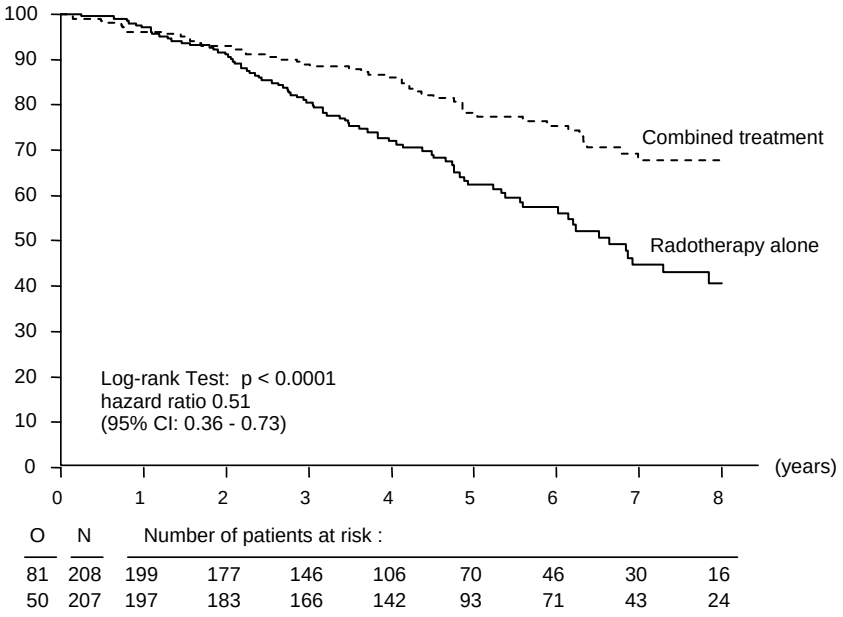
Type of progression	Radiotherapy (N=208)	Combined-treatment (N=207)
Any clinical progression	90	27
Local	15	3
Local and regional	3	0
Distant	56	22
Local and distant	13	2
Local, regional, and distant	3	0

After a median of 65.7 months, progression has occurred in 90 patients in the radiotherapy group and 27 in the combined-treatment group (Table 4.2). In the radiotherapy group, the treatment given at the time of progression was goserelin in 65 cases (72%), surgical castration in seven, another LHRH analogue in five, delayed treatment in five, unspecified treatment in two, no treatment because the patient refused in one case, no treatment because of death from another cause in one case, or no documented treatment (four). In terms of locoregional control, the outcome was better in the combined-treatment group than in the radiotherapy group (5-year cumulative incidence of locoregional failure 16.4% [95% CI 10.8–22.1) *vs* 1.7% [0–3.7]; p<0·0001). There was also a significant difference in favour of combined-treatment in the cumulative incidence of distant metastases (5-year cumulative incidence 29.2% [22.7–35.6] in the radiotherapy group *vs* 9.8% [5.4–14.2]; p<0·0001). There was no difference in the rate of second primary tumours between groups (6% *vs* 5%).

There was a significant difference between the groups in clinical disease-free survival (time to clinical failure or death from any cause; the hazard ratio was 0.34 (0.26–0.46; p<0·0001) with 74% (67–81) of patients in the combined-treatment group clinically disease free at 5 years compared with 40% (32–48) in the radiotherapy-alone group.

Overall survival also differed significantly between groups, with a hazard ratio of 0.51 (95% CI 0.36–0.73); the 5-year survival was 78% (72–84) in the combined-treatment group and 62% (52–72) in the radiotherapy group (p=0·0002; Figure 4.2). 78 deaths were registered in the radiotherapy group and 50 in the combined-treatment group, with 42 and 12 deaths, respectively, due to prostate cancer. 5-year specific survival was 94% (90–98) in the combined-treatment group and 79% (72–86) in the radiotherapy group (p=0.0001; hazard ratio 0.26 [0.15–0.44]).

Figure 4.2
Kaplan-Meier estimates of overall survival by treatment group



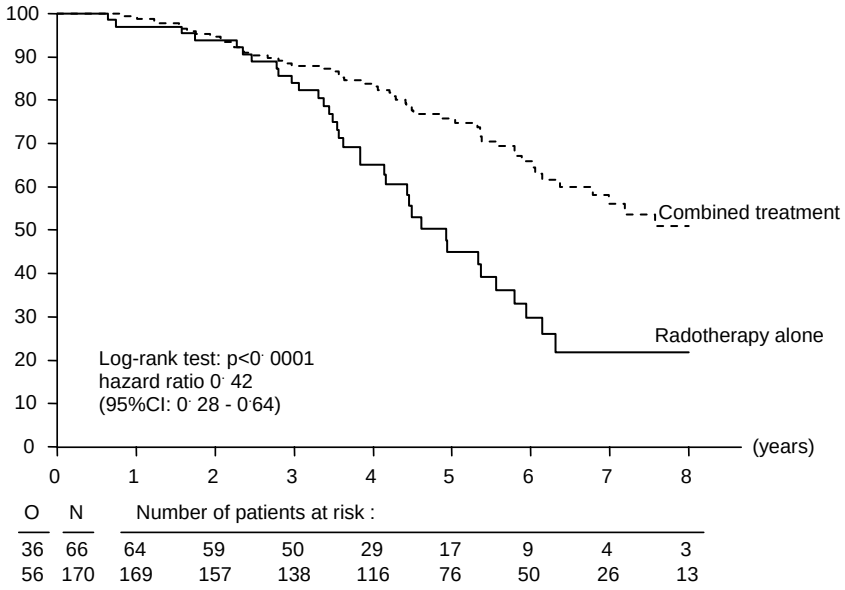
O=number of deaths, N=number of patients

In terms of the biochemical disease-free survival, combined-treatment was also better than radiotherapy alone (hazard ratio 0.42 [0.28–0.64]; p<0.0001; 5-year values 76% [69–83] *vs* 45% [30–60]; Figure 4.3).

The results were similar in the cohort of patients with T3–4 disease (89% of the whole sample); 5-year overall survival was better in the combined-treatment group than in the radiotherapy group (78% *vs* 60%; hazard ratio 0.46 [0.32–0.65]; p=0.0001) as was specific survival (94% *vs* 78%; hazard ratio 0.23 [0.14–0.41]; p=0.0001).

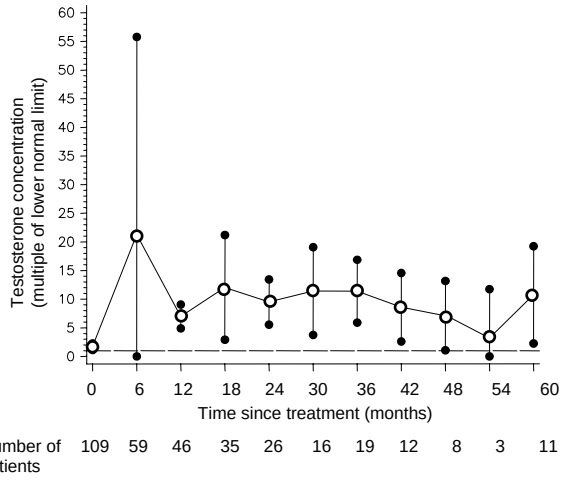
Figure 4.4 shows the testosterone concentration (expressed as multiples of the lower normal limit) after withdrawal of goserelin treatment. In 82 of 109 patients assessed for testosterone concentration after the end of the adjuvant hormonal treatment, the testosterone concentration rose above the lower limit of normal (castration value).

Figure 4.3
Kaplan-Meier estimates of the biochemically defined disease-free survival



O=number of deaths, N=number of patients

Figure 4.4
Testosterone concentration at the end of the combined-treatment



Values are arithmetic means: bars are 95% CIs. The dashed horizontal bar represents the castration value or the lower limit of the normal range

According to the univariate analysis, the prognostic factors for clinical disease-free survival were WHO performance status (0 *vs* >0; p=0.001), associated chronic disease (none or disease other than cardiovascular *vs* cardiovascular disease; p=0.02), G grade (G1–2 *vs* G3; p=0.0003), and initial PSA concentration (≤10 µg/L *vs* >10 µg/L; p=0.016). The effect of N category could not be analysed because only 16 patients were N1. The variables associated cardiovascular disease, WHO performance status, grade G3, and baseline PSA concentration were entered in the multivariate model selection procedure. Associated cardiovascular disease dropped out of the model, leading to the final multivariate model described in table 4.3.

Table 4.3: Multivariate prognostic factors model

	Relative risk estimate (95%CI)	P
WHO performance status		
0*	1.00	
> 0	1.43 (1.01 – 2.03)	0.0459
WHO grade		
G1—2*	1.00	
G3	1.84 (1.33 – 2.54)	0.0002
PSA		
≤ 10 ng/ml*	1.00	
> 10 ng/ml	1.83 (1.17 – 2.88)	0.0085

* Reference category. O/N : observed number of events/number of patients

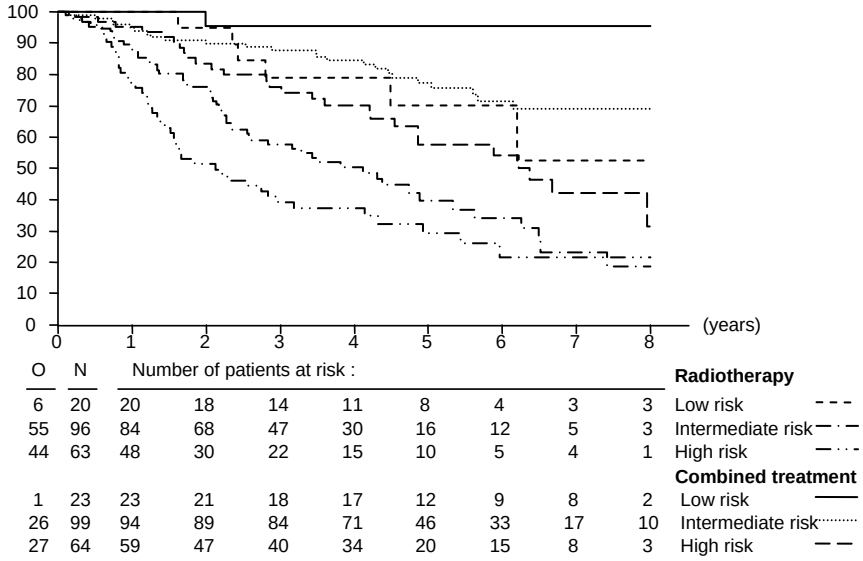
Risk categories can be formed by calculating the prognostic index and classifying patients according to their prognostic indices as shown in table 4.4. Owing to missing values, only 365 patients could be classified in the risk groups: 43 (12%) low risk, 195 (53%) intermediate risk, and 127 (35%) high risk. The 5-year clinical disease-free survival was 83.8% (95% CI: 71.7–96·0), 60.0% (52.4–67.7), and 43.4% (33.6–53.1), respectively, for these categories. Figure 4.5 shows disease-free survival by risk category for both treatment groups. The hazard ratio for combined-treatment versus radiotherapy alone was 0.12 (0.01–1.01; p=0·0508) in the low-risk category, 0.28 (0.18–0.46; p=0·0001) in the intermediate-risk category, and 0.39 (0.24–0.63; p=0·0001) in the high risk category. However, the overall test for a differing treatment effect according to risk category (interaction) was not significant (p=0.20), indicating homogeneous treatment effect in the three risk categories.

Table 4.4: Risk categories

Risk group	Prognostic index*	Interpretation
Low risk group	≤ 0·4	No risk factors or WHO > 0 as the only risk factor
Intermediate risk	> 0·4-0·6	Grade G3 or PSA > 2.5times lower normal limit as only risk factor
High risk	> 0·6	Two or more risk factors

Prognostic index=0+0.36 (if WHO PS>0) + 0.56 (if grade G3)+0.55(if PSA>10 µg/L)

Figure 4.5
Clinical disease-free survival by risk category and treatment group



4.5 Discussion

The combination of 3 years of androgen suppression with external irradiation was associated with better 5-year overall survival of locally advanced prostate cancer than radiotherapy alone. These updated results, with median follow-up of 65.7 months, accord with those reported in 1996, when the median follow-up was 33 months,[12] and in 1997 with median follow-up of 45 months.[8] Androgen suppression provides a method to improve the outcome of external irradiation alone, possibly by elimination of occult

systemic disease. Moreover, androgen suppression and external irradiation seem to have an additive effect on local control by induction of apoptosis.[13–15]

Adjuvant androgen deprivation is also effective with surgery, and in one prospective randomised study,[16] overall survival was better with long-term adjuvant hormonal treatment after radical prostatectomy than after surgery alone. In that study, 98 men who underwent radical prostatectomy and who had pelvic nodal metastases were randomly assigned goserelin or bilateral orchiectomy or to be followed up until disease progression without further treatment. After a median period of 7.1 years, there were significant benefits in overall survival ($p=0.02$), specific survival ($p=0.001$), and progression-free survival ($p<0.001$) in favour of the combined-treatment group. Conversely, short-term neoadjuvant androgen suppression followed by radical prostatectomy was associated with a lower rate of positive margins and lower tumour pathological stage, but there was no difference in overall survival.[17–19] Finally, there is a parallel between breast and prostate cancers: in both disorders the long-term outcome is improved by combination of radiotherapy and hormonal treatment,[20,21] and long-term adjuvant hormonal treatment significantly improves overall survival.[22]

Other clinical studies have addressed the role of androgen suppression with external irradiation. The Radiation Therapy Oncology Group has reported three randomised studies. Protocol 86-10[23,24] compared androgen deprivation plus radiotherapy with radiotherapy alone and included 456 patients with large T2 or T3–4 tumours. Flutamide was given daily and goserelin every 4 weeks, both started 2 months before radiotherapy began and were withdrawn at the completion of radiotherapy. Hormonal therapy increased the 5-year rates of local control ($p<0.002$) and freedom from distant metastases ($p<0.03$), and significantly increased 8-year progression-free survival (including PSA >4 µg/L as failure; $p<0.0001$). Protocol 85-31[5] investigated adjuvant androgen suppression with goserelin in patients classified as T1–2 with regional lymph-node involvement, T3 whatever the regional lymph-node status, or pT3 after prostatectomy. Goserelin was started at the end of the radiotherapy and continued indefinitely. With median follow-up of 4.5 years, there was an increase in the rates of local control ($p<0.0001$) and freedom from distant metastases ($p<0.001$) as well as disease-free survival ($p<0.001$). In patients with centrally reviewed tumours with a Gleason score of 8–10, there was a difference in actuarial 5-year survival in favour of the adjuvant-goserelin group ($p=0.03$). In protocol 92-02,[6] patients with T2c–T4 tumours and PSA concentrations below 150 µg/L received goserelin and flutamide for 2 months before and 2 months during radiation and were randomly assigned no further therapy or 24 additional months of goserelin alone. With median follow-up of 4.8 years, the group assigned long-term androgen suppression had significantly better disease-free survival ($p=0.0001$), local control ($p=0.0001$), time to distant metastasis ($p=0.001$), and time to biochemical failure ($p=0.0001$); disease-specific survival was slightly but not significantly higher ($p=0.07$). These studies of radiation, in combination with androgen deprivation in patients with locally advanced prostate cancer, have been criticised, because there was no group receiving hormone treatment only. A current National Cancer Institute of Canada trial is addressing the role of hormone treatment alone,

comparing maximum androgen blockade with maximum androgen blockade plus pelvic irradiation in stages T3–4 N0 M0.

In our trial, the radiotherapy technique was conventional, far from being optimum. The contribution of radiotherapy to local control can be further improved by three-dimensional conformal radiotherapy. That approach enables physicians to delineate the planning target volume more accurately and to increase the dose to the tumour without increasing acute and late toxic effects. Preliminary results of a study[4] comparing conventional irradiation (70 Gy) and three-dimensional conformal radiotherapy (78 Gy) show that for patients with stage T1–2 disease and PSA concentrations above 10 µg/L, 4-year disease-free survival is 55% with 70 Gy and 93% with 78 Gy (p=0.003). Zelefsky and colleagues[25] have shown that 5-year actuarial PSA-defined relapse-free survival is significantly better in patients with intermediate and unfavourable prognosis receiving more than 75.6 Gy ($p<0.05$).

What is an adequate duration of androgen deprivation therapy? It is not yet known. The period of 3 years of adjuvant hormonal treatment was chosen empirically; shortening of this period would reduce costs and sideeffects of androgen deprivation (eg, hot flushes, fatigue, sexual dysfunction) and may be possible, since patients with locally advanced prostate cancer in the late 1990s had less tumour burden and were younger than those of the mid 1980s. This is the rationale for the EORTC equivalence trial 22961 started in 1997, which is comparing surveillance with hormone therapy (triptoreline) for 2.5 years after external irradiation and 6 months of combined androgen blockade. A period of 6 months of maximum androgen blockade followed by reinstitution of hormone therapy in case of relapse could achieve equivalent survival to the 3-year androgen suppression regimen; after 6 months of maximum androgen blockade, any subsequent tumour growth would still allow hormonal treatment of the proliferating androgen-dependent stem cells,[26] before tumour progression due to overgrowth by androgen-independent cells.
In the future, management of locally advanced prostate cancer will certainly be tailored according to prognostic factors, with a possible escalation of the dose and the addition of chemotherapy[27] to hormonal treatment for high-risk categories.

4.6 References

1. WHO Handbook for reporting results of cancer treatment. Geneva: World Health Organization: 1979.
2. Bagshaw MA, Cox RS, Ray GR. Status of radiation treatment of prostate cancer at Standford University. *Monogr Natl Cancer Inst* 1988; **7:** 47–60.
3. Hanks GE, Corn BW, Lee WR, et al. External beam irradiation of cancer prostate: conformal treatment techniques and outcomes for the 1990s. *Cancer* 1995; **75:** 1972–77.
4. Pollack A, Zagars GK, Smith LG, Antonlak JA. Preliminary results of a randomized dose-esacalation study comparing 70 Gy to 78 Gy for the treatment of prostate cancer. *Int J Radiat Oncol Biol Phys* 1999; **45:** 146.

5. Pilepich MV, Caplan R, Byhardt RW, et al. Phase III trial of androgen suppression using goserilin in unfavorable prognosis carcinoma of the prostate treated with definitive radiotherapy: Report of Radiation Therapy Oncology Group Protocol 85-31. *J Clin Oncol* 1997; **15:** 1013–21.
6. Hanks GE, Lu J, Matchay M, et al. RTOG protocol 92-02. A phase III trial of the use of long term androgen suppression following neoadjuvant hormonal cytoreduction and radiotherapy in locally advanced carcinoma of the prostate.Thirty-sixth Annual Meeting. New Orleans, May 18–21, 2000. *Proc Am Soc Clin Oncol* 2000; **19:** 327a (abstr 1284).
7. Pocock SJ, Simon R. Sequential treatment assignment with balancing for prognostic factors in the controlled clinical trial. *Biometrics* 1975; **31:** 103–15.
8. Bolla M, Gonzalez D, Warde P, et al. Improved survival in patients with locally advanced prostate cancer treated with radiotherapy and goserelin. *N Engl J Med* 1997; **337:** 295–300.
9. ICRU Report 29. Dose specification for reporting external beam therapy with photons and electrons. International Commission On Radiation Units And Measurements: Washington DC, 1978.
10. Dusserre A, Garavaglia G, Giraud JY, Bolla M. Quality assurance of the EORTC radiotherapy trial 22863 for prostatic cancer: the dummy run. *Radiother Oncol* 1995; **36:** 229–34.
11. Freedman LS. Tables of the number of patients required in clinical trials using the logrank test. *Stat Med* 1982; **1:** 121–29.
12. Bolla M, Gonzalez D, Warde P, et al. Immediate hormonal therapy improves locoregional control and survival in patients with locally advanced prostate cancer: results of a randomized phase III clinical trial of the EORTC Radiotherapy and Genitourinary Tract Cancer Cooperative Groups. *Proc Am Soc Clin Oncol* 1996; **15:** 238 (abstr 591).
13. Joon DL, Hasegawa M, Sikes C, et al. Supraadditive apoptotic response of R3327-G rat prostate tumors to androgen ablation and radiation. *Int J Radiat Oncol Biol Phys* 1997; **38:** 1071–77.
14. Zietman AL, Prince EA, Nakfoor BM, et al. Androgen deprivation and radiation therapy; sequencing studies using the Shionogi in vivo tumor system. *Int J Radiat Oncol Biol Phys* 1997; **38:** 1067–70.
15. Zietman AL, Nakfoor BM, Prince EA, et al. The effect of androgen deprivation and radiation therapy on an androgen-sensitive murin tumor: an in vitro and in vivo study. *Cancer J Sci Am* 1997; **3:** 31–36.
16. Messing EM, Manola J, Sarosdy M, et al. Immediate hormonal therapy compared with observation after radical prostatectomy and pelvic lymphadenectomy in men with node-positive prostate cancer. *N Engl J Med* 1999; **341:** 1781–88.
17. Soloway MS, Sharifi R, Wajsman Z, et al. Randomized prospective study comparing radical prostatectomy alone versus radical prostatectomy preceded by androgen blockade in clinical stage B2 (T2bNXM0) prostate cancer. *J Urol* 1995; **154:** 424–28.
18. Goldenberg SL, Klotz LH, Jewett MAS, et al. Randomized, controlled study of neoadjuvant reversible androgen withdrawal therapy with cyproterone acetate in the surgical management of localized prostate cancer. *J Urol* 1995; **153:** 254A (abstr).
19. Van Poppel H, Ridder DD, Elgamal AA, et al. Neoadjuvant hormonal therapy before radical prostatectomy decreases the number of positive surgical margins in stage T2 prostate cancer: interim results of a prospective randomized trial. *J Urol* 1995; **154:** 429–34.

20. Overgaard M, Hansen MB, Overgaard J, et al. Postoperative radiotherapy in high-risk postmenopausal breast-cancer patients given adjuvant tamoxifen. *Lancet* 1999; **353:** 1641–48.
21. Wolmark N, Dignam J, Margolese R, et al. The role of radiotherapy and tamoxifen in the management of node negative invasive breast cancer-1?0 cm treated with lumpectomy. *Proc Am Soc Clin Oncol* 2000; **19:** 70a (abstr 271).
22. Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* 1998; **351:** 1451–67.
23. Pilepich MV, Sause WT, Shipley WU, et al. Androgen deprivation with radiation therapy compared with radiation therapy alone for locally advanced prostatic carcinoma: a randomized comparative trial of the Radiation Therapy Oncology Group. *Urology* 1995; **45:** 616–23.
24. Pilepich MV, Winter K, Roach M, et al. Phase III Radiation Therapy Oncology Group (RTOG) trial 86-10 of androgen deprivation before and during radiotherapy in locally advanced carcinoma of the prostate. *Proc Am Soc Clin Oncol* 1998; **17:** 308a (abstr 1185).
25. Zelefsky MJ, Leibel SA, Gaudin PB, et al. Dose escalation with three-dimensional conformal radiation therapy affects the outcome in prostate cancer. *Int J Radiat Oncol Biol Phys* 1998; **41:** 491–500.
26. Akakura K, Bruchovsky N, Goldenberg SL, et al. Effects of intermittent androgen suppression on androgen-dependent tumors: apotosis and serum prostate-specific antigen. *Cancer* 1993; **71:** 2782–90.
27. Prostate cancer management under scrutiny: one man's metaanalysis is another man's poisson. *J Clin Oncol* 1999; **17:** 3371–73.

Chapter 5:

Prognostic factors of biochemical or clinical progression-free survival in a phase III trial of adjuvant irradiation in a population of patients with pT3N0 prostate cancer after radical prostatectomy: Does adjuvant irradiation benefit to all patients?

5.1 Design and first results of EORTC trial 22911.

5.1.1 Abstract

Background: Local failure after prostatectomy can arise in patients with cancer extending beyond the capsule. We did a randomized controlled trial to compare radical prostatectomy followed by immediate external irradiation with prostatectomy alone for patients with positive surgical margin or pT3 prostate cancer.
Methods: After undergoing radical retropubic prostatectomy, 503 patients were randomly assigned to a wait-and-see policy and 502 to immediate post-operative radiotherapy (60 Gy conventional irradiation delivered over 6 weeks). Eligible patients had pN0M0 tumours and one or more pathological risk factor of: capsule perforation, positive surgical margins, invasion of seminal vesicles. Our primary endpoint was biochemical progression-free survival. Analysis was by intention to treat.
Results: The median age was 65 years (IQR: 61-69). After a median follow-up of 5 years, biochemical progression free survival was significantly improved in the irradiated group (74.0%, 98%CI: 68.7-79.3 versus 52.6%, 46.6-58.5; $p<0.0001$). Clinical progression-free survival was also significantly improved ($p=0.0009$). The cumulative rate of loco-regional failure was significantly lower in the irradiated group ($p<0.0001$). Grade 2 or 3 late effects were significantly more frequent in the post-operative irradiation group ($p=0.0005$), but severe toxicity (grade 3 or higher) were rare, with a 5-year rate of 2.6% in the wait-and-see group and 4.2% in the post-operative irradiation group ($p=0.0726$).
Conclusion: Immediate external irradiation after radical prostatectomy improves biochemical progression free survival and local control in patients with positive surgical margin or pT3 prostate cancer who are at high risk of progression. Further follow-up is needed to assess the effect on overall survival.

5.1.2 Introduction

Radical prostatectomy provides excellent control when prostate cancer is confined to the organ[1-3]. For patients with cancer extending beyond the capsule (pT3) the risk of local failure varies from 10 to 50%[4]. Initial concentration of prostate-specific antigen (PSA), Gleason score of the surgical specimen and positive surgical margins are independent predictors of biochemical relapse[5]. Non randomised studies showed that postoperative radiotherapy eradicates the microscopic disease left in the surgical bed[6] and significantly reduces the local relapse and PSA failure rates without any effect on disease free survival[7-8]. In 1992, the European Organisation for Research and Treatment of Cancer (EORTC) initiated a randomized, multicentre, phase III trial to test the hypothesis that immediate radiotherapy after prostatectomy improves progression-free survival in patients classified as pT3N0M0 who are at risk of local relapse and distant dissemination.

5.1.3 Methods

Trial design and participants

Tumour stage was determined according to the 1983 Union Internationale Contre le Cancer criteria[9]. Before entry, all patients underwent the following examinations: PSA test, bone scan, CT or MRI of the abdomen, chest radiography and a complete blood count.
Eligible patients had previously untreated, histologically proven adenocarcinoma of the prostate with a clinical tumour stage T0-3, nodal stage N0, no distant metastases and pathological stage pT2-3 N0 with at least one of the following risk factors: tumour growth beyond the capsule (capsule perforation), positive surgical margins (including the level of the prostate apex where the capsule is non-existent) or invasion of the seminal vesicles. Patients had to be younger than 76 years, with a WHO performance status[10] of 0 or 1. Radiotherapy began within 16 weeks after surgery. The protocol was approved by local or national ethics review committee for each participating center. Informed consent (written or oral) was obtained from all patients in accordance with national laws.

Procedures

Surgery was done before entry into the study. Surgery consisted of retropubic approach, negative ilio-obturator lymphadenectomy, prostatectomy with total removal of the prostate gland and of the seminal vesicles. A unilateral or bilateral nerve-sparing technique was applied provided the procedure did not increase the risk of macroscopically positive surgical margins.
The surgical specimen was marked with India ink over the entire resection margin. The prostate then had to be sectioned transversely from the distal margin to the bladder neck at 3 to 4-mm intervals in transverse planes perpendicular to the rectal surface, and fixed overnight in 10% formalin. Specific sections were taken from the distal urethral margin,

bladder neck margin and junction of the seminal vesicles with the prostate. Sections were stained with hematoxylin and eosin. Margins were defined as positive if malignant cells were in direct contact with the margins. WHO grading was recommended by the protocol review committee, thus Gleason grades were not assessed. Central review of the specimens was not available.

All patients underwent simulation (definition of target volumes and beam callibration) with an urethrogram and rectal enema. Treatment was given by linear accelerators of 5-25MV using a non-three-dimensional planning with an isocentric technique. The dose was specified at the point of intersection of the beam axes, according to International Commission on Radiation Units and Measurements (ICRU) 29[11]. A dose of 50 Gy was given in 25 fractions over 5 weeks to a volume that included the surgical limits from the seminal vesicles to the apex with a security margin to encompass sub-clinical disease in the periprostatic area. A 10-Gy boost was given in 5 fractions over a week to a reduced volume circumscribing the previous landmarks of the prostate with a reduced security margin. Protocol compliance has been checked by a dummy-run procedure[12]. Irradiation started once patients had recovered from surgery and there were no important voiding problems.

Clinical examinations with digital rectal examinations and PSA tests were done at 4, 8 and 12 months after surgery (randomisation), then every 6 months until the end of the 5th year, then every year until death. Chest radiography and bone scans were done every year or in case of clinical or biochemical suspicion of progression. CT-scans and liver ultrasound were used for confirmation of suspected progression.

After surgery, patients were randomly assigned to receive radiotherapy or to be submitted to a wait-and-see policy until local failure; subsequent treatment (radiotherapy or other) was delayed until biochemical or clinical failure, radiotherapy being the recommended treatment in case of local relapse. Randomisation was centralised at the EORTC data centre. A minimisation technique was used with stratification for institution, capsule invasion, positive margins and invasion of seminal vesicles after verification of all eligibility criteria.

Clinical progression-free survival means survival without any evidence of clinical, sonographic, radiographic or scintigraphic recurrence. Local recurrence is documented by digital rectal examination. Biochemical progression was defined as every increase over the lowest postoperative value to a value more than 0.2 µg/L over the lowest postoperative value measured on three occasions at least 2 weeks apart. Biochemical progression-free survival is counted from the day of randomisation to the day of first biochemical or clinical progression or start of treatment in absence of progression, if any. For the endpoints clinical progression and biochemical progression, deaths in absence of progression were censored and analysed as a competing risk.

Acute side effects of radiotherapy were scored according to the WHO scale. The Late Radiation Morbidity Scoring Scheme of the Radiation Therapy Oncology Group/EORTC was used to assess late adverse effects. Quality of life was not analysed and patients did not assess sexual function.

Statistical analysis

The primary endpoint, local control, was changed by the EORTC protocol review committee to clinical progression-free survival by an amendment in March, 1995, because the potential improvement of local control benefits clinical progression-free survival. For this reason the sample size was further increased to 1000 patients in November, 1998 to ensure 80% power of detecting a difference in the 5-year progression-free survival from 60% to 67.5% with a two-sided logrank test at the 5% significance level (440 events needed). In April, 2003, the independent data monitoring committee accepted biochemical progression-free survival as the primary endpoint, on the grounds of the evolving urological practice, to take into account biochemical relapse as well. In December 2003, when 365 events of biochemical progression-free survival had been recorded, the committee authorised early release of the results based on an O'Brien-Fleming stopping rule requiring statistical significance at the 0.02 level. The present follow-up is not long enough to assess either time to distant metastases or survival.

Analysis was by intention to treat. Both end-points were estimated by Kaplan-Meier methods, and comparisons between treatment groups were made with a logrank test with two-sided significance level of 0.02. Endpoints affected by competing risks were assessed by cumulative incidence and Gray tests[13]. Trial design, conduct, and analysis were done at the EORTC independently of all funding bodies.

Role of the funding source

The funding source has no role in study design, data collection, data analysis, data interpretation, or wirting of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

5.1.4 Results

From November 1992 to December 2001, 1005 patients from 37 institutions entered the trial after radical prostatectomy: 503 were assigned to the wait-and-see policy and 502 to adjuvant radiotherapy. The median follow-up was 5 years for both groups. 37 patients (17 and 20 respectively) were deemed ineligible mainly because of inappropriate disease stage (two and six, respectively), previous or concurrent cancer (eight in each group) and prior treatment (three and two), lack of baseline data (one and three), or incomplete initial work up (three and one). Table 5.1 shows baseline characteristics of patients.

Table 5.1: Baseline characteristics

	Wait-and-see (N=503)	**Irradiation (N=502)**	**Total (N=1005)**
Age (years)*	65 (61-69)	65 (61-69)	65 (61-69)
WHO Performance status (PS)			
PS 0	472 (93.8%)	471 (93.8%)	943 (93.8%)
PS 1	29 (5.8%)	26 (5.2%)	55 (5.5%)
PS 2	1 (0.2%)	2 (0.4%)	3 (0.3%)
Missing	1 (0.2%)	3 (0.6%)	4 (0.4%)
Associated chronic disease	142 (28.2%)	127 (25.3%)	269 (26.8%)
Previous short term neo-adjuvant hormonal treatment	51 (10.1%)	50 (10.0%)	101 (10.0%)
Clinical TNM stage[9]			
Clinical T			
T0	2 (0.4%)	4 (0.8%)	6 (0.6%)
T1	84 (16.7%)	87 (17.3%)	171 (17.0%)
T2	338 (67.2%)	316 (62.9%)	654 (65.1%)
T3	79 (15.7%)	94 (18.7%)	173 (17.2%)
Tx	0 (0.0%)	1 (0.2%)	1 (0.1%)
Clinical N			
N0	493 (98.0%)	486 (96.8%)	979 (97.4%)
Nx	9 (1.8%)	13 (2.6%)	22 (2.2%)
Missing	1 (0.2%)	3 (0.6%)	4 (0.4%)
M category			
M0	499 (99.2%)	494 (98.4%)	993 (98.8%)
M1	0 (0.0%)	1 (0.2%)	1 (0.1%)
Mx	3 (0.6%)	4 (0.8%)	7 (0.7%)
Missing	1 (0.2%)	3 (0.6%)	4 (0.4%)
PSA before surgery (µg/L; Hybritech equivalent)*	12.4 (7.2-20.0)	12.3 (7.2-20.6)	12.3 (7.2 – 20.3)
PSA 3 weeks after surgery but before irradiation*	0.2 (0.1-0.4)	0.2 (0.1-0.3)	0.2 (0.1-0.3)
PSA return to normal after surgery			
<=0·2 µg/L within 3 weeks after surgery	345 (68.6%)	353 (70.3%)	698 (69.5%)
>0·2 µg/L 3 weeks post-surgery but ≤0·2 µg/L later in follow-up prior to relapse or further treatment	95 (18.9%)	98 (19.5%)	193 (19.2%)
Remained >0·2 µg/L	62 (12.3%)	46 (9.2%)	108 (10.7%)
unknown	1 (0.2%)	5 (1.0%)	6 (0.6%)

	Wait-and-see (N=503)	Irradiation (N=502)	Total (N=1005)
Pathological factors			
Pathological N category			
pN0	500 (99.4%)	495 (98.6%)	995 (99.0%)
pN+	0 (0.0%)	2 (0.4%)	2 (0.2%)
pNx	2 (0.4%)	2 (0.4%)	4 (0.4%)
Missing	1 (0.2%)	3 (0.6%)	4 (0.4%)
WHO histopathological grade			
G1	57 (11.3%)	69 (13.7%)	126 (12.5%)
G2	327 (65.0%)	303 (60.4%)	630 (62.7%)
G3	115 (22.9%)	122 (24.3%)	237 (23.6%)
Gx	3 (0.6%)	5 (1.0%)	8 (0.8%)
Missing	1 (0.2%)	3 (0.6%)	4 (0.4%)
Individual risk factors			
Capsule perforation	397 (78.9%)	377 (75.1%)	774 (77.0%)
Invasion of seminal vesicles	128 (25.4%)	128 (25.5%)	256 (25.5%)
Positive surgical margin	317 (63.0%)	312 (62.2%)	629 (62.6%)
Combination of risk factors			
No risk factor (ineligible)	0 (0.0%)	2 (0.4%)	2 (0.2%)
Capsule perforation only	127 (25.2%)	139 (27.7%)	266 (26.5%)
Invasion of seminal vesicles only	19 (3.8%)	23 (4.6%)	42 (4.2%)
Positive margin only	79 (15.7%)	84 (16.7%)	163 (16.2%)
Capsule perforation and invasion of seminal vesicles	40 (8.0%)	26 (5.2%)	66 (6.6%)
Capsule perforation and positive margin	169 (33.6%)	149 (29.7%)	318 (31.6%)
Invasion of seminal vesicles and positive margin	8 (1.6%)	16 (3.2%)	24 (2.4%)
All three risk factors	61 (12.1%)	63 (12.5%)	124 (12.3%)

Data are numbers (%) unless otherwise indicated *Median (IQR)

Treatment information was available for all but five patients (Figure 5.1.1). Five patients (1%) in the wait-and-see group were irradiated. In the 457 irradiated patients, irradiation was initiated a median of 90 days after surgery (IQR 67-105) and lasted a median of 44 days (43-47). A four-field isocentric box technique was used for 320 patients (70.0%). Large fields (>9×9 cm equivalent square-field) were used for 426 patients (92.3%) for the first planning volume, and 247 patients (54.0%) were treated with small (≤9×9cm equivalent square field) for the reduced volume. The target total was 60 Gy: 415 patients (90.8%) received exactly 60 Gy, four (0.9%) received less than 60 Gy and 38 (8.3%) a dose greater than 60 Gy. There were a few minor deviations in dose, fractionation schedule, volume

reduction and dose prescription point[12]. In the wait-and-see group, the recommended dose for salvage irradiation was 70Gy/35fractions/7 weeks, and hormonal treatment was very often luteinizing-hormone releasing hormone analogues.

Irradiation was interrupted as a result of toxic effects in 14 patients (3.1%). For eight of these patients, diarrhoea was one or the sole reason for the interruption. In three patients, frequent passage of urine in association with other toxic effects was a reason for interruption. In the other three patients the reasons were proctitis, cystitis with haematuria and anal pain. Ten of the 14 patients experiencing toxic effects received 60 Gy; two received 64 Gy, one 68 Gy and one 74 Gy; this last patient received a higher dose because of biochemical suspicion of local progression during radiation. The acute adverse effects (worst grade recorded during irradiation) were mild to moderate in most patients (Table 5.2).

Table 5.2: Acute adverse effects from radiation

	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4	Missing
Nausea/vomiting	434 (95.0%)	19 (4.2%)	1 (0.2%)	0 (0.0%)	0 (0.0%)	3 (0.7%)
Diarrhea	174 (38.1%)	175 (38.3%)	81 (17.7%)	24 (5.3%)	0 (0.0%)	3 (0.7%)
Bladder Frequency	151 (33.0%)	205 (44.9%)	79 (17.3%)	15 (3.3%)	2 (0.4%)	5 (1.1%)
Dysuria	229 (50.1%)	173 (37.9%)	47 (10.3%)	5 (1.1%)	0 (0.0%)	3 (0.7%)
Skin	279 (61.1%)	144 (31.5%)	29 (6.3%)	2 (0.4%)	0 (0.0%)	3 (0.7%)
Hematuria	433 (94.7%)	17 (3.7%)	4 (0.9%)	0 (0.0%)	0 (0.0%)	3 (0.7%)

Data are numbers (%) of 547 patients in irradiation group who weree actually irradiated

A median of seven follow-up PSA measurements were available in both groups (range 0-21, 16 patients had no measurement at all). A total of 351 failures (220 in the wait-and-see group and 131 in the post-operative radiotherapy group) were recorded (Table 5.3). Biochemical progression free survival (our revised primary endpoint) was significantly higher in the post-operative irradiation group (Figure 5.2 and Table 5.3) (hazard ratio [HR] = 0.48; 98%CI 0.37-0.62). Kaplan-Meier estimates of 5-year biochemical progression-free-survival rates were 52.6% (98%CI 46.6-58.5) and 74.0% (98%CI 68.7-79.3), respectively. The results were unchanged when adjusting for all baseline factors. The treatment benefit was significant in all pathological risk groups (Table 5.4).

Figure 5.1: Trial Profile

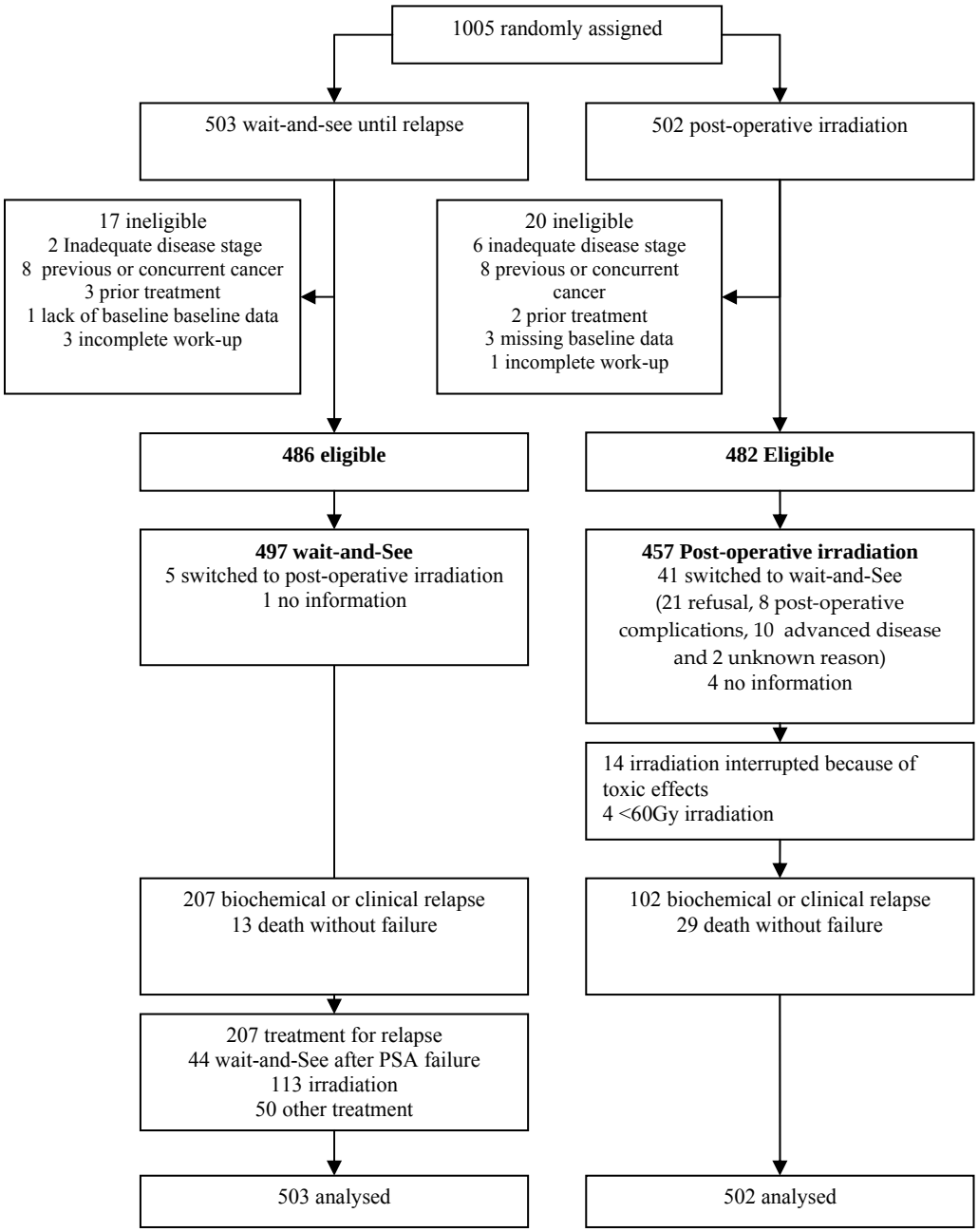


Table 5.3: Events in long term outcome

	Wait-and-See (N=503)	Irradiation (N=502)	Total (N=1005)	Hazard ratio (98%CI)*
Biochemical or clinical progression or death	220 (43.7%)	131 (26.1%)	351 (34.9%)	0·48 (0·37-0·62)
Treated without failure	7	1	8	
PSA failure	188	93	281	
Locoregional failure	11	5	16	
Distant failure	1	3	4	
Death without failure	13	29	42	
Clinical progression or death	113 (22.5%)	75 (14.9%)	188 (18.7%)	0·61 (0·43-0·87)
Loco-regional failure	73	23	96	
Distant failure	18	19	37	
Death without clinical failure	22	33	55	
Death	43 (8.5%)	46 (9.2%)	89 (8.9%)	1·09 (0·67-1·79)
due to prostate cancer	15	8	23	
due to cardiovascular disease	9	10	19	
due to other cancer	12	11	23	
due to other cause	3	12	15	
due to unknown cause	4	5	9	
Second cancer	32 (6.4%)	32 (6.4%)	64 (6.4%)	-

Data are number or number (%). * for irradiation group relative to the wait and see group.

