Methodological Aspects of Prognostic Classifications

Applications in Testicular Cancer

The printing of this thesis was financially supported by the J.E. Jurriaanse Stichting and the Department of Public Health, Erasmus MC - University Medical Center Rotterdam, Rotterdam, The Netherlands.

The studies described in this thesis were financially supported by the Netherlands Organisation for Health Research and Development (ZonMw, projectnumber 904-61-159)

Van Dijk, Merel
Methodological Aspects of Prognostic Classifications: Applications in Testicular Cancer. Thesis Erasmus MC, University Medical Center Rotterdam - with references - with summary in Dutch

ISBN: 978-90-9021505-1

Printed by: Ipskamp, Enschede
Cover photo: www.shutterstock.com
Cover design by: Merel van Dijk

Methodological Aspects of Prognostic Classifications
Applications in Testicular Cancer

Methodologische aspecten van prognostische classificaties
toepassingen in testiscarcinoom

Proefschrift

ter verkrijging van de graad van doctor aan de
Erasmus Universiteit Rotterdam
op gezag van de rector magnificus

Prof.dr. S.W.J. Lamberts

en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 7 maart 2007 om 11.45 uur

door

Merel Roos van Dijk

geboren te Texel

Promotiecommissie

Promotoren: Prof.dr. J.D.F. Habbema
Prof.dr. E.W. Steyerberg

Overige leden: Prof.dr. C.H. Bangma
Prof.dr. M.G.M. Hunink
Prof.dr. K.G.M. Moons

We are all empty houses
Waiting for someone
To open the door and set us free.

One day, my wish comes true.
A man arrives likes a ghost
And takes me away from my confinement.
And I follow, whithout doubts, without reserves,
Until I find my new destiny.

(Director KIM Ki-duk on Bin-Jip)

Contents

1

Introduction

This thesis describes methodological aspects of prognostic classifications in oncology, specifically in testicular cancer. This chapter presents issues in the development of prognostic classifications, background on testicular cancer and the research questions of this thesis.

1.1 Some background on prognostic classifications in oncology

Definition and examples

The prognosis of a patient with cancer is determined by tumour-, patient-, and treatment related factors (Figure 1). Prognosis concerns the further course of disease; in oncology it often refers to survival or recurrence of disease [1], although the (health related) quality of life is also relevant.

Tumour related factors include the extent of disease, as measured by size of the tumour, spread to lymphatic nodes and spread to other organs (metastases). Such characteristics are used in the T(umour) N(ode) M(etastasis) classification system to determine the stage of disease. The TNM classification is based on the tendency of cancers to start small, enlarge, spread beyond the confines of their organs and metastasise first to the lymph nodes and then further via the lymphatic system and blood circulation [2], with increasingly poor prognosis.

Besides factors related to the tumour, patient related factors such as sex, age, co-morbidity, previous diseases and performance status determine a patient's prognosis. Information on co-morbidity is especially relevant among cancers with longer survival such as prostate and breast cancer [3].

Lastly, treatment affects a patient's prognosis. Treatment includes surgery (e.g. resection of the tumour), chemotherapy, and radiotherapy. The surgical method and its successful technical execution (e.g. complete resection) are important for a patient's prognosis. In chemotherapy the agents and doses and the way of administration determine prognosis, while for radiotherapy type of radiation, dose and volume are relevant.

It is often useful to group patients with similar characteristics into a prognostic classification to support an evidence-based estimate of prognosis and to guide individual treatment decisions.

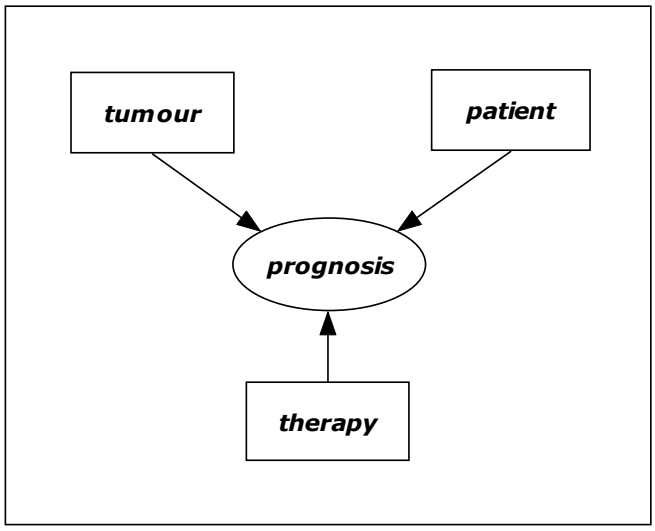


Figure 1 The prognosis of a patient with cancer is influenced by characteristics of the patient, the tumour and therapy [1]

For instance patients with poor prognosis may be considered candidates for more intensive treatment strategies, while patients with a good prognosis may be treated with less burdensome interventions, for example by less toxic chemotherapy regimens [4,5].

In research, prognostic classifications may be used in the design of randomised clinical trials (RCTs) to appropriately target therapies and to increase comparability of patients groups across trials. Prognostic classifications can also be used for fair comparison of outcome of nonrandomised trials and hospital series.

Prognostic groups may be defined by the number of poor prognostic factors, e.g. no poor prognostic factors present (good prognosis) vs. more than one poor prognostic factor present (poor prognosis). Alternatively, prognostic factors can be combined into a scoring system, in which differences in importance between prognostic factors are incorporated by assigning weights to them, e.g. based on regression analysis.

Table 1 Examples of prognostic classification in oncology

Name	Type of cancer	Prognostic factors	Type	Define groups	Use in therapy
NPI[1]	Breast cancer	Lymph-node stage	Tu	Low risk	
		Tumour size	Tu	Medium risk	
		Pathological grade	Tu	High risk	Adjuvant treatment
IPS[2]	Advanced Hodgkin's disease	Serum albumin level	Tu		
		Hemoglobin level	Tu		
		Stage IV disease	Tu	Low risk	Prevent overtreatment
		Leukocytosis	Tu		
		Lymphocytopenia	Tu	High risk	Early intensive CT + ASCT
		Sex	Pt		
		Age	Pt		
IGCC[3]	Advanced testicular cancer	Primary tumour site	Tu	Good prognosis	Lower dose CT
		Non pulmonary visceral metastases	Tu	Inter prognosis	Standard CT
		Tumour markers AFP, HCG, LDH	Tu	Poor prognosis	High dose/dose intense CT

[1] NPI = Nottingham prognostic index (NPI)
[2] IPS = International prognostic factor project
[3] IGCC = International Germ Cell Consensus classification
CT = chemotherapy
ASCT = Autologous stem cell transplantation
Tu = tumour associated factor
Pt = patient associated factor

Table 1 gives a few examples of prognostic classifications. Researchers of the Nottingham City hospital developed the Nottingham Prognostic Index (NPI) for women with breast cancer. The NPI uses tumour related factors, namely lymph-node stage, tumour size and pathological grade, to define high, medium and low risk patients. The aim of the classification was twofold; firstly to aid individual treatment decisions with respect to adjuvant treatment, and secondly for better stratification based on prognosis in the design of clinical trials comparing treatment strategies [6].

The International Prognostic Factors project (IPS) developed a scoring system to predict freedom of progression of disease in patients with advanced Hodgkin's disease. Low and high risk patients were defined based on 2 patient related factors and 5 tumour related factors to distinguish between high and low risk patients. Low risk patients are considered eligible for less intensive treatment to prevent overtreatment, while high risk patients might profit from early intensive chemotherapy with autologous stem cell transplantation [7].

The International Germ Cell Consensus (IGCC) classification, which serves as the central case study in this thesis, distinguishes between good, intermediate and poor prognosis patients with advanced testicular cancer. It is based on the tumour related factors primary tumour site, the presence of metastases and elevated tumour markers[8]. Toxicity in good prognosis patients might be reduced by using a relatively low dose of chemotherapy, while survival of poor prognosis patients might be improved by using high dose or dose intense chemotherapy. In summary, patients with similar tumour- and patient related factors can be grouped in a prognostic classification with the aim to guide treatment decisions.

Methodological aspects in defining prognosis groups

In the prognosis of cancer patients the chance of survival is often the outcome of interest. The IGCC classification, for instance has 5-year survival after diagnosis of cancer as outcome. Ideally all patients are followed until the outcome of interest has or has not occurred, e.g. after 5 years. However, often the outcome is not known for all patients. This is also known as 'censoring'. Censoring can e.g. occur because the investigator stopped the study before all patients had at least 5 years of follow up, or because the patient was lost to follow up. Survival analysis is often used in prognostic studies to appropriately deal with censoring. Survival analysis considers the outcome of interest (e.g. death or recurrence) and the time to the event of interest (survival time) and takes censoring of patients into account. Survival analysis is often reported in terms of survival. The survival probability S(t) is the probability that a patient survives from the time of diagnosis of cancer to a specified time t, for instance 5 years. Survival probability can be estimated using the Kaplan-Meier method. The survival function is defined as follows:

$$S(t_j) = S(t_{j-1})\,(1 - d_j/n_j)$$

where the survival probability $S(t_j)$ at time t is calculated from $S(t_{j-1})$ the survival probability at time t_{j-1}, n_j the number of patients alive just before time t_j, and d_j the number of patients alive just before time t_j. The value of S(t) is constant between events. Therefore the estimated probability is a step function that only changes when an event occurs.