Of the 207 patients who relapsed in the wait-and-see group, 163 (78.7%) were offered an active treatment, and 44 (21.3%) were following a wait-and-see policy for a PSA failure only. Of the 163 treated patients, 56 were treated upon locoregional relapse, 100 upon PSA relapse and seven were treated too early in absence of a relapse (Table 5.5). The first salvage treatment involved irradiation in 113 patients. Other salvage treatment consisted mainly in endocrine therapy. By year 5, 71 (34.3%) patients had started their active salvage treatment. For those in whom treatment was initiated, treatment started a median of 2.2 years after entry on study.

Table 5.4: Treatment effect by postoperative prognostic factors

	Wait-and-See		Irradiation		Treatment effect	
	O / N	% at 5 Years (98% CI)	O / N	% at 5 Years (98% CI)	Hazard Ratio (98% CI)	P
Capsule perforation						
No	43/106	52·2 (38·9-65·5)	27/125	76·4 (66·1-86·7)	0·45 (0·25-0·79)	0·0008
Yes	177/397	52·6 (45·9-59·2)	104/377	73·3 (67·2-79·4)	0·50 (0·37-0·66)	<0·0001
Invasion of seminal vesicles						
No	141/375	59·4 (52·7-66·1)	80/374	78·3 (72·6-84·1)	0·47 (0·34-0·64)	<0·0001
Yes	79/128	32·4 (21·1-43·8)	51/128	61·1 (49·4-72·7)	0·48 (0·31-0·73)	<0·0001
Surgical margins						
Negative	72/186	59·4 (49·5-69·3)	53/190	70·1 (61·0-79·1)	0·66 (0·43-1·01)	0·0207
Positive	148/317	48·3 (41·0-55·7)	78/312	76·2 (69·8-82·6)	0·40 (0·29-0·56)	<0·0001
PSA within 3 weeks post-surgery						
≤0·2µg/L	123/345	59·6 (52·4-66·9)	78/353	78·8 (72·8-84·7)	0·50 (0·36-0·70)	<0·0001
>0·2µg/L	97/157	37·6 (27·9-47·3)	53/144	62·6 (52·1-73·0)	0·46 (0·31-0·68)	<0·0001

Table 5.5: First active salvage treatment for patients who relapsed in the wait-and see group (N=207)

No active treatment (after biochemical failure without clinical failure)	44 (21.3%)
First active treatment for progression	163 (78.7%)
Pelvic radiotherapy	113 (54.6%)
Surgical castration	1 (0.5%)
Hormonal treatment	45* (21.7%)
Other (eg : estracyt, or hormonal treatment with palliative irradiation)	4 (1.9%)
Timing of initiation of the active salvage treatment	
Treated too early	7 (4.3%)
Loco-regional progression (without or after untreated biochemical failure)	52 (31.9%)
Loco-regional progression without biochemical relapse	4 (2.5%)
Biochemical relapse after PSA returned to normal	77 (47.2%)
Biochemical relapse but PSA always remained >0·2 µg/L	23 (14.1%)
Reasons to initiate pelvic radiotherapy as first active salvage treatment	
Treated too early	3 (2.7%)
Loco-regional progression (without or after untreated biochemical failure)	47 (41.6%)
Loco-regional progression without biochemical relapse	4 (3.5%)
Biochemical relapse after PSA returned to normal	48 (42.5%)
Biochemical relapse but PSA always remained > 0·2 µg/L	11 (9.7%)

Data are numbers (%).
* two patients initially treated with irradiation, and given hormonal treatment upon relapse

Most of the 188 failures consisted of loco-regional failure (Table 5.3). Clinical progression free survival was significantly higher in the post-operative irradiation (Figure 5.3 and Table 5.3). The 5-year clinical progression free-survival rates were 77.1% (98%CI: 72.1-82.0) and 86.3% 98%CI: 82.2-90.4) in the wait-and-see and irradiation group, respectively. Time to biochemical failure was longer in the post-operative irradiation group than in the wait-and-see group ($p<0.0001$). The 5-year cumulative incidences of biochemical failure were 21.4% (98%CI: 16.4-26.3) and 44.2% (98%CI: 38.3-50.0) respectively.

The cumulative incidence of loco-regional failure at 5 years were significantly lower in the post-operative irradiation group (5.4%, 2.7 to 8.0) than in the wait-and-see group (15.4%, 11.2 to 19.6; $p<0.0001$; Figure 5.4). Only 59 distant failure events were reported (Table 5.3). The cumulative incidence of this event does not seem to differ by treatment group, but the statistical power is very low and a longer follow-up is needed to assess this endpoint. Most events of clinical failure consisted in loco-regional failure (Table 5.3). At five years, the cumulative incidences of clinical failure were 8.8% (5.4 to 12.2) with post-operative irradiation and 19.0% (14.5 to 23.6) with the wait-and-see policy ($p<0.0001$, Table 5.3). The corresponding values for distant failure were 6.1% (3.2-9.0) and 6.3% (3.4-9.2; Gray $p=0.6689$).

Figure 5.2: Biochemical progression-free survival

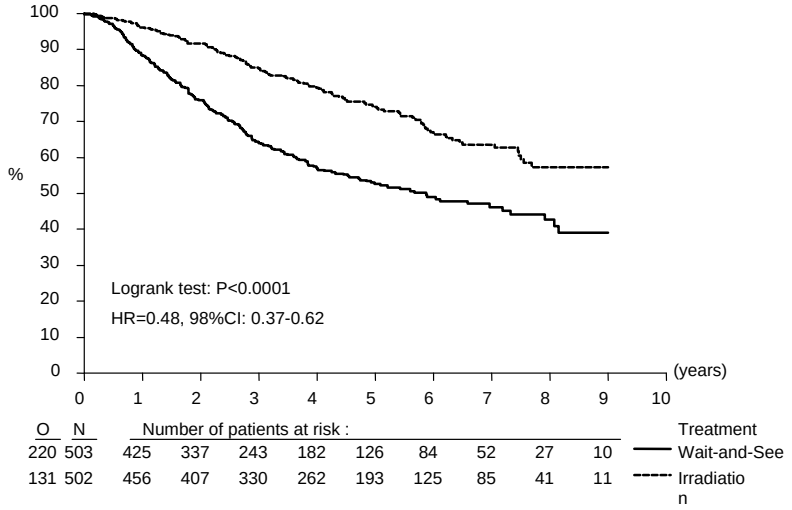


Figure 5.3: Clinical progression-free survival

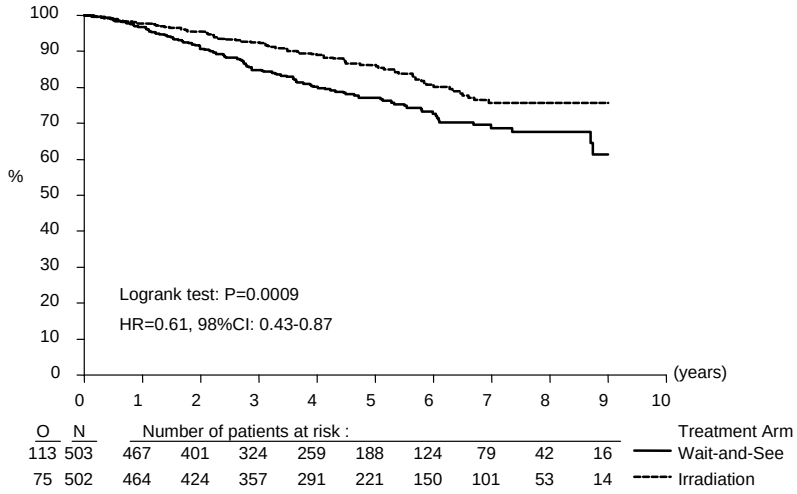
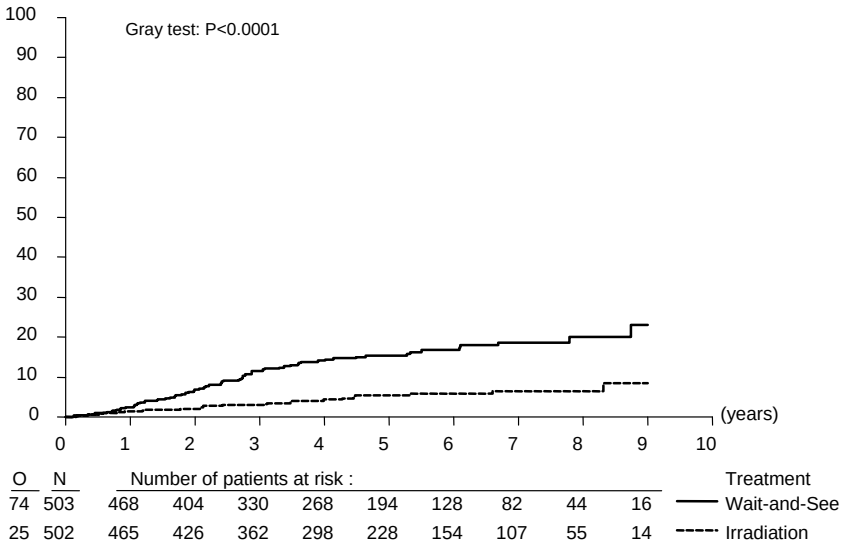


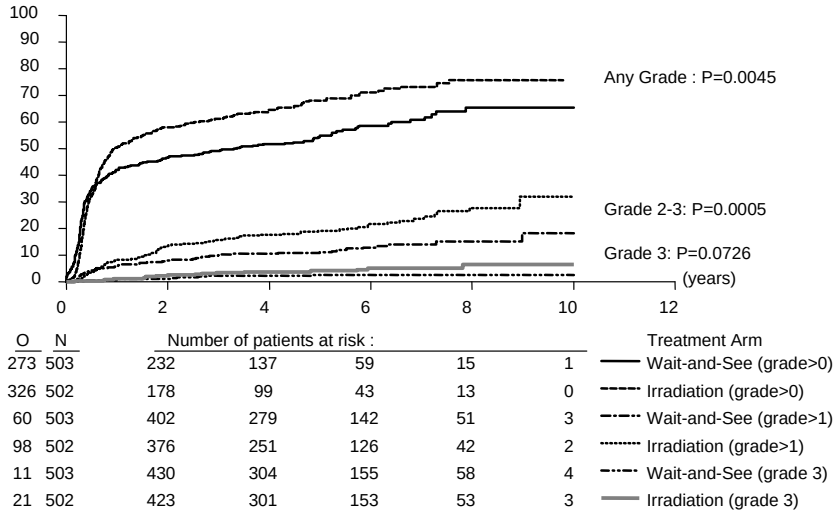
Figure 5.4: Cumulative incidence of loco-regional failure



Although about twice as many patients died of prostate cancer in the wait-and-see group compared with the postoperative irradiation group, longer follow-up is needed to ass this endpoint with sufficient statistical power. However, no significant difference was evident for overall survival (5-year event free rate 93.1%, 98% CI: 90.1-96.2 for wait and see group *vs* 92.3, 89.1-95.5 for irradiation group; logrank p=0.6796; Table 5.3).

Events of grade 3 toxicity were rare and their incidence did not differ significantly between groups. At 5 years, the cumulative incidence of events of grade 3 toxicity was 2.6% (98%CI: 0.8-4.4) in the wait-and-see group and 4.2% (3.4-5.0) in the post-operative irradiation group (p=0.0726, Figure 5.5). No events of grade 4 toxicity were reported. Late effects were more frequent in the post-operative irradiation group, for all types and all grades of late effects or all grade 2 and 3 late effects. Incontinence was not assessed, since it is not mentioned in the Late Radiation Morbidity Scoring Scheme of the RTOG/EORTC. Nevertheless, an interim analysis did not show increased risk for urinary incontinence as a result of post-operative irradiation[14].

Figure 5.5: Cumulative incidence of late complications



5.1.5 Discussion

Our results[15] show a significant improvement in biochemical progression-free survival with immediate post-operative irradiation. Loco-regional failure rate was also significantly reduced with such treatment. Longer follow-up is needed to assess if post-operative irradiation affects the occurrence of distant metastases, survival or both. In the wait-and-see group, deferred post-operative irradiation was given to about half of the relapsing patients: three without evidence of relapse, 51 with clinically documented clinical failure, with or without PSA relapse, and 59 with PSA failure only. Limitations of our analysis include conventional irradiation, the low dose of 60 Gy, a variable postoperative PSA nadir, and variation on the indications and type of salvage treatment in the wait-and-see group. Small retrospectives series provided similar results regarding local control[16-21], one regarding the 5-year freedom from PSA relapse rate[20]. Nevertheless, results similar to ours – from a large randomised trial- have never been reported so far.

The indications of salvage radiotherapy can be divided into three scenarios: clinically palpable or biopsy proven isolated local recurrence (or both); persistently detectable PSA; and delayed PSA rise without any clinically evident disease after initially undetectable postoperative levels. The earlier the initiation of salvage radiotherapy, the better the effect on clinical or biochemical failure free survival[4,22,23], which means that the first two scenarios have the worst outcome. The best situation is the third with a delayed rise of PSA in patients who are likely to have a lower tumoural burden[24]. At the time the trial was running, there were no data based on randomized trials in favour of an immediate salvage

radiotherapy versus a deferred one in case of a rising PSA, and the threshold value of PSA relapse to start with radiotherapy was not known. McCarthy et al.[25] reported that 68% of patients with rising PSA profiles after initially undetectable postoperative levels remained biochemically controlled after radiotherapy, and only 33% of those who had persistent detectable PSA concentrations and were treated with radiotherapy remained free of disease. This is the reason why the results of immediate irradiation might be equivalent to withholding irradiation until the PSA rises to 0.5 and 1 µg/L[26].

Postoperative irradiation was first associated with a rate of moderate or severe late complications of 7-20%, at dose ranging between 60 and 70 Gy[4]. The complication rate was lower in subsequent reports[4,22,27,28] with less grade 3 and 4 side-effects. In this trial, no grade 4 toxic effects were reported and the rate of 5-year grade 3 toxicity was 2.6% in the no further treatment group and 4.2% in the post-operative irradiation group (p=0.0726). The incidence of grade 3 urethral stricture and incontinence, 1.4% (six patients) in both groups is in line with a previous report[14].

10% of patients received a neo-adjuvant hormonal treatment (luteinizing-hormone releasing hormone analogues) for no longer than 3 months, which accounts for a PSA value as low 0.3 µg/L before surgery, without any expected benefit on biochemical recurrence[29].
The clinical tumour-stage distribution of this trial is not typical of contemporary surgical series in which T1c stages are most common. The current rate of positive surgical margins is far lower, as is the median PSA before surgery. For this reason, the results of radical prostatectomy and immediate external irradiation for pathological tumour stage T3 might be even further improved by: (1) treating pT3 patients who were previously clinical stage cT1-2 N0 with a baseline PSA lower than 20 µg/L and a negative post-operative PSA, (2) replacing conventional external irradiation with contemporary conformal radiotherapy and promoting dose escalation up to 64-66 Gy, which improves local control and biochemical disease free survival for patients treated by irradiation alone[30-31].

Long term analysis of the trial is planned in order to confirm if the improvement in biochemical and clinical disease-free survival from immediate post-operative irradiation translates into a survival benefit.

5.1.6 References

1. Consensus statement : the management of clinically localized prostate cancer. National Institutes of Health Consensus Development Panel. Consensus development conference on the management of clinically localized prostate cancer. NCI Monographs 1988;7;3-6.
2. Zincke H, Oesterling JE, Blute ML, Bergstrahl EJ, Myers RP, Barrett DM. Long term results after radical prostatectomy for clinically localized prostate cancer. J Urol 1994;152:1850-57.
3. Blute ML, Bostwick DG, Bergstrahl EJ, Slezack JM, Amling CL, Zincke H. Anatomic site-specific positive margins in organ-confined prostate cancer and its impact on outcome after radical prostatectomy. Urology. 1997;50:733-39.

4. Lennemäs B, Edgren M, Häggman M, Norlén Bo J, and Nilsson S. Postoperative radiotherapy after prostatectomy. A review. Scand J Urol Nephrol. 2003;37:10-15.
5. Kupelian PA, Katcher J, Levin HS and Klein EA. Stage T1-2 prostate cancer: a multivariate analysis of factors affecting biochemical and clinical failures after radical prostatectomy. Int J Radiat Oncol Biol Phys 1997;37:1043-52.
6. Ray Gr, Cassady JR, Bagshaw MA. External beam megavoltage radiation treatment of post-radical prostatectomy residual or recurrent tumour: Preliminary results. J Urol 1975;114:98-101.
7. Vanuytsel L and Van Poppel H. In:Ed Greco C and Zelefsky MJ. Adjuvant radiotherapy following radical prostatectomy. Amsterdam. The Netherlands: Harwood Academic Publishers, 2000:377-84.
8. Perez CA, Beyer DC, Blasko JC et al. Postradical prostatectomy irradiation in carcinoma of the prostate. American College of Radiology. ACR Appropriateness Criteria. Radiology 2000; 215 (Suppl):1419-39.
9. Spiessl B, Bearhrs OH, Hermaneck P, et al.,eds. TNM atlas: illustrated guide to the TNM/pTNM-classification of malignant tumors.3rd ed. Berlin:Springer-Verlag,1989.
10. World Health Organization. WHO Handbook for Reporting Results of Cancer Treatment, Geneva: WHO, 1979.
11. ICRU Report 29. Dose specification for reporting external beam therapy with photons and electrons. Washington DC: International Commission On Radiation Units And Measurements, 1978.
12. Davis B, Reiner B, Dusserre A, Giraud JY, Bolla M. Quality assurance of the EORTC trial 22911. A phase III study of post-operative external irradiation radiotherapy in pathological stage T3N0 prostatic carcinoma : the dummy-run. Radiother Oncol 2002;64:65-73.
13. Gray RJ. A class of K-sample tests for comparing the cumulative incidence of a competing risk. The Annals of Statistics. 1988;16(3):1141-54.
14. Van Cangh PJ, Richard F, Lorge F, et al. Adjuvant radiation therapy does not cause urinary incontinence after radical prostatectomy : Results of a prospective randomized study. J Urol 1998;159:164-66.
15. Bolla M, Van Poppel H, Van Cangh P, et al. Does post-operative radiotherapy after radical prostatectomy improve progression-free survival in pT3N0 prostate cancer (EORTC 22911). Proc Am Soc Clin Oncol. J Clin Oncol 2004;23: 382 (abstr).
16. Zietman AL, Coen JJ, Shipley WU, et al. Adjuvant irradiation after radical prostatectomy for adenocarcinoma of the prostate: analysis of freedom from PSA failure. Urology 1993;42:292-9.
17. Anscher MS, Robertson CN, Prosnitz R. Adjuvant radiotherapy for pathologic stage T3/4 adenocarcinoma of the prostate : ten-year update. Int J Radiat Oncol Biol Phys 1995;33:37-43.
18. Schild SE, Wong WW, Grado G, et al. The results of radical retropubic prostatectomy and adjuvant therapy for pathologic stage C prostate cancer. Int J Radiat Oncol Biol Phys 1996;34;535-41.
19. Elias S, Parker RG, Gallardo D, Law J. Adjuvant radiation therapy after radical prostatectomy for carcinoma of the prostate. Am J Clin Oncol 1997;20:120-124
20. Valicenti RK, Gomella LG, Ismail M, et al. The efficacy of early adjuvant radiation therapy for pT3N0 prostate cancer: a matched analysis. Int J Radiat Oncol Biol Phys 1999;45:53-8.
21. Vargas C, Kestin LL, Weed DW, Krauss D, Vicini FA, and Martinez AA.Improved biochemical outcome with adjuvant radiotherapy after radical prostatectomy for prostate cancer with poor pathologic features. Int J Radiat Oncol Biol Phys 2005;61:714-724.

22. Morris MM, Dallow KC, Zietman AL, et al. Adjuvant and salvage irradiation following radical prostatectomy for prostate cancer. Int J Radiat Oncol Biol Phys.1997;38:731-36.

23. Forman JD, Meetze K, Pontes E, Wood DP, Shamsa F, Rana T et al. Therapeutic irradiation for patients with an elevated post-prostatectomy prostate specific antigen level. J Urol 1997;158:1436-40.

24. Choo R, Hruby G, Hong J, et al. (In)-efficacy of salvage radiotherapy for rising PSA or clinically isolated local recurrence after radical prostatectomy. Int J Radiat Oncol Biol Phys 2002;53:269-76.

25. McCarthy JF, Catalona WJ, Hudson MA. Effect of radiation therapy on detectable serum prostate specific antigen levels following radical prostatectomy; early versus delayed treatment. J Urol 1994;151:1575-78.

26. Nudell DM, Grossfeld GD, Weinberg VK, Roach M III, Caroll PR. Radiotherapy after radical prostatectomy: treatments outcomes and failures patterns. Urology 1999;54:1049-57.

27. Lange PH, Lightner DJ, Medici E, Reddy PK, Vesella RL. The effect of radiation therapy after radical prostatectomy in patients with elevated prostate specific antigen levels. J Urol 1990;144:927-33.

28. Catton C, Gospodarowicz M, Warde P, et al. Adjuvant and salvage radiation therapy after radical prostatectomy for adenocarcinoma of the prostate. Radiother Oncol 2001;59:51-60.

29. Aus G, Abrahamsson PA, Ahlgren G et al. Three-month neoadjuvant hormonal therapy before radical prostatectomy: a 7-year follow-up of a randomized controlled trial. BJU Int. 2002; 90:561-66.

30. Zelefsky MJ, Aschenasy E, Kelsen S, and Leibel SA. Tolerance and early outcome results of postprostatectomy three-dimensional conformal radiotherapy. Int J Radiat Biol Phys 1997;39:327-333.

31. Pollack A, Zagars GK, Smith LG et al. Preliminary results of a randomized radiotherapy dose-escalation study comparing 70 Gy with 78 Gy for prostate cancer. J Clin Oncol 2000;18:3904-3911.

5.2 Assessment of prognostic and predictive factors

5.2.1 Abstract

Background: EORTC trial 22911 demonstrated that immediate postoperative irradiation significantly improved biochemical failure free survival (BPFS) compared to wait-and-see (W&S) until relapse in patients with pT2-3 tumour and pathological risk factors after radical prostatectomy. In this study, we have investigated the heterogeneity of the treatment benefit across defined subgroups of patients.
Methods: Data of 972 eligible patients were used. A risk model was developed in the W&S group. The Logrank test for heterogeneity was applied (α=0.05).
Results: Positive surgical margin (SM+), seminal vesicle invasion (SV+), WHO differentiation grade, pre- and post-operative PSA were independent predictors for BPFS in the W&S group. Men with SV+ were at higher risk of relapse whereas those with SM+ but no capsule infiltration (ECE-) did not seem to differ from those with SM-ECE+ or with SM+ECE+.
Conclusion: Postoperative irradiation improved biochemical progression-free survival in all patient groups. Longer follow-up is needed to assess the endpoint of clinical progression-free survival.

5.2.2 Introduction

Radical prostatectomy is an effective treatment for patients with organ confined prostate cancer[1-5]. However, the recurrence rate after local surgery is largely affected by pathological risk factors such as: extra-prostatic tumor extension (pathological T3), positive surgical margins and high grade disease and by initial PSA. It is estimated that 10 to 15% without and 40 to 60% of patients with these factors will experience biochemical or clinical recurrence following surgery[6-8].

More than two decades ago, postoperative radiotherapy was reported to eradicate the microscopic disease left in the surgical bed[9,10] and retrospective series[11,12] confirmed the significant reduction of local relapse without being able to demonstrate any impact on disease free survival.

In 1992, the European Organisation for Research and Treatment of Cancer (EORTC) initiated a randomized, multi-centre, phase III trial to test the hypothesis that immediate radiotherapy following prostatectomy improves biochemical and clinical progression free survival in patients classified as pT2-3N0M0 [UICC 1993] who are at risk of local relapse and distant dissemination. The first results of the trial (EORTC 22911) were recently published with a median follow-up of 5 years[13]: immediate post-operative irradiation resulted in a statistically significant improvement of the primary endpoint, biochemical progression-free survival (P<0.0001) and of clinical progression-free survival (P=0.0004).

We have now explored whether the observed benefit of immediate irradiation is homogeneous across groups of patients of different absolute risk of clinical or biochemical failure. For that purpose, we first assessed prognostic factors in the patients who were not irradiated immediately to create risk groups. We then assessed whether the prognostic risk groups were also predictive for the treatment effect. Following Freedland et al.[14], we also assessed the prognosis according to the pathological classification in the untreated group and explored the homogeneity of the treatment effect in subsets of patients presenting with various pathological features.

5.2.3 Methods

Patients and trial

The original trial is a randomized multi-centre phase III study comparing immediate post-operative irradiation (60 Gy, external beam) to wait-and-see policy and treatment after progression in prostate cancer patients presenting poor pathological risk factors after radical prostatectomy. Adjuvant hormonal treatment was not foreseen in the protocol. One thousand and five (1005) patients were accrued to the study of which 972 were eligible. Eligible patients were 75 years old or less, with WHO performance status 0-1 and previously untreated pT2-T3N0 prostate cancer (1993 UICC TNM classification). The radical prostatectomy specimen exhibited at least one of the following risk factors: capsule perforation, positive surgical margins (including at the level of the prostate apex where the capsule is non existent) or invasion of seminal vesicles. The study was approved by the ethics committee of each participating institution and was conducted in compliance with the Helsinki declaration. Informed consent was obtained from all patients in accordance with national laws. Randomization was performed centrally by minimization algorithm with stratification for the treating institution, capsule invasion, positive margins and invasion of seminal vesicles.

Endpoints

The primary endpoint is biochemical progression free survival counted from the day of randomization to the day of first biochemical or clinical progression or start of treatment in absence of progression, if any, or death due to any cause. Biochemical progression was defined as every increase over the lowest postoperative value to a value >0.2 ng/ml that is confirmed twice, at minimum 2-week intervals. Local recurrence had to be documented by a digital rectal examination (with or without biopsy) and distant relapse by sonography, radiographic or scintigraphic imaging.

Pathological examination

The surgical specimen had to be marked with India ink over the entire resection margin. The prostate had then to be sectioned transversely from the distal margin to the bladder neck at 3-4 mm intervals in transverse planes perpendicular to the rectal surface, and fixed overnight in 10% formalin. Specific sections were taken from the distal urethral margin, bladder neck margin and junction of the seminal vesicles with the prostate. Sections were stained with hematoxylin and eosin. Margins were defined as positive if malignant cells

were in direct contact with the inked margins. The WHO histopathological grading was used. Gleason grades were not assessed. Central review of the specimens was not available

Statistical methods

Ten baseline factors were screened for prognostic significance in the wait-and-see group: age, presence of other chronic diseases, clinical T category, PSA before surgery and PSA within 3 weeks after surgery (all PSA values were transformed into Hybritech equivalent values), nerve sparing radical prostatectomy, seminal vesicles invasion, capsule perforation, positive surgical margin and WHO histopathological grade. Other collected variables were judged not relevant to the purpose or could not be used because all patients were in the same category (eligibility criteria). The pathological factors were reported by the local pathologist and the Gleason score was not available from that review. Ninety-two patients in the sample had received short term hormonal manipulation prior to surgery (<3months). This variable was not used because its effect represents a selection of patients for short term neo-adjuvant hormonal treatment, which was not foreseen by protocol and was not applied in all centers and because despite this, the patients remained eligible for the study. The continuous variables were transformed into categories using published cut points (PSA) or the median (age). Adjacent levels of discrete variables with small numbers were lumped together.

The Cox proportional hazards model was used for the prognostic factor modeling. A step-wise multivariate selection procedure was applied to select the independent prognostic factors. Statistical significance was set at 0.05 in all analyses. Internal model validation was performed by the bootstrap re-sampling technique that generated 5000 random samples from the original sample on which the model-fitting procedure was independently repeated. This provided a bias-corrected estimate of the area under the receiver operating characteristic curve (ROC AUC)[15] which measures the predictive discrimination of the model: a value of 0.5 indicates no discrimination; a value of 1 indicates perfect discrimination. Model stability was assessed by the frequency of inclusion of each of the component variables into the Cox multivariate models[16] and by the frequency of selection of each admissible multivariate model. The stratified Logrank test and Forest plots were used to test for heterogeneity of the treatment effect between risk groups using meta-analysis methodology[17].

5.2.4 Results

With a median follow-up of 5 years, EORTC Trial 22911[13] showed that post-operative irradiation significantly improved the primary trial endpoint, biochemical progression-free survival (P<0.0001, HR=0.48, 98% CI: 0.37-0.62 based on 351 events) and the secondary endpoint clinical progression-free survival (P=0.0009, HR=0.61, 98%CI: 0.43-0.87 based on 188 events). For 281 of the 351 patients the first failure was a PSA increase (80.0% overall, 85.5% on wait-and-see and 71.0% on post-operative irradiation).

Of the 1005 patients entered in the trial, 972 eligible patients were used for the purpose of the predictive factor analysis (485 irradiated and 487 on wait-and-see) and only the 487 on

wait-and-see were used for identifying prognostic factors. Thirty-three patients were excluded from this analysis for reasons of ineligibility mostly due to previous other cancers or inadequate disease stage. Table 5.6 showing the distribution of the ten considered baseline factors indicates no major difference between the two treatment groups (see Bolla et al.[13] or chapter 5.1 for more details). The ineligible patients were somewhat dissimilar, as could be expected since their disease status was not adequate for inclusion in the study.

Table 5.6: Distribution of the ten baseline factors that were assessed as potential predictors

	Included in analysis set		Excluded (ineligible)	Total
	Wait-and-See (N=487)	Postoperative irradiation (N=485)	(N=33)	(N=1005)
	N (%)	N (%)	N (%)	N (%)
Age				
<=65 y	225 (46.2)	228 (47.0)	21 (63.6)	474 (47.2)
>65y	262 (53.8)	257 (53.0)	12 (36.4)	531 (52.8)
Associated chronic disease				
No	349 (71.7)	357 (73.6)	25 (75.8)	731 (72.7)
Yes	137 (28.1)	125 (25.8)	7 (21.2)	269 (26.8)
Missing	1 (0.2)	3 (0.6)	1 (3.0)	5 (1.4)
Nerve sparing surgical procedure				
No	283 (58.1)	288 (59.4)	18 (54.5)	589 (58.6)
Yes	197 (40.5)	182 (37.5)	15 (45.5)	394 (39.2)
Missing	7 (1.4)	15 (3.1)	0 (0.0)	22 (2.2)
Clinical T Category				
T0-1	83 (17.0)	89 (18.4)	5 (15.2)	177 (17.6)
T2	328 (67.4)	304 (62.7)	22 (66.7)	654 (65.1)
T3	76 (15.6)	91 (18.8)	6 (18.2)	173 (17.2)
Missing	0 (0.0)	1 (0.2)	0 (0.0)	1 (0.1)
WHO G Grade				
G1	57 (11.7)	69 (14.2)	0 (0.0)	126 (12.5)
G2	315 (64.7)	295 (60.8)	20 (60.6)	630 (62.7)
G3	111 (22.8)	114 (23.5)	12 (36.4)	237 (23.6)
Missing	4 (0.8)	7 (1.4)	1 (3.0)	12 (1.2)
Pre-operative PSA				
<=20 ng/ml	365 (74.9)	356 (73.4)	26 (78.8)	747 (74.3)
>20 ng/ml	121 (24.8)	126 (26.0)	6 (18.2)	253 (25.2)
Missing	1 (0.2)	3 (0.6)	1 (3.0)	5 (0.5)
PSA within 3 weeks post-surgery				
<=0.2 ng/ml	334 (68.6)	342 (70.5)	22 (66.7)	698 (69.5)
>0.2 ng/ml	128 (26.3)	123 (25.4)	9 (27.3)	260 (25.9)
Missing	25 (5.1)	20 (4.1)	2 (6.1)	47 (4.7)
Pathological risk factors (individual stratification factors)				
Seminal vesicle invasion	122 (25.1)	122 (25.2)	12 (36.4)	256 (25.5)
Positive surgical margins	306 (62.8)	306 (63.1)	17 (51.5)	629 (62.6)
Capsule perforation	383 (78.6)	365 (75.3)	26 (78.8)	774 (77.0)

	Included in analysis set		Excluded	Total
	Wait-and-See (N=487)	Postoperative irradiation (N=485)	(ineligible) (N=33)	(N=1005)
	N (%)	N (%)	N (%)	N (%)
Pathological stage[1]				
SM+, ECE-	79 (16.2)	82 (16.9)	2 (6.1)	163 (16.2)
SM-, ECE+	125 (25.7)	133 (27.4)	8 (24.2)	266 (26.5)
SM+, ECE+	161 (33.1)	148 (30.5)	9 (27.3)	318 (31.6)
SV+	122 (25.1)	122 (25.2)	12 (36.4)	256 (25.5)
Missing	0 (0.0)	0 (0.0)	2 (6.1)	2 (0.2)

[1]SM = surgical margin, ECE=capsule perforation, SV=seminal vesicle invasion

Prognostic factors for biochemical progression free survival in the wait-and-see arm

In the wait-and-see arm, all baseline factors tested were statistically significant ($P<0.05$) in the univariate analysis except for the presence of associated chronic disease and capsule perforation ($P>0.1$), age and nerve sparing procedure ($P=0.0558$ and $P=0.0633$, respectively). The univariate effect of clinical T category was borderline statistically significant ($P=0.0503$) (Table 5.7). Owing to the likely presence of correlations between factors, the multivariate model selection procedure was initiated with a model that contained all ten variables. After model reduction, five factors were retained as independent predictors of increased risk of clinical or biochemical failure: a PSA>0.2 ng/ml within 3 weeks post-surgery (and prior to irradiation, if any), invasion of seminal vesicles, elevated (>10 ng/ml or >20 ng/ml) PSA prior to surgery, poor tumor differentiation (WHO G grade) and positive surgical margins (Table 5.8). Of note, when this model was applied to the irradiated group, pre-operative PSA and positive surgical margins were no longer statistically significant and the associated hazard ratios were closer to unity ($P=0.2698$ and $P=0.1353$, respectively).

Model validation

The bias-corrected area under the ROC curve for this model with five variables was 0.65. The model was also the most frequently selected reduced multivariate model in the bootstrap validation process. Table 5.9 displays the inclusion frequency of the 10 variables tested in the 5000 Bootstrap samples final multivariate models: pre- and post-operative PSA and seminal vesicles invasion were selected in >85% of the models, WHO G grade and positive surgical margins in >75% of the models, all other variables were selected in less than 30% of the models.

Table 5.7: Univariate analysis of biochemical progression-free survival in the wait-and-see arm

	N	O	Hazard Ratio (95% CI)	P	5-year event-free rate (95% CI)
Age group					
<=65 y[1]	225	87	1.00	0.0558	58.4 (51.3-65.6)
>65y	262	127	1.31 (0.99-1.72)		47.5 (40.5- 54.5)
Associated chronic disease					
No[1]	349	160	1.00	0.3148	51.6 (45.7- 57.5)
Yes	137	54	0.85 (0.63- 1.16)		54.8 (44.9- 64.6)
Nerve sparing procedure					
No[1]	283	136	1.00	0.0633	48.6 (42.0- 55.1)
Yes	197	73	0.76 (0.57- 1.02)		58.7 (50.7- 66.7)
Clinical T Category					
T0-1[1]	8	30	1.27 (1.00- 1.62)	0.0503	61.1 (49.5- 72.7)
T2	32	143			52.6 (46.4- 58.8)
T3	76	41			43.6 (31.1- 56.1)
WHO G Grade[2]					
G1[1]	57	20	1.68 (1.32- 2.13)	<0.0001	63.8 (50.7- 77.0)
G2	315	131			56.8 (50.5- 63.0)
G3	111	63			32.4 (21.9- 43.0)
Pre-operative PSA[2]					
<=10 ng/ml[1]	194	67	1.41 (1.19- 1.67)	0.0001	62.2 (54.2- 70.1)
10-<=20 ng/ml	171	74			50.6 (41.7- 59.4)
>20 ng/ml	121	73			40.4 (31.1- 49.7)
PSA within 3 weeks post surgery					
<=0.2 ng/ml[1]	334	121	1.00	<0.0001	59.2 (53.0- 65.4)
>0.2 ng/ml	128	80	2.09 (1.58- 2.78)		36.3 (27.3- 45.3)
Seminal vesicle invasion					
Not reported[1]	365	139	1.00	<0.0001	59.0 (53.3- 64.7)
Present	122	75	2.10 (1.58- 2.78)		32.6 (22.76-42.4)
Positive surgical margin					
Not reported[1]	181	71	1.00	0.0308	59.0 (50.6 - 67.5)
Present	306	143	1.37 (1.03- 1.83)		48.4 (42.1 - 54.7)
Capsule perforation					
Not reported[1]	104	42	1.00	0.5136	52.5 (41.2 - 63.8)
Present	383	172	1.12 (0.80- 1.57)		52.4 (46.7 - 58.0)

[1] reference category for the calculation of the hazard ratios; [2]: linear trend
O = number of events, N = number of patients

Table 5.8: Multivariate prognostic factor model for biochemical progression-free survival in the wait-and-see arm

Variable Label	Parameter Estimate[1]	SE	Chi-Square	P	Hazard Ratio	95% CI
Positive surgical margins	0.434	0.153	8.05	0.0046	1.54	1.14-2.08
Invasion of the seminal vesicles	0.653	0.153	18.14	<.0001	1.92	1.42-2.59
G Grade: G1* *vs* G2 *vs* G3	0.412	0.123	11.27	0.0008	1.51	1.19-1.92
Pre-operative PSA: ≤10ng/ml* *vs* 10-20 *vs* >20ng/ml	0.314	0.090	12.13	0.0005	1.37	1.15-1.63
PSA within 3 weeks post surgery: >0.2 ng/ml	0.669	0.146	20.90	<.0001	1.95	1.47-2.60

[1]coefficients for building the prognostic index * indicates reference category

Table 5.9: Frequency of inclusion of the individual baseline factors in the final models obtained on the B=5000 bootstrap samples

Variable	Frequency of inclusion among the B=5000 Bootsrap sample models N (%)
Post-operative PSA	4940 (98.8)
Invasion of seminal vesicles	4892 (97.8)
Pre-operative PSA	4385 (87.7)
WHO G grade	4219 (84.4)
Positive surgical margins	3803 (76.1)
Associated chronic disease	1393 (27.9)
Age	1148 (23.0)
Nerve sparing procedure	1088 (21.8)
Clinical T category	940 (18.8)
Capsule perforation	428 (8.6)

Risk groups

A prognostic index (PI) was calculated for each patient by adding the coefficients attached to each of the 5 risk factors in the final model, whenever the risk factor was present for that given patient: PI = 0 + 0.669×[post-operative PSA>0.2ng/ml] + 0.653×[seminal vesicle invasion] + 0.314× [pre-operative PSA>10 ng/ml] + 0.314× [pre-operative PSA>20 ng/ml] + 0.412× [G grade -1] + 0.434× [positive surgical margins].

Risk groups were constituted by classifying the patients on the wait-and-see arm according to tertiles of the prognostic index (low risk: 0 to ≤0.85, intermediate risk: >0.85 to ≤1.55, high risk: >1.55 to ≤3.21). Figure 5.6 shows the biochemical progression-free survival by

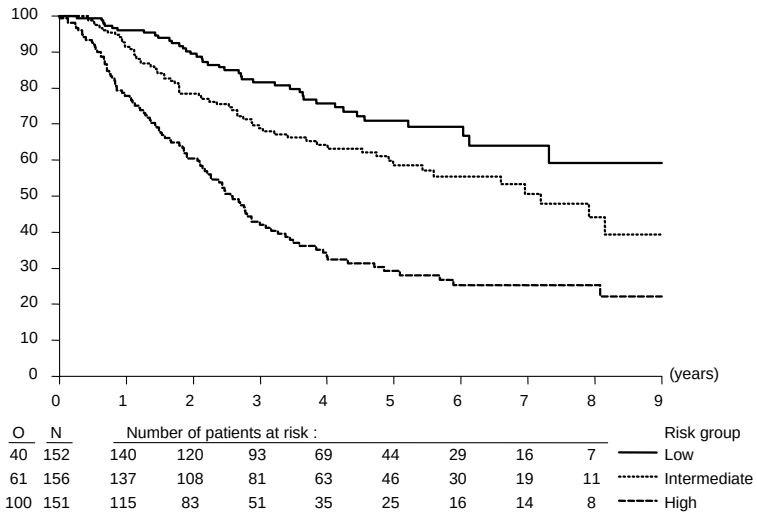
risk group in the wait-and-see arm (development set) and Table 5.10 the 5-year event free rates and hazard ratios.

Table 5.10: Risk groups in the two treatment arms

	Wait-and-See arm				Post-operative irradiation arm			
Risk group	N	O	HR (95% CI)	5-Year event free rate (95% CI)	N	O	HR (95% CI)	5-Year event free rate (95% CI)
Low	152	40	1.00	70.9% (62.4 – 79.4)	146	20	1.00	84.4% (77.4-91.4)
Intermediate	156	61	1.65 (1.10-2.46)	58.7% (49.8 – 67.5)	179	44	1.83 (1.08-3.11)	77.9% (70.8-84.9)
High	151	100	3.63 (2.50-5.26)	29.3% (21.2 - 37.4)	136	51	3.28 (1.95-.50)	61.0% (51.8-70.3)

N = number of patients, O = number of events, CI= confidence interval, HR= hazard ratio

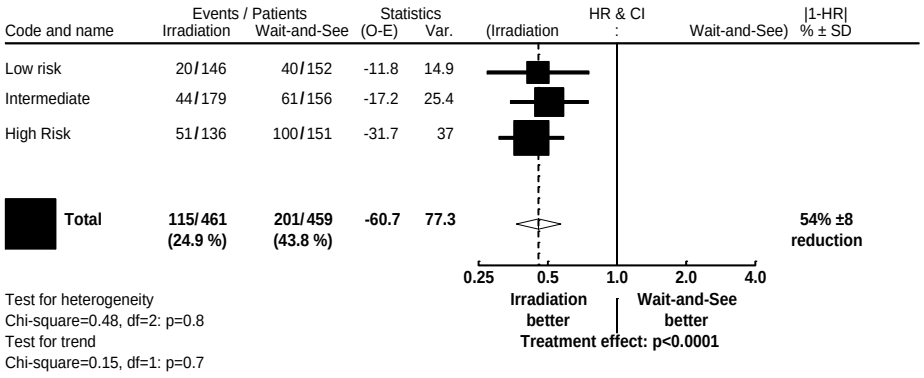
Figure 5.6
Biochemical progression-free survival by risk group in the Wait-and-See arm (development set)



Predictive effect of the prognostic risk groups

The forest plot of the effect of post-operative irradiation on biochemical progression-free survival in the three risk groups (Figure 5.7) indicates no substantial heterogeneity in the magnitude in the hazard ratios for the benefit of postoperative irradiation across the three risk groups (Logrank for heterogeneity P>0.1) for the endpoint of biochemical progression-free survival.

Figure 5.7 Forest plot of the effect of post-operative irradiation on biochemical progression free survival, in the risk groups derived from the multivariate Cox model developed in the wait-and-see arm



The center of the squares is the hazard ratio, bars represent 95% confidence interval, the size of the square is proportionate to the number of events

Predictive effect of the pathological risk factors

The potential predictive value of the individual pathological risk factors was assessed, as well as that of pre- and post-operative PSA and WHO differentiation grade. The three pathological factors were stratified for at randomization. The test for heterogeneity of treatment effects indicated borderline statistically significant heterogeneity in relation to margin status (Figure 5.8, P=0.0568). The hazard ratio was 0.61 (95% CI: 0.43-0.88, P=0.0082) in the margin negative group and 0.40 (95% CI: 0.30-0.52, P<0.0001) in the group with positive margins. The analysis indicated no heterogeneity of the trial results according to pre-operative PSA, post-operative PSA or tumor differentiation (P>0.1).

Prognostic and predictive effect of pathological stage

The biochemical progression-free survival for the groups presenting with positive margins without extra capsular extension (SM+, ECE-), extra capsular extension only (SM-, ECE+),

with both factors (SM+, ECE+) or with seminal vesicle invasion, irrespective of margin status (SV+) is depicted in Figure 5.9 (also see Table 5.6 for the distribution in the two treatment arms). The SV+ subgroup fares much worse than the other three subgroups (P=0.0003). Based on the available data, there was no significant difference between the SM-,ECE+ subgroup and the two subgroups with SM+ (P>0.1). There was no statistically significant heterogeneity in treatment effect across patient groups defined by this classification (Figure 5.10, Logrank P>0.1) although the SM-,ECE+ subgroup appears on the right of the forest plot closer to the unity hazard ratio.

Figure 5.8
Forest plot of the effect of post-operative irradiation on biochemical progression-free survival in subgroups defined by the pathological stratification factors extracapsular extension (ECE), surgical margin (SM) and seminal vesicle invasion (SV)

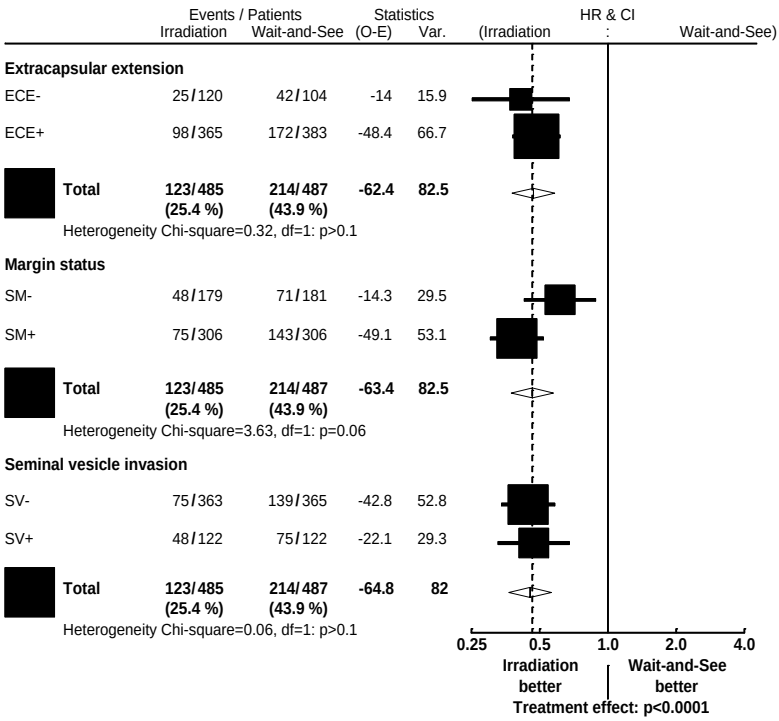
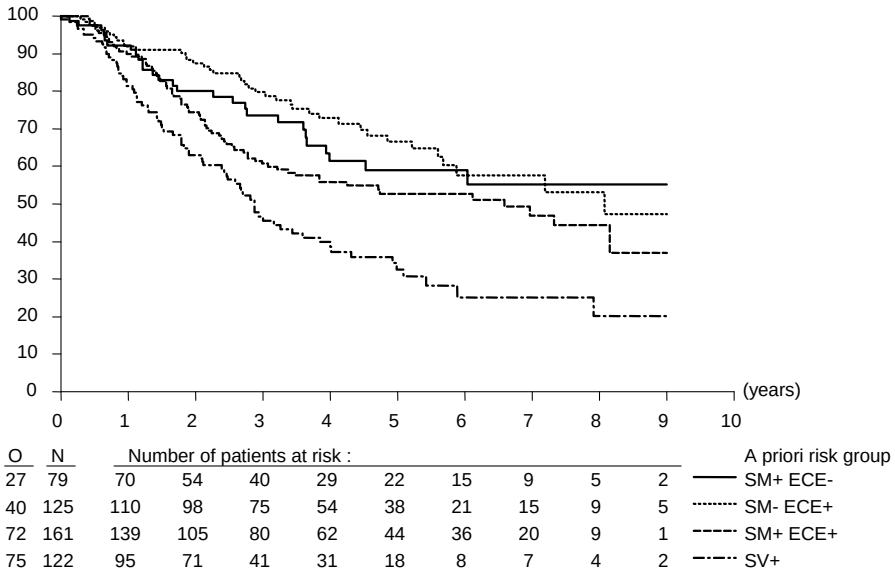


Figure 5.9
Biochemical progression-free survival by pathological stage in the Wait-and-See arm



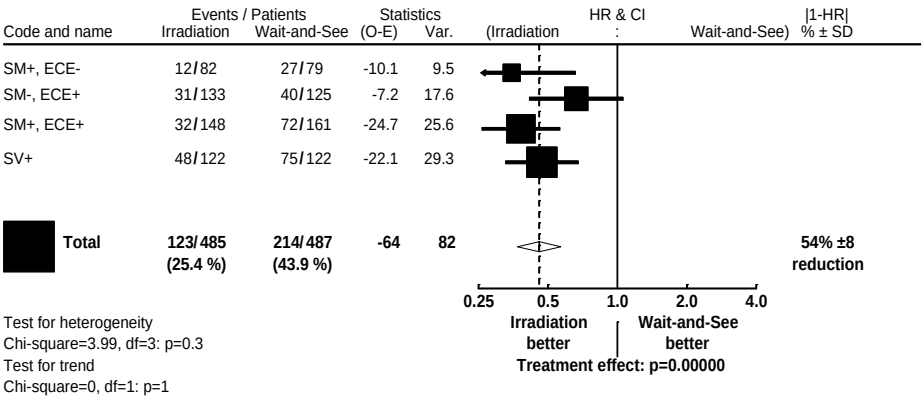
Hazard ratio (HR) for SM-ECE+ vs SM+ECE-: 0.82 (95% CI: 0.50-1.33, P=0.4194), HR for SM+ECE+ vs SM+ECE-: 1.30 (95% CI: 0.83-2.01, P=0.2486). HR for SV+ vs SM+ECE-: 2.25 (95% CI: 1.45-3.48, P=0.0003). Logrank P<0.0001 for the overall comparison of the three groups.

5.2.5 Discussion

Today, the urologists' and radiation oncologists' opinion regarding the value of post-prostatectomy adjuvant radiation in patients presenting with poor post-operative risk factors remains varied[18-20]. The EORTC trial 22911 recently demonstrated a benefit of post-prostatectomy irradiation for preventing biochemical and clinical progression in patients with clinical T0-3N0M0 prostate cancer and at least one pathological risk factor of capsule perforation, positive surgical margins and/or invasion of seminal vesicles[13].

We investigated the homogeneity of the benefit observed in trial 22911 across patient subgroups. We postulated that the benefit of immediate irradiation was greatest in patients who were predicted to be at higher risk of relapse.

Figure 5.10
Forest plot of the effect of post-operative irradiation on biochemical progression-free survival by pathological stage



Our results suggest that the benefit of immediate irradiation as regards biochemical progression-free survival was substantial in all patient subsets, whether formed on trial-based prognostic factors or on pre-defined pathological classification. Statistics supported only the fact that the subset of patients with negative margins may benefit to a somewhat lesser extent than the other subgroups (Heterogeneity P=0.0568). The patients with extra capsular extension as sole risk factor also appeared on the plots to benefit less of immediate irradiation. This statement was not substantiated statistically (heterogeneity P>0.1). In addition, this further subset represented 70% of the margin negative cases and this strongly suggests that the two observed effects are in fact directly related.

For the purpose of assessing the absolute risk of biochemical progression-free survival, we also assessed prognostic factors in the wait-and-see arm. Our results identified pre-operative PSA, post-operative PSA, WHO differentiation grade, invasion of seminal vesicles and positive margins as prognostic factors. Those factors were already identified in validated pre-operative[21-24] as well as post-operative nomograms[8,25] and in numerous other reports[3,26-27 and others]. Our primary objective was not to improve on those models. We could not apply Partin's nomogram in our study, because the Gleason score from the local pathology was not available in the database as the WHO differentiation grade was used in the study. However, a central review pathology exercise is currently ongoing. The charts of all patients entered at the largest recruiting centers will be reviewed. The results of this review will be the object of a subsequent publication. For similar reasons, other factors were not available, which were shown to improve the prediction models such as the proportion of the gland with tumor-bearing high grade histological features and the extent of positive margins[3,27-29]. When applied to the group of patients who were irradiated post-operatively, pre-operative PSA level and the presence of positive surgical margins were no longer statistically

significant. This is obviously related to some predictive value of those two factors, but also suggests that the effect of irradiation superseded the negative impact of these two factors. In a more general perspective, this observation calls attention on the fact that the impact of prognostic factor may be modulated by the (adjuvant) treatments. This ought to be taken into account when designing trials comparing new treatment strategies.

On the basis on the local pathology report, we could identify the pathological stage of all patients. We assessed the biochemical progression-free survival of the non irradiated patients by pathological stage: patients presenting with positive margins only (pT_2), with extra capsular extension (pT3a) with or without positive margins and with invasion of the seminal vesicles (pT3b). Like Freedland and colleagues on the SEARCH database[14], we observed that patients with positive surgical margin without extra capsular extension were at similar risk of failure as men presenting with extra capsular extension with or without positive surgical margins but without invasion of seminal vesicles. Although our results are based on a much smaller subgroup, they are suggestive that men with positive surgical margins following radical prostatectomy might be considered as having pathological stage T3 disease. However, confirmation is required as to the prognosis of this patient subgroup in regard to the risk of distant metastases and disease-related mortality. As indicated earlier, all groups seemed to benefit of immediate irradiation in our series. We also have to observe that current radiation techniques would allow the delivery of an optimized and higher radiation dose and that the overall trial results might have been even more significant in the context of current selection criteria and irradiation standards.

We did not explore the endpoint clinical progression-free survival. The reason is that with currently available 5 years of median follow-up in EORTC trial 22911, little is known as yet as regards the events of distant relapse and death. As yet, the majority of events of biochemical/clinical relapse consisted in biochemical failure, and only 217 events of clinical relapse or death have been reported, most of which consisted in loco-regional failure. A longer follow-up is therefore needed to assess this endpoint and the endpoint of overall survival. The long term follow-up is eagerly awaited, since PSA failure is not yet validated as a surrogate neither for clinical relapse nor for overall survival in the post-operative indication[30].

5.2.6 References

1. Consensus statement : the management of clinically localized prostate cancer. National Institutes of Health Consensus Development Panel. Consensus development conference on the management of clinically localized prostate cancer. NCI Monographs 1988, **7 ,**3-6.
2. Zincke H, Oesterling JE, Blute ML, Bergstrahl EJ, Myers RP, Barett DM. Long term results after radical prostatectomy for clinically localized prostate cancer. *J Urol* 1994, **152,** 1850-1857.
3. Blute ML, Bostwick DG, Bergstrahl EJ, Slezack JM, Amling CL, Zincke H. Anatomic site-specific positive margins in organ-confined prostate cancer and its impact on outcome after radical prostatectomy. *Urology* 1997, **50**, 733-739.

4. Gerber GS, Thisted RA, Chodak GW, *et al.* Results of radical prostatectomy in men with locally advanced prostate cancer: multi-institutional pooled analysis. *Eur Urol* 1997, **32,** 385-390.
5. Holmberg L, Bill-Axelson A, Helgesen F, *et al.* A randomized trial comparing radical prostatectomy with watchful waiting in early prostate cancer. *N Engl J Med* 2002, **347,** 781-789.
6. Kupelian PA, Katcher J, Levin HS, Klein EA. Stage T1-2 prostate cancer: a multivariate analysis of factors affecting biochemical and clinical failures after radical prostatectomy. *Int J Radiation Oncology Biol Phys* 1997, **37,** 1043-1052.
7. J.D. Forman: The metamorphosis of a post-prostatectomy patient: from curable to palliative. *Urol Oncol* 2004, **22,** 48-49.
8. Khan MA, Partin AW, Mangold LA, Epstein JI, Walsh PC. Probability of biochemical recurrence by analysis of pathologic stage, Gleason score, and margin status for localized prostate cancer. *Urology* 2003, **62,** 866-871.
9. George FW, Carlton CE, Dykhuisen RF, Dillon JR. Cobalt-60 telecurietherapy in the definitive treatment of carcinomas of the prostate : a preliminary report. *J Urol* 1965, **158,** 515-518.
10. Ray Gr, Cassady JR, Bagshaw MA. External beam megavoltage radiation treatment of post-radical prostatectomy residual or recurrent tumor: Preliminary results. *J Urol* 1975, **114**, 98-101.
11. Vanuytsel L, Janssens G, Van Poppel H, Rijnders A, Baert L. Radiotherapy for PSA recurrence after radical prostatectomy. *Eur Urol* 2001, **39,** 425-429.
12. Lennemäs B, Edgren M, Häggman M, Norlén Bo J, Nilsson S. Postoperative radiotherapy after prostatectomy. A review. *Scand J Urol Nephrol* 2003, **37**, 10-15.
13. Bolla M, van Poppel H, Collette L, *et al.* Postoperative radiotherapy improved clinical and biocochemical progression free survival after radical prostatectomy: first results of EORTC trial 22911. *Lancet* 2005, **366,** 572-578**.**
14. Freedland SJ, Aronson W, Presti JC Jr, *et al.* Should a positive surgical margin following radical prostatectomy be pathological stage T2 or T3? Results from the SEARCH database. *J Urol* 2003, **169**, 2142-2146.
15. Harrell FE, Lee KL, Mark DB. Multivariate prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996, **15**, 361-387.
16. Sauerbrei W, Schumacher M. A bootstrap resampling procedure for model building: application to the Cox regression model. *Stat Med* 1992, **11**, 2093-2109.
17. Early Breast Cancer Trialists' Collaborative Group. Treatment of early breast cancer, vol 1: worldwide evidence 1985-1990. Oxford, Oxford University Press, 1990.
18. Rodrigues G, D'Souza D, Crook J, Malone S, Sathya J, Morton G. Contemporary management of prostate cancer: a practice survey of Notario genitourinary radiation oncologists. *Radiother Oncol* 2003, **69**, 63-72.
19. Morris SL, Parker C, Huddart R, Horwich A, Dearnaley D. Current opinion on adjuvant and salvage treatment after radical prostatectomy. *Clin Oncol (R Coll Radiol)* 2004, **16**, 277-282.
20. Valicenti RK, Gomella LG, Perez CA. Radiation therapy after radical prostatectomy: a review of the current issues and options. *Semin Radiat Oncol* 2003, **13**, 130-140.
21. D'Amico AV, Whittington R, Malkowicz SB, *et al.* Pretreatment nomogram for prostate-specific antigen recurrence after radical prostatectomy or external-beam radiation therapy for clinically localized prostate cancer. *J Clin Oncol* 1999, **17,** 168-172.

22. Kattan MW, Eastham JA, Stapleton AM, Wheeler TM, Scardino PT. A preoperative nomogram for disease recurrence following radical prostatectomy for prostate cancer. *J Natl Cancer Inst* 1998, **90**, 766-771.
23. Greene KL, Meng MV, Elkin EP, *et al.* Validation of the Kattan preoperative nomogram for prostate cancer recurrence using a community based cohort: results from cancer of the prostate strategic urological research endeavor (capsure). *J Urol* 2004, **171**, 2255-2259.
24. Graefen M, Karakiewicz PI, Cagiannos I, *et al.* A validation of two preoperative nomograms predicting recurrence following radical prostatectomy in a cohort of European men. *Urol Oncol* 2002, **7**, 141-146.
25. Partin AW, Mangold LA, Lamm DM, Walsh PC, Epstein JI, Pearson JD. Contemporary update of prostate cancer staging nomograms (Partin Tables) for the new millennium. *Urology* 2001, **58,** 843-848.
26. Nelson CP, Dunn RL, Wei JT, Rubin MA, Montie JE, Sanda MG. Contemporary preoperative parameters predict cancer-free survival after radical prostatectomy: a tool to facilitate treatment decisions. *Urol Oncol* 2003, **21**, 213-218.
27. Freedland SJ, Terris MK, Csathy GS, *et al.* Preoperative model for predicting prostate specific antigen recurrence after radical prostatectomy using percent of biopsy tissue with cancer, biopsy Gleason grade and serum prostate specific antigen. *J Urol* 2004, **171**, 2215-2220.
28. D'Amico AV, Whittington R, Malkowicz SB, *et al.* Clinical utility of the percentage of positive prostate biopsies in defining biochemical outcome after radical prostatectomy for patients with clinically localized prostate cancer. *J Clin Oncol* 2000, **18**, 1164-1172.
29. Grossfeld GD, Latini DM, Lubeck DP, Mehta SS, Carroll PR. Predicting recurrence after radical prostatectomy for patients with high risk prostate cancer. *J Urol* 2003, **169**, 157-163.
30. Jhaveri FM, Zippe CD, Klein EA, Kupelian PA. Biochemical failure does not predict overall survival after radical prostatectomy for localized prostate cancer: 10-year results. *Urology* 1999, **54**, 884-890.

Chapter 6:

Is baseline quality of life useful for predicting overall survival in patients with metastatic hormone refractory prostate cancer?

6.1 Abstract

Background: Patients with symptomatic metastatic hormone-resistant prostate cancer (HRPC) survive a median of 10 months and are often regarded as a homogeneous group. Few prognostic factors have been identified so far. We examined whether baseline health-related quality of life (HRQOL) parameters assessed by the European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire C30 (EORTC QLQ-C30) were independent prognostic factors of survival and whether they bring extra precision to the predictions achievable with models based on clinical and biochemical factors only.
Methods: Data of 391 symptomatic (bone) metastatic HRPC patients from three randomized EORTC trials were used in multivariate Cox proportional hazards models. The significance level was set at α=.05.
Results: Of the 391 patients, 371 died, most of prostate cancer. Bone scan result, performance status, hemoglobin level, and insomnia and appetite loss as measured by the EORTC QLQ-C30 were independent predictors of survival. This model's area under the receiver operating curve was 0.65 compared with 0.63 without the two HRQOL factors.
Conclusion: Certain HRQOL sores, at baseline, seem to be predictors for duration of survival in HRPC. However, such measurements do not add to the predictive ability of models based only on clinical and biochemical factors.

6.2 Introduction

Symptomatic metastatic hormone-refractory prostate cancer (HRPC) patients are usually considered as a homogeneous group whose prognosis remains poor, with a median overall survival of 10 months.[1] However, there is evidence that within this category of patients, subgroups have differing prognostic profiles and experience different durations of survival.[2-13]

To gain insight into the prognosis of those patients is essential to the physician's daily need to inform and support the patient as effectively as possible and is also of great importance to the patient himself.[14] Patients with a short life expectancy are generally less willing to undergo extensive treatment with possible severe side effects.[14] Finally, prognostic factors and expected survival are useful for designing clinical trials because they influence the choice of therapy and the patient selection and their knowledge can increase the statistical

power of comparisons by reduction of the model variability.[15] To be of use in daily practice, prognostic models should be composed of factors that are easily assessable in clinical practice and should provide precise predictions.

A number of published reports address prognostic factors for survival of (bone) metastatic HRPC patients[2-13] (Table 6.1). The prognostic factors identified so far are biochemical (baseline prostate-specific antigen [PSA], lactate dehydrogenase [LDH], alkaline phosphatase, acid phosphatase, hemoglobin, and creatinine),[2-13] objective (number of bone metastases, duration of response to primary hormonal treatment, Gleason sum, age, and weight),[2,5-7,9,11,13] or subjective (performance status[2,5-8,11-13] and disease-related symptoms[2,4,7,8]). Among the disease-related symptoms, pain,[2,8] anorexia and appetite loss,[2,8] obstructive voiding problems,[2] fatigue,[4] and impairment of physical functioning[7,8] have been identified as prognostic factors in four studies.

There are theoretical arguments to support the hypothesis that health-related quality-of-life (HRQOL) parameters may be important prognostic factors. For example, there are arguments that patients' overall HRQOL is influenced by psychosocial factors, such as stress, social support, emotional expression, and coping strategies,[16-18] and that these psychosocial factors seemed to be of prognostic value for survival in a few studies in cancer.[19-21] HRQOLfactors were reported to be independent prognostic factors in four HRPC studies.[2,4,7,8]

However, HRQOL was measured by means of validated patient selfassessment questionnaires in only one study.[8] In the other three studies, the symptoms were assessed either by the doctor[2] or by means of ad hoc questionnaires.[4,7] Because the use of validated[22,23] patient self-assessment[24,25] questionnaires is a prerequisite for measuring HRQOL reliably, their relevance as prognostic factors remains to be confirmed.

Since 1990, the European Organisation for Research and Treatment of Cancer (EORTC) Genitourinary Group has conducted three randomized clinical trials in patients with (bone) metastatic HRPC. HRQOL was measured in all studies by means of the EORTC Quality of Life Questionnaire C30 (QLQ-C30; version 1.0) before the start of the treatment and during follow-up. The HRQOL data from these studies have been used together with previously identified clinical and biochemical parameters to identify independent prognostic factors for overall survival. The predictive ability of the resulting prognostic model is compared to that of models based on clinical and biochemical factors only.

6.3 Patients and Methods

The data from three randomized EORTC HRPC trials were used (EORTC 30903,[26] EORTC 30921,[27] and EORTC 30944[28]). The trials were conducted in accordance with the Helsinki Declaration and were approved by the local institutions' ethical committees. All patients

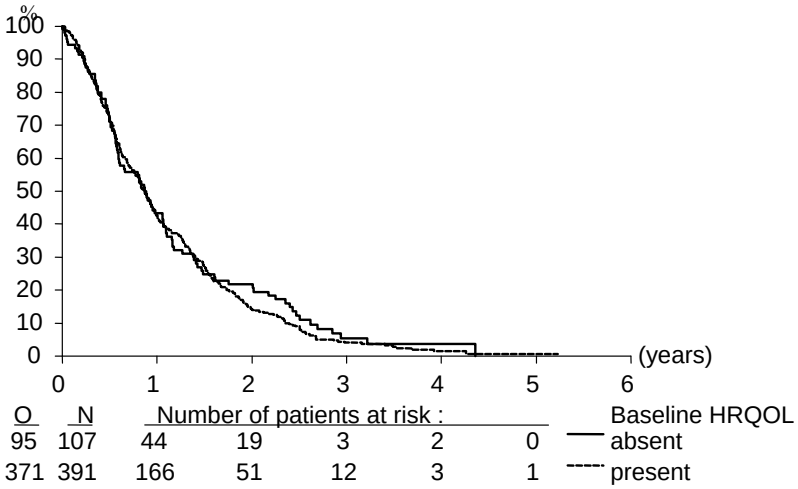
Table 6.1: Prognostic factors in hormone refractory prostate cancer

	N	Objective Factors	Biochemical Factors	Subjective Factors, Assessed by Doctor	Subjective factors, Patient Self-assessment
Emrich[2]	1020	Age, Tumor grade, response status to prior hormonal treatment	Anemia, alkaline phosphatase, acid phosphatase	Performance status, pain, anorexia, obstructive symptoms	
Petrylak[3]	164		Alkaline phosphatase, LDH		
Fossa[4]	58		Hemoglobin, PSA		*Fatigue*
Fossa[5]	152	Duration of response to prior hormonal treatment	Hemoglobin	Performance status	
Fossa[6]	224	Duration of response to prior hormonal treatment	Creatinine, alkaline phosphatase	Performance status	
Fossa[7]	137	Duration of response to prior hormonal treatment	Alkaline phosphatase	Performance status	*Physical functioning*
Tannock[8]	161		Alkaline phosphatase,	Performance status, pain	*Physical* function in one model, *appetite loss* in the other model
Vollmer[9]	348	Weight	Hemoglobin, PSA, PSA velocity		
Kantoff[10]	242		Hemoglobin, alkaline phosphatase, PSA, LDH		
Dowling[11]	133	Age	Hemoglobin, PSA response	Performance status	
Smaeletz[12]	409		Hemoglobin, alkaline phosphatase, albumin, LDH	Performance status	
Halabi[13]	1147	Gleason sum	Hemoglobin,alkaline phosphatase, PSA, LDH	Performance status	

Abbreviations : LDH, lactate dehydrogenase ; PSA, prostate-specific antigen

provided informed consent before randomization. Two hundred one patients with painful metastatic HRPC were entered onto the phase III trial 30903 and randomly assigned to treatment with prednisone (5 mg qid orally [PO]) or flutamide (250 mg tid PO). Two hundred five patients with similar characteristics were randomly assigned in the phase III trial 30921 to receive either one intravenous injection of 150 MBq (4 mCi) of strontium-89 chloride or palliative local radiotherapy with the aim to control pain from osseous metastases. Last, 92 patients were entered onto the phase II trial 30944 and randomly assigned to receive chemotherapy consisting of estramustine phosphate (10 mg/kg/d PO) with or without vinblastine (4 mg/m2 intravenously, weekly for 6 weeks every 8 weeks).

Figure 6.1
Overall survival with (n=391) and without (n=107)
baseline health related quality of life (HRQOL)



O indicates the number of deaths, and N indicates the number of patients

Baseline HRQOL assessment was obtained within 1 week of randomization before the start of the treatment. It was available for 391 (78.5%) of the 498 patients. The group without baseline HRQOL had to be excluded from the analysis. Because the groups with and without baseline HRQOL had similar baseline characteristics (Table 6.2) and overall survival (Figure 6.1), the analyzed subset can be regarded as being representative of the whole. Of the 391 patients included in the analysis, 371 patients (94.9%) have died; 326 of the patients (87.9%) died of prostate cancer. The median survival was 10.4 months (95% CI, 9.2 to 11.5 months), and the 1-year survival rate was 42.7% (95% CI, 37.8% to 47.6%).

The EORTC QLQ-C30 version 1.0 was used to assess the HRQOL in the three trials. This validated measure is one of the most frequently used in cancer clinical trials in Europe and has been granted for use in over 3,000 clinical trials.[29] It consists of 30 items addressing the functioning and symptoms of cancer patients, as well as their general well-being. From the 30 questions, six multi-item function scales are built (physical functioning, role functioning,

emotional functioning, cognitive functioning, social functioning, and global health status/quality of life), along with eight symptom scales (three multi-item scales addressing fatigue, nausea and vomiting, and pain and five single items addressing dyspnea, insomnia, appetite loss, constipation, and diarrhea) and one single-item scale assessing financial difficulties. The scales were transformed linearly to 0 to 100 range according to the EORTC scoring guidelines,[30] so that a higher scale score represents a better level of functioning, whereas for symptom scales, a higher score indicates more symptoms or problems.

Table 6.2: Patient characteristics: Clinical and Biochemical factors

	Baseline HRQOL		
	Absent (N=107) N (%)	Present (N=391) N (%)	Total (N=498) N (%)
Age, years			
<=65	20 (18.7)	94 (24.0)	114 (22.9)
66-75	55 (51.4)	175 (44.8)	230 (46.2)
>75	32 (29.9)	122 (31.2)	154 (30.9)
Median (range)	70.6 (51.6-87.5)	70.8 (34.3-89.3)	70.7 (34.3-89.3)
Previous orchidectomy[a]	57 (53.3)	194 (49.6)	251 (50.4)
Previous treatment with LH-RH[a]	51 (47.7)	212 (54.2)	263 (52.8)
No. of hot spots on initial bone scan			
0-4	22 (20.6)	74 (18.9)	96 (19.3)
5-15	30 (28.0)	130 (33.2)	160 (32.1)
>15/superscan	49 (45.8)	173 (44.2)	222 (44.6)
Missing	6 (5.6)	14 (3.6)	20 (4.0)
Initial pain score			
0, no analgesia	6 (5.6)	28 (7.2)	34 (6.8)
1-2, non narcotic analgesia	52 (48.6)	234 (59.8)	286 (57.4)
3-4, narcotic analgesia	48 (44.9)	128 (32.7)	176 (35.3)
Missing	1 (0.9)	1 (0.3)	2 (0.4)
WHO Performance Status			
WHO Performance Status 0	16 (15.0)	52 (13.3)	68 (13.7)
WHO Performance Status 1	51 (47.7)	222 (56.8)	273 (54.8)
WHO Performance Status 2-3	40 (37.4)	117 (29.9)	157 (31.5)
Alkaline phosphatase, WHO grade			
Grade 0, ≤1.25×UNL[b]	35 (32.7)	108 (27.6)	143 (28.7)
Grade 1, 1.26 – 2.5×UNL[b]	23 (21.5)	110 (28.1)	133 (26.7)
Grade >=2, >2.5×UNL[b]	40 (37.4)	151 (38.6)	191 (38.4)
Missing	9 (8.4)	22 (5.6)	31 (6.2)
Hemoglobin, WHO grade			
Grade 0, ≥11 g/dl	75 (70.1)	318 (81.3)	393 (78.9)
Grade >0, <11 g/dl	26 (24.3)	68 (17.4)	94 (18.9)
Missing	6 (5.6)	5 (1.3)	11 (2.2)
Baseline PSA[c]			
<=10 ×UNL[b]	19 (17.8)	97 (24.8)	116 (23.3)
>10-100 ×UNL[b]	52 (48.6)	177 (45.3)	229 (46.0)
>100 ×UNL[b]	24 (22.4)	96 (24.6)	120 (24.1)
Missing	12 (11.2)	21 (5.4)	33 (6.6)

[a]23 patients had both orchiectomy and LH-RH; [b]upper normal limit; [c]prostate-specific antigen

Table 6.3: Patient characteristics: HRQOL Factors (N=391)

		Scores				
	N (%)	Median	Q1	Q2	Q3	Q4
Function scales[a]						
Global health status/quality of life						
<=50	240 (61.4)	50.0	0	33.3	66.7	100
>50	149 (38.1)					
Missing	2 (0.5)					
Physical Functioning						
<=60	127 (32.5)	60.0	0	40.0	80.0	100
>60	261 (66.8)					
Missing	3 (0.8)					
Role Functioning						
<=50	232 (59.3)	50.0	0	50.0	100	100
>50	153 (39.1)					
Missing	6 (1.5)					
Emotional Functioning						
<=66.7	199 (50.9)	66.7	0	50.0	91.7	100
>66.7	190 (48.6)					
Missing	2 (0.5)					
Cognitive Functioning						
<=83.3	240 (61.4)	83.3	0	66.7	100	100
>83.3	149 (38.1)					
Missing	2 (0.5)					
Social Functioning						
<=66.7	202 (51.7)	66.7	0	50.0	100	100
>66.7	187 (47.8)					
Missing	2 (0.5)					
Symptom scales[b]						
Fatigue						
<=44.4	221 (56.5)	44.4	0	22.2	66.7	100
>44.4	169 (43.2)					
Missing	1 (0.3)					
Nausea/Vomiting						
<=0	245 (62.7)	0.0	0	0	16.7	100
>0	144 (36.8)					
Missing	2 (0.5)					
Pain						
<=50	199 (50.9)	50.0	0	33.3	66.7	100
>50	192 (49.1)					
Dyspnoea[c]						
<=0	198 (50.6)	0.0	0	0	33.3	100
>0	186 (47.6)					
Missing	7 (1.8)					
Insomnia[c]						
<=33.3	254 (65.0)	33.3	0	0	66.7	100
>33.3	134 (34.3)					
Missing	3 (0.8)					

		Scores				
	N (%)	Median	Q1	Q2	Q3	Q4
Appetite Loss[c]						
<=0	199 (50.9)	0.0	0	0	33.3	100
>0	191 (48.8)	(0;0;33.3;100)				
Missing	1 (0.3)					
Constipation[c]						
<=33.3	268 (68.5)	33.3	0	0	66.7	100
>33.3	119 (30.4)	(0;0;66.7;100)				
Missing	4 (1.0)					
Diarrhoea[c]						
<=0	305 (78.0)	0.0	0	0	0	100
>0	82 (21.0)	(0;0;0;100)				
Missing	4 (1.0)					
Financial difficulties[a,c]						
<=0	306 (78.3)	0.0	0	0	0	100
>0	81 (20.7)	(0;0;0;100)				
Missing	4 (1.0)					

HRQOL=health related quality of life, Q=quartile
Numbers are N=number of patients (%)
[a] Best score = 100
[b] Best score = 0
[c] Single-item scales can only take four different values: 0, 33.3, 66.7 or 100

The 15 standard scales of the EORTC QLQ-C30 were used in the analysis together with nine clinical or biochemical items that were recorded at baseline in the three trials. These included the age of the patient, whether the patient had an orchiectomy or had previously received luteinizing hormone-releasing hormone analogs, the number of hot spots on the bone scan (0, <5, 5-15, or >15/superscan), the initial pain score (0 = no analgesics, 1 = occasional use of non-narcotic analgesics, 2 = regular use of non-narcotic analgesics, 3 = occasional use of narcotic analgesics, and 4 = regular use of narcotic analgesics), the baseline WHO performance status, alkaline phosphatase level (WHO grade), hemoglobin level (WHOgrade), and baseline PSA (as multiple of the upper normal limit). Other laboratory measurements (bilirubin, serum creatinine, and blood cell counts) were normal in the vast majority of the patients and were, therefore, of no use to this analysis. The information on visceral and lymph node involvement was not collected in all trials and could not be used for modeling.

The distribution of the discrete variables was assessed, and categories with small numbers (< 25) were pooled together. The quality-of-life scales were dichotomized at the median level. The continuous variables of age and PSA were categorized with cut points at the first and third quartiles. The distribution of the 24 resulting variables is presented in Table 6.3. The Cox proportional hazards model stratified for trial and treatment was used.[31] Preselection for entry of factors into the model-building strategy required significance at the α=.05 level in the univariate analysis. A backward multivariate selection procedure was then applied.[32] Internal model validation was performed by the bootstrap resampling technique with 500 bootstrap samples,33 providing a bias-corrected estimate of the area

under the receiver operating characteristic curve (ROC AUC). The latter measures the predictive discrimination of the model (a value of 0.5 indicates no discrimination and a value of 1 indicates perfect discrimination).[33] The model validation was performed using an S-plus (version 3.3; Statistical Sciences, Seattle, WA) function that is available in the public domain.[34]

6.4 Results

The results of the univariate analysis are listed in Table 6.4. All variables except previous orchiectomy, previous use of luteinizing hormone-releasing hormone, cognitive functioning, diarrhea, dyspnea, pain, and financial difficulties were selected for inclusion in the initial step of the model selection procedure.

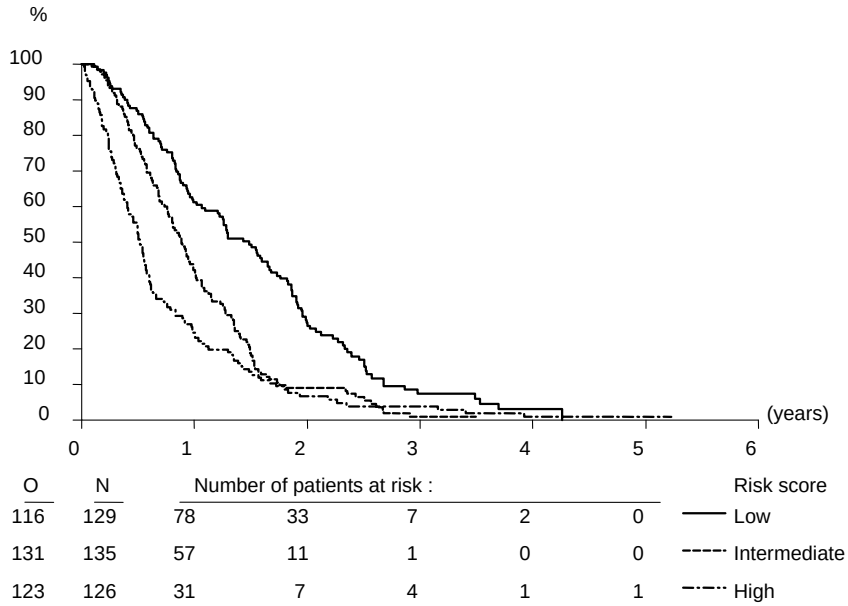
Table 6.4. Univariate Cox regression analysis (N=391)

Variable	P[a]	Hazard ratio *vs* Reference level[b]	95% CI
Clinical/biochemical factors			
Age: ≤ 65 *vs* 66-75 *vs* >75	0.0179	1.19	1.03 - 1.39
Previous orchidectomy: no *vs* yes	0.1781	0.87	0.70 - 1.07
Previous LH-RH: no *vs* yes	0.0843	1.20	0.98 - 1.49
No. of bone lesions on initial bone scan: 0-4 *vs* 5-15 *vs* >15/superscan	<.0001	1.53	1.32 - 1.77
Initial pain score: 0 *vs* 1-2 *vs* 3-4	0.0009	1.36	1.13 - 1.63
Initial WHO Performance Status: 0 *vs* 1 *vs* 2-3	<.0001		
Alkaline Phosphatase: ≤1.25×UNL *vs* 1.26-2.5×UNL *vs* >2.5×UNL	<.0001	1.37	1.19 - 1.57
Hemoglobin: ≥11 g/dl *vs* <11 g/dl	<.0001	2.18	1.65 - 2.88
Initial PSA: ≤10×UNL *vs* >10-100×UNL *vs* >100×UNL	0.0052	1.23	1.06 - 1.43
HRQOL factors			
Global Health Status/QoL: ≤50 *vs* >50	0.0194	0.77	0.62 - 0.96
Physical functioning: ≤60 *vs* >60	0.0001	0.62	0.50 - 0.78
Role functinoning: ≤50 *vs* >50	0.0001	0.64	0.52 - 0.80
Emotional functioning: ≤66.7 *vs* >66.7	0.0006	0.69	0.56 - 0.85
Cognitive functioning: ≤83.3 *vs* >83.3	0.1658	0.86	0.69 - 1.07
Social functioning: ≤66.7 *vs* >66.7	0.0162	0.77	0.62 - 0.95
Fatigue: ≤44.4 *vs* >44.4	0.0018	1.41	1.14 - 1.75
Nausea/vomiting: ≤0 *vs* >0	0.0003	1.50	1.20 - 1.87
Pain: ≤50 *vs* >50	0.0811	1.21	0.98 - 1.49
Dyspnoea: ≤0 *vs* >0	0.0723	1.21	0.98 - 1.49
Insomnia: ≤33.3 *vs* >33.3	0.0008	1.45	1.17 - 1.81
Appetite loss: ≤0 *vs* >0	<.0001	1.70	1.37 - 2.10
Constipation: ≤33.3 *vs* >33.3	0.0007	1.49	1.18 - 1.87
Diarrhea: ≤0 *vs* >0	0.3811	0.88	0.67 - 1.16
Financial Difficulties: 0 *vs* >0	0.5742	1.08	0.83 - 1.40

[a]stratified test (for trend when there are > 2 categories); [b]First value in each comparison is the reference value

At the end of the model selection process, only five variables were retained as independent prognostic factors; these included initial performance status, bone scan result, hemoglobin level, appetite loss, and insomnia (Table 6.5). Age was last to drop out of the model, with P=.0635. Low hemoglobin level (< 11 g/dL) was associated with a 90% increase in the risk of death. This five-factor model was the most frequently selected model during the bootstrap validation. The bias-corrected ROC AUC for the final model was 0.65. The results were similar when hemoglobin, (logarithm of) alkaline phosphatase, and (logarithm of) PSA were considered as continuous variables in the model. The survival probabilities for groups of patients defined by the tertiles of this prognostic model are displayed in Figure 6.2.