The Kaplan-Meier method takes censoring into account by allowing each patient to contribute information for as long as they are known not to have experienced

the outcome. Censoring is assumed to be uninformative, i.e. if patients could have been followed beyond the point in time when they were censored, and they would have had the same survival probability as those not censored at that time.
The Kaplan-Meier survival curve plots the survival probability against time and is used to compare survival of different patients groups, e.g. according to the presence of a prognostic factor. The difference between groups is usually assessed with the log-rank test.

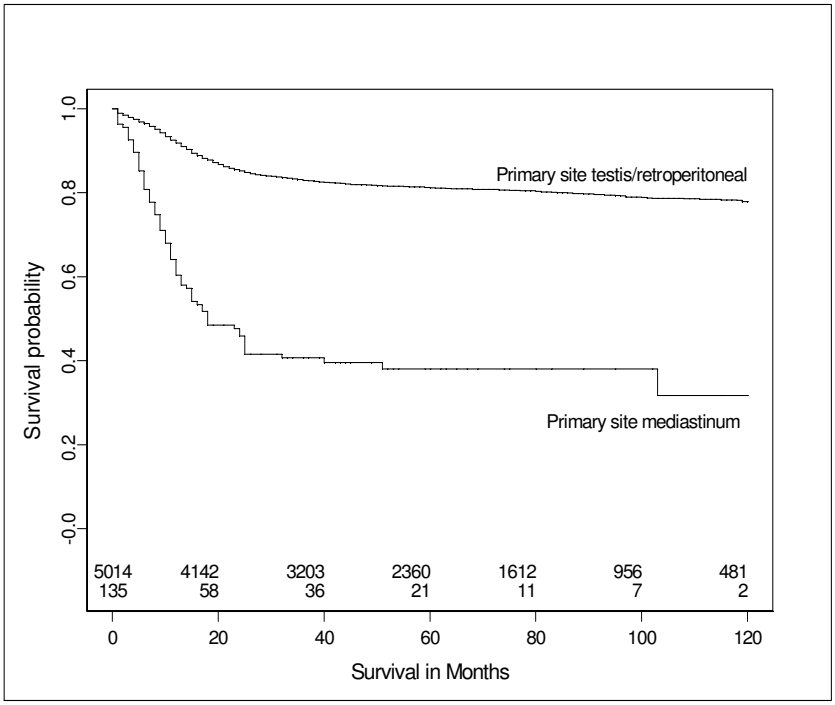


Figure 2 Kaplan-Meier curves showing the difference in survival between nonseminomatous germ cell cancer patients with either testis/retroperitoneum or mediastinum as primary tumour site

Figure 2 shows the survival curves for advanced testicular cancer patients having either testis/retroperitoneum or mediastinum as primary site, with patients with testis as primary site having the better survival. A disadvantage of the Kaplan-Meier method is that only a limited number, categorical prognostic factors can be considered simultaneously. The effect of multiple prognostic factors on survival is

usually assessed with the Cox regression model. The Cox regression model is a regression model that can estimate the combined effects of categorical and continuous prognostic factors on the outcome of interest.
Survival in the Cox regression model is written as

$$S(t) = S_0(t)^{\exp(PI)}$$

where $S_0(t)$ is the baseline survival probability function and PI the prognostic index.

The PI is a linear function of the prognostic factors ($x = x_1, x_2 \ldots x_k$) and the regression coefficients ($\beta = \beta_1, \beta_2 \ldots \beta_k$) :

$$PI = \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

The regression coefficients indicate the size of the effect of the prognostic factors. The effects of prognostic factors can be presented as hazard ratios, exp (β). A hazard ratio above 1 indicates that a prognostic factor is positively associated with the outcome probability, and thus negatively with the length of life.
Survival curves can also be obtained from a Cox regression model as is shown in Figure 3, in which survival of patients with either testis/retroperitoneum or mediastinum as primary site are compared. We note that the survival curves follow the same pattern as the Kaplan-Meier curves in Figure 2, based on assumed proportional hazards of the 2 groups.

Model development

Prognostic classifications should be based on a heterogeneous and sufficiently large sample of patients. To ensure the generalisability of a prognostic classification, the sample on which it was based should be representative of a wider population of patients. Prognostic classifications should therefore preferably be based on patient data from different treatment centres, different treatment settings and different regions/countries [9,10].

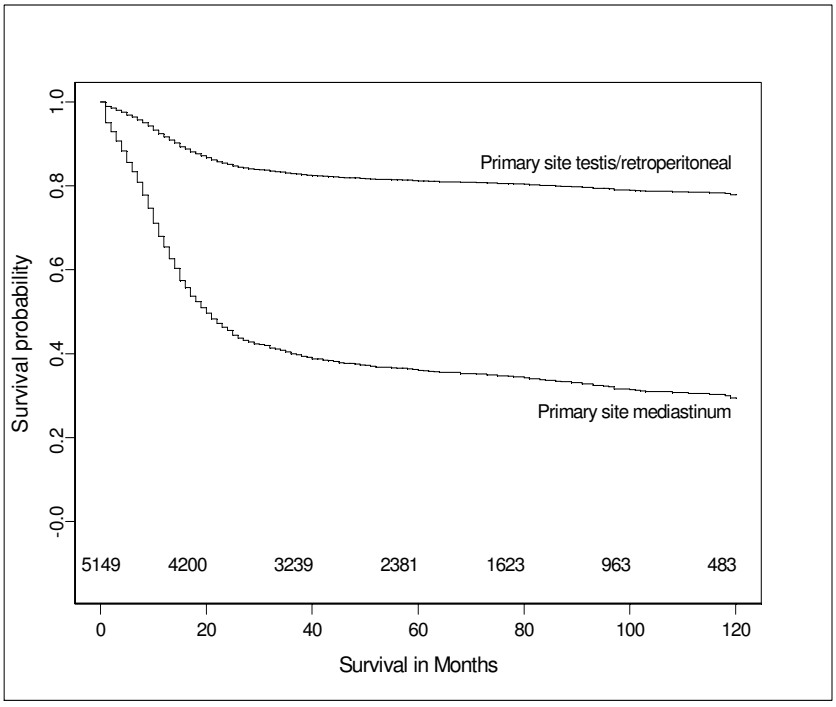


Figure 3 Survival curves based on Cox regression model showing the difference in survival between nonseminomatous germ cell cancer patients with either testis/retroperitoneum or mediastinum as primary tumour site

Preliminary steps

When developing a prognostic model a first step is to investigate the candidate prognostic factors available in the dataset by studying frequency distributions. This also gives insight into the extent of missing values in the data. Often patients with missing values on prognostic factors are excluded (complete case analysis). However, this is statistically inefficient. Furthermore when the excluded patients with missing values differ substantially from patients without missing data, this leads to bias [11,12].

An alternative is to impute missing values with statistical estimates. Imputation leads to a complete data set that can then be analysed using statistical methods for complete data. A simple method of imputation is the use of the mean or the median. A more refined method is to estimate the missing values using regression models, exploiting the correlations with other prognostic factors. Such imputation

methods have the disadvantage that they do not take the uncertainty of the imputation process into account. Imputation methods should include a random component to reflect the fact that imputed values are estimated, especially when the amount of missing data is substantial. This is taken care of by multiple imputation, in which each missing value is imputed several times with different plausible value. The variation among the imputations reflects the uncertainty with which the missing values can be predicted from the observed ones [12,13].

A second step is to decide on the coding of categorical prognostic factors. Categorical variables may be created from continuous variables (e.g. age < 50 or ≥ 50). Although such categorisation improves the interpretability of a model, it can result in a substantial loss of information [14]. Continuous prognostic factors can be included as a linear term, which assumes that the relationship with the outcome is linear. This assumption can be tested by adding nonlinear terms. Examples include simple transformations such as x^2, $\sqrt{x}$, log(x) or exp(x). A more efficient, but mathematically more complex method of transforming continuous variables is the use of restricted cubic splines [15,16].

Data reduction

Often, a selection from many potential prognostic factors has to be made to derive a practically useful prognostic model. Models with a limited number of variables are easier to apply. Also the inclusion of many variables may result in fitting specific patterns in the data ('overfitting') and therefore a poor performance of the model when applied to new patients.

As a general rule of thumb at least 10 events need to be observed for each predictive variable considered [5,15].

A first selection of variables should be based on expert opinion and previous studies. To further limit the number of variables a stepwise selection procedure is often used.