Figure 6.2
Overall survival by tertiles of the prognostic index from the five-variable multivariate model



The median overall survival times were 17.8, 10.6 and 6.2 months in the low-, intermediate- and high-risk groups, respectively. O indicates the number of deaths, and N indicates the number of patients

Table 6.5: Final multivariate model with HRQOL factors

Variable	Parameter Estimate	SE	Hazard Ratio	95% CI	P
Initial WHO Performance Status					
0* *vs* 1	0.11	0.0171	0.89	0.64-1.25	0.0322[a]
0* *vs* 2-3	0.23	0.0191	1.26	0.86 -1.83	
No. of bone lesions on initial bone scan, 0-4* *vs* 5-15 *vs* >15/superscan	0.36	0.078	1.43	1.23 -1.67	<.0001
Hemoglobin, ≥11 g/dl* *vs* <11 g/dl	0.64	0.150	1.90	1.42 -2.55	<.0001
Insomnia[b], ≤33.3* *vs* >33.3	0.37	0.121	1.45	1.15 -1.84	0.0020
Appetite loss[c], ≤0* *vs* >0	0.38	0.121	1.47	1.16 - 1.86	0.0015

* indicates the reference category
[a] df=2
[b] answer to the question "Have you had trouble sleeping?"
[c] answer to the question "Have you lacked appetite?"

To assess the increase in predictive ability of our model in comparison to a model that does not include theHRQOL factors, the model-building process was repeated using only the nine clinical and biochemical factors. Only the three factors of WHO performance status, bone scan result, and hemoglobin were retained in the model (Table 6.6). Age was again borderline significant (P=.0692). The bias-corrected ROC AUC for this model was 0.63. The bias-corrected ROC AUC for a model using only HRQOL factors was 0.59.

Table 6.6: Final multivariate model without the HRQOL factors

Variable	Parameter Estimate	SE	Hazard Ratio	95% CI	P
Initial WHO Performance Status					
0* *vs* 1	0.02	0.0168	1.02	0.74-1.42	0.0107[a]
0* *vs* 2-3	0.39	0.0185	1.48	1.03-2.13	
No. of bone lesions on initial bone scan, 0-4* *vs* 5-15 *vs* >15/superscan	0.38	0.076	1.46	1.25-1.69	<.0001
Hemoglobin, ≥11 g/dl* *vs* <11 g/dl	0.65	0.147	1.92	1.44-2.57	<.0001

* indicates the reference category
[a] df=2

6.5 Discussion

In this multivariate analysis of overall survival in 391 (bone) metastatic HRPC patients, only the symptom items of appetite loss and insomnia from the EORTC QLQ-C30 were retained as independent prognostic factors of overall survival, in addition to three clinical or biochemical factors (the hemoglobin level, the number of bone metastases on the initial bone scan, and the initial WHO performance status). The factor WHO performance status had a P value of .0322 only, and the group with a performance status of 1 did not

significantly differ from the relatively small group (13.3%) with a performance status of 0. None of the HRQOL function scales were selected in the multivariate model. The model evaluation requires that a bone scan, a blood sample, and the completed EORTC QLQ-C30 questionnaire be available. The ROC AUC for this model is 0.65. This value means that, with these five factors, one is able to predict which of two patients will live longer in 65% of the cases.

Of the two HRQOL factors identified as independent prognostic factors in our analysis, only appetite loss was identified earlier as an independent prognostic factor.8 Several factors that proved to be of independent prognostic value in earlier prostate cancer studies, such as physical functioning,[7,8] pain,[2,8] and fatigue,[4] were not identified as such in our analysis. Neither was global health status/quality of life, a factor that was identified as an important prognostic factor in more general cancer patient populations.[35,36] In our data, physical functioning and role functioning were substantially correlated with WHO performance status. Each became statistically significant whenever WHO performance status was forced out of the model. The model validation, however, retained the model with WHO performance status as providing the best prediction. It is surprising that, in our study, global health status/ quality of life was not statistically significant even in absence of the factor of WHO performance status. Fatigue was also associated with appetite loss and insomnia; however, this factor was not statistically significant even in absence of the other two factors.

This evaluation confirms that, in HRPC, there is an association between the duration of survival with HRPC and the patient's subjective assessment of their wellbeing. The statistical significance of HRQOL factors in multivariate prognostic models indicates that these factors are useful in predicting overall survival on average over groups of patients.

However, HRQOL assessments by means of questionnaires are not always easy to obtain. They require extra efforts for the patient and add an extra burden on the shoulders of the clinical or nursing staff at the hospital.[17,19] Therefore, before considering a prognostic model including such factors for routine use in clinical practice, it is worthwhile to assess the amount of improvement to the accuracy of the prediction of the survival of individual patients that such factors bring when they are added to prediction models based on clinical factors only.[15,37]

Smaletz et al[12] and Halabi et al[13] developed risk scores by means of nomograms on two large databases of 409 and 1,101 HRPC patients, respectively. In each, the nomograms were built on the basis of seven clinical and biochemical factors and did not include HRQOL. Smaletz et al[12] used Karnofsky performance status, hemoglobin, alkaline phosphatase, albumin, LDH, baseline PSA, and age. PSA and age were not statistically significant predictors of survival in the multivariate analysis. Internal and external validation of their nomogram achieved a ROC AUC of 0.71 and 0.67, respectively. In Halabi et al,[13] the nomogram used the presence of visceral disease, Gleason sum, Eastern Cooperative Oncology Group performance status, hemoglobin, alkaline phosphatase, LDH, and baseline

PSA. Visceral disease was not statistically significant in the multivariate analysis. The ROC AUC for the model of Halabi et al was 0.68 on the validation set.

It is of note that the patient population in Halabi et al had a somewhat better prognosis than our study population, with a median survival of 13 months, and a better performance status. Of the factors used in those two models, ours includes only WHO performance status and hemoglobin level. Alkaline phosphatase and PSA were not statistically significant in our multivariate model. However, when the factor of bone involvement was forced out of the model, alkaline phosphatase was significant and accounted for about half the effect assigned to the bone scan in our model. Albumin and LDH, visceral disease, and Gleason score were not collected in the EORTC studies. Without these variables, we could not assess directly the extra precision that the addition of HRQOL would bring when added to the models of Halabi et al or Smaletz et al. Therefore, we compared the ROC AUC of the five-factor model containing HRQOL factors with that of the best prediction model that could be developed on our database when using only the clinical and biochemical factors.

This simplified model included only WHO performance status, hemoglobin, and the number of hot spots on the bone scan and achieved a ROC AUC of 0.63. This was lower than the ROC AUCs achieved by the two nomograms, which could be anticipated from important factors in the models of Halabi et al and Smaletz et al not being available in the EORTC database. Nevertheless, it demonstrates that the addition of the HRQOL component to the three-factor model adds only 1% to the model predictive ability. This is very little and is less than what is obtained by adding clinical information such as Gleason sum, alkaline phosphatase, or LDH in the two nomograms.

From this, we conclude that HRQOL factors, even if they have been shown to be independent prognostic factors in several studies including ours, are not particularly useful for predicting the overall survival of individual HRPC patients once clinical and biochemical factors are taken into account. Given the extra burden they require and the little extra precision they bring to these predictions, we suggest that models based on clinical and biochemical factors should be used whenever all the factors are available from routine practice. When they are not available, only a rough prediction may be obtained on the basis of HRQOL factors only. However, our study shows that other prognostic factors should be investigated to further increase the predictive ability of the existing prediction models proposed by Halabi et al[13] and Smaletz et al.[12] In that search, the statistical significance of the new factors should not be the only focus, and there must be a careful assessment of the predictive ability of the models and of the extra precision to the predictions that is achieved when the new factors are added to existing models.

6.6 References

1. Eisenberger MA: Chemotherapy for endocrine resistant cancer of the prostate. Prog Clin Biol Res 359:155-164, 1990
2. Emrich LJ, Priore RL, Murphy GP, et al: Prognostic factors in patients with advanced stage prostate cancer. Cancer Res 45:5173- 5179, 1985

3. Petrylak DP, Scher HI, Li Z, et al: Prognostic factors for survival of patients with bidimensionally measurable metastatic hormone-refractory prostatic cancer treated with single-agent chemotherapy. Cancer 70:2870-2878, 1992
4. Fossa SD, Paus E, Lindegaard M, at al: Prostate-specific antigen and other prognostic factors in patients with hormone-resistant prostatic cancer undergoing experimental treatment. Br J Urol 69:175-179, 1992
5. Fossa SD, Waehre H, Paus E: The prognostic significance of prostate specific antigen in metastatic hormone-resistant prostate cancer. Br J Cancer 66:181-184, 1992
6. Fossa SD, Dearnaley DP, Law M, et al: Prognostic factors in hormone-resistant progressing cancer of the prostate. Ann Oncol 3:361-366, 1992
7. Fossa SD: Quality of life after palliative radiotherapy in patients with hormone resistant prostate cancer: Single institution experience. Br J Urol 74:345-351, 1994
8. Tannock IF, Osoba D, Stockler MR, et al: Chemotherapy with mitoxantrone plus prednisone or prednisone alone for symptomatic hormone-resistant prostate cancer: A Canadian randomized trial with palliative end points. J Clin Oncol 14:1756-1764, 1996
9. Vollmer RT, Kantoff PW, Dawson NA, et al: A prognostic score for hormone-refractory prostate cancer: Analysis of two Cancer and Leukemia Group B studies. Clin Cancer Res 5:831-837, 1999
10. Kantoff PW, Halabi S, Conaway MR, et al: Hydrocortisone with or without mitoxantrone in men with hormone refractory prostate cancer: Results of the Cancer and Leukemia Group B 9182 study. J Clin Oncol 17:2506-2513, 1999
11. Dowling AJ, Czaykowski PM, Krahn MD, et al: Prostate specific antigen response to mitoxantrone and prednisone in patients with refractory prostate cancer: Prognostic factors and generalizability of a multicenter trial to clinical practice. J Urol 163:1481-1485, 2000
12. Smaletz O, Scher HI, Small EJ, et al: Nomogram for overall survival of patients with progressive metastatic prostate cancer after castration. J Clin Oncol 20:3972-3982, 2002
13. Halabi S, Small EJ, Kantoff PW, et al: Prognostic model for predicting survival in men with hormone-refractory metastatic prostate cancer. J Clin Oncol 21:1232-1237, 2003
14. Weeks JC, Cook EF, O'Day SJ, et al: Relationship between cancer patients' predictions of prognosis and their treatment preferences. JAMA 279:1709-1714, 1998
15. Korn EL, Simon R: Measures of explained variation for survival data. Stat Med 9:487-503, 1990
16. Cotton SP, Levine EG, Fitzpatrick CM, et al: Exploring the relationships among spiritual well-being, quality of life, and psychological adjustment in women with breast cancer. Psychooncology 8:429-438, 1999
17. Eton DT, Lepore SJ, Helgeson VS: Early quality of life in patients with localized prostate carcinoma: An examination of treatment-related, demographic, and psychosocial factors. Cancer 92:1451-1459, 2001
18. Shapiro SL, Lopez AM, Schwartz GE: Quality of life and breast cancer: Relationship to psychosocial variables. J Clin Psychol 57:501- 519, 2001
19. Garssen B: Psycho-oncology and cancer: Linking psychosocial factors with cancer development. Ann Oncol 13:172-175, 2002 (suppl 4)
20. Butow PN, Coates AS, Dunn SM: Psychosocial predictors of survival: Metastatic breast cancer. Ann Oncol 11:469-474, 2000

21. Tross S, Herndon J, Korzun A, et al: Psychological symptoms and disease-free and overall survival in women with stage II breast cancer. Cancer and Leukemia Group B. J Natl Cancer Inst 88:661-667, 1996
22. Penson DF, Litwin MS, Aaronson NK: Health related quality of life in men with prostate cancer. J Urol 169:1653-1661, 2003
23. Young T, de Haes H, Fayers P, et al: Guidelines for Assessing Quality of Life in Clinical Trials (ed 2). Brussels, Belgium, EORTC Quality of Life Study Group Publications, 2002
24. Litwin MS, Lubeck DP, Henning JM, et al: Differences in urologist and patient assessments of health related quality of life in men with prostate cancer: Results of the CaPSURE database. J Urol 159:1988-1992, 1998
25. da Silva FC, Fossa SD, Aaronson NK, et al: The quality of life of patients with newly diagnosed M1 prostate cancer: Experience with the EORTC clinical trial 30853. Eur J Cancer 32A:72- 77, 1996
26. Fossa SD, Slee PHT, Brausi M, et al: Flutamide versus prednisone in patients with prostate cancer symptomatically progressing after androgen ablative therapy: A phase III study of the European Organization for Research and Treatment of Cancer Genitourinary Group. J Clin Oncol 19:62-71, 2001
27. Oosterhof GON, Roberts JT, de Reijke ThM, et al: Strontium(89) chloride versus palliative local field radiotherapy in patients with hormonal escaped prostate cancer: A phase III study of the European Organisation for Research and Treatment of Cancer, Genitourinary Group. Eur Urol 44:519-526, 2003
28. Albrecht W, Van Poppel H, Horenblas S, et al: Randomized phase II trial assessing estramustine and vinblastine combination chemotherapy versus estramustine alone in patients with progressive hormone escaped metastatic prostate cancer. Br J Cancer 90:100-105, 2004
29. Fayers P, Bottomley A: Quality of Life Research within the EORTC: The EORTC QLQC30, on behalf of the EORTC Quality of Life Group. Eur J Cancer 38:125-133, 2002 (suppl 4)
30. Fayers P, Aaronson NK, Bjordal K, et al: EORTC QLQ-C30 Scoring Manual (ed 3). Brussels, Belgium, EORTC, 2001, pp 1-86
31. Cox DR: Regression models and life tables. J R Stat Soc B 34:187-220, 1972
32. Collett D: Modelling Survival Data in Medical Research. London, United Kingdom, Chapman and Hall, 1994
33. Harrell FE, Lee KL, Mark DB: Multivariate prognostic models: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15:361- 387, 1996
34. Harrel FE: Design: S-Plus function for biostatistical/ epidemiological modelling, testing, estimation, validation, graphics, prediction, and typesetting for storing enhanced model design attributes in the fit, 1994. http://biostat.mc .vanderbilt.edu/twiki/bin/view/Main/Design
35. Dancey J, Zee B, Osoba D, et al: Quality of life scores: An independent prognostic variable in a general population of cancer patients receiving chemotherapy. Qual Life Res 6:151-158, 1997
36. Coates A, Porzsolt F, Osoba D: Quality of life in oncology practice: Prognostic value of EORTC QLQ-C30 scores in patients with advanced malignancy. Eur J Cancer 33:1025-1030, 1997
37. Harrell FE, Califf RM, Pryor DB, et al: Evaluating the yield of medical tests. JAMA 247:2543-2546, 1982

Chapter 7:

Statistical aspects of the analysis of phase III clinical trials: why trials of maximal androgen blockade in metastatic prostate cancer rarely demonstrate statistically significant differences?

7.1 Abstract

Background: The meta-analysis of maximal androgen blockade (MAB) concluded that there is no survival advantage of MAB over castration alone. However, the results from the largest trials yield conflicting results.
Methods: The design and results of three trials were examined.
Results: Most studies were planned to detect an over-optimistic difference in survival and immature data were published. The survival curves show that statistical assumptions are not fulfilled. Excluding from the meta-analysis, all trials where a negative impact of disease flare on survival could not be excluded results in no difference in survival between MAB and castration.
Conclusions: Trials of MAB should be planned to detect differences of no more than 5-10% in median survival. The analyses should only be carried out on mature data and should take into account the possibility of a negative impact on survival due to disease flare if no anti-androgen has been given initially with an LH-RH agonist.

7.2 Introduction

Since Labrie *et al.*[1] in 1982 reported excellent response rates with maximum androgen blockade (MAB) in stage M1 prostate cancer, many trials have been undertaken to try to establish the advantage of MAB over castration alone. The meta-analysis by the Prostate Cancer Trialists Collaborative Group published in The Lancet in 1995[2] and updated in 2000[3] identified 27 trials which randomized a total of 8275 patients. Of these 27 trials, the first NCI Intergroup trial INT-0036[4-5] that compared leuprolide plus flutamide to leuprolide alone demonstrated a statistically significant benefit of MAB in terms of survival. Of the other 26 trials, only 2 (EORTC 30853[6-8] and PONCAP[9]) demonstrated a statistically significant or borderline significant benefit in terms of duration of survival. The other 24 trials were inconclusive. The meta-analysis also failed to demonstrate any statistically significant benefit of MAB versus castration in terms of survival. Later, the NCI Intergroup trial INT-0105 that compared orchiectomy with or without flutamide failed to confirm a difference of 25% in overall survival that was observed in the earlier trial NCI INT-0036[10]. In view of this conflicting evidence, the value of MAB in advanced prostate cancer remains controversial.

Several authors have attempted to explain the lack of statistical significance of individual trials and of the meta-analysis. Klotz[11] investigated the methodology of the first meta-analysis and questioned whether the inclusion of 13% non-metastatic patients, the short follow-up in some of the studies and the inclusion of small unpublished trials of poor methodological quality did not 'wash out' a potential advantage of MAB in the meta-analysis.

Amongst others, Studer and Mills[12] questioned the significant advantage for MAB found in the NCI INT-0036 trial. They urged caution in interpreting the results of the trial and point out the fact that patients in the LH-RH agonist only group may have suffered from disease flare, this even repeatedly when questioning the compliance to the LH-RH agonist daily subcutaneous injections as used in trial NCI-INT 0036 Since the flare effect does not occur after orchiectomy, studies with MAB versus orchiectomy give a potentially clearer indication of the true benefit of MAB. It is also of note that no trial using a depot LH-RH showed a statistically significant difference in survival between LH-RH plus short term anti-androgen and LH-RH plus long term anti-androgen.

Several reports[4,13,14] based on retrospective subgroup analyses suggest a larger benefit of MAB in patients with minimal disease and performance status 0-2. MAB may thus be more effective in patients with small disease burden than in patients with more extensive disease and specific trials would be needed to assess the true value of MAB in the good prognosis patients.

Finally Blumenstein[15] considered the lack of power of several studies of MAB at the time of their publication. The lack of power could be a statistical explanation of the absence of a statistically significant difference in many individual studies.

In this paper, we consider a number of statistical issues related to the design and analysis of trials of MAB that could help in understanding the lack of significance in most studies. The issue of the power is considered but other conditions underlying the clinical and the statistical design as well as the analysis of the trials, such as the proportional hazards assumption, are also examined. We discuss the meta-analysis by the Prostate Cancer Trialists Collaborative Group and the three major MAB trials (NCI-0036, EORTC 30853 and NCI-0105). We re-analyze EORTC 30853 and the meta-analysis to exemplify our findings.

7.3 Material

Meta-Analysis
Design: The meta-analysis by the Prostate Cancer Trialists' Collaborative Group[2,3] considered all published or unpublished randomized trials of MAB versus castration that began before 1991. Individual patient data were collected and the overall survival data were analysed using a stratified 2-sided Logrank test[16].
Results: Thirty-six trials were identified. Individual patient data could be obtained for 27 trials which included 12 trials of MAB with flutamide. No statistically significant difference

was found in overall survival between the MAB and the castration arm (P =0.11 for all the 27 trials) based on 5932 deaths in 8275 patients. The observed median ratio was 1.04 (hazard ratio = 0.96) with 95% confidence interval (95% CI) from 0.91 to 1.01.There was a slight benefit in favor of MAB in the 12 trials using Flutamide (P = 0.02) corresponding to a hazard ratio of 0.92 (95% CI: 0.87 – 0.92).

NCI INT-0036

Design: The NCI Intergroup trial INT-0036 compared daily subcutaneous injections of leuprolide with or without flutamide[4,5] in a randomized double blind trial in patients with previously untreated stage D2 prostate cancer. The target sample size was 600 patients (300 in each treatment arm). The statistical design aimed at detecting a 40% improvement in median survival time from 3 years to 4.2 years at a significance level of 0.05.

Results: A total of 617 patients were registered in the trial, of which 14 were ineligible. The analysis was carried out on 603 eligible patients using a stratified two-sided Logrank test when 398 patients had died. The test led to a significant difference in overall survival with two-sided P-value of 0.035. The median survival was 28.3 months on the leuprolide plus placebo group and 35.6 months on the leuprolide plus flutamide group (MR=1.26).

EORTC 30853

Design: The European Organization for Research and Treatment trial EORTC 30853 compared Goserelin plus flutamide to orchiectomy[6-8] in M1 prostate cancer. The initial statistical design had time to disease progression as the main endpoint and required 226 patients and 192 progressions. The trial design was amended to consider overall survival as primary endpoint. The sample size was increased to 300 patients in order to observe 192 deaths, which would provide 80% power of detecting a 50% difference in median survival using a two-sided logrank test at the 0.05 significance level.

Results : The trial was first published in 1990[6] with a median follow-up of 1.5 years. At the time of the first publication 107 of 327 patients (33%) had died and there was no statistically significant difference in overall survival (P = 0.89, two-sided Logrank test). The trial was then published in 1993[7] when 223 patients had died (68%). At the time of that publication, the median survival was 27.1 months on orchiectomy and 34.4 months on MAB. The difference between the overall survival curves was statistically significant (P = 0.02). The long term results of the trial were finally published in 1998[8] with a median follow-up of approximately 7.2 years. When that analysis was carried out, 259 of 327 (79%) patients had died, the median duration of survival on orchiectomy was 27 months on orchiectomy and 34 months on MAB, leading to a statistically significant test with P = 0.04 and a hazard ratio (HR) of 0.77 (95% CI: 0.60 – 0.99) in favor of MAB compared to orchiectomy, or equivalently to a median ratio (MR) of 1.30.

NCI INT-0105

Design : The NCI Intergroup trial INT-0105 compared orchiectomy with flutamide to orchiectomy plus placebo in metastatic prostate cancer[10]. This double blind trial aimed at detecting a 25% improvement in median survival with 90% power using a one-sided Logrank test assuming 28.3 months median survival on the orchiectomy plus placebo group. The required sample size was estimated to be 1248 patients equally distributed between

both arms. However, at the time of the publication, a two-sided test was reported instead of the planed one sided test. The same planned sample size provides 83% power to the two-sided Logrank test.
Results : A total of 1387 patients entered in the study. At the time of the publication 948 patients (68%) had died. The median duration of survival was 29.9 months on the placebo group and 33.5 in the MAB group. The estimated death hazard ratio was 0.91 (MR = 1.10) in the MAB group compared to the orchiectomy group (90% CI for HR: 0.81 – 1.01, corresponding to a one-sided test; 95% CI: 0.79 – 1.03 corresponding to a two-sided test). The two-sided Logrank test was not significant (P = 0.14).

7.4 Statistical and Clinical Issues

The issue of statistical power

When analyzing clinical trials with overall survival as the main endpoint, the power of the statistical test is determined by the number of observed events and not by the number of patients entered in the trial. The power of a statistical test is the probability that the test will give a statistically significant P-value (traditionally $P<0.05$), conditional on the size of the actual difference in efficacy between the 2 treatment groups.

The power of the statistical test increases with increasing treatment effect and with increasing number of events. The power of the 3 trials considered above was calculated based on the number of events observed at the time of the final publication of the trial, under a range of possible alternatives covering a 4% improvement in median survival to a 50% improvement in median survival (Table 7.1). The number indicated in **bold** represents the power computed based on the observed difference, that indicated in *underlined italic* represents the power against the alternative specified in the protocol. It is of note that EORTC 30853 was designed to detect a difference of 50% (MR=1.50) in median overall survival and INT-0036 to detect a difference of 40% (MR=1.40).

The relative difference in median survival between MAB and medical or surgical castration as estimated from the meta-analysis is approximately 4% (MR=1.04). As can be seen in the table all the individual trials have a power much less than 50% of detecting a difference of that magnitude. From this table, it can also be seen that the EORTC trial, despite being statistically significant, had only 55% power of detecting a difference of the size observed in that trial (namely, 30%). Also, the most recent trial from the NCI International group, which led to a non significant P-value, had only 31% power of detecting a difference of the order of magnitude of 10% which was actually observed in that trial. This shows that, even if chance sometimes yields statistical significance in spite of low power, as in EORTC trial 30853, lack of power is generally the main reason why individual trials are unable to detect a moderate difference in treatment effect should one exist. Any effect often turns out to be of smaller magnitude than that expected at the time of designing the protocol.

Table 7.1: Post-hoc power of the 3 trials against a range of alternatives ranging from 4% to 30% improvement in median overall survival for a 2-sided logrank test at the 0.05 significance level.

	Post-hoc power			
Median ratio (hazard ratio)	**INT-0036 (398 deaths)**	**EORTC 30853 (259 deaths)**	**INT-0105 (948 deaths)**	**Meta-analysis (5864 deaths)**
1.50 (0.67)	97.88	*89.59*	99.99	99.99
1.40 (0.71)	*91.39*	76.49	99.92	99.99
1.30 (0.77)	77.96	**55.53**	98.01	99.99
1.25 (0.80)	**60.12**	43.18	*92.80*	99.99
1.20 (0.83)	44.18	30.96	79.93	99.99
1.15 (0.87)	28.50	20.13	57.46	99.96
1.10 (0.91)	15.63	11.63	**31.07**	95.42
1.05 (0.95)	7.03	5.85	11.33	46.33
1.04 (0.95)	5.83	5.00	8.75	**32.33**

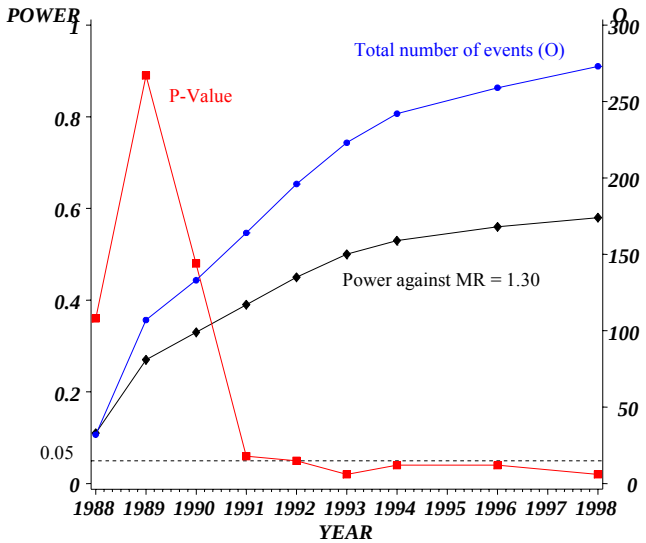
In **bold** the post-hoc power (i.e. the statistical power of the test statistic based on the data available at the time of the final analysis) to detect a difference in median survival of the magnitude of that **observed** at the time of the final analysis; in *underlined italic*, the post-hoc power for detecting a difference in median survival of the magnitude *pre-specified in the protocol* (i.e. for which the trial was designed).

The impact of increasing power over time on statistical significance: the case of EORTC trial 30853

Many formal and informal interim analyses of EORTC 30853 were carried out from the closure of the trial in May 1988 onwards. Searching back in the archives of the EORTC, we were able to retrieve the survival curves as they were at the time of each of the 8 analyses. We also re-analyzed the data in November 1998. A summary of the results of these analyses in terms of overall survival is given in Table 7.2 and the results in the table are shown in Figure 7.1.

From Table 7.2 and Figure 7.1, it is seen that the power of the trial to detect a difference of 30% increases as the total number of observed events increases. The power of the statistical test thus increased with time. Also, it is seen that the hazard ratio fluctuated very much in the early times, when the number of events was still small to moderate, and was associated with large fluctuations in the P-value of the test. From 1993 onwards, the power becomes larger than 50%, the median ratio tends to stabilize around 1.30 and the P-value fluctuates less, always remaining below 0.05. At the time of the final analysis (1996) the trial was able to detect a difference smaller than that expected in the protocol (MR = 1.30 instead of MR = 1.50) despite limited statistical power. The survival curves as they appeared in 1990, 1993 and 1996 are displayed in Figure 7.2 a), b) and c), respectively.

Figure 7.1
Power and p-value foor EORTC 30853 as a function of year of analysis and data maturity



The right vertical axis indicates the total number of events observed (line with dots), left vertical axis indicates Power and P-value corresponding to the line with diamonds (power) and the line with squares (P-value). The power is seen to increase with increasing number of events. The P-value fluctuates, then stabilizes around P=0.04.

This example shows how the release of immature data (i.e.: when the number of events is less than that specified in the protocol) can be misleading, as the results may still change dramatically with increasing number of events and statistical power. It is to be noted that the opposite pattern of changes may be seen with some trials leading to statistically significant differences early on, which then disappear with increasing follow-up[17]. In trials of MAB however where the difference in efficacy is usually not seen shortly after randomization, this latter situation is less likely. Nevertheless, it remains that trials should only be published when the data are mature for final analysis. In addition, it has to be kept in mind that multiplying the number of analyses being performed dramatically increases the risk to find a statistically significant difference just by chance.

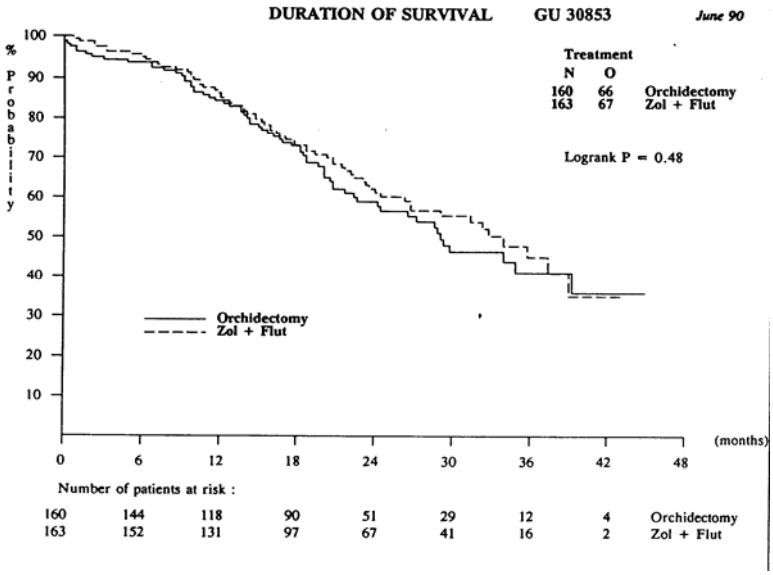
Table 7.2: Summary of results of all interim and final analyses of EORTC trial 30853

Date[a]	Events[b] Orch./MAB	P-value[c]	HR[d]	MR[e]	95%CI[f] for MR	Power[g] (1.30)
May 88	18 / 14	0.36	0.72	1.38	0.69-2.78	0.11
August 89	52 / 55	0.89	0.97	1.03	0.70-1.49	0.27
June 90	66 / 67	0.48	0.88	1.13	0.81-1.59	0.33
September 91	88 / 76	0.06	0.75	1.33	0.98-1.82	0.39
August 92	103/ 93	0.05	0.76	1.32	1.00-1.75	0.45
May 93	116/107	0.02	0.73	1.37	1.05-1.79	0.50
June 94	123/119	0.04	0.77	1.30	1.01-1.67	0.53
August 96	131/128	0.04	0.77	1.30	1.01-1.67	0.56
November 98	139/134	0.02	0.75	1.33	1.04-1.69	0.58

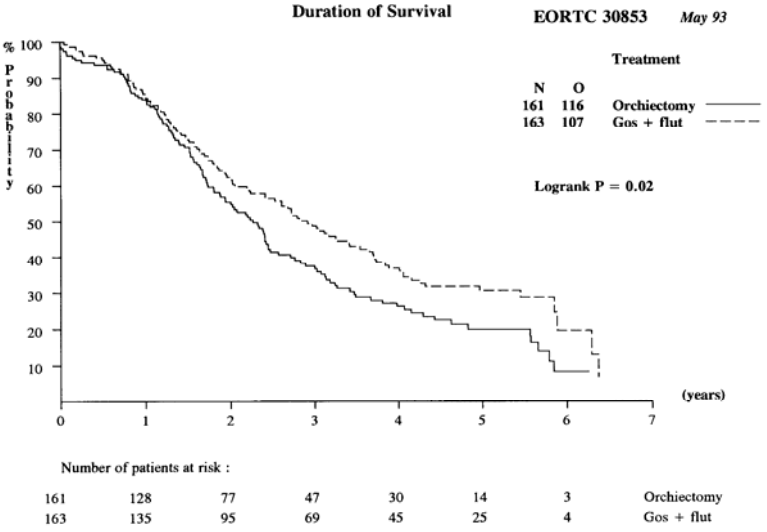
[a] Time of the analysis
[b] Number of deaths observed in the orchiectomy arm / in the MAB arm
[c] p-value of the Logrank test comparing overall survival between the 2 groups
[d] Hazard ratio (HR) for overall survival comparing MAB to orchiectomy (HR<1 favour MAB)
Median ratio (MR) of overall survival (ratio of median overall survival on MAB to median overall survival on orchiectomy, MR>1 favour MAB)
[f] 95% confidence interval (CI) for the median ratio
[g] Power of the test statistic for detecting a relative difference of 30% in overall survival (i.e.: for detecting MR=1.30 or equivalently, HR=0.77)

Figure 7.2
Overall survival in EORTC Trial 30853

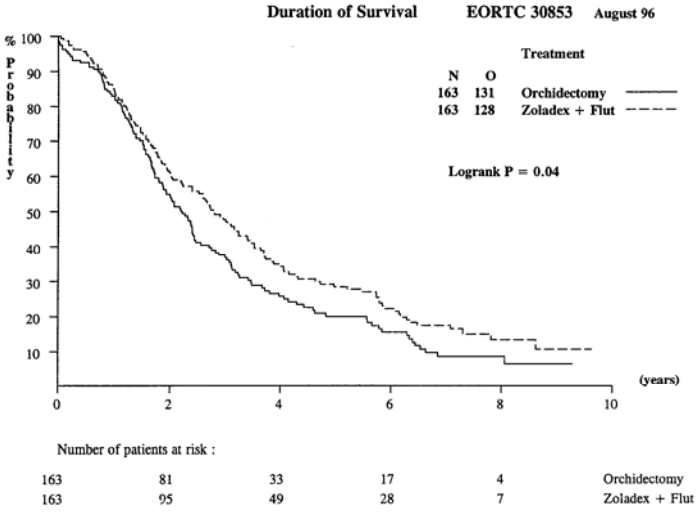
(a) Two years after closure to patient entry (1990)



(b) Five years after closure to patient entry (1993)



(c) Eight years after closure to patient entry (1996)

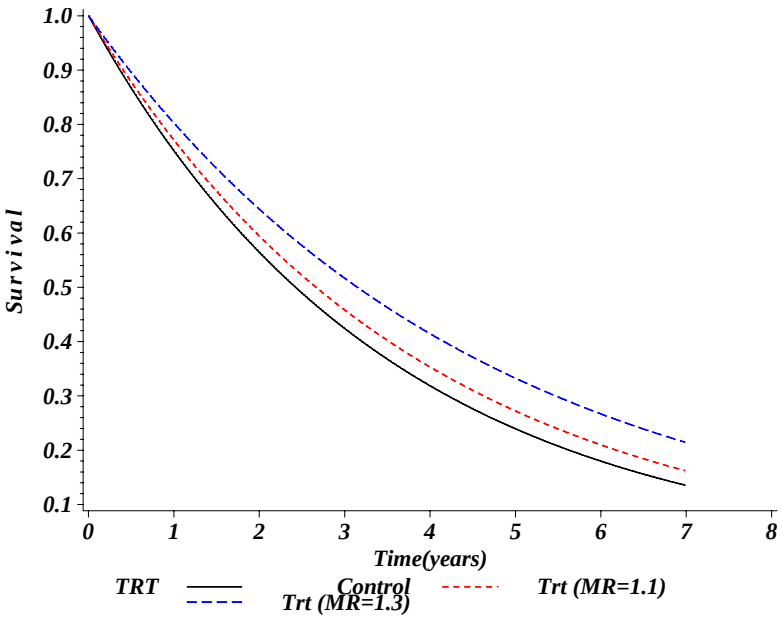


N stands for the number of patients, O for the number of deaths observed, P is the P-value for the Logrank test comparing the two curves.

How statistical assumptions are violated in trials of MAB versus castration

One aspect of clinical trials which is often overlooked is the set of statistical assumptions that underlie both the statistical design of clinical trials with survival as the endpoint and their analysis. In summary, the main assumption is the so called 'proportionality assumption' by which it is assumed that the ratio of the hazards of death on both treatment arms is constant over time. It is of note that the Logrank test is most powerful (most likely to give statistical significance) when that assumption is fulfilled. Also the Cox proportional hazards regression model, which is often used in analyzing survival data in clinical trials requires, as the name indicates, the proportionality assumption to be fulfilled. In addition, when computing sample size for a given hazard ratio, exponential survival distributions are assumed, which results in the hazard in each treatment group being constant over time. Theoretical distribution functions that fulfil both assumptions are displayed in Figure 7.3.

Figure 7.3
Theoretical survival following exponential distributions

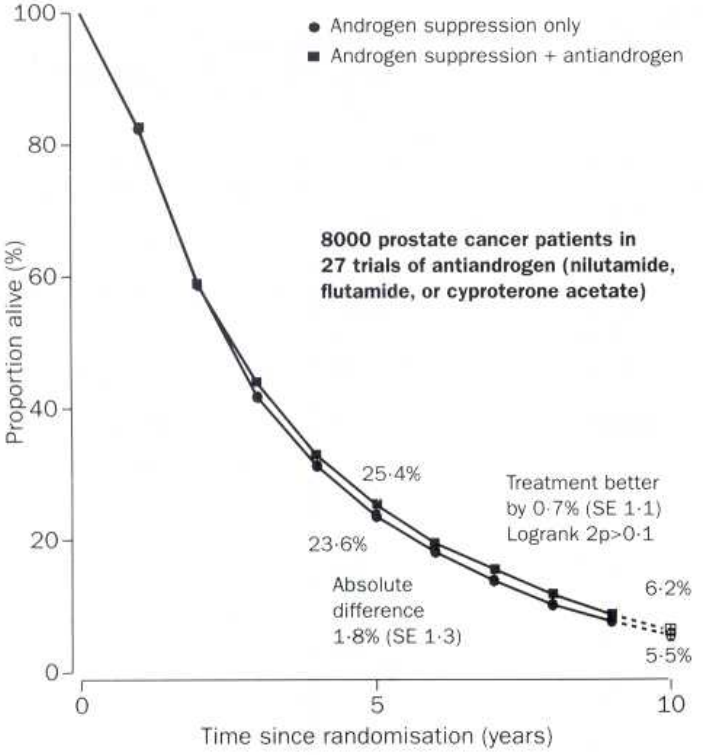


The solid line represents a control group, the short dashed line an experimental treatment with a median survival 1.10 times that of the control, the long dashed line an experimental treatment with median survival 1.30 times that of the control.

As is immediately observed when comparing the survival curves in Figure 7.2 to the theoretical curves in Figure 7.3, the proportionality assumption is not fulfilled in the EORTC trial of MAB versus castration. Indeed, it is observed that the two curves remain close together during an initial period lasting approximately 1.5 years after randomization

and that the treatment effect appears only later on. This characteristic is also observed in the other two trials: in NCI-0036, the initial period is seen to last roughly 6 months and in NCI-0105 it lasts about 24 months. Also in the meta-analysis, the figure of overall survival (Fig. 7.4, reproduced by permission) shows that the difference in treatment effect appears only from year 2 following the start of the treatment.

Figure 7.4
Overall survival in the meta-analysis of 27 trials by the PCTCG



Reprinted with permission from Lancet, 2000; 355:1491-98. Prostate Cancer Trialists' Collaborative Group. Maximum Androgen Blockade in Advanced Prostate Cancer: an Overview of the Randomized Trials. Lancet, 2000; 355:1491-98

Re-analysis of Trial EORTC 30853

In order to confirm that the proportionality assumption is not fulfilled in trial EORTC 30853, the data were reanalyzed in a Cox proportional hazards regression model which allowed a different hazard ratio in the first time period (time 0 to year 1.5 following randomization) and in the second time period (year 1.5 onwards) based on the data as of November 1998. From that model, it was estimated that the hazard ratio in the first time period was 0.90 (95% confidence interval: 0.59-1.36) which was not statistically significant (P = 0.608). For the second time period, the hazard ratio was estimated to be 0.69 (95%

confidence interval: 0.52-0.93) which was statistically significant (P = 0.013). This analysis suggests that the hazard ratio is not constant over time as it is 0.90 in the beginning of the follow-up period and 0.69 at a later time point and therefore suggests that for this trial, the proportionality assumption may indeed have not been fulfilled.

Impact of disease flare on the results of the meta-analysis.