This statistical method considers step by step the additional predictive value of variables, by adding (forward stepwise selection) or deleting (backward stepwise selection) potential prognostic factors. The variable is selected if the additional predictive value is statistically significant, usually using a significance level of 5%. Stepwise selection results in models which are easy to interpret as the number of prognostic factors is limited. However the use of stepwise selection may lead to the exclusion of important prognostic factors. Moreover, the values of the

regression coefficients in the selected model are too large, and their uncertainty is underestimated (e.g. too small standard error of regression coefficients).
Once a set of prognostic factors is selected, we need to consider the additivity assumption, which is made by most regression models. The additivity assumption is that the effect of one prognostic factor does not modify the effect of another prognostic factor. This assumption can be tested by including interaction terms [17,18].

Alternative models

Instead of using regression models which assume a linear relationship between variables and (a transformation of) outcome, one could also use nonlinear models such as recursive partitioning, and neural networks [19,20].
Recursive partitioning is based on splitting groups of patients into smaller groups differing in prognosis. The partitioning algorithm starts with the prognostic factor that best discriminates between two groups according to statistical criteria. Splitting continues for each subgroup using all available prognostic factors. The same prognostic factor may be used more than once. Splitting continues until the subgroups reach a specified minimum size or until no further difference in prognosis can be made. The use of recursive partitioning directly results in a tree with groups differing in prognosis. If the number of identified groups is large, groups with similar predicted outcome can be combined.
Figure 4 shows an example of a regression tree for patients with advanced testicular cancer, which uses the prognostic factors visceral metastases, tumour location and abdominal metastases to identify 5 groups differing in survival.
The resulting tree models are attractive because they are easy to apply and interpret. They have few restrictions, which makes them suitable for finding interactions between prognostic factors. Furthermore trees may show more resemblance with the way clinicians make decisions than linear models. On the other hand, this flexibility makes trees 'data hungry'. Use of relatively small datasets will lead to unstable tree models, and optimism in the performance of the model [21,22].
In artificial neural networks there are some latent, or ‘hidden’ intermediary variables between the prognostic factors and the outcome variable. The most common model is the three-layer model, in which the prognostic factors (input) do not act directly on the outcome variable (output), but channel their influence in a series of latent variables. It is the relative importance of these unobservable

variables, which determines outcome. An advantage of artificial neural networks is that they are very flexible and can include complex relationships between prognostic factors and outcome. However, inclusion of complex relationships can lead to overfitting, limiting the generalisability to new patients. Furthermore artificial neural networks are difficult to interpret, as the importance of prognostic factors is often not clear.

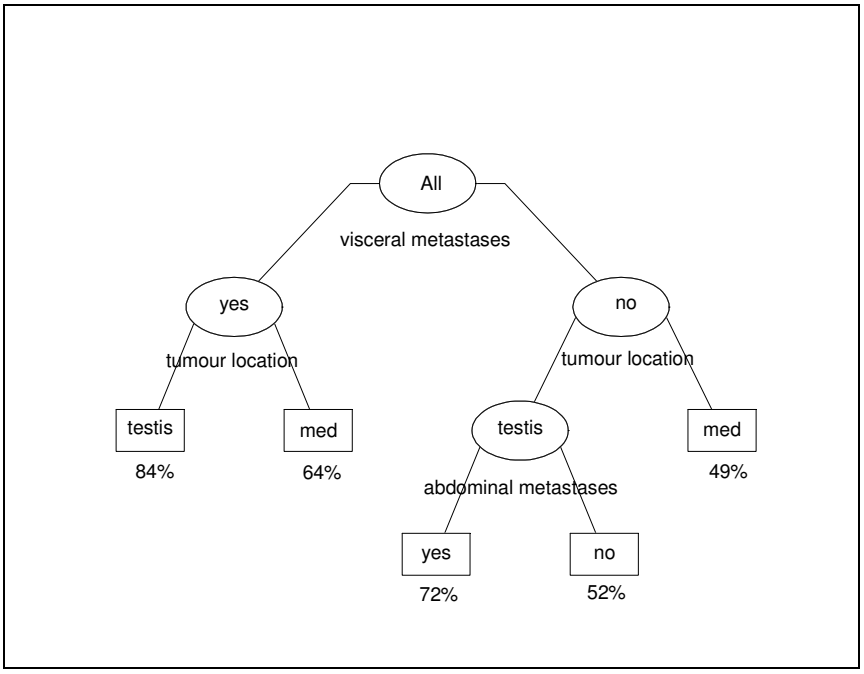


Figure 4 Example of regression tree obtained through recursive partitioning for patients with nonseminomatous germ cell cancer [23]

Model performance

An important aspect of a prognostic classification is its performance, i.e. its ability to distinguish between patients with different outcome.

Differences in survival curves between groups of patients give an important first impression. A more formal method is to use an index of concordance, which compares predicted survival with observed survival to determine discriminative

ability. The value of the concordance measure varies between 0.5 and 1 for sensible classifications.
Alternatively, measures such as the model-chi square, the Brier score and the degree of explained variance of the model (R^2) can be used.[24].

Validation and updating of prognostic classifications

Before a prognostic classification is applied to newly diagnosed patients it should be validated. The prognostic classification might perform well on the data in which it was developed, but may be too optimistic for generalisation to other settings. Three types of validity can generally be distinguished: apparent, internal and external validity. Apparent validity refers to the performance of a model on the patients in which the model was developed. Internal validity refers to the performance of a model in a population of similar patients. There are several methods of assessing internal validity. A method that is often used is split sample, in which e.g. 2/3 of the data are used for the development of the model and 1/3 for validation. Drawbacks of this method are that sample size is limited for both model development and model validation and that due to chance, substantial differences could occur between the development and the validation set. More efficient are crossvalidation methods or resampling techniques such as bootstrapping [25]. With bootstrapping random samples are drawn with replacement from the data. These bootstrap samples are similar in structure to the original data. Each bootstrap sample is representative of the underlying population from which the original dataset was drawn. In each bootstrap sample the development process of the model is repeated and then evaluated in the original data. The difference in performance between the development and testing situations indicates the degree of optimism.
External validity refers to the performance of a model in new patients, differing for instance in time, setting or region of treatment [26]. An external validation study may show that patient outcome has improved over time, e.g. because of improved treatment strategies. This may motivate an update of a prognostic classification.

Prognostic groups from Cox models

To define prognostic groups cutoffs can be applied to the prognostic index (PI) from a Cox regression model. Groups can be made based on percentiles of patients (e.g. 10% patients with lowest predicted probability), optimisation of

model performance (maximising a chi-square statistic over different cutoffs of the PI) or maximising the difference in patient outcome between groups (separation of survival curves).
Ideally cutoffs are based on decision analytic techniques in which the expected gain in survival due to treatment (benefit) and the burden (harm) due to treatment are weighed to determine at what risk which treatment might be beneficial. Ideally benefit is based on results from RCTs.
The optimal cutoff point is where the benefit of treatment equals the harm of treatment. Patients whose risk of cancer mortality is above the threshold should be treated with the alternative treatment, while patients with a risk below the threshold should receive standard treatment.

Examples

Table 2 summarises the methodological aspects of the examples presented in table 1.
The Nottingham prognostic index was constructed with Cox regression analysis on 387 patients, because of missing values in 113 patients. Based on the results of Cox regression analysis a simplified prognostic index was proposed and cutoffs determined based on stratification of patients using only lymph-node stage. This model has later been validated and modified, identifying three prognosis groups differing in survival [6,27].
The International prognostic factor project also used Cox regression analysis to identify seven prognostic factors for patients with advanced Hodgkin disease [7]. Prognosis groups differing in survival were identified based on the number of adverse prognostic factors present. Although 5141 patients from 25 centres were available only 1618 patients were included in the final analysis due to missing data. Internal validation was done on a subset of 2643 patients with incomplete data.
The IGCC classification dismissed 2154 of 5202 patients due to missing values. Prognostic factors were identified using Cox regression analysis. Prognosis groups were defined by the presence of good, intermediate or poor prognostic factors but not by the number of prognostic factors. Hence, effectively a 'max function' was applied. The IGCC classification was both internally and externally validated [8].

Table 2 Methodological aspects of examples of prognostic classifications in oncology

Name	Type of cancer	Patients	Missing [1]	Model	Weights	Definition groups	Validation
NPI[2]	Breast cancer	500	113	Cox regression	Simplified regression weights	Cutoffs on prognostic index	external
IPS[3]	Advanced Hodgkin disease	5141	3253	Cox regression	Equal weights	Number of adverse prognostic factors	internal
IGCC[4]	Advanced testicular cancer	5202	2154	Cox regression	Simplified regression weights	Max function on number of adverse prognostic factors	internal and external

[1] Number of patients excluded because of missing values
[2] NPI = Nottingham prognostic index (NPI)
[3] IPS = International prognostic factor project
[4] IGCC = International Germ Cell Consensus classification
CT = chemotherapy
ASCT = Autologous stem cell transplantation

1.2 Defining prognosis groups in advanced testicular cancer

Clinical background advanced testicular cancer

Although testicular cancer accounts for only 1% of all cancers in men, it is the most common cancer in young adult men [26]. An estimated 8000 new cases of testicular cancer occurred in 2005 in the US [26]. In 1998 there were 482 new cases in the Netherlands [28].