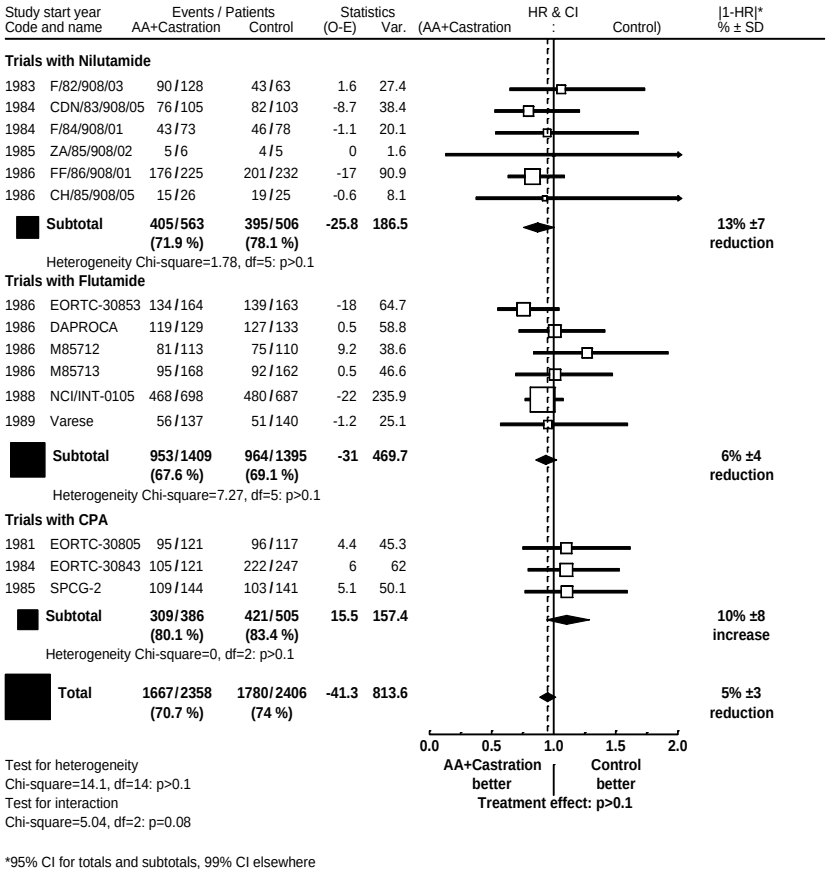
In order to assess the impact of disease flare on the results of the meta-analysis, the meta-analysis data was re-analysed by excluding all trials where no short term anti-androgen was given at the start of the LH-RH treatment. The eleven trials comparing orchidectomy to orchidectomy plus anti-androgen were included in the analysis, as well as the four trials comparing ochidectomy/LHRH to orchidectomy/LHRH plus anti-androgen in which short term anti-androgen was given at the start of the LHRH treatment. A meta-analysis of these 15 trials leads to a hazard ratio of 0.95 (MR = 1.05) with 95% confidence interval (0.89 – 1.02) and to a P-value of 0.15 for the corresponding stratified 2-sided Logrank test as displayed on the Forest plot represented in Figure 7.5. These results show that excluding the impact of disease flare from the meta-analysis results in no significant benefit of MAB over castration in terms of survival. The results in the subset of trials using flutamide are very similar with a hazard ratio of 0.94 (95% CI: 0.86 – 1.02, MR=1.06) and a P-value of 0.15.

7.5 Discussion

From the above review, it appears that trials of MAB versus castration in general suffer from several statistical and methodological drawbacks.

First of all, as also noted by Blumenstein[15], most studies suffer from a lack of power to detect relative differences in efficacy of the order of magnitude seen in the meta-analysis (4%) or even of the size seen in INT-0036 (25%) and in EORTC 30853 (30%). Several trials also were first published based on immature data, like trial 30853, at a time when the observed number of events did not provide large enough statistical power for detecting the survival difference hypothetized in the protocol, thus resulting in the publication of potentially misleading conclusions of the absence of a difference. As wisely noted by Shellhammer in his review of combined androgen blockade in M1 prostate cancer[18], early analyses of data, even from trials with sample sizes large enough to achieve adequate statistical power at maturity can reach invalid conclusions, as seen with trial 30853. The EORTC has now changed its publication policy in order to prevent publication of immature data. The trial now gets published only when the data have reached maturity, as defined in the protocol, and interim results are released only upon the advice of an independent data monitoring committee. Secondary endpoints are also never published before maturity of the primary endpoint is obtained.

Figure 7.5
Forest plot of the meta-analysis including 15 trials where the possibility of disease flare is excluded



The codes used to represent the trials are those of the publication of the meta-analysis. The middle of the box represents the point estimate of the hazard ratio in each trial. The horizontal bars represent their 99% confidence intervals. The diamonds represent the estimate of the overall effect (middle of the diamond, dotted vertical line) with its associated 95% confidence interval (width of the diamond). The solid vertical line represents the hazard ratio corresponding to no difference in effect. Squares located to the left of the vertical axis favor MAB, those located to the right favor castration alone.

Second, it is seen in several trials and in the meta-analysis that the difference in survival between MAB and castration, if any, does not appear in the first period following randomization but after 6 months to 2 years. Thus, the treatment effect varies over time. This implies that the main assumption underlying the sample size calculation, the so-called 'proportionality assumption' is not fulfilled, which may result in an inaccurate estimation of the number of patients and events required to ensure the specified statistical power of the comparison. This also implies that the statistical test, usually the Logrank test or a Wald test in a Cox proportional hazard regression model, does not have maximal power. Finally, the conclusions of the analysis, which are summarized by a unique estimate of the average treatment effect over the whole follow-up period, are inaccurate as the treatment effect is different in the first and in the second part of the follow-up.

Explanations for this change in the treatment effect over time which appears in many studies may be sought. One possible explanation relates to the observed difference in effect in the patients with minimal and extensive disease which was observed in trials NCI INT-0036 and EORTC 30853, but was not observed in NCI INT-0105. In both studies[12,13], a subgroup analysis of the patients with minimal disease and good performance status revealed a more striking effect than in patients with more extensive disease. From this, it may be hypothetized that the group of patients represented in the survival curves is a mixture of these two populations. The patients with more extensive disease would fail earlier and there would be no difference in treatment efficacy between MAB and castration in this group of patients. This would lead to the two survival curves remaining close together early in the follow-up period. Patients with minimal disease would fail later and benefit from MAB, thus leading to the observed separation of the survival curves at a later time in the follow-up. However, this effect has not been confirmed in NCI INT 0105, a trial in which randomization was stratified by disease extent. Also, the conclusions for patients with minimal disease in the early studies were based on analyses of small subgroups of patients and therefore may be subject to selection bias and large random variations[19,20]. One way to validate the hypothesis that patients with minimal disease benefit more from MAB would be to use the data of the meta-analysis of the Prostate Cancer Trialists Collaborative Group and to perform separate meta-analyses in the two subgroups. This would however require that the information on disease extent and performance status be available for all studies, which is not the case. Even if this analysis were feasible, the results would need to be confirmed by an adequately designed prospective randomized trial. Other subset analyses by age group or metastatic status reported in the meta-analysis do not suggest any benefit to be likely to be seen in any subset of patients.

A last point of debate regarding the meta-analysis and trials that compared MAB to LH-RH alone relates to the flare effect which is observed in the first weeks of treatment with an LH-RH and to a potential lack of compliance to the LH-RH treatment. As discussed by Schellhammer[18], Schröder[21] as well as by Eisenberger in the publication of NCI INT-0105[10], LH-RH used without an anti-androgen lead to increased production of testosterone during the first weeks of treatment, which may be associated with exacerbation of the disease in around 20% of the patients. In the trial of Crawford et al.[4,5], disease progressed more rapidly in patients treated with leuprolide alone during the first three months. It is

possible that the differences in efficacy observed in that study may be partly explained by the ability of flutamide to prevent the initial episode of flare[4,12]. Another consideration is the possibility that poor compliance to the required daily self-injections of leuprolide may have led to repeated episodes of flare during the follow-up as this phenomenon may occur after an interruption of only a few days of the daily LH-RH treatment[22]. This effect would have been buffered by the combination with the anti-androgen flutamide[18]. Thus, any survival difference could also be attributed to a possible shortening of survival in patients treated by LHRH only without coverage against disease flare[12]. Also, as this flare-up phenomenon may be particularly detrimental in patients with large volume metastases[23], this effect might also be related to differences in treatment efficacy seen in patients with extensive versus minimal disease in the first NCI international study. As we have shown, one possibility to answer the question of the impact of the flare phenomenon is through a re-analysis of the meta-analysis data that excludes all trials where LH-RH was used without being combined with an anti-androgen in the first weeks of treatment. Indeed, this re-analysis showed that when the impact of disease flare is excluded, the data do no exhibit any survival benefit of MAB over castration even for trials using Flutamide.

Finally, it also emphasizes that of the 15 trials that were considered, only EORTC 30853 showed a statistically significant benefit of MAB over castration whereas the overall result shows no difference between the 2 treatment approaches. It cannot be excluded that the result of EORTC 30853 may be a false positive result related to the random variation (statistical type I error) and there would be no actual difference in terms of overall survival between MAB and castration.

The body of evidence regarding the relative efficacy of MAB over castration shows that the size of the benefit, if any, is of little clinical relevance. This is the conclusion of our review, but also of that of the Blue Cross and Blue Shield Association[24] and of the recent review by Laufer et al.[25]. In addition, the few trials which assessed quality of life using a validated questionnaire[9,26] showed that MAB is associated with more side effects and a decreased quality of life as compared to castration alone. In the PONCAP trial[9] patients treated with bicalutamide monotherapy complained less frequently about loss of libido or erectile dysfunction than those treated with flutamide plus goserelin. They also appeared to have a better social functioning, vitality, emotional well-being and physical capacity. For trial NCI-INT 0105[26], Moinpour et al. report that patients treated with flutamide had more diarrhea at 3 months and a worse emotional functioning at 3 and 6 months as compared to patients treated by orchiectomy alone, even though quality of life seemed to improve on both arms as compared to baseline.

In terms of costs, few reports are available that have compared the costs of MAB to those of surgical and medical castration[24,27,28]. Nicol et al.[27] found that surgical castration was about four times cheaper than an LHRH analogue and five times cheaper than the combination of an LHRH and flutamide. Hillner et al.[28] estimated the cost effectiveness of MAB using flutamide and found that the incremental cost per life-year gained varied between $ 20,000 and $ 30,000. As indicated by Neymark[29], these results were however very sensitive to changes in the costs and duration of flutamide use. The Blue Cross and Blue Shield

Association[24] review concluded that MAB with an LHRH must increase efficacy by 20 percent compared to orchiectomy before this drug combination is considered cost effective and by 10% when MAB combined an anti-androgen with orchiectomy. These values are not within the 95% confidence limits given by the meta-analysis, therefore MAB is unlikely to be cost-effective as compared to orchiectomy.

7.6 Conclusion

From this review we conclude that trials in metastatic prostate cancer should be initially planned to detect differences in median survival of not more than 5% to 10% and should therefore accrue a large number of patients. The analyses should only be carried out once the observed number of events ensures a high power of the statistical test. Furthermore, our re-analysis of the meta-analysis data by excluding all trials where an LH-RH analogue was used without short term anti-androgen to omit any negative impact of disease flare on survival shows no statistically significant survival benefit of MAB over castration even when using flutamide. The available body of evidence does not show any clinically relevant benefit of MAB over castration alone in terms of survival and seems associated with more side effects and a lesser quality of life. The associated costs are also higher than those of castration, particularly than those associated with orchiectomy.

7.7 References

1. Labrie F, Dupont A, Belanger A, Cusan L, Lacourciere Y, Monfette G, Laberge JG, Emond JP, Fazekas AT, Raynaud JP, Husson JM: New hormonal therapy in prostatic carcinoma: combined treatment with an LHRH agonist and an anti-androgen. Clin Invest Med 1982;5:267-275.
2. Prostate Cancer Trialists' Collaborative Group. Maximum androgen blockade in advanced prostate cancer: an overview of 22 randomized trials with 3283 deaths and 5710 patients. Lancet 1995;346:265-269.
3. Prostate Cancer Trialists' Collaborative Group. Maximum androgen blockade in advanced prostate cancer: an overview of the randomized trials. Lancet 2000;355:1491-98.
4. Crawford DE, Eisenberger MA, McLeod DG, Spaulding JT, Benson R, Dorr AF, Blumenstein BA, Davis MA, Goodman PJ. A controlled trial of leuprolide with and without flutamide in prostatic carcinoma. New Engl J Med 1989;321:419-424.
5. Crawford DE, Blumenstein BA, Goodman PhJ, Davis MA, Eisenberger MA, Mc Leod DG, Spaulding JT, Benson R, Dorr FA. Leuprolide with or without flutamide in advanced prostate cancer. Cancer 1990;66(Suppl 1):33-38.
6. Denis L, Robinson M, Mahler C, Smith Ph, Keuppens F, Carneiro de Moura JL Bono A, Newling D, Sylvester R, De Pauw M, Verweylen K, Ongena P and Members of the EORTC-GU Group. Orchidectomy versus Zoladex plus Eulexin in patients with metastatic prostate cancer (EORTC 30853), J Steroid Biochem Mol Biol 1990;37:951-959.
7. Denis LJ, Whelan P, Carneiro de Moura JL, Newling D, Bono A, De Pauw M, Sylvester R and Members of the EORTC GU Group and Data Center. Goserilin acetate and flutamide versus bilateral orchiectomy: a phase III EORTC trial (30853). Urology 1993;42:119-130.

8. Denis LJ, Keuppens F, Smith PH, Whelan P, Carneiro de Moura JL, Newling D, Bono A, Sylvester R for the EORTC Genito-Urinary Tract Cancer Cooperative Group and the EORTC Data Center. Maximal androgen blockade: final analysis of EORTC phase III trial 30853. Eur Urol 1998;33:144-151.
9. Boccardo F, Pace M, Robagotti A , Guarneri D, Decensi A, Oneto F, Martorana G, Giuliani L, Selvaggi F, Battaglia M, Delli Ponti U, Petracco S, Cortellini P, Ziveri M, Ferraris V, Bruttini GP, Epis R, Comeri G, Gallo G, and other participants in the Italian Prostatic Cancer Project (PONCAP) Study Group. Goserelin acetate with or without flutamide in the treatment of patients with locally advanced or metastatic prostate cancer. Eur J Cancer 1993;29A:1088-1093.
10. Eisenberger MA, Blumenstein BA, Crawford DE, Miller G, McLeod DG, Loehrer PJ, Wilding G, Sears K, Culkin DJ, Thompson IM Jr, Bueschen AJ, Lowe BA. Bilateral orchiectomy with or without flutamide for metastatic prostate cancer. New Engl J Med 1998;339:1036-1042.
11. Klotz LH, Newman T. Does maximal androgen blockade (MAB) improve survival? Can J Urol 1996;3:246-250.
12. Studer UE, Mills RD. Arguments agianst the long-term use of combined androgen blockade. Eur Urol 1998;34(suppl 3):29-32.
13. Sylvester RJ, Denis L, de Voogt H. The importance of prognostic factors in the interpretation of two EORTC metastatic prostate cancer trials. Eur Urol 1998;33:134-143.
14. Eisenberger M, Crawford ED, Wolf M, Blumenstein B, Mc Leod DG, Benson R, Dorr FA, Benson M, Spaulding JT. Prognostic factors in stage D2 prostate cancer: important implications for future trials: Results of a cooperative intergroup study (int 0036). Semin Oncol 1994;21:613-619.
15. Blumenstein BA. Some statistical considerations for the interpretation of trials of combined androgen therapy. Cancer 1993;72(Suppl 12):3834-40.
16. Early Breast Cancer Trialists' Collaborative Group. Systemic treatment of early breast cancer by hormonal, cytotoxic, or immune therapy. 133 randomized trials involving 31 000 occurrences and 24 000 deaths among 75 000 women. Lancet 1992;339:1-15,71-85.
17. Sylvester R, Bartelink H, Rubens R. A reversal of fortune: Practical problems in the monitoring and inerpretation of an EORTC breast cancer trial. Stat in Med 1994;13:1329-1335.
18. Shellhammer PF Combined androgen blockade for the treatment of metastatic prostate cancer. Urology 1996; 47:622-628.
19. Fitzpatrick JM. Combined androgen blockade is the treatment of choice for patients with advanced prostate cancer: the argument against. Eur Urol 1996;29(suppl 2):37-39.
20. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. JAMA 1991;266: 93-98.
21. Schröder FH. Hormonal manipulation of prostatic cancer. Too soon for total androgen blockade? Brit Med J 1991;303:1489-90.
22. Klijn JG, de Voogt HJ, Schröder FH, de Jong FH. Combined treatment with buserelin and cyproterone acetate in metastatic prostatic carcinoma. Lancet 1985;2:493.
23. Boccon-Gibod L. Maximum androgen blockade in 1996. Eur Urology 1996;30(suppl 1):15-18.
24. Agency for Health Care Policy and Research, Rockville, MD. Relative effectiveness and cost-effectiveness of methods of androgen suppression in the treatment of advanced prostatic cancer. Summary, Evidence Report/Technology Assessment: Number 4, January 1999. http://www.ahrq.gov/clinic/prossumm.htm.
25. Laufer M, Denmeade SR, Sinibaldi VJ, Carducci MA, Eisenberger MA. Complete androgen blockade for prostate cancer: what went wrong? J Urol 2000; 164:3-9.

26. Moinpour CM, Savage MJ, Troxel A, Lovato LC, Eisenberger M, Veith RW, Higgins B, Skeel R, Yee M, Blumenstein B, Crawford ED, Meyskens FL Jr. Quality of life in advanced prostate cancer: results of a randomized therapeutic trial. JNCI 1998;90:1537-44.
27. Nicol DL, Heathcote PS, Kateley GD, Lloyd S. Advanced prostate cancer. The role of high priced hormone therapy. Med J Aust 1993;159:16-9.
28. Hillner BE, Mc Leod DG, Crawford ED, Bennet CL. Estimating the cost effectiveness of total androgen blockade with flutamide in M1 prostate cancer. Urology 1995; 45:633-40.
29. Neymark N: Assessing the economic value of anticancer therapies. Recent Results in Cancer Research 148. Springer-Verlag Berlin 1998.

Chapter 8:

General discussion

8.1 General framework of (cancer) clinical research

The goal of clinical research is to identify effective new treatments for specific diseases and to determine the relative benefits of competing therapies with the aim of establishing the optimum treatment regimen for each indication. For cancer, the relative benefit of new treatment strategies is most often small, but medically and/or biologically important. Therefore, properly conducted clinical trials following the principles of scientific experimentation are particularly needed in this field in order to reliably assess the efficacy and safety of new, potentially toxic treatments.

A "clinical trial" is any form of planned experiment which involves patients and is designed to elucidate the most appropriate treatment of future patients with a given medical condition Its essential characteristic is that is uses results obtained on a limited sample of patients to make inferences about how treatment should be conducted in the general population of patients who will need to be treated in the future[1].

8.1.1 Phases of clinical development

Once a drug has successfully passed the stage of animal studies, it goes through three stages of clinical testing in humans. These three stages correspond to three types of clinical trials, each with specific design and objectives, namely phase I, phase II and phase III clinical trials.

Phase I trials are small scale studies (around 20 patients) that are usually carried out in a single or a very limited number of centers and involve patients with very advanced disease for whom no effective therapy is available. The entry criteria are usually broad so that patients with various cancers types may enter a same study. The main objectives of phase I trials are to study the treatment mechanism of action and to identify the best dose for use in humans in further stages of treatment development.

Once a drug has gone through phase I testing, one or several early phase II trials are carried out in order to further characterize the safety profile and the activity of the given regimen in specific tumor types. Phase II trials commonly involve 20 to 50 patients with advanced cancer of a specified type. They aim at determining the probability that a patient responds to

the treatment and also at further characterizing the serious acute side effects. Phase II trials help determine whether further (phase III) studies in the given tumor type are warranted.

After phase II trials have demonstrated that a given drug or combination of drugs is active in a specific disease group, the efficacy of the new treatment and its benefit to the patient must be assessed. This is done by comparing the long term clinical outcome obtained with the new treatment with that achieved with the best available standard treatment (which may be no treatment or placebo). This is done by means of phase III clinical trials in which large numbers of patients with specific disease characteristics are randomized, treated with the new or the standard treatment, and followed-up until the event of interest according to a very precise protocol. (Confirmed) randomized clinical trials are at the top of the hierarchy of evidence concerning treatment efficacy[2].

8.1.2 Phase III clinical trials and sample size requirements

Phase III clinical trials are comparative in nature. They are designed to assess as reliably and as precisely as possible the relative efficacy of a new treatment in comparison to a 'standard' treatment. Randomization (allocation of patients to treatments by chance) is an essential feature of phase III trials as it is the only means to ensure the comparability of the groups of patients being treated. To eliminate sources of confounding and selection bias, randomization should be unpredictable to the investigators, so that they are unlikely to select patients to preferentially receive one treatment over the other. In order to minimize other sources of variability and bias, a clinical trial requires a well written and detailed protocol that details the characteristics of the eligible patients, the treatment procedures, the follow-up scheme and follow-up examinations as well as the reporting and statistical analysis procedures. Most importantly, to avoid data dredging and statistical bias, the study protocol must clearly state the study objectives (primary and secondary) and the endpoints that will be used for comparing the two treatments. In phase III cancer clinical trials, a time-to-event endpoint, such as survival or progression-free survival is generally used as the primary trial endpoint.

To be useful and to provide valid conclusions, phase III clinical trials ought to be designed to provide sufficient statistical power to detect the specified alternative hypothesis, were it to show a defined difference in treatment efficacy or to demonstrate non-inferiority of the new treatment within defined boundaries. The target alternative hypothesis (for which the power calculations are done) should reflect realistic expectations and should be clinically meaningful at the same time: for difference trials, the target difference should not be overoptimistically large but should be large enough to reflect clinically relevant differences; for non-inferiority trials, the non-inferiority boundary should be small enough to exclude any clinically relevant difference.

For cancer trials in general, the expected relative benefit of a new treatment is relatively small: usually no more than a 30% relative increase in median time to the event. Similarly, the non-inferiority boundaries in non-inferiority studies are also very small: usually no more

than a 20% relative loss over the best available treatment, this in order not to waste a too large part of the gain in efficacy that was obtained when the current standard treatment was established. In addition, for time-to-event endpoints, the statistical power is driven by the number of events (deaths or progressions) that will be available at the time when the final analysis is carried out. The sample size of the trial and the duration of follow-up are driven by the power requirements. For cancer clinical trials, the power requirements typically induce trials of relatively large sample size and/or of long duration.

8.2 Clinical trials in prostate cancer

Several aspects of prostate cancer render clinical trials in this disease more difficult than trials in other solid tumors.

8.2.1 Aspects relevant to phase II trials

Tumor response is usually used as primary endpoint of phase II studies in solid tumors. For prostate, the tumor is encapsulated inside the organ, so that the size of the prostate does not reflect the size of the disease and the measurement of the tumor volume inside the prostate can be inaccurate. It is thus very difficult and unreliable to use traditional endpoints that rely on tumor measurements (such as WHO or RECIST) for phase II trials in localized disease.

In addition, prostate cancer spreads primarily to the bone and less than 10% of patients with metastatic disease primarily present with non bony visceral metastases[3,4]. Bone metastases are also difficult to measure: the appearance of new lesions can be assessed on the bone-scan but the measurement of changes in existing lesions is unreliable[3]. On the other hand, the presence of visceral metastases is associated with a poorer prognosis, thus restriction of entry to trials to patients with measurable disease only would exclude a large proportion of patients and would result in a biased patient selection, and would thus limit the applicability of the trial results to the general metastatic population. Tumor measurements are thus difficult to use as endpoint of phase II clinical trials in an (unselected) metastatic population.

This has led to the development of a consensus guideline for using the marker PSA as endpoint in phase II trials in hormone refractory prostate cancer[4]. Caution is still needed however because the expression of PSA is regulated differently by different agents and because effects of an agent on PSA may be independent of its effects on cell growth[5]. A sound understanding of the mechanics of treatment activity is thus needed to define appropriately the endpoint to use for phase II studies in prostate cancer.

Phase II clinical trials not being the focus of this thesis, we shall not further detail this aspect of clinical research and rather consider the difficulties that more directly affect phase III trials.

8.2.2 Aspects relevant to phase III trials

Prostate cancer differs in several ways other cancer types. First of all, the untreated history of localized prostate cancer can span over a decade or more whereas in most other cancer types, the disease history is less protracted. Second, unlike most other cancers, prostate cancer may remain asymptomatic for years and many patients with prostate cancer ultimately do not die of this disease[6]**.** Consequently, active treatment against prostate is often deferred until the occurrence of symptoms or biochemical signs of disease activity or even until clinical evidence of progression. In other cancer types, treatment is rarely deferred. Third, due to the slow growing nature of the disease, the control of prostate cancer progression is often equivalent to cure: delaying the progression of the disease is often sufficient to allow the patients to continue and terminate their life without or with minimal burden from the disease. This also is uncommon to most cancers that are often more aggressive and that are treated with the aim to achieve complete remission. Finally, (non-metastatic) prostate cancer is a disease that affects men of relatively advanced age in whom co-morbidities induce a substantial risk of non-prostate cancer related death[7].

In other cancers, the TNM staging system is used to classify the patients into risk groups according to their disease presentation at diagnosis and to select them for entry into clinical trials with a specific goal (cure, prevention of disease progression or prolongation of survival). In prostate cancer, the treated and untreated disease history should be taken into account when selecting patients for entry into trials and a classification based on signs and symptoms as they present at the time of diagnosis may not be as relevant than the past disease history of the given patient. These considerations led Scher[8] to develop a dynamic model of clinical states for prostate cancer to which a specific therapeutic objective is attached. Scher defines five states of prostate cancer that we here adapt to the European setting:

1. Patients with no diagnosis of cancer but who are at high risk of developing the disease. We shall not detail this patient group that is considered in prevention trials, as this is not the focus of this thesis

2. Patients with localized disease that can be cured by local therapy solely directed to the prostate (e.g. surgery or radiation therapy). In this group, the primary therapeutic objective is cure. After successful local treatment the most relevant outcome is the probability of disease recurrence. Traditionally, the endpoint used in adjuvant trials in this disease group has been time to objective clinical progression. More recently however, the risk of disease progression was shown to severely increase after increase in PSA (PSA rising above the undetectable level after radical prostatectomy, or a rise in PSA from nadir after irradiation). PSA increase is thus a strong prognostic factor for disease recurrence after curative treatment. In consequence, time to first documentation of PSA increase after local treatment has been used as an endpoint in this disease group. Formal statistical demonstration that PSA increase is a surrogate endpoint for clinical disease progression in this disease state is however still lacking. In this disease group, the risk of death due to the disease remains low and the survival is long, so that trials attempting to target a survival endpoint require a very long follow-up and a large sample size, due to the good prognosis and to the substantial competition from non-

cancer related causes that lead to information loss regarding prostate cancer survival. We discuss in chapter 5 an example of a study in this disease state.

3. Rising PSA after primary therapy. This state can be reached after deferred therapy or after local treatment with or without neo-adjuvant hormone therapy. Two categories of patients must be distinguished: patients who relapse early following local treatment or with a short PSA doubling time in whom the PSA rise is indicative of the presence of (subclinical) distant disease spread and patients with a late PSA rise or a slow PSA doubling time after local treatment that is indicative of local relapse only. In this second group, salvage local treatment can be envisaged, with the goal to cure. In the first patient group with signs of non localized disease, a systemic (hormonal) therapy is usually envisaged, with the aim to prevent clinical or symptomatic disease progression. Indeed, once clinical or symptomatic metastases have become clinically evident, the risk of death from prostate cancer becomes significant, it is thus important to prevent or slow down disease progression.
4. Metastatic disease. At the point when clinically evident metastases have developed, the major risk of death is from prostate cancer and no longer from co-morbid conditions. The therapeutic objectives may vary depending whether the identified disease spread was to the nodes or to distant sites but the objective is always to delay disease progression and to prolong survival. In this thesis we especially investigate clinical trials and endpoints in this disease group (chapters 2, 3, 4 and 7)
5. Hormone unresponsive disease. Nowadays, this group comprises both patients whose disease progressed after endocrine treatment that was initiated upon evidence of distant metastases but also patients that became hormone unresponsive after receiving endocrine treatment for a rising PSA but in absence of clinical evidence of metastases. The second group is of better prognosis than the first group. In both patient groups, the therapeutic objective consists primarily in prolonging life but in the group with metastatic hormone refractory disease, palliation of the disease symptoms and improving quality of life is an important secondary objective because the disease burden to the patient may be important. We study in chapter 6 a group of 400 hormone refractory patients with (primarily bone) metastases.

As appears from the above discussion, specifics of phase III clinical trials in prostate cancer relate to the definition of the patient groups, to the variety of treatment pathways that induce heterogeneity of prognosis and therapeutic goals, to the long natural disease history that induces long observation times to the clinically relevant endpoint, to the presence of strong markers of prognosis that are used for clinical decision making and to the co-morbid conditions that induce a substantial amount of competing risks of death.

In this thesis we utilize existing clinical trial data from past studies, most of which involve endocrine treatments against prostate cancer, to challenge a number of pathways of improvement of the efficiency of phase III clinical trial designs in prostate cancer. We shall discuss here our findings.

8.3 Is PSA response or PSA progression prognostic for the duration of survival in a population of painless good risk metastatic prostate cancer treated with anti-androgens?

In chapter 2, we explored a database of 310 patients who were randomly allocated between endocrine treatment with Flutmamide 250 mg t.i.d. or Cyproterone Acetate 100 mg t.i.d. within EORTC protocol 30892, a multicenter phase III clinical trial attempting to demonstrate a differential treatment effect onto overall survival. Patients in the study were "good risk" metastatic patients: the patients were diagnosed with clinical evidence of metastases but no pain and at least two out of three factors that were regarded as "good risk" features: WHO performance status 0, alkaline phosphatase within the normal range, T category < T4. The trial had demonstrated no survival difference between the treatment groups[9]. The median overall survival was 3.2 years and 64.5% of the patients who had died, died of their disease.

In this study, we investigated the association *at the individual patient level* between intermediate endpoints defined on the basis of PSA (PSA response, PSA progression) and the final study endpoint, overall survival, in an attempt to elucidate the value of PSA changes as *prognostic factor* for the final endpoint.

8.3.1 Rationale

The rationale for this evaluation stems from the observation that PSA response had been assessed as a prognostic factor for overall survival in metastatic prostate cancer by a number of authors who almost unanimously found a significant correlation with the clinical outcome[10-14]. A number of reports also investigated the value of PSA progression as a prognostic factor[15-17]. However, a wide range of definitions of the PSA endpoint were used in these studies whereby PSA response (progression) was defined either on the basis of a threshold value or on the basis of a relative decrease from baseline (increase from nadir). Moreover, not all evaluations used appropriate statistical techniques for assessing the association (Landmark analysis for PSA response, time-dependent covariate effect for PSA progression) and this may have biased the results. This is why we re-analyzed the data from trial 30892.

8.3.2 PSA endpoints considered

We assessed a range of definitions of the PSA endpoint and used appropriate statistical techniques. Doing so, we confirmed that a PSA decline from baseline to a value less than or equal to 1 ng/ml within a Landmark period of 6 months was strongly associated with improved prognosis: the median survival for this group was as long as 6.4 years, or double median of the unselected group, whereas the median survival of the patients whose PSA did not decline to a value below or equal to 10 ng/ml (2.5x normal value) was only 2.3 years (or

roughly two thirds of the median of the unselected group). More interesting was the analysis of time to PSA progression.

We assessed five definitions PSA progression:

1. A confirmed one hundred percent increase over nadir value to > 4 ng/ml
2. Two successive doublings over nadir value to > 4 ng/ml
3. A first 100% increase over nadir value to > 4 ng/ml
4. A first 50% increase over nadir value to > 4 ng/ml
5. A first 20% increase over nadir value to > 4 ng/ml

Definitions 3, 4 and 5 do not require a confirmation of the PSA elevation at a subsequent measurement. Definition 3 is the unconfirmed variant of definition 1. We also assessed the value of the confirmed variants of definitions 4 and 5.

8.3.3 Findings

We observed a significant prognostic association between PSA progression and survival *irrespective of the definition and also irrespective of confirmation*: the hazard of death increased 3-fold or more after PSA progression, irrespective of the definition. We then went-on to studying the sensitivity and specificity of the definitions for predicting survival and cancer specific survival at 5 years. We observed that the specificity of any definition was lower than 60% for predicting overall survival and lower than 70% for predicting specific survival. Specificity is however of lesser importance than the sensitivity of the measures, considering that it there is less damage to unnecessarily change treatment in a patient than to fail to do so when necessary. We thus based our selection of the most powerful PSA progression endpoint on maximizing sensitivity and opted for definition 5, which led to 76% and 86% sensitivity for predicting overall and cancer-specific survival, respectively. Noticeably, this definition does not require confirmation of the PSA increase but the performance of the confirmed variant remained similar (62% and 76% sensitivity respectively). With the above definition 5 taken as our best choice, 70% of the patients in the sample experienced an event of PSA progression at a median of 1.42 years after entry on study. Interestingly, the median survival time *after PSA progression* was nearly two years. This suggest that this endpoint, were it proven to be a surrogate endpoint for overall survival, could reduce the overall trial duration by about 2 years thus shortening the follow-up time by more than 50%.

Nevertheless, we could not demonstrate surrogacy in this study, neither by means of the so-called “Prentice criteria” because the trial failed to demonstrate statistically significant differences, nor by means or the meta-analytic approach, because the number of observations in the trial is too small. However, this study highlighted a number of problems which are worth discussing further.

8.3.4 Problems and limitations of this analysis

First, our analysis was a retrospective review of existing data that had been prospectively collected at the time of the study. PSA was indeed measured in the trial but it was not planned to be studied as a trial endpoint. PSA tests were scheduled at entry and at every follow-up visit *until treatment was stopped* but the PSA tests were not strictly mandatory. This forced us to exclude from the analysis a small number of patients without (baseline) PSA measurements.

Second, as PSA was measured only until treatment was stopped, we missed the PSA values from progression to death. This again led to the impossibility to obtain the confirmation of the PSA increase in a number of cases who stopped treatment immediately after their first PSA increase. This might have affected the assessment of the confirmed definitions to some extent.

On average, only 8 PSA measurements were available per patient (range 1-26). The frequency of measurements was low and several months may have elapsed between two PSA measures. This may have affected our ability to assess the nadir PSA because the nadir may have been reached in between two registered measurements. This in turn may have affected our assessment of the date of PSA progression because the definitions relate to the nadir value. In addition, infrequent PSA measurements may have affected the date of PSA progression.

Because PSA was not a trial endpoint, no central laboratory was set up for analyzing the PSA samples in this study. This led to some complications ad the normal values differed between centers. Before statistical analysis, the PSA values were standardized by the local institution upper normal value and all were then transformed back to a range referring to a normal value of 4 ng/ml. These modifications may also have induced some heterogeneity because the Data Center may not have been informed in due time of changes in the kit used at some institutions and thus the upper normal values may have been inaccurate for some observations.

Despite the above shortcomings, we were able to demonstrate the presence of a strong association between the PSA endpoints and survival, even if there may be some imprecision in the estimation of the actual association.

Nevertheless, we had to recognize that the study sample was atypical as it consisted of a subset of metastatic patients that was already pre-selected on the basis of good risk features. In addition, the study treatments were atypical: the anti-androgen monotherapies tested in this study are no longer considered standard treatment for this patient group. Thus, the main limitation of this prognostic factor study resides in the difficulty to generalize the results to an unselected group of metastatic prostate cancer patients treated with a current standard treatment involving castration.

The other limitation was that we could not assess the association *at the trial level* between the treatment effect on the PSA endpoints and the survival endpoint, that is, we could not demonstrate surrogacy with this database.

For these reasons, we searched for other databases of studies involving an unselected patient population, in which reliable PSA measurements were available and that were large enough to undertake a formal validation exercise of the PSA endpoints as potential *surrogate* for the survival endpoint. Thanks to the active collaboration of AstraZeneca Pharmaceuticals, we could locate and obtain individual patient data from three of their major studies of Casodex in metastatic cancer.

8.4 How to demonstrate surrogacy?

8.4.1 Prognostic factor

We already expressed in the introduction how the notion that a marker be a prognostic factor of disease outcome differs from the notion that that same marker be a surrogate for the same outcome taken as endpoint in randomized trials.

To establish a marker as a prognostic factor for the endpoint of interest, a strong association *at the patient level* between the marker and the true endpoint must be demonstrated. Such an association means that for every individual patient, the evolution of the marker is associated with improved (decreased) risk of experiencing the true endpoint. This notion does not relate at all to treatment comparisons but only to association between a marker and a disease outcome. Ideally, prognostic factors ought even to be established in an untreated population[18,19] in whom the natural evolution of the disease can be observed. However, it is also accepted[20] that prognostic factors can be established for the disease evolution observed with a given (standard) treatment. Prognostic factors are useful to for informing on decision to take as regards the management of an individual patient.

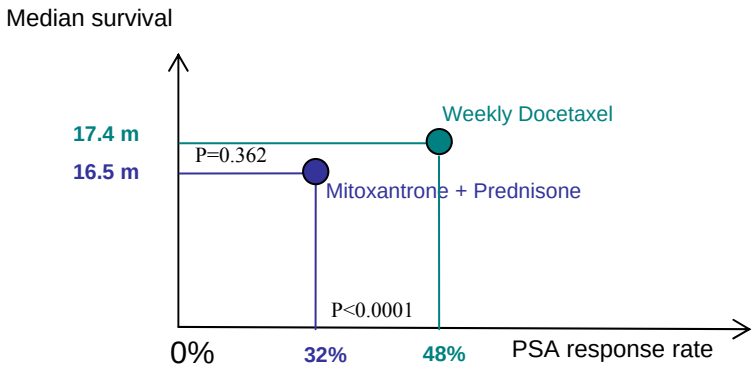
8.4.2 A prognostic factor does not make a surrogate: an example

Contrary to the concept of prognostic factor, the concept that a marker s a surrogate for the clinical endpoint of interest refers directly to the setting of randomized treatment comparison. This concept requires that a strong association exists between treatment effects affecting the marker-based endpoint and the true endpoint. Obviously, this concept refers to groups of patients, as a same patient cannot be managed according to two distinct policies in a same given disease state. Thus surrogacy is a notion that refers to associations *at the trial level (ie: considering groups of patients)* between the observed differential effect of two distinct treatment strategies onto the putative surrogate endpoint and the differential effect of the same two treatments onto the true endpoint.

That a marker be a strong prognostic factor is not sufficient to establish that marker as a surrogate endpoint. Indeed, although association between the marker evolution and the true endpoint may be present in all patients, the strength or shape of that association may vary across patients and/or in relation to the treatment. Therefore proven prognostic factor association at the patient level does not necessarily induce association between the endpoints when comparing groups of patients[21] and thus is not sufficient to prove surrogacy**.**

To illustrate this situation, let us consider the recently published secondary results of the Aventis Tax-327 study. This study compared a weekly and a three-weekly schedule of docetaxel plus prednisone to mitoxantrone and prednisone in hormone refractory prostate cancer[22]. In this study like often in this disease state, patients who achieved a PSA response had a 60% reduction in mortality risk compared with non responders (HR=0.40, 95% CI: 0.31-0.51). The reduction of the PSA by 50% or more from baseline value, which was defined as a PSA response, is a strong prognostic factor for survival. Now considering PSA response as an endpoint and as a putative surrogate for overall survival, we observe with the authors that the weekly docetaxel arms resulted in a response rate of 48% which was significantly different from the 32% response rate that was obtained with standard arm mitoxantrone plus prednisone (P<0.0001). However, the median overall survival on the weekly Taxotere arm amounted 17.4 months and did not differ significantly from the 16.5 months median survival achieved with the standard treatment (P=0.362, Figure 8.1). Thus in this study, PSA response although it was a strong prognostic factor for survival at the patient level, did not appear to be reliable as a surrogate for survival when comparing the weekly Docetaxel treatment to mitoxantrone plus prednisone. Of note, in this study, the PSA response rate and median survival with the 3 weekly Docetaxel schedule were respectively 45.4% and 18.9 months.

Figure 8.1
A prognostic factor does not make a surrogate endpoint – the Tax 327 trial



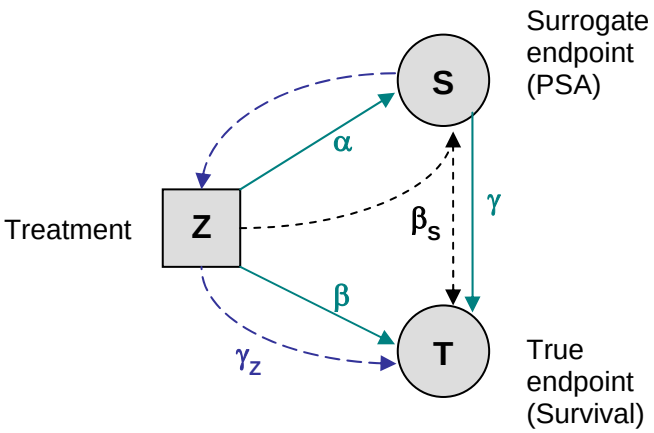
We and others have shown that PSA-derived endpoints are significant prognostic factors for overall survival in metastatic prostate cancer, but this is not enough to establish their use as surrogate endpoints.

8.4.3 Validation using data from a single trial: the Prentice Criteria and associated problems

Traditionally, the "Prentice Criteria"[23] were used for the purpose of demonstrating surrogacy on the basis of data from one single trial.

Expressing the criteria in a similar way to Buyse et al.[24] we will consider statistical tests of association between a treatment Z, a putative surrogate endpoint (PSA denoted S) and a true endpoint (T) **(**also see Figure 8.2 for further definition of symbols**)**. To be a surrogate according to Prentice, PSA should fulfill the following conditions, where H0 represents the null hypothesis of a statistical test and H1 represents the alternative hypothesis that the tested effect is non null.

Figure 8.2
The Prentice Criteria



(a) In a model assessing the impact of the treatment onto the PSA endpoint, there must be a statistically significant treatment effect onto the PSA endpoint (H1: $\alpha \neq 0$)

(b) In a model assessing the impact of the treatment onto survival, there must be a statistically significant treatment effect on survival (H1: $\beta \neq 0$)

(c) In a model assessing the impact of the PSA endpoint onto survival, the PSA endpoint must be a statistically significant prognostic factor for survival (H1: $\gamma \neq 0$)

(d) In a survival model that includes both the treatment and the PSA endpoint as explanatory variables, the treatment effect on survival must be *non significant.* (H1: $\beta_S \neq 0$)

Obviouly, Prentice's condition (b) requires that the trial used for the assessment shows statistically significant benefit for the primary endpoint. In metastatic prostate cancer, very few studies have demonstrated statistically significant benefits of new treatments. For example only 3 of the 27 trials (EORTC 30853, NCI-INT0036 and PONCAP trial) included in the meta-analysis of maximal androgen blockade in metastatic patients showed statistically significant benefit at the conventional 5% significance level and the meta-analysis itself did not demonstrate a statistically significant benefit[25]. The opportunities to demonstrate surrogacy of PSA have thus been scarce. In addition, condition (b) also prevents the use of studies that aimed at demonstrating non-inferiority or absence of treatment difference.