Germ cell tumours account for 95% of testicular cancer, with 10% of all germ cell tumours arising from extragonadal primary sites, such as the mediastinum and retroperitoneum. Germ cell tumours are distinguished according to histology in seminomas and nonseminomas. A tumour is diagnosed as a seminoma if the tumour contains pure seminomatous tissue and the serum level of alpha-fetoprotein, a marker of nonseminomatous tumours, is normal. Non-seminomatous tumours consist of any combination of embryonal carcinoma, teratoma, choriocarcinoma, yolk sac, or seminoma cell types.

After diagnosis of testicular cancer treatment usually starts with orchidectomy to remove the primary tumour. Further treatment depends on clinical staging which is determined by assessments of primary tumour, lymph node, the presence of distant metastasis and tumour markers alpha-fetoprotein (AFP), human chorionic

gonadotrophin (HCG) and lactate dehydrogenase (LDH). When disease is limited to the testis, epididymis, or spermatic cord, this is labelled stage I disease. When distant disease is suspected, further treatment may consist of chemotherapy or retroperitoneal lymph node dissection (RPLND) [29]. Approximately 30% of patients with stage I disease have occult metastatic disease detected at RPLND or surveillance. Since many predictors of occult metastatic disease are known various simple classification schemes are available to select patients for more intensive treatment [30]. Advanced disease includes metastases in the regional nodes (stage II), and the presence of distant metastasis or elevated serum tumour markers (stage III) and is treated with chemotherapy or resection of residual masses [31,32].

Since the 1970s long term cure rates of patients with advanced germ cell tumours have increased to over 80%, because of the ability of cisplatin-based chemotherapy to cure advanced disease [33-36]. Because of the high overall cure rate, interest has shifted from increasing the overall cure rate to reducing treatment related toxicity for patients with a good prognosis [37]. On the other hand poor prognosis patients should be identified, who are eligible for new treatment regimens such as dose-intensification and high-dose chemotherapy with stem cell support [38,39].

Methodological aspects in defining prognosis groups in advanced testicular cancer: an historical overview

Several prognostic studies had been conducted before the IGCC classification to identify groups of patients with nonseminomatous germ cell cancer that differ in prognosis (Table 3). These studies established that prognosis is not only related to anatomic spread of disease, but also to the primary site (extragonadal or gonadal) and to the extent of production of the serum tumour markers AFP, HCG, LDH.

Bosl and colleagues (1983) used a logistic regression model to predict the probability of complete response (CR) of 171 patients treated at the Memorial Sloan Kettering cancer center (MSKCC) [40]. The total number of metastatic sites and pretreatment levels of tumour markers LDH and HCG were the most important prognostic factors, which predicted response correctly for 83% of patients. Patients with a predicted probability ≥ 0.5 were considered as good prognosis, while patients with a predicted < 0.5 were considered as poor prognosis. The predictive ability has been supported by the results of 300 patients treated in trials conducted by the MSKCC since 1982. CR rates were 51 and 40%

in two poor risk trials and 96 and 93% in a randomised trial comparing treatments for good risk patients. The prognostic importance of tumour markers LDH and HCG and the number of metastatic sites has been confirmed in an additional 100 patients [41].

Table 3 Overview of prognostic models developed to classify advanced testicular cancer patients before the IGCC classification

Name (year)	N	Statistical model	Prognostic factors	Prognosis groups		n	outcome
MSKCC (1983)	171	Logistic regression	Log (AFP) Log (LDH) Total number of metastases	Good	p(CR) > 0.5	136	84% CR
				Poor	p(CR) < 0.5	35	20% CR
Indiana (1986)	137	Logistic regression	Indiana staging system Elevated tumour markers	Good	minimal and moderate	51	96% CR
				Poor	advanced	86	58% CR
EORTC (1987)	154	Logistic regression	Presence trophoblastic elements primary tumour AFP ≥1000 ng/ml Presence lung metastases Size and number of lung metastases	Good	high p(CR)	97	89-100% CR
				Inter	inter p(CR)	46	41% CR
				Poor	low p(CR)	11	18% CR
MRC (1992)	795	Cox regression model	Liver, bone or brain metastases AFP > 1000 kU/L or HCG > 1000 IU/L Mediastinal mass > 5 cm > 20 lung metastases	Good	no adverse prognostic factors	528	92% OS
				Poor	≥ 1 adverse prognostic factor	267	68% OS

MSKCC = Memorial Sloan Kettering Cancer Center
EORTC = European Organisation for Research and Treatment of Cancer
MRC = Medical Research Council
p(CR) = probability of complete response; OS = overall survival
Indiana staging system: minimal: elevated HCG or AFP, cervical nodes, unresectable, but nonpalpable disease, minimal pulmonary disease; moderate: palpable abdominal mass only, moderate pulmonary metastases; advanced: advanced pulmonary metastases, palpable abdominal mass plus pulmonary metastases, hepatic, osseous, or CNS metastases

The Indiana University system also used logistic regression to extend the Indiana staging system which classifies patients as having minimal, moderate or advanced disease [42]. Fifty-one patients with minimal or moderate disease had a response rate of 96% and formed the good prognosis group. The group of 86 patients with

advanced diseases had a response rate of 58%. The number of elevated tumour markers further subdivided this group into subgroups with estimated response rates of 73, 65 and 45%.

The European Organisation for Research and Treatment of Cancer (EORTC) identified three prognostic groups (n=154) using the prognostic factors presence of trophoblastic elements in primary tumour, the value of AFP, presence of lung metastases and the size and number of lung metastases using logistic regression. The definition of good, intermediate and poor prognosis groups was based on the probability of a CR.

The models developed by the MSKCC, Indiana University and the EORTC each have some disadvantages. First, the generalisability of these models was limited because of the small number of patients in all three studies, and because of the homogeneity of the group of patients in the studies by MSKCC and Indiana University (single center cohorts). Furthermore all three studies used CR as outcome instead of the more informative outcome, survival.

The Medical Research Council Working Party on Testicular Cancer was the first to combine individual patient data of different treatment centres. They analysed survival of 458 patients treated between 1976 and 1982 at six British centres. With the prognostic factors tumour volume, degree of elevation of tumour markers AFP and HCG three prognosis groups were identified with 3-year survival of 91, 74 and 47% respectively [43].

This study was succeeded by the Second Medical Research Council study, which assessed prognosis in a more recent treatment era (1982-1986). Individual patient data treated in 13 centres in the UK and Norway were combined resulting in a large and heterogeneous database of 795 patients [44].

Using Cox regression analysis they selected the prognostic factors presence of liver, bone or brain metastases, AFP ≥ 1000 kU/L and/or HCG ≥10.000 IU/L, mediastinal mass more than 5 cm and more than 20 lung metastases on which a simple prognostic classification was based. Good prognosis patients (n=528) had no adverse prognostic factors and a 5-year survival of 92%. Poor prognosis patients (n=267) had one or more adverse prognostic factor and a 3-year survival of 68%.

The IGCC Classification: methodological merits and limitations

The coexistence of classifications differing in type, complexity and ability to separate good from poor prognosis complicated international collaboration in randomised trials and made comparison of nonrandomised studies impossible. Therefore the International Germ Cell Cancer Collaborative Group (IGCCCG) was formed which resulted in the development of the International Germ Cell Consensus Classification (IGCC classification) [8].

The IGCCCG analysed data from previously conducted trials from the British Medical Research Council (MRC), the EORTC, groups from the United States (MSKCC, New York, NY; Indiana University Hospital; University of Texas MD Anderson Cancer Center), and national germ cell groups from Canada, Australia, New Zealand, Spain, France, Denmark, and Italy.

Patients participating in these trials had either nonseminomatous (n=5202) or seminomatous (n=660) germ cell cancer and were treated between 1975 and 1990. All patients were treated with cisplatin- (or carboplatin-) containing chemotherapy. Five readily available prognostic factors were selected from a wider set following Cox regression analyses. The site of primary tumour was categorised as testis or retroperitoneal vs. mediastinal. The presence of non-pulmonary visceral metastases (NPVM) was defined as disease at any nonpulmonary visceral site (liver, bone, brain, kidney, skin or gastrointestinal). The pretreatment levels of tumour markers alpha-fetoprotein (AFP), human chorionic gonadotrophin (HCG) and lactate dehydrogenase (LDH). AFP and HCG were analysed as ratios more than each institution's upper limit of normal. Due to interlaboratory variations in assays, LDH levels were considered only as ratios, formed by dividing absolute values by the upper limit of the normal range for each institution. The tumour markers were categorised into three categories (good, intermediate, poor).

The prognostic factors were combined into three prognostic groups for patients with nonseminomatous germ cell tumours (NSGCT) with either good, intermediate or poor prognosis (Table 4). The good prognosis group was characterised by the absence of adverse prognostic factors, the intermediate prognosis group by the presence of any intermediate tumour marker, and the poor prognosis group by the presence of any of the poor prognostic factors mediastinal primary site, NPVM, AFP poor, HCG poor or LDH poor. The classification can be seen as a max function where the good, intermediate and poor prognosis groups have a maximum score of zero, one or two respectively.