In essence, Prentice's condition (d) states that the treatment effect on survival should be mediated in full by the treatment effect on the intermediate PSA endpoint. This is medically unlikely, especially for overall survival, as other factors than PSA affect the mortality due to other causes than prostate cancer. For specific survival, it is also known that some patients will progress without increased expression of PSA[26]**.**

Furthermore, condition (d) raises a conceptual difficulty and a technical impossibility: the criterion requires that the test for the effect of the treatment on survival be *non significant* after adjustment for PSA. However, a *non significant statistical test* does not prove that the null hypothesis of no treatment effect is true: a non significant statistical test may be the consequence of lack of power or lack of chance[27].

8.4.4 Freedman's proportion explained

In order to move away from the problem of having to prove the absence of a treatment effect in Prentice's condition (d), Freedman[28] proposed to calculate the proportion of the treatment effect mediated by the surrogate. This quantity is known as the "Proportion Explained" (PE). With the above notations, the PE is defined as the following ratio:

$$PE = \frac{\beta - \beta_S}{\beta} = 1 - \frac{\beta_S}{\beta}$$

where β represents the treatment effect on the survival endpoint by univariate analysis
and β_S represents the treatment effect on the survival endpoint as assessed in a multivariate model with adjustment for both treatment effect and the effect of PSA on survival.

A number of problems surrounding PE have been identified in the literature and were nicely summarized by Molenberghs[29].

First of all, when the (unadjusted) treatment effect onto the survival endpoint (β) is close to zero (a situation that is frequent in cancer clinical trials), PE will tend to be unstable. Second, the confidence intervals surrounding PE will also tend to be very large and may span outside the [0,1] interval or even be unbounded. Finally, it can easily be shown that

PE is in fact *not a proportion*: whenever the adjustment for the surrogate (PSA) in the model of the treatment effect onto survival results in a change of the direction (sign) of the treatment effect, the ratio (β_S/β) is negative, and thus PE is greater than one! We shall now show an example of this using data from EORTC trial 22911 which was presented in chapter 5.

8.4.5 An example to illustrate the problems related to the Proportion Explained: EORTC trial 22911

To illustrate our arguments against the proportion explained (PE), we re-considered the EORTC trial 22911 which we studied in chapter 5. In that study which randomized pathologically high risk patients between post-prostatectomy irradiation and no further immediate treatment, biochemical (or clinical) progression-free survival (BPFS) was used as the primary trial endpoint, after an amendment and the original primary trial endpoint clinical progression-free survival (PFS).

Biochemical progression was defined as every increase over the lowest postoperative value that is confirmed twice, at minimum 2-week intervals. This study is the only one testing the value of immediate adjuvant irradiation after radical prostatectomy in pT3 prostate cancer. Thus a meta-analytic validation is not possible. The Prentice criteria and the relative effect can however be calculated: the study showed a statistically significant treatment effect onto BPFS (the putative surrogate in this example: (P<0.0001, HR=0.48 based on 351 events) and onto PFS (the true endpoint in this example: P=0.0004, HR=0.61 based on 188 events). The results of the four Prentice's criteria are summarized in Table 8.1. The results show that all four conditions are fulfilled since α, β and γ are statistically significantly different from zero whereas β_S is not statistically significant.

Table 8.1: Validation by the Prentice Criteria in EORTC trial 22911

Effect of	HR (95% CI)	P-value	Parameter of interest
(a) Treatment on BPFS	0. 48 (0.39-0.60)	P<0.0001	α= -0.73267
(b) Treatment on PFS	0.61 (0.46-0.82)	P=0.0011	β = -0.48813
(c) BPFS on PFS	11.3 (8.2-15.5)	P<0.0001	γ = 2.42399
(d) Joint model:			
BPFS on PFS	11.6 (8.3-16.2)	P<0.0001	γ_Z = 2.45520
Treatment on PFS	1.09 (0.80-1.48)	P=0.5799	β_S=0.08678

Of note, the sign of the treatment effect changed when the model was adjusted for BPFS: $\beta<0$ and $\beta_S>0$. Thus the proportion explained (PE) in this example takes a value above one showing that PE is indeed not a proportion.

$$PE = 1 - \frac{\beta_S}{\beta} = 1 - \frac{0.08678}{-0.48813} = 1.17778$$

8.4.6 Buyse and Molenberghs Relative Effect

To move away from the problems related to the proportion explained, Buyse and Molenberghs[30] suggested the use of another quantity for the validation of surrogate endpoints: the "Relative Effect" (RE). RE is defined as the ratio of the effects of treatment upon the surrogate endpoint (PSA) and upon the true endpoint (survival). Using the same notations as above, RE is defined as the following ratio:

$$RE = \frac{\beta}{\alpha} = \frac{Treatment\ effect\ on\ true\ endpo\,\text{int}}{Treatment\ effect\ on\ surrogate}$$

where β represents the treatment effect on the survival endpoint by univariate analysis
and α represents the treatment effect on the PSA endpoint by univariate analysis

Using this measure, a surrogate is said to be perfect at the population level is RE equals 1. Confidence intervals for RE (like for PE) can be derived mathematically. Unlike the ratio PE, the quantity in the denominator of RE is the treatment effect on the surrogate endpoint. That quantity α, unlike β (the treatment effect onto the survival endpoint), is expected to be reasonably large in real situations. Therefore RE can usually be estimated more precisely than PE.

In the EORTC trial 22911 described above, the Relative Effect takes value
$RE = \frac{\beta}{\alpha} = \frac{-0.48813}{-0.73267} = 0.666$ with 95% confidence limits: 0.327-0.956.

The point estimate of the relative effect indicates a moderate association between the endpoints but the confidence interval around this value is very wide and spans from a low association (0.327) to a very strong association (0.956). The results are therefore not extremely convincing that BPFS can be used as a surrogate for PFS. This conclusion is in line with the views of the Independent Data Monitoring Committee of trial 22911 who, in light of clinical judgment, accepted to change the primary trial endpoint from PFS to BPFS and that the BPFS results be published *with the proviso* that the follow-up be continued and that long term definitive evidence for the clinical endpoints be published at a later date.

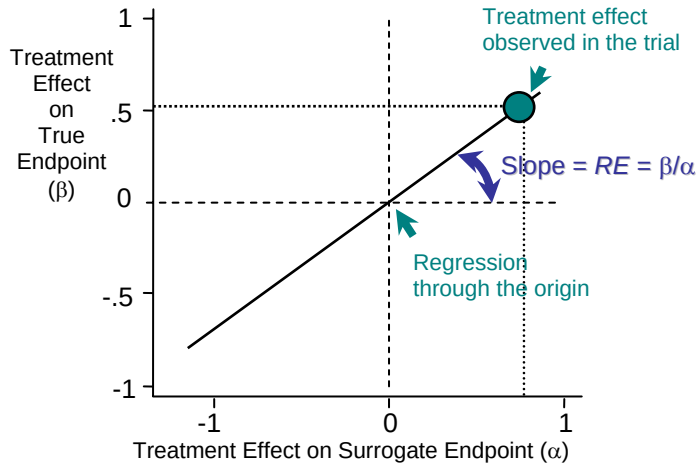
8.4.7 Moving away from statistical testing: RE as a prediction model estimated from a single observation

The idea underlying the relative effect is the idea of a prediction model: the ratio RE can be viewed as the slope of a regression line of the treatment effects on the true endpoint over the treatment effect on the surrogate endpoint (Figure 8.3).

The regression line would be used to predict the treatment benefit expected on the true endpoint (β) from the treatment benefit observed on the surrogate endpoint (α). The precision of the prediction would increase with increasing precision in the estimation of RE. The precision in the estimation of RE is directly proportionate to the size of the clinical trial.

Thus, with RE, one has moved away from the hypothesis testing paradigm of the Prentice criteria to an estimation paradigm.

Figure 8.3
Prediction using data from one single trial: the Relative Effect (RE)



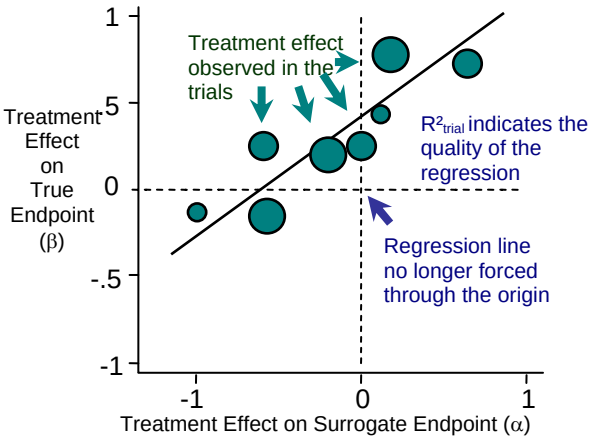
Whenever the validation exercise is carried out using data from only one single trial, the regression line is undefined unless one adds the constraint that is should pass through the origin. To force the regression through the origin makes a very strong assumption: one assumes that the relationship between the treatment effects on the surrogate and the true endpoint would be constant for all trials testing the therapeutic question under consideration. This assumption may not hold true in practice and *cannot be tested using data from one single trial*.

In the example of trial 22911, the point estimate of RE was 0.666 with 95% confidence interval (0.327 – 0.956). Assuming that in a new study, the relationship between the treatment effect on biochemical progression free survival and on clinical progression free survival would be *exactly as in trial 22911*, and if one had observed the treatment effect on BPFS in this new trial (say a relative risk reduction α of 60% for BPFS), one could infer that the effect of the treatment on PFS in that new study would be roughly two thirds as large as the effect seen on BPFS (thus a relative risk reduction of roughly 40%) but the confidence interval would tell that the actual predicted relative risk reduction would be between one third and 95% that is a relative risk reduction between 20% and 57%.

8.4.8 Validation using data from several trials: the meta-analytic method

The shortcoming that the validation of surrogacy using RE relies on a strong unverifiable assumption can be solved whenever data from several trials are available: with several observation units (e.g. trials) the regression line is no longer forced to pass through the origin and the variability in the association of treatment effects across trials can be integrated in the assessment. This is the idea underlying the meta-analytic approach that we used in chapter 3. With the meta-analytic method (Figure 8.4), the squared correlation coefficient of the regression line, that we shall denote R^2_{trial} measures the quality of the regression line. Indeed R^2_{trial} is a measure of the strength of the linear association between the treatment effects on the surrogate and on the true endpoint. R^2_{trial} can also be interpreted as a measure of uncertainty in predicting the treatment effect on true endpoint (survival) from the treatment effect on the surrogate (PSA) endpoint. Thus R^2_{trial} is a reasonable measure of surrogacy at the trial level.

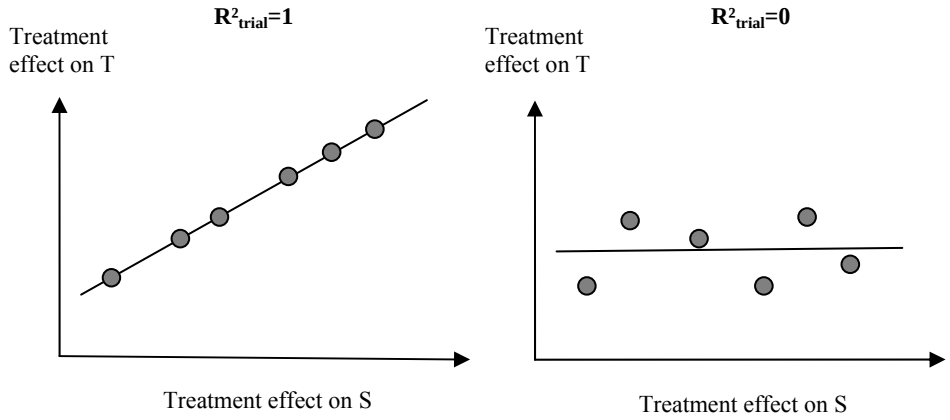
Figure 8.4
Prediction using data from several trials: the meta-analytic validation method



A perfect surrogate would have a value close to unity[31,32]. Indeed, in a situation where all observations would lie exactly on the regression line, R^2_{trial} would take the value 1, indicating that the treatment effect on survival would be perfectly predictable from the effect observed on PSA (Figure 8.5). If there would be no association between the treatment effects, that is if the regression line was horizontal, R^2_{trial} would take the value 0.

With the meta-analytic method, a valid surrogate is a surrogate that allows a precise prediction of the treatment effect on the true endpoint from the treatment effect observed on the surrogate endpoint: the higher R^2_{trial} and the narrower the confidence interval around it (i.e.: the higher the confidence in the estimated value of R^2_{trial}), the better the quality of the prediction and the more useful the surrogate.

Figure 8.5
Interpretation of R^2_{trial}



8.5 Is PSA as surrogate endpoint for survival in metastatic prostate cancer receiving endocrine therapy involving Casodex? – An application of the meta-analytic validation method

AstraZeneca Pharmaceuticals shared with us the data from 2161 patients with primary diagnosis of metastatic prostate cancer who had been treated within their (bicalutamide) Casodex development program and for whom PSA data were available. Because the trials were designed to show non-inferiority of Casodex monotherapy compared to castration or of casodex-based combined androgen blockade versus flutamide-based blockade, the Prentice method could not have been used to demonstrate surrogacy using this dataset. However, since a large number of observations were available, we could split the data set

into 21 units of analysis on the basis of the trial and of the patient's country of residence to utilize the meta-analytic validation approach.

8.5.1 Putative surrogate PSA endpoints considered

We considered overall survival as the true endpoint of interest. We successively assessed as potential surrogates the following PSA endpoints: PSA response (a decline of the PSA by 50% or more relative to baseline, to a level below the normal), PSA normalization, time to PSA progression (TTPP), and the complete series of PSA measurements ('PSA profile') as potential surrogates for overall survival. We considered two alternative definitions of time to PSA progression. The first one is that which was shown to be the most sensitive in chapter 2, namely, PSA progression was defined as a PSA value above normal (4 ng/ml) representing a first increase equal or larger than 20% above the nadir (TTPP1). The definition had been used in other trials[33] and was defined as a PSA value above 2.5 times the normal range (10 ng/ml) representing a first increase equal or larger than 50% above the moving average (based on three consecutive measurements) nadir.

8.5.2 Findings

Our analyses confirmed the known value of PSA as a prognostic factor of survival (individual-level association). However at the trial-level, the association between the treatment effect on any PSA-based endpoint and the treatment effect on overall survival was in general low ($R^2_{trial} < 0.69$ with wide confidence intervals). Logically, the greatest association was observed when the complete series of PSA measurements on each patient was used as surrogate. It would however be difficult to use this endpoint in clinical trials because the treatment effect on this longitudinal endpoint cannot be summarized easily and would thus be impractical in real life.

The next best endpoint was TTPP-2: the time to PSA progression defined as a confirmed 50% relative increase above the previously observed nadir. The meta-analysis using this endpoint resulted in $R^2_{trial}= 0.66$ (SE=0.13) with associated 95% confidence interval ranging from 0.30 to 0.85. The association at the individual patient level (prognostic factor association) was of similar magnitude. Sensitivity analyses using prostate-cancer survival as the true endpoint led to similar results.

8.5.3 Interpretation of the results

Is this value of R^2_{trial} (0.66 with SE=0.13) large enough to warrant the use of TTPP-2 as surrogate for survival in future trials of endocrine treatment in metastatic prostate cancer?

The choice of the threshold for R^2_{trial} required for a valid surrogate is still a matter of debate[24]. Nevertheless, we argue that the precision of the prediction of the treatment effect on OS from the effect on the PSA-based endpoints, indicated by the R^2_{trial} values observed

in our study, is insufficient to claim that TTPP-2 is a statistically valid surrogate endpoint for overall survival.

To illustrate this problem, let us consider a new trial with time to PSA progression as primary endpoint (defined as TTPP-2) where data analysis occurs after 400 events and yields a hazard ratio of 0.75 for PSA progression (with 400 events, SE(log(HR) would be of the order of 0.10 resulting in $P<0.01$). Without adjusting for the estimation error in the parameters of the prediction model and using our results, one could predict with approximately 95% confidence that the corresponding survival hazard ratio would lie within the interval [0.48-1.12]. Adjustment for the estimation error would widen the confidence interval even further.

From this example we concluded that non null treatment effects on survival would potentially be identifiable only in very large new trials showing a very large effect on the PSA endpoint (e.g. HR around 0.50 with SE=0.10). A precise enough prediction of the treatment effect onto overall survival necessitates that the hazard ratio for the PSA endpoint be estimated very precisely. Thus, with the information at hand, a trial based on the PSA endpoint would not require fewer patients than survival trial.

As an example of a successful validation exercise using a similar method. D. Sargent et al.[34] reported at ASCO 2004 the results of a validation study of disease-free survival as a surrogate for overall survival for adjuvant colorectal cancer studies. Using data from 12915 patients entered in 15 randomized studies, they could demonstrate a correlation coefficient of 0.90 (95%CI: 0.81-0.96) between the treatment effects on disease-free survival and on overall survival. With these results, they concluded that disease-free survival at 3 years is a valid surrogate for overall survival in this disease. Their results enable to predict that an observed hazard ratio of 0.80 on disease-free survival would translate into a hazard ratio of 0.83 for overall survival, with a 95% confidence interval ranging 0.76-0.90. Since all the trials involved 5-Fluoruracil-based chemotherapy, they also indicated that the correlation might be less for studies involving targeted agents or other non-cytotoxic agents that delay rather than prevent disease-progression.

8.6 How to use PSA to shorten phase III trials in prostate cancer?

From our results in the Casodex database, we concluded that PSA endpoints do not completely fulfill the statistical requirements for being surrogates of long term clinical outcomes in hormonally treated advanced prostate cancer. We suggest however, that PSA could still be used to shorten the duration of phase III trials.

8.6.1 Using PSA for early registration pending subsequent confirmation on the basis of the survival endpoint

For a new trial where early registration on the basis of a PSA endpoint would be envisaged, we suggest that the trial be designed with time to PSA progression or biochemical progression-free survival as the intermediate endpoint and survival as the definitive endpoint.
The trial sample size should be determined to demonstrate a very large effect on the PSA endpoint with great precision precision, for example to demonstrate the presence of a hazard ratio of the order of 0.50 on the PSA endpoint, with a standard deviation of the order of 0.10. The trial sample size calculation ought not to be on power considerations, as these would necessarily result in very small sample size due to the large target effect, but should be based on the required precision of the estimation of the treatment effect onto the PSA endpoint.

An interim analysis plan should be set up with plan for one or several interim looks at the PSA endpoint as well as to the safety data and one longer term analysis on the survival endpoint. At the interim looks, the trial results on the PSA endpoint could be used to estimate a prediction of the survival treatment effect using the regression results from former meta-analytic validation exercises. Whenever the prediction interval for the survival hazard ratio would exclude the null effect, the trial results could be submitted for early registration on the basis of the PSA results.

In the light of the fact PSA is unlikely to capture all the potential (negative) effects of the treatment on survival and because PSA did not qualify as a surrogate endpoint, we would not recommend to stop the study follow-up when the significant PSA results become available. On the contrary, we would recommend that the follow-up should be continued to later document long term safety of the treatments and their impact on survival.

We thus suggest that on the proviso that longer term analysis of the trial be planned to confirm the findings, time-to-PSA progression, could potentially serve as a basis for accelerated drug approval, together with other trial data documenting safety and other measures of patient benefit such as Quality of Life.

8.6.2 Using PSA to stop a study early for futility

If a clinical trial is set up to test a new treatment the effect of which is known (from preclinical and early phase studies) to be expressed or mediated by PSA, one could also build an early stopping rule to stop the study if no effect is seen on the PSA.

Indeed, in such circumstances, a treatment effect on the PSA endpoint might be seen as a prerequisite for an ultimate effect on survival, even if a significant effect on PSA would not necessarily translate into a significant survival benefit.

Thus along the lines proposed by Royston and Parmar[35] one could design a study with survival as the final endpoint, but with planned interim looks at the PSA endpoint and a decision to stop the study for futility if an insufficient benefit or a negative effect of the experimental treatment (e.g. HR>0.5) was seen on the PSA endpoint. The assumption underlying this design is that a significant treatment effect on the final endpoint (survival) would be very unlikely if no beneficial effect was seen on the intermediate endpoint (PSA). The information on the correlation between the treatment effects on the PSA endpoint and on survival that is necessary to the set up of stopping rules according to Royston and Parmar is readily available from the value of R^2_{trial} obtained from the meta-analytic validation.

Thus although perfect surrogacy of PSA as a substitute for survival endpoints remains unproven, PSA could be useful to shortening the duration of phase III trials using the above approaches to interim looks and decisions to stop or release the trial results early.

8.7 Can our validation results in endocrinally treated M1 disease justify the use of PSA-endpoints as surrogate endpoint in other disease states and/or in studies involving other than endocrine treatments?

One could now question whether our results could serve to justify the use of PSA in other states of prostate cancer or in trials testing other treatments than hormone therapy. On strict statistical grounds, a validation exercise ought to be repeated in each considered setting. This of course limits a lot the usefulness of the validation exercise because the validation itself requires that large amounts of data from past studies be available, whereas surrogate endpoints are needed to speed up drug development.

To our knowledge, two other validation exercises similar to ours were carried out in prostate cancer. The first involved a group of patients with hormone refractory disease that were treated with Liarozole or an anti-androgen[33]**.** The second used AstraZeneca's Early Prostate Cancer Program[36]**:** over 8000 patients localized or locally advanced disease who were treated with Casodex 150 mg daily versus placebo in addition to standard care (radical prostatectomy, radiotherapy or watchful waiting).

8.7.1 Meta-analytic validation in hormone refractory disease treated with Liarozole

In the first study, Buyse et al.[37] tested the endpoint of PSA response and TTPP-2 using the same definitions that we used. They showed that despite a strong association at the individual patient level, none of the endpoints qualified as a surrogate for overall survival: R^2_{trial} was < 0.45 for all tested PSA endpoints. One of the reasons for the lack of association may relate to the mode of action of liarozole which is an imidazole-like compound that causes elevation of retinoic acid, postulated to have anti tumour activity and which effect may not be mediated by PSA. Other reasons for the lack of association might be that the

patient population was very advanced and that PSA expression might be affected by tumor dedifferentiation. This suggests at least that surrogacy of PSA endpoints might not be transportable between treatments with very different modes of actions or which effect on PSA is expected to differ substantially.

8.7.2 Meta-analytic validation in localized or locally advanced disease treated with Casodex

In the second study, Newling et al.[36] regarded time to objectively confirmed disease progression or death from any cause (PFS) as the true endpoint, and time to PSA progression defined as a 50% increase over baseline level, or objective progression or death (BPFS) as the putative surrogate. They report a value of R^2_{trial} of 0.65 (95% CI: 0.55-0.92) for the whole group of 8000 patients and of 0.52 (95% CI: 0.37-0.89) in a sensitivity analysis including only the patients from the two out of three studies carried out in Europe. Their conclusions were similar to ours: a large positive treatment effect onto BPFS is likely to reflect a clinically important benefit from Casodex as regards clinical PFS. However, it is to be noted that part of the association may be induced by the overlap in their definition of PFS and of BPFS, and that if BPFS was used as the trial endpoint, the observation time would not be reduced for the patients in whom the two endpoints are identical.

Although the final endpoint is clinical progression-free survival and not overall survival, there is similarity between their findings and ours. Their study as well as ours involved Bicalutamide-based treatment, in a setting adapted to the disease stage and the final endpoints that were considered for the validation reflected the most clinically relevant clinical event in the respective patient group.

The similarity of the findings in the two Casodex trials (the value of R^2_{trial} was identical) may suggest that the validation results are not very much affected by the disease stage as long as the mode of action of the treatments being assessed are the same.

8.7.3 The Tax 327 study in hormone refractory patients: PSA response would have led to misleading conclusions as regards the value of the weekly Docetaxel arm

The validation studies discussed so far are the only ones that are based on the meta-analytic approach. Other published studies used the Prentice criteria[38-41] but as explained above, fulfillment of the Prentice operational criteria is not sufficient to prove surrogacy (they may be fulfilled even if the marker is not a good a surrogate). Thus the criteria are mostly useful to disregard a marker as being a surrogate whenever it does not fulfil the criteria. To this respect, the data from Aventis Tax-327 trial of Docetaxel plus Prednisone versus Mitoxantrone plus predisone in advanced hormone escaped prostate cancer[22,42] are interesting. Indeed, the study demonstrated a significant survival advantage of the 3-weekly Docetaxel arm over the control arm ($P<0.01$) but showed no significant survival benefit of

the weekly Docetaxel arm over the control (P=0.08). However, if one had based the study on PSA response (defined as a confirmed decline by 50% from baseline), one would have concluded that both experimental arms were significantly better than the control: the PSA response rate was 45% and 48% on the 3-weekly and the weekly Docetaxel arms, versus 32% on the control arm, and both comparisons were significant at the 0.001 level. This shows that in this particular setting, PSA response ought not to be used a surrogate for survival. Whether this would invalidate the use of PSA endpoints in other disease states where treatment with Docetaxel is being tested remains unanswered.

8.7.4 The RTOG trial 92-02: the validation by Prentice Criteria failed

Sandler et al.[39] showed at ASCO 2003 that in the RTOG trial 92-02 that compared short term versus long term androgen deprivation in addition to irradiation for T2c-T4 prostate cancer, time to PSA failure (defined using the ASTRO definition) was not a surrogate for cancer-specific survival: the PSA endpoint failed the fourth Prentice criteria. In that study, time to PSA failure was longer on the long term androgen deprivation arm but the survival time after PSA failure was shorter. The authors postulated that some patients who had biochemical failure on the long term androgen deprivation arm, may have already had hormone insensitive disease at the time of PSA relapse and thus decreased responsiveness to salvage treatment. They concluded that time to PSA failure should not be used as a surrogate endpoint in trials that test endocrine treatment of differing duration.

In summary, we concluded that PSA could possibly be used as an intermediate endpoint under specific conditions (large expected effect, large sample size, knowledge of the treatment mode of action, or as an indicator for early stop for futility) but that the evidence is not strong enough to support its use as the final endpoint in phase III prostate cancer trials. Surrogacy is likely not to be extrapolated between treatments that differ in their mode of action. Extrapolation between disease stages when the trial designs are similar might be possible. Surrogate endpoint validation requires large amounts of data, and this limits the development of surrogate endpoints.

8.8 Why is the knowledge of prognostic factors of outcome and predictive factors of treatment benefit important to the design of trials?

The question whether the results of a study are convincing enough is related to the issue of the comparability of the treatment groups. Lack of balance in important prognostic factors may bias the results of clinical trials, especially in cancers where the impact of prognostic factors outweighs by far the expected treatment benefits. The knowledge of prognostic factors is therefore very important for the purpose of stratifying the randomization into clinical trials[2] in order to ensure that these factors are well balanced between the treatment groups.

Whenever important prognostic factors are present, analyses that adjust for prognostic factors can be performed. Simulation studies recently performed by Anderson et al.[43] showed that statistical power may be improved by including any strongly predictive covariates in the analysis, regardless of their use in the randomization. They also showed that there can be a substantial loss in power when highly predictive covariates were not accounted for in the analysis, especially if they were used as stratification factors in the randomization. Thus knowledge of important prognostic factors can improve the efficiency of the statistical analysis of the trials.

Another aspect relates to heterogeneity in the magnitude of treatment benefit across groups of patients: statistical analysis of this heterogeneity may suggest that not all patients benefit from the tested therapy or that they do not all benefit to the same extent. Some baseline factors may indeed be *predictive* for the magnitude of the benefit from the tested treatments. Careful exploratory investigations of the results ought to be carried out in order to identify such predictive factors. Because such analyses induce subgroup analyses, appropriate methodology involving an overall test for the presence of a factor-by-treatment interaction ought to be carried out, in order to limit the dangers of spurious findings generated by multiple comparisons in small patient subgroups[2]**.** The identification of predictive factors is complex, because prognostic factors may or may not be predictive factors and conversely, and because of the complex associations between treatment effects and prognostic and predictive factors. The knowledge of predictive factors is however obviously useful to the clinical treatment decision-making, but is also useful to inform on the patient selection for entry into subsequent phase III trials testing similar treatments.

In chapters 4 and 5, we revisited two EORTC clinical trials and assessed both prognostic factors and predictive factors of the primary outcome.

8.9 Prognostic factors of clinical progression-free survival in a phase III trial of adjuvant hormone treatment versus nil in a population of high risk T3-4 or G3 prostate cancer patients treated with external beam irradiation: Does adjuvant hormone therapy benefit to all patients?

8.9.1 The trial

EORTC trial 22863 compared external beam irradiation with three years of adjuvant endocrine treatment with the LH-RH analog Zoladex to external beam irradiation alone in a group of patients with high risk (T3-4 N0 WHO grade G1-G3 or T1-2 N0 G3) prostate cancer. The trial demonstrated a statistically significant benefit of adjuvant treatment onto the trial primary endpoint (clinical progression free survival: HR=0.34, 95% CI: 0.26-0.46) and on overall survival (HR=0.51, 95% CI: 0.36-0.73).

8.9.2 Prognostic factors

In chapter 4, we identified three independent prognostic factors for an increased risk of clinical progression or death: WHO performance status >0, WHO differentiation grade G3 and baseline PSA >10 ng/ml. We then formed three risk groups on the basis of these factors: a low risk group to which belonged all patients with no risk factors or WHO PS>0 as the only risk factor, a high risk group for the patients with 2 or 3 risk factors and an intermediate risk group for all other patients. The 5-year clinical disease-free survival was strongly dependent on the risk group: it was 83.8% in the low risk group, 60.0% in the intermediate risk group and 43.4% in the high risk group. This demonstrates that prognostic factors may indeed severely affect the outcome.

8.9.3 Assessment of heterogeneity

Because we thought that the benefit of an additional treatment, here adjuvant endocrine therapy, might be proportionate to the absolute risk of death; i.e. that poor risk patients might benefit from adjuvant treatment to a larger extent than good risk patients. Therefore we decided to test whether the risk classification was also predictive of the benefit from adjuvant treatment or in other words, to assess if the treatment benefit was homogeneous over the risk groups or if some patient groups benefited more or less from the treatment. For that purpose, we tested the presence of a treatment by risk group interaction effect by means of a Logrank test for interaction. That test showed that there was no statistical evidence of heterogeneity in the treatment effect across the risk groups (P=0.20). In consequence, one can conclude that the treatment benefited all patients groups. Thus in this trial, the prognostic factors were not predictive for the treatment effect.

8.10 Prognostic factors of biochemical progression-free or clinical progression-free survival in a phase III trial of adjuvant irradiation in a population of patients with pT3N0 prostate cancer after radical prostatectomy: Does adjuvant irradiation benefit to all patients?

In chapter 5 we considered EORTC trial 22911, that we already referred to earlier.

The patients entering this trial had to present post-prostatectomy with at least one out of three pathological risk factors of: positive surgical margins (SM+, including at the level of the prostate apex where the capsule is non existent), capsule perforation (ECE+), or invasion of seminal vesicles (SV+). The patients were thus pT_2 with positive margin or pT_3. The three factors were however used as separate stratification factors in the minimization program that was used for the randomization of patients in the trial. The balance of the treatment arms within the subgroup defined by each pathological factor taken separately was thus guaranteed.

8.10.1 Assumptions underlying our approach to predictive factors identification

The trial involves a group of patients treated with radical prostatectomy alone and a group that received immediate adjuvant irradiation. We were primarily interested in finding *predictive factors* for the benefit of adjuvant treatment.
We assumed that the potential benefit of irradiation was proportionate to the patient's absolute risk of biochemical or clinical failure. For that reason, we first identified the prognostic factors of the risk of relapse in the treatment arm without adjuvant irradiation.

Another reason to this approach is that, assuming that a factor X is a predictive factor for the benefit of the treatment, it implies that the observed prognostic effect of that same factor X would differ between treatment arms. Thus, the identification of prognostic factors would be confounded by the differential treatment effect, if one was considering the whole group for the prognostic factor analysis.

8.10.2 Prognostic factors in the control arm

After model reduction, five factors were retained as independent predictors of increased risk of clinical or biochemical failure in the control group: a PSA>0.2 ng/ml post-surgery (within 3 weeks after surgery or prior to irradiation, if any), invasion of seminal vesicles, elevated PSA prior to surgery, poor tumor differentiation (WHO G grade) and positive surgical margins (SM+). Of note, when this model was applied to the irradiated group, pre-operative PSA and positive surgical margins were no longer statistically significant and their associated hazard ratios were close to unity (P=0.2873 and P=0.3188, respectively). This may be related to interaction (predictive) effects of these factors for the treatment benefit.

8.10.3 Search for predictive factors

We then assessed if the treatment benefit was homogeneous across three risk groups defined on the basis of the above model. The analysis indicates no substantial heterogeneity in the magnitude in the hazard ratios for the benefit of postoperative irradiation across the three risk groups (Logrank for heterogeneity P>0.1). Thus, when they are grouped together, the five factors are not predictive for the treatment benefit as regards clinical/biochemical progression-free survival.

Because the trial design was based on the assumption that pathological risk factors were likely more important, we then assessed the homogeneity of the treatment effect for each of the three pathological risk factors (SM status, ECE status, SV status). This analysis is more robust than any other, because the factors were stratified for at the time of randomization, so that there ought to be a fair balance between the treatment arms within the subgroups.

The analyses showed borderline statistically significant heterogeneity in relation to margin status (P=0.0699). The hazard ratio was 0.64 (95% CI: 0.50-0.91, P=0.0128) in the SM- subgroup and 0.43 (95% CI: 0.33-0.55, P<0.0001) in the SM+ subgroup. In respect to the fact that tests for interaction often lack statistical power, we consider that these results suggest that the SM- patients might benefit less from immediate adjuvant irradiation than the SM+ patients. Nevertheless, the treatment benefit was statistically significant even in that subgroup.

Finally, we assessed the prognostic and predictive value of pathological stage, defined according to the TNM by the combination of the three pathological factors; positive margins without extra capsular extension (SM+, ECE- i.e. pT_2 with SM+), extra capsular extension only (SM-, ECE+ i.e. pT_{3a} with SM-), both factors (SM+, ECE+ i.e. pT_{3a} with SM+) or with seminal vesicle invasion, irrespective of margin status (SV+ i.e. pT_{3b}).

The prognostic factor analysis in the control group indicated that the pT_{3b} subgroup fares much worse than the other three subgroups (P=0.0014) but that there appeared to be little difference between pT_2 and the pT_{3a} subgroups. The absence of difference between these subgroups may however be due to lack of statistical power in relation to the small event numbers.

Again there was no statistically significant heterogeneity in treatment effect across patient groups defined by this classification (Logrank P>0.1).

8.10.4 Conclusions

We concluded that the benefit of immediate irradiation was substantial in all patient subsets, whether formed on trial-based prognostic factors or on pre-defined pathological classification. Statistics supported only the fact that the subset of patients with negative margins may benefit to a somewhat lesser extent than the other subgroups.

A similar analysis ought to be carried out for the endpoint of clinical disease free survival, but longer follow-up is needed to enable subgroup analyses. One could also argue that the prognostic factor model we used was not optimal, as Gleason scores were not available for all patients in this study. The surgical specimens from only a subset of the institutions that participated in this study were reviewed by a central pathologist (Dr. Th. Van der Kwast at Erasmus University) who assessed the Gleason score amongst other pathological features. Further research will be performed to investigate the prognostic importance of those pathological features in the subgroup of patients whose specimens were reviewed by the central pathologists.

8.11 Is baseline quality of life useful for predicting overall survival in metastatic hormone refractory prostate cancer?

8.11.1 Patient population

In chapter 6, we assessed a group of 400 patients with metastatic hormone refractory (HRPC) prostate cancer. These patients were treated within three randomized clinical trials that were designed in the first half of the 1990's. We shall first observe that to be eligible to the trials, the presence of objective evidence of progressive metastatic disease was mandatory. Thus the group of patients that we consider has a worse prognosis than that of patients who are HRPC according to current definitions[44]. Indeed, nowadays patients with biochemical failure despite castrate testosterone levels as the only sign of relapse after hormone manipulation (without clinical evidence of metastases) are also regarded as hormone refractory[4]**.** This change in definition subsequent to the introduction of PSA into clinical practice prolonged the median survival in this disease state. This tendency is also reflected in the change in therapeutic objectives of current clinical trials testing new chemotherapeutic agents in this patient group.

Thus, we recognize that the patient group in our study is not representative of the HRPC population as it is defined today but represents a subgroup that is more advanced state in the hormone refractory cascade. This subgroup itself is however regarded as a heterogeneous group of diseases[3,45] The understanding of this heterogeneity is critical to the evaluation of biomarkers and the design of clinical trials in this disease state.

8.11.2 Objectives

In consequence of the very advanced disease status of the patients, the clinical objective of the three trials considered in our analysis was primarily symptom palliation and the anti-cancer therapeutic objective was limited. Despite this survival was the primary endpoint in these studies but health related quality of life was assessed as an important secondary endpoint. In trial 30903, the patients were treated with either the anti-androgen Flutamide or with Prednisone. In trial 30921, the patients with symptomatic bone metastases were treated with Strontium-89 Chloride or with local palliative irradiation. Trial 30944 was a phase II randomized trial in which patients were randomized between Estramustine Phosphate chemotherapy with or without Vinblastine. All trials were negative as regards overall survival but trial 30903 demonstrated some advantage of Prednisone in terms of quality of life parameters[46].

Having recognised the importance of health related quality of life (HRQOL) as an endpoint in this disease, we questioned whether the patient's quality of life at entry on study could

inform about the patient's long term survival. We developed prognostic factor models that included both the HRQOL parameters assessed at baseline by means of the EORTC QLQ-C30 and previously identified clinical and biochemical parameters.

8.11.3 Findings and validation of results

Our final prognostic factor model retained two quality of life scales (appetite loss and insomnia) in addition to performance status, alkaline phosphatase level and haemoglobin level. The patient's age was not retained in the final model but was borderline statistically significant.
Because many of the factors included in the initial model were highly correlated, we validated the model using Bootstrap re-sampling technique[47]**.** This method assesses model stability by repeating the model development process in a large number of randomly selected replicates of the original dataset. This procedure provides an estimation of the frequency with which a given variable or group of variables are retained in the final model. The procedure confirmed that our final model was the most stable of all potential models. Thus our analysis demonstrates that for groups of patients, those with greater appetite loss and more sleeping problems fare worse in terms of survival, all other clinical and biochemical conditions being equal.

We nevertheless challenged our final model against other models that included only the biochemical and clinical parameters, in order to assess whether the HRQOL factors that were retained in our model would improve the prediction of the survival for *an individual patient*. For this purpose, we calculated a measure of predictive ability of the model (the C-index[47]) that is a bias-corrected estimate of the area under the Receiver-Operating-Curve. This further analysis showed that despite that the factors were predictive at the population level, they added little to the ability of the models to predict the outcome of specific patients.

8.11.4 Conclusion

This study nicely distinguishes the value of prognostic factors for assessing the average outcome of groups of patients from the value of prognostic factors for predicting the outcome at the individual patient level. Factors may be relevant at the population level, without necessarily improving the precision of the predictions at the individual level.

For the purpose of designing clinical trials, the population level is most relevant and one could argue therefore that HRQOL parameters ought to be used for the stratification of metastatic HRPC patients into clinical trials. Doing this would certainly improve the compliance to HRQOL assessments in the studies. However, given the extra burden that HRQOL assessments require and the little extra precision they bring to these predictions, we recommend that prognostic models based on clinical and biochemical factors should be used to estimate the patient's prognosis, whenever all the factors are available from routine

practice. We do support however the use of HRQOL measurements as subsidiary endpoints in clinical trials in this disease.

The biochemical and clinical variables that were measured in the trials that we used in our analysis were very simple variables that are very commonly measured in day-to-day clinical practice. More sophisticated methods have recently become available such as micro-array data analysis, which rapidly increase our understanding of the biology of this disease. These new methods may hopefully identify new, potentially much more powerful, prognostic markers.

8.12 Aspects of the analysis of phase III clinical trials: why trials of maximal androgen blockade in metastatic prostate cancer rarely demonstrate statistically significant differences?

In chapter 7, we revisited the meta-analysis of trials of Maximal Androgen Blockade (MAB) versus castration in metastatic prostate cancer and three of the major trials in that meta-analysis. The body of evidence regarding the relative efficiency of MAB over castration shows that the magnitude of the benefit, if any, is of little clinical relevance (the meta-analysis showed a survival benefit of roughly +4% only[25]). The clinical trials indicated also that MAB may be associated with more side effects and impaired quality of life[48] and increased costs[49,50]**.** Nevertheless, by revisiting these studies we could illustrate a number of the pitfalls that might affect the statistical analysis of phase III clinical trials in cancer and that are not necessarily limited to MAB treatments.

In view of the actual treatment benefit estimated from the meta-analysis, we could illustrate the fact that the target difference had been largely overestimated at the time of designing the trials. In consequence, all trials were by large underpowered to detect treatment benefits of a realistic magnitude.

8.12.1 Interim looks and early publication

Using EORTC trial 30853 that was analyzed a number of times without correction for the interim looks we illustrated how the trial results change dramatically with increasing number of events and statistical power. We showed how the release of immature data (i.e.: when the number of events is less than that specified in the protocol) can be misleading with either significant results becoming non significant or the opposite, as in trial 30853. This illustrates the need to carefully account for any interim looks preferably at the design stage, but other wise at the analysis stage, by using appropriate measures to control the overall type I (false positive) error rate.

8.12.2 The proportionality assumption

The trials of MAB helped us to illustrate a problem that is not uncommon in cancer trials that assess long term survival endpoints: the violation of an important hypothesis underlying the statistical design of time-to-event trials that assume the ratio of the hazards of death on both treatment arms is constant over time. This assumption was violated in the trials of MAB because the treatment effect appeared only after several years. In prostate cancer trials, the proportionality assumption is also often violated in trials of immediate versus deferred therapeutic approach in which a large treatment difference occurs early on but tends to vanish at longer follow-up times (lemon shape). This situation may also occur in studies that include heterogeneous patient subgroups, some of which respond to the treatment and some of which don't. In such trials, the resulting overall survival curves are from a mixture of two or more survival distributions of differing shape, and do not generally satisfy the proportionality condition.

Whenever non proportionality of the effects is expected, this ought to be taken into account at the time of designing the study. Indeed, in such circumstances, overall measures of treatment effect such as the hazard ratio may not be relevant and traditional analysis methods such as the Logrank test or the Cox proportional hazard regression model will not be powerful or appropriate. Specific statistical methods for planning and analyzing non proportional hazards exist and ought to be used instead. The shape of the treatment effect should also be taken into account when designing early stopping rules. Indeed, in a situation where lemon-shape treatment effects are expected, one ought not to implement stopping rules that would stop the trial early on the basis of evidence of a large difference early on since these large differences are not expected to be sustained at later follow-up times.