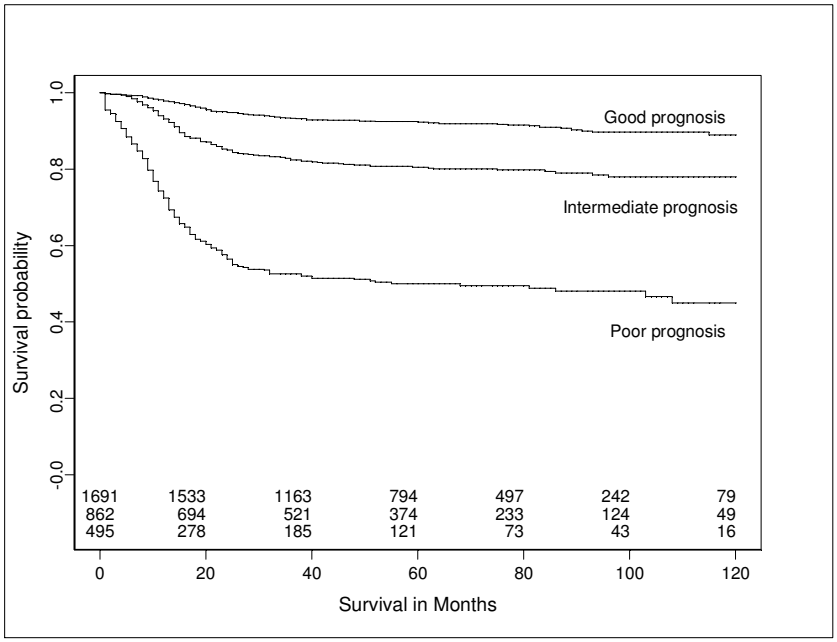


Figure 5 Survival of patients with nonseminomatous germ cell cancer from the IGCCC database, with either good, intermediate or poor prognosis

In the published report 56, 28 and 16% of patients were allocated to the good, intermediate and poor prognosis groups, with a 5-year survival of 92, 80 and 48% respectively (Figure 5) [8]. For patients with seminomatous germ cell tumours only a good and intermediate prognosis group were identified. They represent 90% and 10% of seminomas respectively and have a 5-year survival of 86 and 72%. Good prognosis was defined as any primary site, no NPVM and normal tumour markers, while intermediate prognosis is defined as any primary site, the presence of NPVM and normal tumour markers.

The IGCC classification has quickly become the standard for the design and selection of patients for RCTs and the analysis of observational studies, such as Phase I/II trials and hospital series. The findings of the IGCCCG have been incorporated into the AJCC staging system for testicular cancer, extending the TNM system to a TNMS system including serum tumour markers.

The IGCC classification has several advantages over its predecessors. The international collaboration of several research groups resulted in a heterogeneous

study population with thousands of patients. The use of patient data of several institutions and countries ensures the generalisability of the IGCC classification.
The selection of prognostic factors was based on previous studies and analysis was performed on a sufficiently large sample size. This resulted in a classification with a limited number of well-established prognostic factors, also identified in previous studies.

Table 4 The International Germ Cell Consensus (IGCC) Classification [8]

GOOD PROGNOSIS	
NONSEMINOMA	SEMINOMA
Testis/retroperitoneal primary site And No nonpulmonary visceral metastases And AFP good and HCG good and LDH good 56% of nonseminomas 5 year PFS 89% 5 year OS 92%	Any primary site And No nonpulmonary visceral metastases And Normal AFP, any HCG, any LDH 90% of seminomas 5 year PFS 82% 5 year OS 86%
INTERMEDIATE PROGNOSIS	
NONSEMINOMA	SEMINOMA
Testis/retroperitoneal primary site And No nonpulmonary visceral metastases And AFP intermediate or HCG intermediate or LDH intermediate 28% of nonseminomas 5 year PFS 75% 5 year OS 80%	Any primary site And Nonpulmonary visceral metastases And Normal AFP, any HCG, any LDH 10% of seminomas 5 year PFS 67% 5 year OS 72%
POOR PROGNOSIS	
NONSEMINOMA	SEMINOMA
Mediastinal primary site Or Nonpulmonary visceral metastases Or AFP poor or HCG poor or LDH poor 16% of nonseminomas 5 year PFS 41% 5 year OS 48%	No patients classified as poor prognosis

Tumour markers alpha-fetoprotein (AFP)/ human chorionic gonadotrophin (HCG)/lactate dehydrogenase (LDH): Good - AFP < 1000 ng/ml, HCG < 5000 iu/l, LDH < 1.5 x upper limit of normal; Intermediate - AFP 1000 – 10000 ng/ml, HCG 5000 - 50000 iu/l, LDH 1.5 - 10 x N; Poor - AFP > 10000 ng/ml, HCG > 50000 iu/l, LDH > 10 x N

The IGCC classification has some limitations. The IGCC classification did not consider differences in importance between intermediate tumour markers and differences in importance between poor prognostic factors. Furthermore no distinction is made between the number of intermediate tumour markers in the intermediate prognosis group and the number of poor prognostic factors in the poor prognosis group. Better discrimination might be achieved by incorporating differences in predictive strength and testing specific interaction terms.

All prognostic factors included in the IGCC classification had missing values, especially the tumour marker LDH (37%). Data analysis was performed on 3048 complete cases, discarding data of 2154 patients. Exclusion of patients because of missing data was statistically inefficient, and could have led to bias in the survival estimates of the prognostic groups in the IGCC classification.

The survival estimates of the IGCC classification were based on patients treated between 1975 and 1990. Some patients, treated before 1985, were treated with carboplatin containing chemotherapy, which is now considered inferior to cisplatin containing chemotherapy. Survival may have increased since 1990 due to further improvements in treatment strategies. The survival estimates of the IGCC classification may therefore not be generalisable to patients diagnosed nowadays.

The group of poor prognosis patients is especially of interest, since their survival could be improved by alternative, more intensive treatment strategies such as high-dose chemotherapy with stem cell support or dose intensification.

Further subgrouping of poor prognosis patients would allow for a more precise identification of individuals patients at high risk, and the use of risk-adapted treatment strategies.

The German Testicular cancer group developed a regression tree model to identify subsets of patients within a group of poor prognosis patients, as defined by the IGCC classification, treated in three clinical trials (n=332). Three subgroups were identified with 2-year survival varying from 49% to 84% [23]. Validation of this model in a different group of patients is necessary before this model can be used in clinical practice and the design of clinical trials.

The criteria used to define 'poor prognosis' in the IGCC classification were prognosis (a low estimated survival) and sample size (a large enough group for clinical trials). Ideally, when considering a more intensive treatment the toxic side effects or burden due to this treatment ('harm') should also be taken into account and weighed against the expected gain in survival ('benefit'). By specifying harm

and benefit of a treatment, treatment cutoffs can be determined through decision analysis.

1.3 Research questions

In this thesis we study methodological aspects of defining prognosis groups. We use the IGCCC classification for nonseminomatous germ cell cancer for illustration.
Firstly we evaluate the validity of the IGCC classification; is the model underlying the classification correctly specified or can performance be improved, and can the survival estimates be generalised to newly diagnosed patients. Secondly, we look at alternative ways of defining prognostic groups; we specifically look at poor prognosis patients, as this group has most to gain from alternative treatment strategies.

Validity of the IGCC classification

1. Are the assumptions made in the development of the IGCC classification valid with regard to the inclusion of prognostic factors, or can discriminative ability be improved?
The IGCC classification considers all prognostic factors to be equally important, and makes no distinction between number of prognostic factors within a prognosis group. Incorporating differences in importance and number of prognostic factors may result in better discriminative ability (Chapter 2).

2. What is the effect of missing values on survival estimates of the IGCC classification?
In the IGCC classification many patients had missing values, and they were excluded from analysis. Currently imputation techniques are available to more appropriately deal with this problem (Chapter 3).

3. Has survival of patients with advanced testicular cancer improved since the introduction of the IGCC classification?
The IGCC classification included patients treated between 1975 and 1990, and updating of estimates may be necessary (Chapter 4).

Definition of the poor prognosis group for advanced testicular cancer

4. Is regression tree analysis an appropriate method for further subgrouping within poor prognosis patients?

Further subgrouping within the poor prognosis group has been considered with regression tree analysis, but the validity of this subgrouping needs further study (Chapter 5).

5. At what risk of cancer-mortality should patients with advanced testicular cancer be treated with high-dose chemotherapy?

In the IGCC classification, the creation of prognostic groups was loosely based on percentiles of prognosis, with 16% labelled 'poor prognosis patients' being candidates for more intensive therapy, and 56% labelled 'good prognosis patients' being candidates for less intensive therapy. Decision-analytic approaches need to be considered for better support of cutoffs in the IGCC classification (Chapter 6).