8.12.3 Impact of inappropriate treatment delivery

In our study, we also performed a sensitivity analysis of the meta-analysis data that excluded all the trials where disease-flare due to the absence of anti-androgen coverage at the beginning of medical castration treatment might have reduced the expected MAB-treatment benefit. Our analysis showed that even in absence of such a flare-effect, there would have been no clinically relevant benefit of MAB over castration alone in terms of overall survival.

8.13 Conclusions

From this research we concluded to the following

- Irrespective of the definition chosen, PSA progression is not very specific for predicting survival in metastatic prostate cancer patients under endocrine treatment. The sensitivity of the marker can be increased by relaxing the criterion of PSA progression by removing the requirement of the confirmation of the PSA increase and by utilizing a low threshold of PSA increase

- The evidence that PSA progression is a surrogate endpoint for overall survival in endocrinally treated metastatic prostate cancer is moderate. Our results showed a correlation of 0.66 between treatment effects onto PSA progression and onto overall survival. Non null treatment effects on survival would potentially be identifiable only in very large new trials showing a very large effect on the PSA endpoint (e.g. HR around 0.50 with SE=0.10). In such circumstances however, we recommend that the sample size of the trial should not be based on power considerations on the basis of the PSA endpoint itself but requirements of precision of the estimation of the treatment effect onto the PSA endpoint. We would not recommend stopping the study follow-up when statistically significant PSA results become available but recommend that the follow-up should be continued to later document long term safety of the treatments and their impact on survival. With the proviso that longer term analysis of the trial be planned to confirm the findings, time-to-PSA progression, could potentially serve as a basis for accelerated drug approval, together with other trial data documenting safety and other measures of patient benefit such as Quality of Life. We also suggest that a PSA endpoint could be used to stop the trial early on for futility in case no apparent effect on the PSA endpoint is shown and if there is strong evidence that the treatment effect is at least in part expressed through PSA changes.

- From our review of the evidence from other validation studies of PSA as surrogate for long term clinical endpoints in other disease states and in trials using other types of treatments we concluded that proven surrogacy for a given treatment is likely not to be extrapolated to trials assessing new treatments that seriously differ in their mode of action from the treatments that were used for the demonstration of surrogacy. Extrapolation between disease stages when the trial designs are similar might be possible. We recommend a thorough understanding of the treatment biological mechanisms of action and of the marker is mandatory before defining a marker-based endpoint and testing it in clinical trials.

- Validation of PSA response as a surrogate for survival in the recent trials of Docetaxel in hormone refractory prostate cancer ought to be attempted. Similarly, there may be enough data available to attempt a validation in studies of adjuvant hormonal treatment after irradiation, but the definition of the PSA endpoint in these trials will be

problematic because a same definition may not apply to the treatment arms with and without hormonal treatment.

- We also exemplified the difficulties related to the use of the Prentice criteria for the validation of surrogate endpoints. We recommend that reports demonstrating that all four conditions are fulfilled should not be regarded as definitive proof of surrogacy.

- We illustrated in two studies, the appropriate methodology to assess predictive factors of treatment benefit by retrospective analysis of existing clinical data. We recommend that attention be paid to the confounding effect between prognostic factors and predictive factors when undertaking such evaluations, because factors that are predictive of the treatment effect will influence the prognostic analysis whenever pooled treatment arms are used in the analysis. We also recommend that treatment effects in subgroups of patients be considered only if there is evidence of heterogeneity of results from a single statistical test for interaction, so as to avoid multiple testing problems that increase the chance of finding spurious statistical findings.

- We recommend that important known prognostic factors should be used for stratifying patients into clinical trials and ought to be used as well for model adjustment at the time of analysis. Prospective studies should be undertaken to confirm the true predictive value of putative predictive factors or classification. We recommend the designs as suggested by Sargent, Conley, Allegra and Collette[20] for this purpose.

- We studied baseline health related quality of life as prognostic factor for overall survival in a population of hormone refractory prostate cancer patients. We illustrated the relevance of the health related quality of life parameters for the estimation of the average prognostic of groups of patients and the lack of relevance of the same factors for predictions at the individual patient level. We recommend that Bootstrap validation be undertaken for every prognostic factor study and that the predictive ability of the models be described in the reports. We concluded that health related quality of life parameters should not be utilized for stratifying patients into clinical trials. We have hope that new prognostic markers of disease status and outcome will become available in the future, thanks to the emergence of sophisticated tissue analyses such as micro-arrays.

- We reviewed a number of pitfalls in the design and analysis of phase III clinical trials recommend great care in defining the target treatment difference to test and that the trial sample size ought to be large enough to provide power to detect the minimum clinically relevant difference. At the planning stage as well as at the time of analysis, the shape of the expected treatment difference over time should be investigated so that appropriate statistical methods be planned and used whenever the proportional hazard assumption is violated. Finally, trial results ought not to be reported prematurely and any interim look at the data ought to be planned and adjusted for in the analysis by correction of the significance level, so as to protect the overall risk of false positive and negative results.

8.14 Practical recommendations for designing future trials using a PSA intermediate endpoint

Although PSA endpoints do not completely fulfill the statistical requirements for being surrogates of long term clinical outcomes in hormonally treated advanced prostate cancer, we would suggest that the following approach to the design of new trials.

For a new trial where early registration on the basis of a PSA endpoint would be envisaged, we suggest that the trial be designed with time to PSA progression or biochemical progression-free survival as the intermediate endpoint and survival as the definitive endpoint. The trial sample size should be determined to demonstrate a very large effect on the PSA endpoint (HR of the order of 0.50) with great precision precision (SE<0.10). The sample size calculation should be based on the required precision of the estimation of the treatment effect onto the PSA endpoint rather than on statistical power considerations. An interim analysis plan should be set up with interim looks at the PSA endpoint as well as to the safety data and one longer term analysis on the survival endpoint. At the interim looks, the trial results on the PSA endpoint would be used to infer prediction for the survival treatment effect using the results from former meta-analytic validation exercises. Whenever the prediction interval for the survival hazard ratio would exclude the null effect, the trial results could be submitted for early registration on the basis of the PSA results. The trial follow-up ought however to be continued in order to document survival treatment effects at a subsequent analysis, as well as long term safety results.

In addition or separately from the above, an early stopping rule for futility could be set up, with the possibility to stop the study in case an insufficient or detrimental treamtent effect would be seen on the PSA endpoint. The assumption underlying this design is that a significant treatment effect on survival endpoint would be very unlikely if no beneficial effect was seen on the PSA endpoint. The information on the correlation between the treatment effects on the PSA endpoint and on survival that is necessary to the set up of such stopping rules is readily available from the value of R^2_{trial} obtained from the meta-analytic validation.

Thus although perfect surrogacy of PSA as a substitute for survival endpoints remains unproven, PSA could be useful to shortening the duration of phase III trials using the above approach to interim looks and decisions to stop or release the trial results early.

8.15 References

1. Pocock SJ 'Clinical trials: a practical approach', John Wiley and Sons, Chichester, 1983
2. Green SB. Patient heterogeneity and the need for randomized clinical trials. Control Clin Trials 1982; 3: 189-198.

3. Scher HI, Masumdar M. Kelly WK. Clinical trials in relapsed prostate cancer: Defining the target. J Natl Cancer Inst 1996; 88(22): 1622-1634.
4. Bubley GJ, Carducci M, Dahut W, Dawson N, Daliani D, Eisenberger M, Figg WD, Freidlin B, Halabi S, Hudes G, Hussain M, Kaplan R, Myers C, Oh W, Petrylak DP, Reed E, Roth B, Sartor O, Scher H, Simons J, Sinibaldi V, Small EJ, Smith MR, Trump DL, Wilding G: Eligibility and response guidelines for phase II clinical trials in androgen-independent prostate cancer: recommendations from the Prostate-Specific Antigen Working Group. J Clin Oncol 1999;17:3461-3467.
5. Dixon SC, Knopf KB, Figg WD. The control of prostate-specific antigen expression and gene regulation by pharmacological agents. Pharmacol Rev 2001;53: 73-91.
6. Johansson JE, Andren O. Andersson SO, Dickman PW, Holmberg L, Magnuson A, Adami HO. Natural history of early localized prostate cancer. JAMA 2004; 291(22): 2713-2719.
7. Newschaffer CJ, Otani K, McDonald MK, Penberth LT. Causes of death in elderly prostate cancer patients and in a comparison nonprostate cancer cohort. J Natl Cancer Inst 2000;92: 61-621.
8. Scher HI, Heller G. Clinical states in prostate cancer: toward a dynamic model of disease progression. Urology 2000; 55: 323-327.
9. Schroeder FH, Whelan P, de Reijke ThM, Kurth KH, Pavone-Macaluso M, Mattelaer J, van Velthoven RF, Debois M, Collette L. Metastatic prostate cancer treated by flutamide versus cyproterone acetate: Final analysis of the European Organisation for Research and Treatmetn of Cancer (EORTC) protocol 30892. Eur Urol 45: 457-464, 2004
10. Dijkman GA, Janknegt RA, de Reijke ThM, Debruyne FM: Long-term efficacy and safety of nilutamide plus castration in advanced prostate cancer, and the significance of early prostate specific antigen normalization. International Anandron Study Group. J Urol 1997;158:160-163.
11. Oosterlinck W, Mattelaer J, Casselman J, Van Velthoven R, Derde MP, Kaufman L.: PSA evolution: a prognostic factor during treatment of advanced prostatic carcinoma with total androgen blockade. Data from a Belgian multicentric study of 546 patients. Acta Urol Belg 1997;65:63-71.
12. Reynard JM, Peters TJ, Gillatt D. Prostate-specific antigen and prognosis in patients with metastatic prostate cancer - a multivariate analysis of prostate cancer mortality. Brit J Urol 1995;75: 507-515.
13. Cooper EH, Armitage TG, Robinson MRG, Newling DW, Richards BR, Smith PH, Denis L, Sylvester R: Prostatic specific antigen and the prediction of prognosis in metastatic prostatic cancer. Cancer 1990;66:1025-1028.
14. Matzkin H, Soloway MS, Schellhammer PF , Chodak G, Smith JA, Caplan R, Kennealey GT: Prognostic factors in stage D2 prostate cancer treated with a pure non-steroidal anti-androgen. Cancer 1993;72:1286-1290,.
15. Miller JI, Ahmann FR, Drach GW, Emerson SS, Bottaccini MR: Serum PSA levels predict duration of remission and survival post hormone therapy of metastatic prostate cancer. J Urol 1991; 145: 384A.
16. Newling DW, Denis L, Vermeylen K. Orchiectomy versus goserelin and flutamide in the treatment of newly diagnosed metastatic prostate cancer. Analysis of the criteria of evaluation used in the European Organization for Research and Treatment of Cancer - Genitourinary Group Study 30853. Cancer 1993;72(suppl 2): 3793-3798.
17. Strohmaier WL, Keller T, Bichler KH. Follow-up in prostate cancer patients: which parameters are necessary? Eur Urol 1999;35: 21-25.

18. Simon R and Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer 69: 979-85, 1994
19. Hayes DF, Trock B, Harris AL. Assessing the clinical impact of prognostic factors: when is "statistically significant" clinically useful? Breast Cancer Res and Treatment 52: 305-319, 1998
20. D.J. Sargent, B.A. Conley, C. Allegra, L. Collette. Clinical trial designs for predictive marker validation in cancer treatment trials. J Clin Oncol (23):2020-2027, 2005
21. Baker SG, Kramer BS. A perfect correlate does not a surrogate make. BMC Med Res Methodol 2003; 3(1):16
22. Roessner M, de Wit R, Tannock IF, Yateman N, Yao S, Yver A, Eisenberger MA on behalf of the Tax327 investigators. Prostate-specific antigen (PSA) response as surrogate endpoint for overall survival (OS): Analysis of the TAX 327 Study comparing decetaxel plus prednisone to mitoxantrone plus prednisone in advanced prostate cancer. Proc Am Soc Clin Oncol 2005, J Clin Oncol 2005; 23:391s, abstract 4554.
23. Prentice RL. Surrogate endpoints in clinical trials: definitions and operational criteria. Stat Med. 1989;8:431-440.
24. Buyse M, Molenberghs G, Burzykowski T, Renard D and Geys H. Statistical validation of surrogate endpoints: problems and proposals. Drug Inf J. 2000; 34: 447-454.
25. Prostate Cancer Trialist's Collaborative Group. Maximum androgen blockade in advanced prostate cancer: an overvie of the randomized trials. Lancet 2000; 355(9214): 1491-98.
26. Kelloff GJ, Coffey DS, Chabner BA, Dicker AP, Guyton KZ, Nisen PD, Soule HR and D'Amico AV. Prostate-specific antigen doubling time as a surrogate marker for evaluation of oncologic drugs to treat prostate cancer. Clin Cancer Res. 2004;10(11):3927-3
27. Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ 1995; 311: 485.
28. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Stat Med. 1992; 11:167-178.
29. Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. Control Clin Trials. 2002; 23: 607-625.
30. Buyse M, Molenberghs G. Criteria for the validation of surrogate end-points in randomized experiments. Biometrics 1998;54: 1014-29.
31. Buyse M, Molenberghs G, Burzykowski T, Renard D and Geys H. The validation of surrogate endpoints in meta-analyses of randomized experiments. Biostatistics. 2000;1: 49-68.
32. Burzykowski T, Molenberghs G, Buyse M, Geys H, Renard D. Validation of surrogate endpoints in multiple randomized clinical trials with failure-time endpoints. J Roy Stat Soci Series C (Appl Stat). 2001;50: 405-22.
33. Debruyne FJM, Murray R, Fradet Y, Johansson JE, Tyrrell C, Boccardo F, Denis L, Marberger JM, Brune D, Rassweiler J, Vangeneugden T, Bruynseels J, Janssens M, de Porre P for the Liarozole Study Group. Liarozole – a novel treatment approach for advanced prostate cancer: results of a large randomized trial versus cyproterone acetate. Urology 1998;52:72-81.
34. Sargent DJ, Wieand S, Benedetti J, Labianca R, Haller DG, Shepherd LE, Seitz JF, Francini G, De Gramont A, Goldberg RM. Disease-free survival (DFS) vs. overall survival (OS) as a primary endpoint for adjuvant colon cancer studies: Individual patient data from 12,915 patients on 15 randomized trials. J Clinl Oncol, 2004; 22(14S): 342. (abs. 3502)
35. Royston P, Parmar M and Qian W. Novel designs for multi-arm clinical trials with survival outcomes with an application in ovarian cancer. Stat Med 2003; 22: 2239 – 56.

36. Newling D, Carroll K, Morris T. Is prostate-specific antigen progression a surrogate for objective clinical progression in early prostate cancer? Proc Am Soc Clin Oncol 2004, J Clin Oncol 2004; 22 (14S): 4652.

37. Buyse M, Vangeneugden T, Bijnens L, Renard, D., Burzykowski, T., Geys, H., and Molenberghs, G. Validation of Biomarkers ad Surrogates for Clinical Endpoints. In Biomarkers in Clinical Drug Development (Bloom JC, ed) New-York: Springer-Verlag, 2003. p. 149-168.

38. D'Amico A, Moul JW, Carroll PR, Sun L. Lubeck D, Chen MH. Surrogate endpoint for prostate cancer specific mortality after radical prostatectomy or radiation therapy. J Natl Cancer Inst 2003; 95(18): 1376-1382.

39. Sandler HM, Pajak TF, Hanks GE, Porter AT, DeSilvio M, Shipley WU. Can biochemical failure (ASTRO definition) be used as a surrogate endpoint for prostate cancer survival in phase III localized prostate cancer clinical trials? Analysis of RTOG protocol 92–02 (abstract 1529). J Clin Oncol 2003;22:381.

40. D'Amico AV, Moul J, Carroll PR, Sun L, Lubeck D, Chen MH. Surrogate endpoint for prostate cancer specific mortality in patients with nonmetastatic hormone refractory disease. J Urol 2005; 173: 1572-6.

41. Crawford ED, Pauler DK, Tangen CM, Hussain MHA, Small EJ, Taplin EJ, Burch PA, Greene GF, Lara P, Petrylak DP. Three-month change in PSA as a surrogate endpoint for mortality in advanced hormone-refractory prostate cancer (HRPC): Data from Southwest Oncology Group Study S9916. Proc Am Soc Clin Oncol 2004; 22: 14S, abstract 4505.

42. Tannock IF, de Wit R, Berry R, Horti J, Pluzanska A, Chi KN, Oudard S, Théodore C, James ND, Turesson I, Rosenthal MA, Eisenberger MA. Docetaxel plus prednisone or mitoxantrone plus prednisone for advanced prostate cancer. N Engl J Med 2004; 351 (15): 1502-1512.

43. Anderson GL, LeBlanc M, Liu PY, Crowley J. On use of covariates in randomization and analysis of clinical trials. In Handbook of Statistics for Clinical Trial. J. Crowley, 2nd ed, 2005.

44. Oefelein MG, Agarwal PK, Resnick MI. Survival of patients with hormone refractory prostate cancer in the prostate specific antigen era. J Urol 2004; 171:1525-28.

45. Shah RB, Mehra R, Chinnaiyan AM, Shen R, Ghosh D, Zhou M, Macvicar GR, Varambally S, Harwood J, Bismar TA, Kim R, Rubin MA, Pienta KJ. Androgen-independent prostate cancer is a heterogeneous group of diseases: lessons from a rapid autopsy program. Cancer Res. 2004 Dec 15;64(24):9209-16.

46. Fossa SD, Slee PHT, Brausi M, Horenblas S,. Hall RR,. Hetherington JW, Aaronson N,. de Prijck L and Collette L. Flutamide versus Prednisone in patients with prostate cancer symptomatically progressing after androgen ablative therapy: a phase III study of the European Organisation for Research and Treatment of Cancer Genitourinary Group. J Clin Oncol 19(1): 62-71, 2001.

47. Harrell FE, Lee KL, Mark DB. Multivariate prognostic models: issues in developing models,evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 15: 361-387, 1996.

48. Moinpour CM, Savage MJ, Troxel A, Lovato LC, Eisenberger M, Veith RW, Higgins B, Skeel R, Yee M, Blumenstein B, Crawford ED, Meyskens FL Jr. Quality of life in advanced prostate cancer: results of a randomized therapeutic trial. JNCI 1998;90:1537-44.

49. Nicol DL, Heathcote PS, Kateley GD, Lloyd S. Advanced prostate cancer. The role of high priced hormone therapy. Med J Aust 1993;159:16-9.

50. Hillner BE, Mc Leod DG, Crawford ED, Bennet CL. Estimating the cost effectiveness of total androgen blockade with flutamide in M1 prostate cancer. Urology 1995; 45:633-40.

Summary

Prostate cancer is a significant health problem in most industrialized Western countries where it is the most commonly diagnosed cancer in men after middle age. It is also a leading cause of death from cancer among men diagnosed with cancer in the US and in Europe. Prostate cancer is a relatively slow growing disease and patients diagnosed early often benefit from long survival time.

Phase III clinical trials to assess the clinical benefit of new treatment options often require large patient numbers and long follow-up. In diseases with a long natural history, such as prostate cancer, the final result of comparative trials with survival endpoints is often not known before five to ten years after the study start. With increasing knowledge of the genetic background of cancer and cancer progression and of the mechanisms of signal transduction, a large number of new compounds are under development. In addition, there is increasing public pressure for promising new drugs to receive marketing approval as rapidly as possible, in particular for life threatening diseases such as cancer. For these reasons, there is an urgent need to find ways of speeding up the new drug development process by optimizing the design of (especially phase III) clinical trials.

Additionally, prostate cancer is a rather heterogeneous disease and many prognostic factors (such as Gleason score, number of positive biopsies ...) severely affect the natural course of the disease. Some factors that relate to the patient's prognosis (e.g. for instance pathologic stage after prostatectomy) may also act as predictive factors for the benefit of new treatments, thus inducing extra heterogeneity into clinical trials.

In this thesis, we study various means by which the design and reporting of phase III clinical trials in advanced prostate cancer could be made more efficient. We argue that in order to improve the efficiency of clinical trials in prostate cancer, a thorough understanding of *prognostic factors* of outcome is needed, as well as an exploration of potential *predictive factors* that might affect treatment benefit and cause heterogeneity of results. In addition, the development of *surrogate endpoints* of the long term clinical outcome is needed in order to speed up the drug development process by reducing the observation time required to assess the final endpoint.

We utilize existing clinical trial data from past studies, most of which involve endocrine treatments against prostate cancer, to challenge a number of pathways of improvement of the efficiency of phase III clinical trial designs in prostate cancer:

First, we observed that the long duration of phase III clinical trials mostly results from the use of a long term clinical endpoint (clinical progression, survival) especially when studying a slowly developing disease like prostate cancer. Therefore, to replace that long term endpoint by another that could be measured earlier, more conveniently or more frequently and that would adequately reflect the benefit of new treatments on the true endpoint would seem attractive. Such replacement endpoints are called "surrogate" endpoints. In this study,

we questioned whether changes in the Prostate Specific Antigen (PSA) could be used as surrogate for long term clinical outcomes such as progression-free or overall survival.

The notion of surrogacy relates directly to the ability of an intermediate endpoint (here PSA) to figure, early in time, the eventual impact of the treatment on the longer term clinical endpoint (here progression-free or overall survival). This notion is often confounded with that of the intermediate outcome (here PSA) being a prognostic factor for the long term outcome. In the **Introduction**, we exemplify the difference between these two notions and introduce the methodology that can be used to assess the true validity of a putative surrogate endpoint.

The most known method of surrogate endpoint validation is by means of the Prentice Criteria that are rooted in the paradigm of statistical hypothesis testing. We discuss that this method relies on serious non testable assumptions. We also discuss that the Prentice criteria are actually not sufficient nor necessary to demonstrate surrogacy when the true endpoint is the time to an event. We also describe and discuss a new validation method, the "meta-analytic" validation method, which was developed by the statisticians from Limburgs Universitair Centrum, Diepenbeek, Belgium.

In C**hapter 2**, we use data from an EORTC trial in metastatic prostate cancer (EORTC 30892) to determine the most sensitive definition of time to PSA progression for patients treated endocrinologically. We conclude that irrespective of the definition chosen, PSA progression is not very specific for predicting survival in metastatic prostate cancer patients under endocrine treatment but that the sensitivity of the marker can be increased by relaxing the criterion of PSA progression by removing the requirement of the confirmation of the PSA increase and by utilizing a low threshold of PSA increase such as 20% or 50%.

In **Chapter 3**, we use data from a large database of metastatic prostate cancer patients provided by AstraZeneca Pharmaceuticals and the meta-analytic validation method to assess the value of the time to PSA progression defined in chapter 2 and of other variants of this endpoint as surrogate endpoint for overall survival. We conclude that the evidence that PSA progression is a surrogate endpoint for overall survival in endocrinologically treated metastatic prostate cancer is moderate: our results showed a correlation of 0.66 between treatment effects onto PSA progression and onto overall survival. We however suggested ways to use PSA to shorten the duration of clinical trials by using it at the stage of interim analysis to decide either on early stop of the trial for futility in case no effect is seen on the PSA or to decide on release of PSA results or accelerated registration of the trial results with the proviso that long term survival data be made available at a later stage.

The further chapters 4 to 6 of this thesis are devoted to a second aspect of phase III clinical trials: the heterogeneity in the patient population and in the study results.

Heterogeneity of the patient groups entered in clinical trials and of their responsiveness to the tested treatment modalities can result in increased variability in the estimation treatment effect. In addition, when comparing two or more treatment modalities in the frame of a

randomized clinical trial, lack of balance of important prognostic factors between treatment groups may result in severe bias since difference in patient prognosis between groups will be confounded with the treatment effect and may either artificially dilute or inflate it, thus leading to inflation of the risk of a false negative (type II) and of a false positive (type I) trial conclusion, respectively. Thus the knowledge of prognostic factors is extremely important to the choice and implementation of relevant stratification factors in future randomized trials.

Besides heterogeneity of the patient population, the treatment results themselves may also be heterogeneous across groups of patients characterized by *predictive* factors. Predictive factors are factors that modulate the magnitude of the expected treatment benefit. In a clinical trial, a predictive factor is characterized as causing a statistically significant interaction with the treatment effect. The identification of predictive factors of treatment benefit is important because they may limit the applicability of the overall trial results to subgroups of patients. For planning future studies, the knowledge of predictive factors may be helpful to defining appropriate selection criteria. By selecting the group of patients that is most sensitive to the experimental treatment, one increases the likelihood of finding significant treatment benefits whilst sparing unnecessary treatment to patients that are less or not responsive to that treatment.

In **Chapter 4**, we revisit an EORTC trial (EORTC 22863) in locally advanced prostate cancer that established the use of adjuvant hormonal treatment after irradiation as a standard of care in this patient group. We identified prognostic factors and assessed that these factors were not predictive for the treatment benefit, so that the overall trial result can apply to the whole population selected to the study.

In **Chapter 5**, we present the results of EORTC trial 22911 that compared immediate postoperative irradiation to a wait-and-see policy after radical prostatectomy in a high risk population of patients with localized disease and poor pathological risk factors. We demonstrated how to approach the issue of assessing the presence of predictive factors in a trial where strong prognostic factors of the natural disease history are present. We could identify that the patients with extracapsular extension as sole risk factor seemed to benefit less from the treatment than the patients' subgroups with positive margins or invasion of the seminal vesicles. The results however demonstrated a statistically significant benefit from immediate postoperative irradiation in all subgroups. The results of this study were also revisited in the discussion to exemplify the limitations of the Prentice Criteria as a surrogate endpoint validation method.

In **Chapter 6**, we turn to the (bone) metastatic hormone refractory patient population. We pooled the data from three EORTC studies that were carried out in the 1990's and that assessed health related quality of life (HRQOL) in addition to clinical endpoints. We use the baseline HRQOL data together with previously identified clinical and biochemical parameters to identify if baseline HRQOL is an independent prognostic factors for overall survival. We further compared the precision of the predictions from the resulting prognostic model to that of models based on clinical and biochemical factors and concluded that

although HRQOL parameters were significant in the prognostic model, they do not seem to improve the prediction of the expected survival of an individual patient.

In **Chapter 7**, we consider the trials that compared maximal androgen blockade to castration in metastatic prostate cancer to identify further pitfalls in the design and analysis of these trials that may have limited their ability to demonstrate a survival benefit. From this review we recommend great care in defining the size of the target treatment effect to be detected. This is to guarantee that the trial sample size be large enough to have power to detect the minimum clinically relevant difference. At the planning stage as well as at the time of analysis, the shape of the expected treatment difference over time should be investigated so that appropriate statistical methods be planned and used whenever the proportional hazard assumption is violated. Finally, this review exemplifies that trial results ought not to be reported prematurely and any interim look at the data ought to be planned and adjusted for in the analysis by correction of the significance level, so as to protect the overall risk of false positive and negative results and to avoid publishing misleading results.

In **Chapter 8**, the general discussion, we critically review the results presented in chapters 2-7 in terms of statistical requirements and of clinical relevance and in the light of evidence from other existing publications in this area. In particular, we discuss the methodology of surrogate endpoint validation and its practical limitations. We also address the issue of the transportability of surrogacy of an endpoint across disease stages and across treatment modalities. We conclude with some recommendations for the use of PSA as endpoint and with suggestions for the actual design of future trials in prostate cancer using PSA data.

Samenvatting

Prostaatkanker is een belangrijk gezondheidsprobleem in de westerse wereld aangezien het in deze landen de meest gediagnosticeerde kanker is bij mannen boven de 50 jaar. Prostaatkanker is tevens één van de belangrijkste doodsoorzaken onder mannen die gediagnosticeerd worden met kanker in de Verenigde Staten en Europa. Prostaatkanker is een relatief langzaam groeiende tumor en een vroege diagnose resulteert vaak in een lange overlevingstijd.

Fase III klinische onderzoeken waarbij het nut van nieuwe behandelingen wordt onderzocht, vereisen grote aantallen patiënten en een lange opvolging. Bij ziekten met een lang natuurlijk beloop, zoals prostaatkanker, zullen de uitkomsten van vergelijkende studies met overleving als eindpunt pas vijf tot tien jaar na de start van de studie bekend kunnen worden. De nieuwe inzichten met betrekking tot de genetische achtergrond van kanker, progressie van de kanker en de mechanismen van de signaal transductie hebben geleid tot de ontwikkeling van een groot aantal nieuwe medicijnen. Daarnaast is er een toegenomen druk vanuit de bevolking om veelbelovende medicijnen zo snel mogelijk goed te keuren voor algemene toepassing, in het bijzonder bij levensbedreigende ziekten zoals kanker. Daarom is het noodzakelijk nieuwe methoden te vinden waarbij de opzet van klinische studies (voornamelijk de Fase III studies) geoptimaliseerd wordt, teneinde de duur van deze studies te verkorten.

Daarnaast is prostaatkanker een tamelijk heterogene tumor en meerdere prognostische factoren, (zoals de Gleason score, aantal positieve prostaatbiopten, ...) hebben een duidelijke invloed op het natuurlijke beloop van de ziekte. Factoren die betrekking hebben op de prognose voor de patiënt (bijvoorbeeld het pathologische stadium na een radicale prostatectomie) kunnen ook een reden zijn waardoor nieuwe behandelingen wel of niet aanslaan en dit voegt bijkomende heterogeniteit toe aan de klinische studies waarbij deze nieuwe behandelingen worden getest.

In dit proefschrift bestuderen we verschillende methoden om opzet en rapportering van Fase III klinische studies bij gevorderde prostaatkanker efficiënter te maken. Wij argumenteren dat hiertoe een grondig begrip van de *prognostische factoren* met betrekking tot de klinische uitkomst, alsook een onderzoek naar mogelijke *voorspellende factoren* die een invloed hebben op het verschil tussen de behandelingen en daardoor heterogeniteit in de studieresultaten veroorzaken, noodzakelijk is. Tevens is het nodig om *surrogaat eindpunten* voor de lange termijn resultaten te ontwikkelen om het ontwikkelingsproces van nieuwe medicijnen te bespoedigen door inkorting van de observatietijd nodig om het eindpunt van de studie waar te nemen.

Bestaande gegevens van al afgesloten klinische studies werden gebruikt, de meeste studies bestudeerden de hormonale behandelingen bij prostaatkanker, om een aantal mogelijkheden te testen voor een verbetering van de efficiëntie in de opzet van klinische Fase III studies bij prostaatkanker:

Ten eerste werd geconstateerd dat de lange duur van de Fase III klinische studies meestal veroorzaakt werd door een klinisch eindpunt met een lange duur, zoals klinische progressie of overleving, en dit geldt met name bij een langzaam groeiende tumor zoals prostaatkanker. Het lijkt daarom aantrekkelijk om eindpunten met een lange observatie termijn te vervangen door een ander eindpunt, dat eerder, gemakkelijker of frequenter kan gemeten worden en dat daarenboven het voordeel van een nieuwe behandeling met betrekking tot het eigenlijke eindpunt adekwaat weergeeft. Dergelijke eindpunten worden "surrogaat eindpunten" genoemd. In dit proefschrift onderzoeken we of veranderingen in serumspiegels van het Prostaat Specifiek Antigeen (PSA) gebruikt kunnen worden als surrogaat voor lange termijn klinische eindpunten zoals overleving of progressievrije overleving.

Het begrip surrogaat heeft direct betrekking op het vermogen van een vroeg eindpunt (PSA in dit geval) om eerder aan te geven wat de invloed is van de te onderzoeken behandeling op het uiteindelijke klinische eindpunt (in dit geval overleving of progressievrije overleving). Dit begrip wordt vaak verward met de prognostische waarde van de vroege uitkomst (PSA in dit geval) met betrekking tot de uiteindelijke uitkomst. In de **Inleiding** tonen we het verschil tussen deze beide begrippen en bespreken de methodologie welke gebruikt kan worden om de validiteit van mogelijke surrogaat eindpunten te bepalen.

De bekendste methode voor het valideren van surrogaat eindpunten is met behulp van de Prentice criteria, die gebaseerd zijn op de principes van hypothese toetsen (en niet op schatters). We bediscussiëren dat deze methode zich baseert op niet te testen veronderstellingen. Tevens zijn de Prentice criteria in feite niet voldoende, noch noodzakelijk om een surrogaat eindpunt aan te tonen wanneer het uiteindelijke eindpunt een mogelijk gecensureerde variabele is (tijdsduur tot een gebeurtenis). Een nieuwe validatie methode wordt beschreven en bediscussieerd, de 'meta-analytische' validatie methode, welke ontwikkeld werd door statistici van het Limburgs Universitair Centrum, Diepenbeek, België.

In **Hoofdstuk 2** worden gegevens van EORTC studie 30982 in gemetastaseerde prostaatkanker gebruikt om de meest gevoelige definitie vast te stellen voor de tijd tot PSA progressie voor patiënten die hormonaal behandeld werden. We concluderen dat onafhankelijk van de gekozen definitie, PSA progressie niet zeer specifiek is om de overleving te voorspellen voor patiënten met een gemetastaseerde prostaatkanker, die hormonaal behandeld worden. De gevoeligheid van de marker kan echter verhoogd worden indien de voorwaarde van PSA progressie ruimer gesteld wordt door het weglaten van de bevestiging van een PSA toename en door een lage drempelwaarde voor de PSA stijging, zoals 20% of 50%, te kiezen.

In **Hoofdstuk 3** gebruiken we gegevens uit een grote databank van patiënten met een gemetastaseerde prostaatkanker ter beschikking gesteld door AstraZeneca Pharmaceuticals. Hierop testen we de meta-analytische validatie methode om de waarde vast te stellen van de tijd tot PSA progressie zoals gedefinieerd in Hoofdstuk 2 en van andere variaties van dit

eindpunt als surrogaat eindpunt voor overleving. Geconcludeerd wordt dat het bewijs dat PSA progressie een surrogaat eindpunt is voor overleving bij patiënten met een gemetastaseerde prostaatkanker die hormonaal behandeld worden slechts matig is: de resultaten tonen een correlatie van 0.66 tussen effect van de behandeling op de PSA progressie en op de algehele overleving. Er worden echter mogelijkheden aangegeven om PSA te gebruiken om de duur van de klinische studies te bekorten door het te gebruiken ten tijde van een tussentijdse analyse waarbij eventueel het besluit genomen kan worden om enerzijds het onderzoek te staken wegens gebrek aan effectiviteit of anderzijds de behandeling vrij te geven onder bepaalde voorwaarden of versnelde registratie van de studieresultaten onder de voorwaarde dat de lange termijn overlevingsresultaten op een later tijdstip ter beschikking gesteld zullen worden.

De hoofdstukken 4 tot 6 hebben betrekking op een tweede aspect van Fase III klinische studies: de heterogeniteit van de patiënten populatie en van de studieresultaten.

De heterogeniteit van patiënten die in klinische studies worden ingesloten en de reactie van deze patiënten op de te testen behandelingen kunnen resulteren in een toegenomen variabiliteit van de schatting van het effect van de te testen behandeling. Indien er dan tevens sprake is van een onevenwicht van belangrijke prognostische factoren tussen de te vergelijken patiëntengroepen in het kader van een gerandomiseerde studie, kan er ernstige bias ontstaan, omdat verschil in prognose tussen de te vergelijken groepen patiënten en verschil in behandelingsuitkomst verward zullen worden. Het gevolg zal kunstmatige onder- or overschatting van het effect van de behandeling zijn, aldus leidend tot respectievelijk een te hoog risico op een vals negatieve (type II fout) en een vals positieve (type I fout) conclusie van de studie. Zodoende is kennis van de prognostische factoren uitermate belangrijk voor de keuze en implementatie van relevante stratificatie factoren in toekomstige gerandomiseerde studies.

Naast heterogeniteit van de patiëntenpopulatie kunnen de behandelingsresultaten zelf ook heterogeen zijn verspreid over patiëntengroepen gekarakteriseerd door *voorspellende* factoren. Voorspellende factoren beïnvloeden de grootte van het te verwachten behandelresultaat. In een klinische studie wordt een voorspellende factor gekarakteriseerd als een factor welke een statistisch significante interactie heeft met het verschil tussen de gerandomiseerde behandelingen. Het onderkennen van voorspellende factoren voor het behandelresultaat is belangrijk omdat deze de toepasbaarheid van de studie resultaten tot deelgroepen van patiënten kunnen beperken. Bij het opzetten van nieuwe studies kan de kennis van voorspellende factoren nuttig zijn om daarmee goede selectie criteria te definiëren. Indien een selecte groep patiënten, die naar verwachting goed zal reageren op de te onderzoeken behandeling, in de studie wordt ingesloten, wordt de kans op een goed resultaat vergroot, terwijl onnodige behandeling gespaard word voor patiënten die minder of niet zullen reageren op de te onderzoeken behandeling.

In **Hoofdstuk 4** wordt een EORTC studie (EORTC 22863) besproken, die de rol van het gebruik van adjuvante hormonale therapie na bestraling in patiënten met een lokaal uitgebreide prostaatkanker als standaard behandeling aantoonde. We identificeerden

prognostische factoren en stelden vast dat deze factoren niet voorspellend waren voor het verschil tussen de behandelingen, waaruit geconcludeerd kan worden dat de uitkomsten van deze studie gelden voor de gehele studie populatie.

In **Hoofdstuk 5** worden de resultaten van EORTC studie 22911 gepresenteerd, waarin de onmiddellijke postoperatieve bestraling vergeleken wordt met een afwachtend beleid na een radicale prostatectomie bij patiënten met een hoog risico gelokaliseerd prostaatkanker en slechte pathologische kenmerken. In deze studie met sterke prognostische factoren voor het natuurlijk beloop van de ziekte wordt getoond hoe het vraagstuk van de mogelijke aanwezigheid van voorspellende factoren kan worden aangepakt. Aangetoond werd dat patiënten met een extra-capsulaire uitbreiding van de ziekte als enige risico factor (pT2) minder voordeel hadden van de behandeling in vergelijking met patiënten deelgroepen die positieve snijvlakken of invasie van de vesicula seminalis hadden. De resultaten van de studie toonden echter een statistisch significant voordeel van een onmiddellijke post operatieve bestraling voor alle deelgroepen. De resultaten van deze studie werden ook gebruikt in de discussie om de beperkingen van de Prentice criteria als een validatie methode voor een surrogaat eindpunt te tonen.

In **Hoofdstuk 6** worden patiëntengroepen besproken met een (bot) gemetastaseerde prostaatkanker welke hormoon refractair is. De gegevens van drie EORTC studies welke in de negentiger jaren werden uitgevoerd, werden samengenomen. Naast klinische eindpunten onderzochten deze studies ook de ziekte gerelateerde levenskwaliteit. De initiële ziekte gerelateerde levenskwaliteit gegevens werden samen met eerder geïdentificeerde klinische en biochemische parameters gebruikt om aan te tonen of de initiële ziekte gerelateerde levenskwaliteit een onafhankelijke prognostische factor is voor overleving. Tevens werd de nauwkeurigheid van de voorspellingen vanuit het opgestelde prognostische model vergeleken met modellen gebaseerd op klinische en biochemische factoren en geconcludeerd werd dat hoewel ziekte gerelateerde levenskwaliteit parameters in het prognostische model significant zijn, deze parameters de voorspelling voor het overleven van een individuele patiënt niet verbeteren.

In **Hoofdstuk 7** bestuderen we de studies welke de maximale androgeen blokkade vergeleken met castratie bij patiënten met een gemetastaseerde prostaatkanker om valkuilen te identificeren in de opzet en analyse van deze studies, waardoor mogelijkerwijs in deze studies geen overlevingsvoordeel kon worden aangetoond. We bevelen aan zeer voorzichtig te zijn bij het bepalen van de alternatieve hypothese. Hiermee wordt gegarandeerd dat de steekproefgrootte van de studie voldoende onderscheidingsvermogen levert om het minimale klinisch relevante verschil te ontdekken. Ten tijde van de opzet van de studie en ook bij de analyse van de studie resultaten, moet de te verwachten evolutie in de tijd van het relatieve risico (*hazard ratio*) tussen de behandelingen onderzocht worden teneinde de juiste statistische methode te plannen en toe te passen in geval de hypothese van proportionele relatieve risico's (*proportional hazards*) niet houdbaar blijkt. Tot slot illustreert deze bespreking dat de resultaten van klinische studies niet te vroeg gepresenteerd mogen worden en dat iedere tussentijdse analyse van te voren gepland en onder de juiste correctie van de significantiedrempel, teneinde het risico op vals positieve en vals negatieve

uitkomsten op de vooropgestelde waarde te houden en daarmee ook de publicatie van misleidende resultaten te vermijden.

In de algemene discussie, **Hoofdstuk 8,** kijken we kritisch terug op de resultaten welke gepresenteerd werden in de hoofdstukken 2 tot 7 wat betreft statistische voorwaarden en klinische relevantie ook met betrekking tot eerder gepubliceerde gegevens op dit gebied. Met name wordt ingegaan op de methodologie van de validatie van surrogaat eindpunten en de praktische beperkingen daarvan. Ook wordt de problematiek van de bredere toepasbaarheid van surrogaat eindpunten over meerdere ziekte stadia en behandelmethoden besproken. We sluiten af met enige aanbevelingen voor het gebruik van PSA als eindpunt en suggesties voor de concrete opzet van nieuwe klinische studies met gebruik van PSA data.

Résumé

Le cancer de la prostate représente un problème de santé majeur dans la plupart des pays industrialisés occidentaux. C'est le cancer le plus fréquemment diagnostiqué chez les hommes de 50 ans et plus et la cause principale de décès dû au cancer parmi les hommes européens et américains. Le cancer de la prostate se développe relativement lentement de sorte qu'un diagnostique précoce amène le plus souvent une longue survie malgré le cancer.

Les essais cliniques de phase III évaluant le bénéfice de nouvelles options thérapeutiques nécessitent en général un grand nombre de patients et une longue durée de suivi. Pour des maladies dont l'évolution naturelle est lente, tels le cancer de la prostate, les résultats des essais comparatifs qui utilisent la survie comme critère d'évaluation ne sont bien souvent connus pas moins de cinq à dix ans après le début de l'étude. Notre connaissance accrue des mécanismes génétiques sous jacents au cancer et à son évolution, ainsi que notre connaissance nouvelle des signaux de transduction ont favorisé le développement d'un grand nombre de nouveaux composés thérapeutiques. De plus et en particulier pour des maladies potentiellement fatales telles le cancer, la pression du public pour que la distribution de nouveaux médicaments soit autorisée le plus rapidement possible s'accroît de jour en jour. Pour ces raisons, il est urgent de découvrir des moyens d'accélérer le processus de développement des nouveaux traitements en optimisant le design des essais cliniques (en particulier des essais de phase III).