References

1. Hermanek P, Hutter RV, Sobin LH. Prognostic grouping: the next step in tumor classification. J Cancer Res Clin Oncol 1990;116(5):513-6.
2. Dorn HF. Activities of the International Union against Cancer. Acta Unio Int Contra Cancrum 1958;14(2):77-81.
3. Piccirillo JF, Tierney RM, Costas I, Grove L, Spitznagel EL, Jr. Prognostic importance of comorbidity in a hospital-based cancer registry. Jama 2004;291(20):2441-7.
4. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. Br J Cancer 1994;69(6):979-85.
5. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361-87.
6. Haybittle JL, Blamey RW, Elston CW, et al. A prognostic index in primary breast cancer. Br J Cancer 1982;45(3):361-6.
7. Hasenclever D, Diehl V. A prognostic score for advanced Hodgkin's disease. International Prognostic Factors Project on Advanced Hodgkin's Disease. N Engl J Med 1998;339(21):1506-14.
8. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. J Clin Oncol 1997;15(2):594-603.
9. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. Ann Intern Med 1999;130(6):515-24.
10. Hukkelhoven CW, Rampen AJ, Maas AI, et al. Some prognostic models for traumatic brain injury were not valid. J Clin Epidemiol 2006;59(2):132-43.
11. Little R, Rubin DB. Statistical Analysis with Missing Data. Second Edition. New York: John Wiley and Sons, 2002.
12. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley, 1987.
13. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999;18(6):681-94.
14. Royston P, Altman DG, Sauerbrei W. Dichotomizing continuous predictors in multiple regression: a bad idea. Stat Med 2006;25(1):127-41.
15. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. Stat Med 1984;3(2):143-52.
16. Harrell FE, Jr., Lee KL, Pollock BG. Regression models in clinical studies: determining relationships between predictors and response. J Natl Cancer Inst 1988;80(15):1198-202.

17. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med Decis Making 2001;21(1):45-56.
18. Harrell FE, Jr. Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer, 2001.
19. Cross SS, Harrison RF, Kennedy RL. Introduction to neural networks. Lancet 1995;346(8982):1075-9.
20. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Belmont, California: Wadsworth, 1984.
21. Marshall RJ. The use of classification and regression trees in clinical epidemiology. J Clin Epidemiol 2001;54(6):603-9.
22. Crichton NJ, Hinde JP, Marchini J. Models for diagnosing chest pain: is CART helpful? Stat Med 1997;16(7):717-27.
23. Kollmannsberger C, Nichols C, Meisner C, Mayer F, Kanz L, Bokemeyer C. Identification of prognostic subgroups among patients with metastatic 'IGCCCG poor-prognosis' germ-cell cancer: an explorative analysis using cart modeling. Ann Oncol 2000;11(9):1115-20.
24. Schumacher M, Graf E, Gerds T. How to assess prognostic models for survival data: a case study in oncology. Methods Inf Med 2003;42(5):564-71.
25. Efron B, Tibshirani RJ. An introduction to the bootstrap. London: Chapman & Hall Inc., 1993.
26. Jemal A, Murray T, Ward E, et al. Cancer statistics, 2005. CA Cancer J Clin 2005;55(1):10-30.
27. Sauerbrei W, Hubner K, Schmoor C, Schumacher M. Validation of existing and development of new prognostic classification schemes in node negative breast cancer. German Breast Cancer Study Group. Breast Cancer Res Treat 1997;42(2):149-63.
28. Visser O, Coebergh JWW, van Dijck JAAM, Siesling Se. Incidence of cancer in the Netherlands 1998. Utrecht: Vereniging van Integrale Kankercentra, 2002.
29. Bosl GJ, Motzer RJ. Testicular germ-cell cancer. N Engl J Med 1997;337(4):242-53.
30. Vergouwe Y, Steyerberg EW, Eijkemans MJ, Albers P, Habbema JD. Predictors of occult metastasis in clinical stage I nonseminoma: a systematic review. J Clin Oncol 2003;21(22):4092-9.
31. Steyerberg EW, Gerl A, Fossa SD, et al. Validity of predictions of residual retroperitoneal mass histology in nonseminomatous testicular cancer. J Clin Oncol 1998;16(1):269-74.

32. Steyerberg EW, Keizer HJ, Fossa SD, et al. Prediction of residual retroperitoneal mass histology after chemotherapy for metastatic nonseminomatous germ cell tumor: multivariate analysis of individual patient data from six study groups. J Clin Oncol 1995;13(5):1177-87.
33. Bosl GJ, Motzer RJ. Testicular germ-cell cancer. N Engl J Med 1997;337(4):242-53.
34. Hartmann JT, Kanz L, Bokemeyer C. Diagnosis and treatment of patients with testicular germ cell cancer. Drugs 1999;58(2):257-81.
35. Sonneveld DJ, Hoekstra HJ, van der Graaf WT, et al. Improved long term survival of patients with metastatic nonseminomatous testicular germ cell carcinoma in relation to prognostic classification systems during the cisplatin era. Cancer 2001;91(7):1304-15.
36. Steele GS, Richie JP, Stewart AK, Menck HR. The National Cancer Data Base report on patterns of care for testicular carcinoma, 1985-1996. Cancer 1999;86(10):2171-83.
37. de Wit R, Roberts JT, Wilkinson PM, et al. Equivalence of three or four cycles of bleomycin, etoposide, and cisplatin chemotherapy and of a 3- or 5-day schedule in good-prognosis germ cell cancer: a randomized study of the European Organization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group and the Medical Research Council. J Clin Oncol 2001;19(6):1629-40.
38. Bokemeyer C, Kollmannsberger C, Meisner C, et al. First-line high-dose chemotherapy compared with standard-dose PEB/VIP chemotherapy in patients with advanced germ cell tumors: A multivariate and matched-pair analysis. J Clin Oncol 1999;17(11):3450-6.
39. Bokemeyer C, Oechsle K, Hartmann JT, et al. Treatment-induced anaemia and its potential clinical impact in patients receiving sequential high dose chemotherapy for metastatic testicular cancer. Br J Cancer 2002;87(10):1066-71.
40. Bosl GJ, Geller NL, Cirrincione C, et al. Multivariate analysis of prognostic variables in patients with metastatic testicular cancer. Cancer Res 1983;43(7):3403-7.
41. Bajorin DF, Geller NL, Bosl GJ. Assessment of risk in metastatic testis carcinoma: impact on treatment. Urol Int 1991;46(3):298-303.
42. Birch R, Williams S, Cone A, et al. Prognostic factors for favorable outcome in disseminated germ cell tumors. J Clin Oncol 1986;4(3):400-7.
43. Prognostic factors in advanced non-seminomatous germ-cell testicular tumours: results of a multicentre study. Report from the Medical Research Council Working Party on Testicular Tumours. Lancet 1985;1(8419):8-11.
44. Mead GM, Stenning SP, Parkinson MC, et al. The Second Medical Research Council study of prognostic factors in nonseminomatous germ cell tumors. Medical Research Council Testicular Tumour Working Party. J Clin Oncol 1992;10(1):85-94.

2

Survival of Patients with Nonseminomatous Germ Cell Cancer: A Review of the IGCC Classification by Cox Regression and Recursive Partitioning

M.R. van Dijk, E.W. Steyerberg, S.P. Stenning, E. Dusseldorp, and J.D.F. Habbema

Br J Cancer 2004;90(6):1176-83

Abstract

Background The International Germ Cell Consensus (IGCC) classification identifies good, intermediate and poor prognosis groups among patients with metastatic nonseminomatous germ cell tumours (NSGCT). It uses the risk factors primary site, presence of nonpulmonary visceral metastases and tumour markers alpha-fetoprotein (AFP), human chorionic gonadotrophin (HCG) and lactate dehydrogenase (LDH). The IGCC classification is easy to use and remember, but lacks flexibility. We aimed to examine the extent of any loss in discrimination within the IGCC classification in comparison with alternative modelling by formal weighing of the risk factors.

Methods We analysed survival of 3048 NSGCT patients with Cox regression and recursive partitioning for alternative classifications. Good, intermediate and poor prognosis groups were based on predicted 5-year survival. Classifications were further refined by subgrouping within the poor prognosis group. Performance was measured primarily by a bootstrap corrected c-statistic to indicate discriminative ability for future patients.

Results The weights of the risk factors in the alternative classifications differed slightly from the implicit weights in the IGCC classification. Discriminative ability, however, did not increase clearly (IGCC classification, c=0.732; Cox classification, c=0.730; Recursive partitioning classification, c=0.709). Three subgroups could be identified within the poor prognosis groups, resulting in classifications with five prognostic groups and slightly better discriminative ability (c=0.740).

Conclusion In conclusion, the IGCC classification in three prognostic groups is largely supported by Cox regression and recursive partitioning. Cox regression was the most promising tool to define a more refined classification.

2.1 Introduction

Testicular germ cell tumours (seminomatous and nonseminomatous) are the most common cancers among young adult men. Since the 1970s, long term cure rates of patients with germ cell tumours have increased to over 80%, because of the ability of cisplatin-based chemotherapy to cure advanced disease [1-4]. Owing to the high overall cure rate, interest has shifted from increasing the overall cure rate to reducing treatment-related toxicity for patients with a good prognosis [5]. On the other hand, high risk patients, eligible for more intensive treatment, for example, stem cell support or high-dose chemotherapy, should be identified [6,7].

Several classifications have been proposed in the past to distinguish patients according to prognosis, by identifying and combining the main prognostic factors for progression-free survival (PFS) and overall survival [8-10]. The coexistence of classifications differing in type, complexity and ability to separate good from poor prognosis complicated international collaboration in randomised trials and made comparison of nonrandomised studies impossible. International collaboration by the International Germ Cell Cancer Collaborative Group resulted in the development of the International Germ Cell Consensus Classification (IGCC classification), which is widely applied and easy to use and remember [11].

For the IGCC classification, readily available risk factors were selected from a wider set following Cox regression analyses, namely primary site, presence of nonpulmonary visceral metastases (NPVM) and elevation of the tumour markers alpha-fetoprotein (AFP), human chorionic gonadotrophin (HCG) and lactate dehydrogenase (LDH). All variables were categorical, since no major differences in performance were found compared to using continuous variables [12]. In Table 1, how the risk factors were combined into three prognostic groups for patients with nonseminomatous germ cell tumours (NSGCT) with either good, intermediate or poor prognosis are shown. The good prognosis group is characterised by the absence of adverse risk factors. The intermediate prognosis group is defined by the presence of any intermediate tumour marker, that is, one or more intermediate tumour markers are present. The poor prognosis group is

characterised by the presence of any of the poor risk factors mediastinal primary site, NPVM, AFP poor, HCG poor or LDH poor, that is, one or more poor risk factors are present. The classification can be seen as a max function where the good, intermediate and poor prognosis groups have a maximum score of zero, one or two, respectively.

In the IGCC classification, all intermediate tumour markers and all poor risk factors were required only to be sufficiently bad to be classified as intermediate and poor prognosis, respectively, that is, differences in importance between intermediate tumour markers and differences in importance between poor risk factors are not taken into account. Furthermore, no distinction is made between the number of intermediate tumour markers in the intermediate prognosis group and the number of poor risk factors in the poor prognosis group. Better discrimination might be achieved by incorporating differences in predictive strength and testing specific interaction terms.

Furthermore, it is difficult to adjust the current classification for changes in treatment strategy. A more flexible scoring system could more easily identify subgroups for the identification of very high risk patients eligible for novel chemotherapy approaches such as high-dose chemotherapy or the use of novel cytotoxic agents [6,13]. We however note that an important consideration in developing the IGCC classification was that all the prognostic groups should be large enough to make randomised trials of new treatments for each prognostic group feasible [11].

The aim of this study was to reconsider steps taken in the development of the IGCC classification, and to investigate alternative classifications based on Cox regression and recursive partitioning [14] that may discriminate better and be more suitable to identify more subgroups.

Table 1 The International Germ Cell Consensus (IGCC) Classification [15]

GOOD PROGNOSIS
NONSEMINOMA
Testis/retroperitoneal primary site = 0 And No nonpulmonary visceral metastases = 0 And AFP good = 0 and HCG = 0 good and LDH good = 0 Max = 0
INTERMEDIATE PROGNOSIS
NONSEMINOMA
Testis/retroperitoneal primary site = 0 And No nonpulmonary visceral metastases = 0 And AFP intermediate =1 or HCG intermediate =1 or LDH intermediate =1 Max = 1
POOR PROGNOSIS
NONSEMINOMA
Mediastinal primary site =2 Or Nonpulmonary visceral metastases =2 Or AFP poor = 2 or HCG poor =2 or LDH poor =2 Max = 2

Tumour markers alpha-fetoprotein (AFP)/ human chorionic gonadotrophin (HCG)/lactate dehydrogenase (LDH): Good - AFP < 1000 ng/ml, HCG < 5000 iu/l, LDH < 1.5 x upper limit of normal; Intermediate - AFP 1000 – 10000 ng/ml, HCG 5000 - 50000 iu/l, LDH 1.5 - 10 x N; Poor - AFP > 10000 ng/ml, HCG > 50000 iu/l, LDH > 10 x N

2.2 Materials and methods

Patients

Centres participating in the International Germ Cell Collaborative Group provided retrospective data of 5202 adult male patients with NSGCT. All patients were treated between 1975 and 1990 with cisplatin-based chemotherapy. Data were collected on age, primary site, date of diagnosis, levels of serum AFP, HCG and LDH, nodal disease in the abdomen, mediastinum, and neck, lung metastases, spread to other visceral sites like liver, bone and brain and on treatment details like previous therapy. For the development of the IGCC classification, patients without missing data on the risk factors primary site, NPVM, tumour markers AFP, HCG and LDH and the outcome survival were selected (n=3048) [11].

Outcome and IGCC risk factors

The outcome measures were PFS and overall survival from the start of the chemotherapy. The risk factors in the IGCC classification were primary site (testis/retroperitoneal vs. mediastinum), presence of NPVM (yes/no) and tumour markers AFP, HCG and LDH. Each tumour marker had three categories; good, intermediate and poor with specific cutoff points on the continuous tumour markers (see Table 1) [11]. The same risk factors and categories were used to construct the alternative classifications based on Cox regression and recursive partitioning.

Statistical analyses

The IGCC classification makes no clear distinction between the intermediate tumour markers and between the poor risk factors and is represented by a max score. One way to assess this assumption is by evaluating whether the weights in the IGCC classification were optimally allocated to the risk factors. We hereto varied the IGCC weights (1/2) over the levels of the risk factors and compared all possible combinations with respect to performance. Performance was quantified by the difference in minus twice the log likelihood (model χ^2) [16].

We used the Cox regression to study the univariable and multivariable effects of the IGCC risk factors on the overall survival, expressed as Hazard ratios and regression coefficients.

The Cox regression model formed the basis of classification '5R'. We multiplied the multivariate regression coefficients by 10 and rounded them to obtain weights. A sum score was calculated by multiplying the weights with individual patient characteristics and adding the resulting individual scores [17]. We calculated the estimated 5-year survival rate for each score.

The IGCC classification can be viewed as implying that the risk factors are strongly dependent, that is, that there are interactions between risk factors. There is, for example, no distinction made between patients with one poor risk factor or three poor risk factors. To test whether and which interactions were present, we added all two-way interactions between the IGCC risk factors in a Cox regression model. Important interactions were selected through stepwise backward selection ($P<0.05$). Since interactions based on small number of patients give unreliable regression coefficients, the interaction terms were defined as linear. The resulting model forms the basis of classification '5Ri'. A sum score based on a regression model with interactions is, however, more difficult to calculate and interpret. Therefore, a table was constructed with 5-year survival estimates for all possible combinations of the IGCC risk factors based on the Cox regression model with linear interactions. The number of patients on which each survival estimate was based is given to indicate the reliability of the survival estimates.

An alternative and visually more attractive way of exploring and presenting interactions between risk factors is by growing a tree through recursive partitioning [14,18,19] that we used to construct classification '5T'. A binary tree is built by the following process: first the risk factor that best splits the data into two groups, leading to the largest decrease in prediction error, is determined (recursive partitioning or splitting method). Splitting continues until the subgroups reach a minimum size or until no improvement can be made (stopping rule). The full tree, which is often too complex and overfit, is pruned using crossvalidation. All trees within one standard error of the lowest crossvalidated prediction error

are considered as equivalent. From these equivalent trees, the simplest is chosen as final tree [14].

As a splitting method, the exponential scaling method was used [20,21]. The splitting process stopped when a minimum of five patients per groups was reached or when there was no further decrease in prediction error. We used 10-fold crossvalidation to determine the optimal tree size. Modelling was performed with S-plus version 2000 using the RPART library that contains a recursive partitioning method for survival data.

The RPART library (rpart2.zip) and manual (rpart2doc.zip) can be found at http://www.stats.ox.ac.uk/pub/SWin.

Prognostic groups

In all classifications, three prognostic groups were identified using the estimated 5-year survival by sum score (classification 5R), combination of risk factors (5Ri) or binary tree (5T). Subgroups with a 5-year survival higher than 90% were considered as good prognosis, between 65 and 89% as intermediate prognosis, and lower than 65% as poor prognosis.

Furthermore for each classification, we explored the possibility of identifying more subgroups. For the IGCC classification, this was carried out by allowing weights to vary from zero to four (instead of zero to two), and comparing all possible combinations on performance. For classifications 5R, 5Ri and 5T, we changed the cutoff points on estimated 5-year survival. A 5-year survival rate higher than 90% was considered as good prognosis, 75-89% as intermediate prognosis, 60-74% as good-poor prognosis, 40-59% as intermediate-poor prognosis, and lower than 40% as poor-poor prognosis [13]. Survival of the five groups of the IGCC classification and classifications 5R, 5Ri and 5T was presented by Kaplan-Meier curves.