D'autre part, le cancer de la prostate est une maladie assez hétérogène et de nombreux facteurs de pronostique (tels le score du Gleason, le nombre de biopsies positives, etc..) affectent fortement l'évolution naturelle de la maladie. Certains facteurs pronostiques (tels par exemple, le stade pathologique après prostatectomie) peuvent également agir comme facteurs prédictifs du bénéfice de nouveaux traitements, augmentant encore l'hétérogénéité présente dans les résultats des essais cliniques.

Dans la présente dissertation, nous étudions différentes méthodes par le biais desquelles nous pouvons rendre plus efficaces le design des études cliniques de phase III dans le cancer de la prostate de stade avancé et améliorer la façon dont leurs résultats sont présentés. Nous prétendons qu'afin d'améliorer l'efficacité de ces essais, il est nécessaire de comprendre en profondeur les facteurs de pronostique affectant le devenir des malades et qu'il est également nécessaire d'investiguer les facteurs prédictifs potentiels qui pourraient affecter le bénéfice relatif d'un nouveau traitement et causer ainsi de l'hétérogénéité dans les résultats de l'essai. De plus, nous maintenons que le développement de critères d'évaluation de remplacement (« critères surrogate ») est nécessaire afin d'accélérer le processus de développement des thérapies en réduisant le temps d'observation requis pour évaluer le résultat.

Nous utilisons les données d'études cliniques passées, dont la plupart concernent le traitement hormonal contre le cancer de la prostate, pour éprouver les pistes que nous envisageons pour améliorer les essais de phase III dans le cancer de la prostate

Nous observons tout d'abord que la cause majeure de la durée excessive des essais de phase III réside principalement dans le fait que le critère d'évaluation utilisé dans les essais est un critère à long terme (tel le temps jusqu'à la progression ou la survie). Ceci est particulièrement marquant lorsque la maladie étudiée évolue lentemen comme le cancer de la prostate. Dès lors, remplacer un critère à long terme par un autre critère qui pourrait être mesuré plus tôt, plus facilement ou plus fréquemment et qui reflèterait de façon adéquate le bénéfice des nouveaux traitements sur le critère de survie nous semble une option intéressante. De tels critères de remplacement sont appelés critères «surrogate». Dans notre recherche, nous avons évalué si des changements de l'antigène spécifique de la prostate (PSA) pourraient être utilisés comme critère «surrogate» en lieu et place de critères d'évaluation clinique à long terme tels la survie ou la survie sans rechute.

Pour qu'un critère intermédiaire (ici le PSA) soit un «surrogate» pour un critère définitif à long terme (ici la survie ou la survie sans rechute) il faut que ce critère intermédiaire soit capable d'indiquer de façon précoce le bénéfice qu'apportera finalement le nouveau traitement en termes du critère définitif. Cette notion est souvent confondue avec celle de facteur de pronostique. Dans l'**Introduction**, nous expliquons la différence entre ces deux notions et nous présentons la méthodologie qui doit être utilisée pour démontrer qu'un critère intermédiaire est un surrogate fiable.

La méthode la plus connue de validation des critères «surrogate» est celle de Prentice, qui est fondée sur le paradigme de test statistique d'hypothèses. Nous expliquons aussi que cette méthode repose sur des hypothèses fortes mais non vérifiables. Nous expliquons aussi que les critères de Prentice ne sont ni nécessiares, ni suffisants pour démontrer qu'un critère intermédiaire est un surrogate fiable lorsque le critère d'évaluation final est le temps jusqu'à un événement. Nous décrivons et discutons aussi les avantages d'une nouvelle méthode de démonstration, «la méthode méta analytique de validation des surrogates», qui a été développée par les statisticiens du Limburgs Universitair Centrum de Diepenbeek en Belgique et qui repose sur le paradigme statistique d'estimation.

Au **Chapitre 2**, nous utilisons les données d'une essai de phase III de l'EORTC dans le cancer de la prostate métastatique (EORTC 30892) afin de déterminer la définition la plus sensible (au sens statistique) du temps jusqu'à la progression du PSA pour les malades sous traitement endocrinien. Nous concluons que quelle que soit la définition choisie, le temps jusqu'à la progression du PSA n'est pas très spécifique pour prédire la durée de survie, mais que la sensibilité du critère peut être améliorée d'une part en ajustant la définition du critère de progression en PSA et en supprimant la nécessité de confirmer l'augmentation du marqueur, et d'autre part en utilisant un critère souple tel que une augmentation relative de 20% ou 50% au lieu d'un doublement.

Au **Chapitre 3**, nous utilisons une très grande base de données sur les cancers métastatiques qui nous a été gracieusement fournie par AstraZeneca pour appliquer la méthode de validation méta analytique et vérifier si la définition du temps jusqu'à la progression du PSA définie au chapitre 2 et d'autres variantes de ce critère sont des surrogates fiables de la

survie globale des malades. Nous concluons à une satisfaction modérée des critères de validation dans les cancer de la prostate métastatique sous traitement endocrinien: nos résultats ont mis en évidence une corrélation de 0.66 entre les effets du traitement sur le critère de progression du PSA et les effets du traitements sur la survie. Nous suggérons cependant des moyens d'utiliser tout de même le PSA afin de réduire la durée totale des essais de phase III. Nous suggérons que le PSA peut être utilisé au cours des analyses intérimaires d'un essai afin de décider soit d'arrêter l'étude façon précoce pour futilité (si aucun effet sur le PSA n'apparaît) soit de diffuser les résultats ou de soumettre un dossier d'enregistrement accéléré sur base des résultats en PSA, sous la condition expresse que le suivi de l'essai soit néanmoins poursuivi afin de documenter par la suite les résultats de survie à long terme.

Dans les chapitres suivants, chapitres 4 à 6, nous envisageons un second aspect des essais cliniques de phase III: l'hétérogénéité dans la population de malades et dans les résultats d'une étude.

La présence d'hétérogénéité dans les caractéristiques du groupe de malade participant à un essai clinique ou de leur faculté de répondre aux traitements testés peut augmenter la variabilité de l'estimation de l'effet du traitement. De plus, lorsque l'on compare deux traitements ou plus, un déséquilibre dans la distribution des facteurs de pronostique entre les groupes peut être résulter en un biais important dans l'estimation de l'effet du traitement. En effet, l'effet du déséquilibre pronostique sera confondu avec celui du traitement et pourra artificiellement soit augmenter soit diminuer celui-ci, ce qui à son tour résultera en une augmentation du risque de fausse conclusion positive (type I) ou négative (type II). Dès lors, la connaissance des facteurs de pronostique est essentielle à la sélection des facteurs de stratification appropriés dans le design des essais ultérieurs.

En sus de l'hétérogénéité de la population de malade, les effets du traitement peuvent eux-mêmes être hétérogènes pour des groupes de malades caractérisés par des facteurs *prédictifs*. Sont prédictifs des facteurs qui modulent la taille de l'effet attendu du traitement. Dans un essai clinique, un facteur prédictif est caractérisé une interaction statistique importante avec l'effet du traitement. L'identification de tels facteurs est également très utile car ces facteurs peuvent limiter l'applicabilité des résultats globaux de l'étude à certains sous groupes de malades. Dès lors, leur connaissance est critique pour la planification de nouvelles études car elle permet de définir les critères de sélection de façon appropriée. De plus, en sélectionnant le groupe de malade le plus sensible au traitement expérimental, l'on peut augmenter la probabilité de démontrer un bénéfice significatif du nouveau traitement, tout en épargnant un traitement inutile aux malades qui ne sont pas ou sont moins sensibles à ce traitement.

Dans le **Chapitre 4** nous ré analysons l'essai 22863 de l'EORTC qui a établi l'usage d'un traitement hormonal adjuvant après irradiation comme traitement standard dans les cancers localement avancés. Nous identifions des facteurs de pronostique et établissons que ces facteurs n'étaient pas prédictifs. De la sorte, nous confirmons que les résultats globaux de

l'étude s'appliquent à l'ensemble de la population sélectionnée dans l'étude (soit les cancers T3-4 et/ou G3 N0 M0).

Au **Chapitre 5**, nous présentons les résultats de l'étude EORTC 22911 qui comparait une irradiation postopératoire immédiate à une approche de surveillance dans une population de malades ayant un cancer localisé mais présentant après prostatectomie radicale des facteurs pathologiques indiquant un risque de rechute élevé. Nous expliquons au moyen de cette étude la méthodologie à utiliser lorsque l'on désire établir la présence de facteurs prédictifs lorsque des facteurs pronostiques affectant fortement l'évolution naturelle de la maladie sont également présents. Nous avons identifié que les patients ayant une extension extra capsulaire comme seul facteur pathologique de risque (pT2) semblent bénéficier du traitement adjuvant dans une moindre mesure que les patients présentant avec des marges opératoires positives ou avec une invasion des vaisseaux séminaux. Les résultats démontrent cependant que même ce sous groupe pT2 bénéficie de l'irradiation postopératoire de façon statistiquement significative. Nous avons ensuite utilisé les résultats de cette étude dans la discussion pour donner un exemple des problèmes liés à la méthode de validation de Prentice citée plus haut.

Au **Chapitre 6**, nous envisageons une population de malades présentant un cancer de la prostate métastatique (principalement osseux) mais réfractaires au traitement hormonal. Nous avons groupé les données de trois études EORTC qui ont été conduites dans les années nonante et dans lesquelles la qualité de vie a été mesurée en sus des critères d'évaluation cliniques. Nous évaluons si la qualité de vie mesurée lors de l'entrée du malade dans l'essai est un facteur de pronostique statistiquement indépendant d'autres facteurs pronostiques cliniques et biochimiques connus pour prédire la durée de survie des malades. Nous comparons la précision des prédictions obtenues sur base du modèle incluant les critères de qualité de vie que nous avons construit à celui de modèles existants se basant strictement sur les critères cliniques et biochimiques. Nous conclusons que bien que des critères de qualité de vie à l'entrée soient des facteurs de pronostique statistiquement significatifs, ils n'améliorent pas la précision des prédictions de la durée de survie d'un malade individuel et ne méritent donc pas d'être utilisé dans la pratique clinique journalière.

Au **Chapitre 7**, nous revenons à des études qui ont comparé le blocage androgénique complet à la castration seule pour les malades métastatiques sensibles au traitement hormonal. Nous identifions dans le design et l'analyse de ces études des problèmes qui pourraient avoir limité leur faculté de démontrer un bénéfice en survie. Nous recommandons une très grande attention dans la définition de la différence d'effet « cible » que l'essai doit démontrer. Ceci est essentiel afin de garantir que la taille de l'échantillon soit suffisamment grande et permette de détecter la plus petite différence d'intérêt clinique. Au moment de la définition de l'essai, ainsi qu'au moment de l'analyse, la façon dont cette différence évolue au cours du temps doit également être étudiée afin de planifier et d'utiliser des méthodes statistiques appropriées lorsque l'hypothèse de proportionnalité des effets est violée. Enfin, nous mettons en évidence que les résultats d'une étude ne doivent pas être divulgués prématurément et que toute analyse intérimaire doit être planifiée et prise en compte au moment de l'analyse finale par l'usage de méthodes de correction de l'erreur de

type I. Ceci, afin de protéger le risque global de faux positif et de faux négatifs, et afin d'éviter de publier des résultats erronés.

Au **Chapitre 8**, nous discutons de façon critique les résultats présentés aux chapitres 2 à 7 selon des critères statistiques d'une part, et cliniques d'autre part, et à la lumière d'autres publications dans ce domaine. En particulier, nous discutons la méthode méta analytique de validation des critères «surrogate» et ses limites pratiques. Nous discutons également si la validation d'un critère intermédiaire peut être généralisée à d'autres stades de la maladie ou à d'autres modalités de traitement que ceux présents dans la validation. Nous concluons par quelques recommandations pour l'utilisation du PSA comme critère d'évaluation et suggérons comment planifier des études cliniques dans le cancer de la prostate dans le futur, en utilisant le PSA.

Curriculum Vitae

Laurence Collette is born on July 18th 1969 in Namur (Namen), Belgium. After graduating from college at the Institut Sainte Marie à Huy (Belgium) she started studying German languages at the Catholic University in Louvain-la-Neuve, Belgium in 1987. After having passed the first year, in 1988 she turned to studying Mathematics at the same university and graduated in June 1992 with mention "La Plus Grande Distinction". In 1992 and 1993, she followed the Master of Sciences in Biostatistics at the Limburgs Universitair Centrum, Diepenbeek, Belgium and graduated in January 1994. Since September 1994 until March 1995, she worked at assistant to the professor of statistics at the Biometrics Unit of Faculty of Agronomy of the Catholic University of Louvain-la-Neuve.

In March 1995, she joined the European Organisation for Research and Treatment of Cancer (EORTC) Data Center. The first three years she worked as fellow statistician on the design and analysis of clinical trials. In March 1998, she became full time employee at the EORTC Data Center, now employed as Senior Biotatistician. From the very start, she was been involved with clinical trials in prostate cancer and worked with the Radiotherapy and Genito-Urinary Groups. Over the years, she also worked in the field of head and neck cancer, meta-analysis, fungal infections and all cancers treated with irradiation. She has made a number of contributions to Oncology and Urology Congresses (EAU, AUA, ECCO, ASCO) and contributed to over 60 publications in major peer reviewed journals.

In 1997, she met Prof. Schröder and in February 2004, she registered as a PhD student at Rotterdam University for the present research project. EORTC offered her the possibility to carry out this project in conjunction with her regular work at the organization.

She also acts as a reviewer for several scientific journals (European Journal of Cancer, European Urology, Journal of Urology, Nature Clinical Practice Urology, Journal of Clinical Oncology, The Lancet) and since 2001 was appointed as an independent expert on the Clinical Trials Advisory and Awards Committee for Cancer Research UK.

List of publications

Biostatistics

D.J. SARGENT, B.A. CONLEY, C. ALLEGRA, **L. COLLETTE**. Clinical trial designs for predictive marker validation in cancer treatment trials. *J Clin Oncol (*23):2020-2027, 2005.

R. SYLVESTER, M. VAN GLABBEKE, **L. COLLETTE**, S. SUCIU, B. BARON, C. LEGRAND, T. GORLIA, G. COLLINS, C. COENS, L. DECLERCK, P. THERASSE. Statistical methodology for phase III cancer clinical trials: advances and future perspectives. *Eur J Cancer* 38:S162-S168, 2002

R. SYLVESTER AND **L. COLLETTE**. When do statistics lie? *Urooncology* 1: 185-194, 2001

R. SYLVESTER, **L. COLLETTE** and L. DUCHATEAU. The role of meta-analysis in assessing cancer treatments. *Eur J of Cancer*, 36: 1351-58, 2000.

L. DUCHATEAU, **L. COLLETTE**, R. SYLVESTER and J.P. PIGNON. Estimating number of events from the Kaplan Meier curve for incorporation in literature based meta-analysis: what you don't see you can't get ! Biometrics, *Biometrics* 56 (3): 887-893, 2000.

R. SYLVESTER, **L. COLLETTE**. Statistical considerations of chemoprevention clinical trials in prostate cancer. *Eur Urol* 35: 519-522, 1999.

Prostate Cancer

W. FIEDLER, N. TCHEN, J. BLOCH, P. FARGEOT, R. SORIO, JB. VERMORKEN, **L. COLLETTE**, D. LACOMBE AND C. TWELVES Open label Phase II study of Sabarubicin (MEN-10755) in patients with progressive hormone refractory prostate cancer: a study of the European Organisation for Research and Treatment of Cancer New Drug Development Group. *Eur J Cancer 2005, in press*

C.N. STERNBERG P. WHELAN J. HETHERINGTON B. PALUCHOWSKA P.H.TH.J. SLEE K. VEKEMANS P. VAN ERPS C. THEODORE O. KORIAKINE T. OLIVER D. LEBWOHL M. DEBOIS A. ZURLO, **L. COLLETTE** for the Genitourinary Tract Group of the EORTC. Phase III Trial of Satraplatin, an Oral Platinum plus Prednisone vs. Prednisone alone in Patients with Hormone-Refractory Prostate Cancer. Oncology 68:2–9; 2005.

F.H. SCHROEDER, P. WHELAN, TH. M. DE REIJKE, K.H. KURTH, M. PAVONE-MACALUSO, J. MATTELAER, R.F. VAN VELTHOVEN, M. DEBOIS, **L. COLLETTE**. Metastatic prostate cancer treated by flutamide versus cyproterone acetate: Final analysis of the European Organisation for Research and Treatmetn of Cancer (EORTC) protocol 30892. *Eur Urol* 45: 457-464, 2004

F.H. SCHROEDER, K.H. KURTH, S.D. FOSSA, W. HOEKSTRA, P.P.M. KARTHAUS, M. DEBOIS**, L. COLLETTE** AND MEMBERS OF THE EORTC GU GROUP. Early versus delayed endocrine treatment in pN1-3M0 prostate cancer without local treatment – Results of EORTC 30846 – a phase III study. *J Urol*, 172, 923–927, 2004

F ATAMAN F, A ZURLO, X ARTIGNAN, G VAN TIENHOVEN , LE BLANK, P WARDE, JB DUBOIS, W JEANNERET, F KEUPPENS, J BERNIER, A KUTEN, **L COLLETTE,** M PIERART, M BOLLA. Late toxicity following conventional radiotherapy for prostate cancer: analysis of the EORTC trial 22863. *Eur J Cancer*, 40(11):1674-81, 2004.

W. ALBRECHT, **L. COLLETTE**, C. FAVA, O.B. KORIAKINE, P. WHELAN, U.E. STUDER, TH.M. DE REIJKE, P.J. KIL AND L.A. REA. Intermittent Maximal Androgen Blockade in Patients with Metastatic Prostate Cancer - An EORTC Feasibility Study *Eur Urol* 44: 505-11, 2003.

G.O.N. OOSTERHOF, J.T. ROBERTS, TH. M. DE REIJKE, S.A. ENGELHOLM, S. HORENBLAS, H. VON DER MAASE, N. NEYMARK, M. DEBOIS AND **L. COLLETTE** AND MEMBERS OF THE EORTC-GU GROUP. Strontium-89 Chloride versus palliative local field radiotherapy in patients with hormonal escaped prostate cancer: a phase III study of the European Organisation for Research and Treatment of Cancer Genitourinary Group. *Eur Urol* 44: 519-26,2003.

P.J. KIL, H.M. GOLDSCHMIDT, B.J. WIEGGERS, O.B. KARIAKINE, U.E. STUDER, P. WHELAN, J. HETHERINGTON, T.M. DE REIJKE, J.W. HOEKSTRA and **L. COLLETTE**. Tissue Polypeptide-Specific Antigen (TPS) Determinations before and during Intermittent Maximal Androgen Blockade in Patients with Metastatic Prostatic Carcinoma. *Eur Urol* 43(1): 31-8, 2003

TH.M. DE REIJKE, **L. COLLETTE.** EORTC prostate cancer trials: what have we learnt? *Crit Rev Oncol Hematol* 43: 159-165, 2002.

A. ZURLO, **L. COLLETTE**, G. VAN TIENHOVEN, L.E. BLANK, P. WARDE, J.B. DUBOIS, W. JEANNERET, G. STORME, J. BERNIER, A. KUTEN, M. PIERART, M. BOLLA. Acute toxicity of conventional radiation therapy for high risk prostate cancer in EORTC trial 22863. *Eur Urol* 42: 125-132, 2002.

M. BOLLA, TH. M. DE REIJKE, A. ZURLO, **L. COLLETTE.** Adjuvant hormone therapy in locally advanced and localized prostate cancer. Three EORTC Trials. *Front Radiat Ther Oncol* 36: 81-86, 2002

H. VAN POPPEL, **L. COLLETTE**, Z. KIRKALI, M. BRAUSI, W. HOEKSTRA, D.W. NEWLING, M. DECOSTER and members of the EORTC GU group. Quality control of radical prostatectomy: a feasibility study. *Eur J Cancer* 37: 884-891, 2001.

S.D. FOSSA, P.H.TH SLEE, M. BRAUSI, S. HORENBLAS, R.R. HALL, J.W. HETHERINGTON, N. AARONSON, L. DE PRIJCK and **L. COLLETTE**. Flutamide versus Prednisone in patients with prostate cancer symptomatically progressing after androgen ablative therapy: a phase III study of the European Organisation for Research and Treatment of Cancer Genitourinary Group. *J Clin Oncol* 19(1): 62-71, 2001.

F.H. SCHRÖDER, **L. COLLETTE**, P. WHELAN, TH. M. DE REIJKE and members of the EORTC Genito-Urinary Group. Prostate cancer treated by antiandrogen: is preservation of sexual function possible?. *Br J Cancer*, 82 (2): 283-290, 2000.

M. BOLLA, D. GONZALEZ, P. WARDE, J.B. DUBOIS, R. MIRIMANOFF, G. STORME, J. BERNIER, A. KUTTEN, C. STERNBERG, J. MATTELAER, J. LOPEZ TORRECILLA, R PFEFFER, CL CUTAJAR, M. PIERART, **L. COLLETTE**, T. GIL AND H. BARTELINK. Adjuvant hormonal therapy improves logal regional control and survival in patients with locally advanced prostate cancer treated with radiotherapy. *N Engl J Med* 337 (5):295-299, 1997

Bladder Cancer

C.N.STERNBERG, P.DE MULDER, J.H.SCHORNAGEL, C.THEODORE, S.D.FOSSA, A.T.VAN OOSTEROM, J.A.WITJES, M.SPINA, C.J. VAN GROENINGEN, , B. DUCLOS, J.T. ROBERTS, C. DE BALINCOURT**, L.COLLETTE** and the EORTC Genito-Urinary Cancer Group. Seven year update of an EORTC phase III trial of high dose intensity M-VAC chemotherapy and G-CSF versus classic M-VAC in advanced urothelial tract tumors (EORTC protocol 30924). *Eur J Cancer, in press*

M. BRAUSI M., **L. COLLETTE**, KH KURTH, APM VAN DER MEIJDEN, W OOSTERLINCK, JA WITJES, D NEWLING ,C BOUFFIOUX, RJ SYLVESTER. Variability in the recurrence rate at first follow-up cystoscopy after TUR in stage TaT1 transitional cell carcinoma of the bladder: a combined analysis of seven EORTC studies. *Eur Urol* 41:523-531, 2002

C.N. STERNBERG, P.H.M. DE MULDER, J.H. SCHORNAGEL, C. THÉODORE, S.D. FOSSA, A.T. VAN OOSTEROM, F. WITJES, M. SPINA, C.J. VAN GROENINGEN, C. DE BALINCOURT AND **L. COLLETTE** for the European Organization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group. Randomized phase III trial of high-dose intensity Methotrexate, Vinblastine, Doxorubicin and Cisplatin (MVAC) chemotherapy and recombinant human granulocyte colony-stimulating factor versus classic MVAC in advanced urotherelial tract tumors: European Organization for Research and Treatment of Cancer Protocol n°30924. *J Clin Oncol* 19(10): 2638-2646, 2001.

A. VAN DER MEIJDEN, R. SYLVESTER. **L. COLLETTE**, A. BONO, AND F. TEN KATE. The role and impact of review pathology in stage TaT1 bladder cancer: a combined analysis of five EORTC trials. *J Urol*, 164:1533-1537, 2000

PHM. DE MULDER, C. THEODORE, A. SELLA, O. KORIAKINE, CN STERNBERG, **L. COLLETTE** AND C. DE BALINCOURT. Phase II EORTC trial with 5-fluorouracil, cisplatin and interferon-α as second line treatment of advanced transitional cell carcinoma of the urothelial tract. *Annals Oncol* 11: 1391-94, 2000.

J.A. WITJES, A.P.M. VAN DER MEIJDEN, **L. COLLETTE**, R. SYLVESTER, F.M.J. DEBRUYNE, A. VAN AUBEL, W.P.J. WITJES and the EORTC GU group and the Dutch South East Cooperative Urological Group. Long term follow-up of an EORTC randomized prospective trial comparing intravesical bacille Calmette-Guérin-RIVM and Mitomycin C in superficial bladder cancer. *Urology* 52(3): 403-410, 1998.

Germ-Cell Cancer

S.D. FOSSA, B. PALUCHOWSKA, A. HORWICH, G. KAISER, PHM. DE MULDER, O. KORIAKINE, AT. VAN OOSTEROM, L. DE PRIJCK, **L. COLLETTE**, R. DE WIT. An EORTC phase II trial of C-BOP/BEP intensive induction chemotherapy for intermediate and poor risk metastatic germ cell tumours. *Br J Cancer, in press*

S.D. FOSSA, R. DE WIT, J.T. ROBERTS, P.M. WILKINSON, P.H.M. DE MULDER, G.M. MEAD, P. COOK, L. DE PRIJCK, S. STENNING, N.K AARONSON, A. BOTTOMLEY and **L. COLLETTE**. Quality of Life in Good Prognosis Patients With Metastatic Germ Cell Cancer: A Prospective Study of the European Organization for Research and Treatment of Cancer Genitourinary Group/Medical Research Council Testicular Cancer Study Group (30941/TE20). *J Clin Oncol*, 21(6):1107-18, 2003.

Y. VERGOUWE, EW STEYREBERG, R. DE WIT, J.T. ROBERTS, HJ. KEISER, **L. COLLETTE**, S.P. STENNING AND J.D.F. HABEMA. External validity of a predition rule for residual mass histology in testicular cancer: an evaluation for good prognosis patients. *Br J Cancer* 88: 843-847, 2003.

R. DE WIT, J.T. ROBERTS, P.M. WILKINSON, P.H.M. DE MULDER, G.M. MEAD, S.D. FOSSA, P. COOK, L. DE PRIJCK, S. STENNING AND **L. COLLETTE**. Equivalence of three or four cycles of Bleomycin, Etoposide and Cisplatin and of a 3- or 5-day schedule in good prognosis germ cell cancer: a randomized study of the European Organisation for Research and Treatment of Cancer Cooperative Group and the Medical Research Council. *J Clin Oncol* 19(6):1629-1640, 2001

L. COLLETTE, R.J. SYLVESTER, S.P. STENNING, S.D. FOSSA, G.M. MEAD, R. DE WIT, P.H.M. DE MULDER , N. NEYMARK, E. LALLEMAND and S.B. KAYE for the European Organization for Research and Treatment of Cancer Genito-Urinary Tract Cancer Collaborative Group and the Medical Research Council Testicular Cancer Working Party. Impact of the treating institution on survival of patients with "poor prognosis" metastatic nonseminoma. *J Nat Cancer Inst* 91(10): 839-846, 1999.

R. DE WIT, **L. COLLETTE**, R. SYLVESTER, P.H.M. DE MULDER, D.T. SLEIJFER, W.W. TEN BOKKEL HUININK, A.T. VAN OOSTEROM, E. BOVEN AND G. STOTER. Serum Alpha-fetoprotein surge after the initiation of chemotherapy for non-seminomatous testicular cancer has an adverse prognostic significance. *Br J Cancer* 78(10):1350-1355, 1998.

R. DE WIT, G. STOTER, D.TH. SLEIJFER, JP. NEIJT, W.W. TEN BOKKEL HUININK, L. DE PRIJCK, **L. COLLETTE** AND R. SYLVESTER. Four cycles of BEP versus four cycles of VIP in patients with intermediate-prognosis metastatic testicular non seminoma: a randomised study of the EORTC Genitourinary Tract Cancer Cooperative Group. *Br J Cancer*, 78(6):828-832. 1998

S.B. KAYE, G.M. MEAD, S. FOSSA, M. CULLEN, R. DE WIT, I. BODROGI, C. VAN GROENINGEN, R. SYLVESTER, **L. COLLETTE**, S. STENNING, L. DE PRIJCK, E. LALLEMAND, P. DE MULDER. Intensive induction - sequential chemotherapy (BOP/VIP-B) compared to treatment with BEP/EP for "poor prognosis" metastatic non-seminomatous germ cell tumour: a randomized EORTC/MRC trial. *J Clin Oncol*, 16(2):692-701, 1998

S.D. FOSSA, SB KAYE, GM MEAD, M. CULLEN, R. DE WIT, J. BODROGI, CJ VAN GROENINGEN, P.H.M DE MULDER, S. STENNING, L. DE PRIJCK and **L.COLLETTE** on behalf of the EORTC GU Group/MRC Testicular Cancer Working Party. Filgrastim (G-CSF) during combination chemotherapy for patients with poor prognosis metastatic testicular germ cell malignancy (A phase III trial of the EORTC GU Group/MRC Testicular Cancer Working Party). *J Clin Oncol* 16(2):716-724, 1998.

R. DE WIT, G. STOTER, S.B. KAYE, D.TH. STEIJFFER, W.G. JONES, W.W. TEN BOKKEL HUININK, L.A. REA, **L. COLLETTE,** R. SYLVESTER. The importance of Bleomycin in Combination Chemotherapy for Good Prognosis Testicular Non-Seminoma; A randomized Study of the EORTC Genitourinary Tract Cancer Cooperative Group. *J Clin Oncol* 15(5): 1837-43, 1997.

Breast Cancer

P. POORTMANS, F. ATAMAN, JB DAVIS, H. BARTELINK, JC HORIOT, M. PIÉRART, **L. COLLETTE**. Quality assurance in the EORTC phase III randomised 'boost vs. no boost' trial for breast conserving therapy: Comparison of the results of two individual case reviews performed early and late during the accrual period. *Radioth Oncol, 2005. In press*

P POORTMANS, H BARTELINK, JC HORIOT, H STRUIKMANS, W VAN DEN BOGAERT, A FOURQUET, J JAGER, W HOOGENRAAD, P RODRIGUS, C WARLAM-RODENHUIS, **L COLLETTE**, M PIERART; EORTC RADIOTHERAPY AND BREAST CANCER GROUPS. The influence of the boost technique on local control in breast conserving treatment in the EORTC 'boost versus no boost' randomised trial. *Radiother Oncol*;72(1):25-33, 2004.

C VRIELING, **L COLLETTE**, A FOURQUET, WJ HOOGENRAAD, JC HORIOT, JJ JAGER, S BING OEI, HL PETERSE, M PIERART, PM POORTMANS, H STRUIKMANS, W VAN DEN BOGAERT, H BARTELINK, and on behalf of the EORTC Radiotherapy, Breast Cancer Groups. Can patient-, treatment- and pathology-related characteristics explain the high local recurrence rate following breast-conserving therapy in young patients? *Eur J Cancer,* 39(7): 932-944, 2003

H. BARTELINK, J.C. HORIOT, P. POORTMANS, H. STRUIKMANS, W. VAN DEN BOGAERT, I. BARILLOT, A. FOURQUET, J. BORGER, J. JAGER, W. HOOGENRAAD, **L. COLLETTE**, M. PIERART. Recurrence rates after treatment of breast cancer with standard radiotherapy with or without additional radiation. *N Engl J Med* 345 (19): 1378-1987, 2001

H. BARTELINK, **L. COLLETTE**, A. FOURQUET, J.C. HORIOT, J.J. JAGER, J.L. PETERSE, M. PIERART, P.M. POORTMANS, H. STRUIKMANS, W. VAN DEN BOGAERT. Impact of a boost dose of 16 Gy on the local control and cosmesis in patients with early breast cancer: The EORTC "boost versus no boost" trial. *Int J Radiat Oncol Biol Phys* 48 (3): 111-111, 2000.

C. VRIELING, **L. COLLETTE**, A. FOURQUET, W.J. HOOGENRAAD, J.C. HORIOT, J.J. JAGER, M. PIERART, P.M. POORTMANS, H. STRUIKMANS, B. MAAS, E. VAN LIMBERGEN and H. BARTELINK. The influence of patient, tumor and treatment factors on the cosmetic results after breast-conserving therapy in the EORTC 'boost vs. no boost' trial. EORTC Radiotherapy and Breast Cancer Cooperative Groups. *Radiother Oncol* 55(3):219-32, 2000.

C. VRIELING, **L. COLLETTE**, E. BARTELINK, J.H. BORGER, S.J. BRENNINGKMEYER, J.C. HORIOT, M. PIERART, P. POORTMANS, H. STRUIKMANS, E. VAN DER SCHUEREN, J. VAN DONGEN, E. VAN LIMBERGEN AND H. BARTELINK on behalf of the EORTC Radiotherapy and Breast Cancer Cooperative Groups. Validation of the methods of cosmetic assessment after breast-conserving therapy in the EORTC "Boost versus no boost" trial*. Int. J. Radiation Oncology Biol. Phys* 45 (3): 667-676, 1999.

C. VRIELING, **L. COLLETTE**, A. FOURQUET, W. HOOGENRAAD, JC. HORIOT, J.J. JAGER, M. PIERART, P.M. POORTMANS, H. STRUIKMANS, M. VAN DER HULST, E. VAN DER SCHUEREN and H. BARTELINK on behalf of the EORTC Radiotherapy and Breast Cancer Cooperative Groups. The influence of the boost in breast-conserving therapy on cosmetic outcome in the EORTC "boost versus no boost" trial. *Int J Radiat Oncol Biol. Phys* 45 (3): 677-685, 1999.

Other Cancers

M.J. VAN DEN BENT, D. AFRA, O. DE WITTE, M. BEN HASSEL, S. SCHRAUB, K. HOANG-XUAN, P.-O. MALMSTRÖM, **L. COLLETTE**, M. PIÉRART, R. MIRIMANOFF AND A.B.M.F. KARIM for the EORTC Radiotherapy and Brain Tumor Groups and the UK Medical Research Council. Long term results of EORTC study 22845: a randomized trial on the efficacy of radiation therapy of low-grade astrocytoma and oligodendroglioma in the adult. *Lancet* 366: 985-90, 2005.

JF BOSSET, G. CALAIS, L. MINEUR, PH MAINGAON, L. RADOSEVIC-JELIC , A. DABAN N, E. BARDET, A. BENY, A. BRIFFAUX, **L. COLLETTE** FOR THE EORTC RADIOTHERAPY GROUP. Enhanced tumorocidal effect of preoperative radiotherapy on rectal cancer by adding chemotherapy : results from the 22921 EORTC randomized trial. *J Clin Oncol* 23(24): 5620-5627, 2005.

List of co-authors

ALBRECHT (Walter)
Dept of Urology
Krankenanstalt Rudolfstiftung
Vienna, Austria

BERNIER (Jacques)
Ospedale San Giovanni
Bellinzona, Switzerland

BLANK (Leo C.E.M)
Dept of Radiotherapy
Academisch Medisch Centrum
Amsterdam, The Netherlands

BOLLA (Michel)
Dept de Radiotherapie
CHR de Grenoble-la-Tronche
Grenoble, France

BOSSET (Jean-Francois)
Dept de Radiotherapie & Oncologie
CHR de Besançon – Hopital Jean Minjoz
Besançon, France

BOTTOMLEY (Andrew)
EORTC Data Center, Quality of Life Unit
Brussels, Belgium

BURZYKOWSKI (Tomasz)
Dept Biostatistics
Limburgs Universitair Centrum
Diepenbeek, Belgium

CARROLL (Kevin)
Statistical Expert for Global Oncology
AstraZeneca Pharmaceuticals
Macclesfield, United Kingdom

CUTAJAR (C.L.)
Dept of Surgery/Urology
St Luke's Hospital
Guardamangia, Malta

DA POZZO (Luigi Filippo)
Dept of Urology
Istituto Scientifico H. San Raffaele
Milano, Italy

DENIS (Louis)
Oncology Centre Antwerp
Antwerpen, Belgium

DE REIJKE (Theo M.)
Dept. of Urology
Academisch Medisch Centrum
Amsterdam, The Netherlands

DUBOIS (Jean Bernard)
CRLC Val D'Aurelle
Montpellier, France

FOSSA (Sophie D.)
Dept of Oncology
Norwegian Radium Hospital
Oslo, Norway

HAUSTERMANS (Karin)
Radiation Oncology
Universitaire Ziekenhuis Gasthuisberg
Leuven, Belgium

KUTEN (Abraham)
Dept of Oncology
Rambam Medical Center
Haïfa, Israel

LOPEZ TORRECILLA (Jose)
Dept Radiation Oncology
Hospital Genreal Universitario
Valencia, Spain

MARECHAL (Jean-Marie)
Dept d'Urologie
Hôpital Edouard Herriot
Lyon, France

MATTELAER (Johan)
Dept Urology
CAZK Groenhinghe
Kortrijk, Belgium

MIRIMANOFF (René Olivier)
Service de Radio-Oncologie
Centre Hospitalier Universitaire Vaudois
Lausanne, Switzerland

MORRIS (Tom)
Product Medical Director
AstraZeneca Pharmaceuticals
Macclesfield, United Kingdom

NEWLING (Don)
Medical Director Urology
AstraZeneca Pharmaceuticals
Macclesfield, United Kingdom

OOSTERHOF (Gosse O.N.)
Dept of Urology
Academisch Ziekenhuis Maastricht
Maastricht, The Netherlands

PFEFFER (Raphael)
Dept of Oncology
Chaim Sheba Medical Center
Tel-Hashomer, Israel

PIERART Marianne
EORTC Data Center
Brussels, Belgium

SCALLIET (Pierre)
Dept Radiotherapie
Cliniques Universitaires Saint Luc
Brussels, Belgium

SCHROEDER (Fritz H.)
Dept of Urology
ERASMUS MC
Rotterdam, The Netherlands

STERNBERG (Cora)
Dpt Medical Oncology
San Camillo and Forlanini Hospitals
Roma, Italy

STORME (Guy A.)
Dept of Oncology
Academisch Ziekenhuis VUB
Brussels, Belgium

STUDER (Urs E.)
Dept of Urology
Inselspital
Bern, Switzerland

SYLVESTER (Richard)
EORTC Data Center, Biostatitics Unit
Brussels, Belgium

VAN ANDEL (George)
Dept of Urology
Onze Lieve Vrouw Gasthuis
Amsterdam, The Netherlands

VAN CANGH (Paul)
Dept of Urology
Cliniques Universitaires Saint Luc
Brussels, Belgium

VAN POPPEL (Hein)
Dept of Urology
U.Z. Gasthuisberg
Leuven, Belgium

VAN VELTHOVEN (Roland)
Dept Urology
Institut Jules Bordet
Brussels, Belgium

VEKEMANS, Kris
Dept Urology
Virga Jesse Hospital
Hasselt, Belgium

VERBAEYS (Anthony)
Dept of Urology
Universiteit Gent
Gent, Belgium

WARDE (Padraig)
Dpt Radiation Oncology
Princess Margaret Hospital
Toronto, Canada

Acknowledgements

Since 1995 I have enjoyed working at the EORTC data center, and especially serving as a statistician for the EORTC Genito-urinary and Radiotherapy Groups.

First of all I would like to thank Richard Sylvester, my colleague and Director of Biostatistics from whom I learned most if not all I know today about cancer clinical trials. Richard, I really sincerely enjoy collaborating with you and I will always remain thankful for the everyday collaboration you offer to me.

From the EORTC, I thank the General Director Professor Françoise Meunier and the Data Center Director Doctor Patrick Therasse for giving me the opportunity to carry out this PhD research project in the frame of my daily work at the Data Center. I hope that this research has also offered some useful visibility to the EORTC Data Center.

I want to thank my colleagues Statisticians at the EORTC Data Center, especially Stefan Suciu for his kind support in the days of doubt and anxiety. I am also especially grateful to Marianne Piérart who does the data management of the radiotherapy trials and to the Data Management team of the Genito-Urinary Group.

Over the years, I learned from my colleague urologists and radiation therapists from the EORTC Genito-Urinary and Radiation Groups, all I know today about prostate cancer and other genitor-urinary cancers, about radiation techniques and about the hectic life of the large community of researchers in oncology and urology. I am indebt to all of them for this and for having welcomed me in their Groups when I was still almost a baby...

In particular, I am extremely grateful to Professor Fritz Schröder without whom certainly this thesis would never have been completed, nor even started. Professor Schröder, thank you so much for pushing me to start this thesis, for pushing me to insist when papers were turned down by journals and especially thank you for being the precious example that you are for me of a researcher always curious to learn and to understand and of a human being full of positive philosophy.

I want also to take this opportunity as well to thank Professor Michel Bolla, for the long lasting collaboration on radiation prostate cancer trials, for his kindness and his understanding, and for sharing with me his profound knowledge of prostate and breast cancer health care problems. I am also sincerely thankful to Professor Urs Studer, Professor Hein Van Poppel and Professor Charles Mahler, to Professor Sophie Fossa and to Doctor Cora Sternberg from whom I learned so much, over the years along our collaborations not only about prostate cancer and PSA, but also more generally about life, its fascination and complexity. I am also very grateful to Dr. Theo de Reijke for collaborating with me on the EORTC protate cancer DOG of the GU group and for translating the summary of this thesis into Dutch with my colleague Jan Bogaerts.

I am also indebt to all EORTC members and institutions who contributed patients and data for the EORTC trials 30892, 30903, 30921, 30944, 22863 and 22911 and without whom the publications presented here would not have been possible.

I take this opportunity to thank AstraZeneca Pharmaceuticals for permission to use the data from the Casodex Development Program in advanced prostate cancer patients and in particular to Dr. Tom Morris, Mr. Kevin Carroll and Professor Don Newling for their continuous support and their collaboration in working on the PSA surrogacy issue.

From the side of statistics, I am warmly thankful to Dr. Tomasz Burzykowski and the Surrogate Endpoints Group of the Limburs Universitair Centrum, for leading me through the arcanes of surrogate endpoints and surrogate endpoints validation and for sharing their expertise with me for analyses of the Casodex Data Base. Tomasz, really without you, this would not have been possible.

I also want to acknowledge all my friends, especially Muriel Debois and Yolande Lievens for supporting me along this project and for sharing with me the bad days as much as the good ones! Yolande, Muriel, Catherine, Isabelle, Marie-Hélène, Pierre, Soeren, Arnaud, Xavier, Philippe, Stéphane and all my friends, you know how much I appreciate you. However, this will not prevent me from thanking you again for being who you are and how you are!

I am very grateful to my parents who offered me the invaluable chance to study at the University and to pursue post-graduate studies: I know I owe you so much!

Most of all, I am indebt to Professor Paul Mainguet, without whom I certainly would not have been here today. Professor Mainguet, there are no words to tell how much you represent to me.

Finally, I would like to thank the organizations and industries that offered me the financial support for printing the printing document.

Laurence Collette, 2005