Performance

The classifications were evaluated by their ability to distinguish between patients differing in survival. An indication of the discriminative ability is the difference in 5-year survival rates between the good, intermediate and poor prognosis groups. A c-statistic was also calculated for both the three and five group classifications. For binary outcomes, the c-statistic is similar to the area under the ROC curve [22]. The c-statistic for survival data indicates the probability that for a randomly chosen pair of patients, the one having the higher predicted survival is the one who survives longer [22]. Overall performance of the three and five group classifications was measured by model χ^2. When a model is developed and evaluated on the same data, the performance of the model is usually too optimistic. The optimism can be quantified with statistical methods, known as internal validation techniques [23].To estimate and correct for the optimism in discriminative ability, the steps taken in the Cox regression and recursive partitioning were internally validated by taking random bootstrap samples (100) [24,25].

2.3 Results

The median follow up time of surviving patients was 50 months. Disease progression occurred in 680 patients, and 533 patients died. The 5-year PFS was 78% (95% CI 76-79%) and the 5-year overall survival 82% (95% CI 81-84%). Most patients had as primary site testis or retroperitoneum (97%), no NPVM (92%), and good AFP, HCG and LDH levels (84, 87 and 67%, respectively) (Table 2). All risk factors were predictors of survival as indicated by the Hazard ratios ranging from 2.1 to 6.2, where the tumour marker AFP was the weakest risk factor in the univariable analysis.

Alternative classifications

The regression-based weights of the risk factors in classification 5R, and the cutoff points on the resulting sum score are presented in Table 3, with the weights and cutoff points of the IGCC classification.

Table 2 Characteristics of 3048 NSGCT patients on the IGCC risk factors

IGCC risk factors	Number of patients (%)		5-year survival	95% CI	HR	95% CI
Primary site						
Testis/ Retroperitoneal	2947	(97)	84%	82-85%	1	-
Mediastinum	101	(3)	37%	27-47%	6.1	4.7-7.9
NPVM						
No	2808	(92)	85%	84-86%	1	-
Yes	240	(8)	49%	42-55%	4.6	3.8-5.6
AFP						
Good	2559	(84)	85%	84-87%	1	-
Intermediate	349	(12)	71%	66-76%	2.1	1.7-2.6
Poor	140	(5)	56%	47-65%	3.6	2.7-4.7
HCG						
Good	2656	(87)	86%	84-87%	1	-
Intermediate	238	(8)	65%	58-71%	3.0	2.3-3.8
Poor	154	(5)	48%	39-56%	5.0	3.9-6.4
LDH						
Good	2036	(67)	89%	88-91%	1	-
Intermediate	977	(32)	68%	65-71%	3.3	2.8-3.9
Poor	35	(1)	51%	34-67%	6.2	3.9-10.1
Total number subjects	3048	(100)	82%	81-84%	-	-

NPVM = nonpulmonary visceral metastases

The weights suggest that differences between risk factors were present. Tumour marker AFP had a much lower weight in the multivariate analysis than tumour markers HCG and LDH. As a result, a poor AFP level (score 3) is not sufficient to be classified as poor prognosis in classification 5R. Also, the combination of two or three intermediate tumour markers, which would lead to an intermediate prognosis in the IGCC classification, results in a score of over 10 and thus in classification in the poor prognosis group in classification 5R. The presence of risk factor NPVM (score 7) alone was not sufficient to be classified as poor prognosis, in contrast with the IGCC classification. Patients would only be classified as poor prognosis when other risk factors besides NPVM or AFP are present.

We identified four significant interactions in the Cox regression model; between AFP and primary site ($P<0.001$), AFP and NPVM ($P<0.01$), HCG and NPVM ($P<0.003$) and HCG and LDH ($P<0.01$). The regression coefficients all had negative signs, indicating that the effect of the risk factors together was smaller than the sum of their separate effects. For all 108 combinations of the IGCC risk factors, we present 5-year survival estimates from the Cox regression model with interactions (Appendix). Patients with testis as primary site and good or intermediate tumour markers had the highest estimated survival (55-92%).

Table 3 Weights, coding of variables, and cutoff on the max function of the IGCC classification and the sum score of the regression-based classification 5R

Classification		IGCC	5R
Risk factors in the model	Coding of risk factors	implicit weights	regression weights
Primary site	Testis/ retroperitoneal	0	0
	Mediastinum	2	15
NPVM	No	0	0
	Yes	2	7
AFP	Good	0	0
	Intermediate	1	2
	Poor	2	3
HCG	Good	0	0
	Intermediate	1	9
	Poor	2	11
LDH	Good	0	0
	Intermediate	1	7
	Poor	2	9
Cutoff points	Good	Max 0	Sum 0
	Intermediate	1	2-10
	Poor	≥2	≥11

NPVM = nonpulmonary visceral metastases

Patients with mediastinum as primary site and NPVM had the worst estimated survival (0-64%). Since the number of patients with more than one poor risk factor was limited, the survival estimates for these patients were less reliable. Recursive partitioning resulted in a tree with seven subgroups with 5-year survival ranging from 35 to 91% (Figure 1), forming the basis of classification 5T. Tumour marker LDH was the principal determinant of 5-year survival, making a split between good LDH (N=2036) and intermediate/poor LDH (N=1012). The majority of the 'good LDH' subgroup consists of patients with no risk factors (N=1865) with an observed 5-year survival of 91% (95% CI 90-93%). Furthermore, a subgroup of 29 patients with primary site mediastinum had a 5-year survival of 55% (95% CI 34-72%) and patients with intermediate or poor HCG (N=142) had a 5-year survival of 70% (95% CI 61-77%). Within the subgroup intermediate/poor LDH, four further subgroups were identified with the

risk factors NPVM, primary site and HCG, with 5-year survival ranging from 35 to 80%.

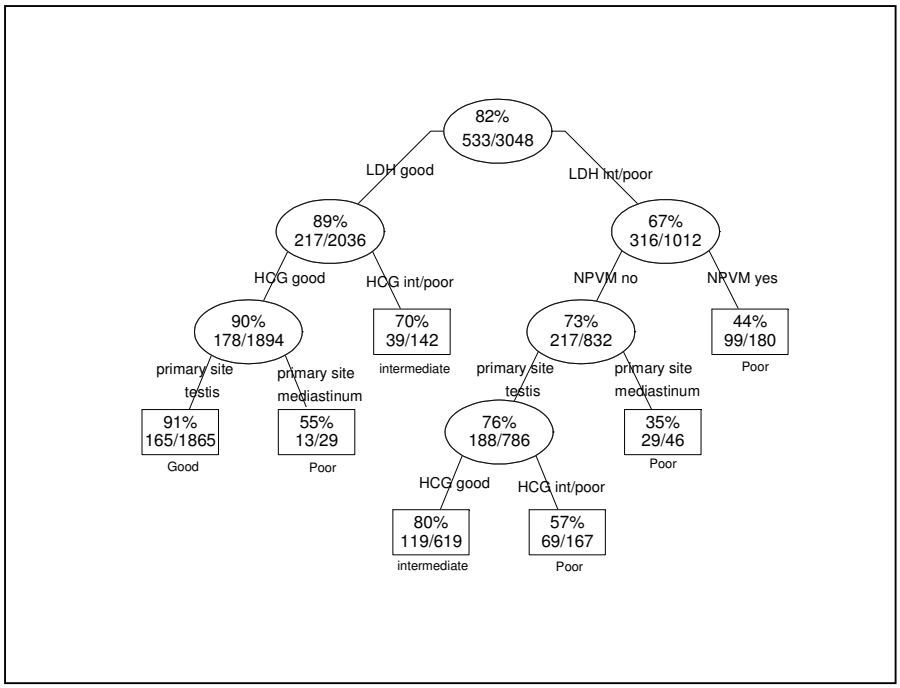


Figure 1 The final tree fitted by recursive partitioning, using the exponential scaling method. The 5-year survival rates, events and total number of observations per subgroup are given. The resulting subgroups are displayed in rectangulars and determine classification 5T

Performance

The 5-year survival rates for the good, intermediate and poor prognosis groups were comparable for the IGCC classification and classifications 5R, 5Ri and 5T (Table 4). The c-statistic of the IGCC classification was 0.732. The apparent c-statistics of classifications 5R, 5Ri and 5T were 0.732, 0.735 and 0.718, respectively. Validation showed minor optimism in the c-statistic in the Cox regression models (0.002). More optimism was present in the classification 5T, with the c-statistic decreasing from 0.718 to 0.709. Classification 5R did not show

an improvement in model χ^2 compared to the IGCC classification (model χ^2 402 and 401, respectively, 2 d.f.). Classifications 5Ri did show a statistically significant increase in overall performance over the IGCC classification (model χ^2 422, 2 d.f.). Classification 5T had a worse overall performance with a model χ^2 of 374 (2 d.f.).

Identification of more subgroups

Within the max score, different weights did not lead to an improvement in overall performance over the weights of the IGCC classification (model χ^2 402, 2 d.f.). The following weights were allocated to derive a max function with five prognostic groups in the IGCC classification with the score varying between 0 and 4; primary site mediastinum (4), NPVM (3), AFP good/intermediate/poor (0/1/2), HCG good/intermediate/poor (0/2/3) and LDH good/intermediate/poor (0/1/3). The 5-year survival varied from 37 to 92% for the five groups of the IGCC classification, from 34 to 92% for classification 5R, from 36 to 92% for classification 5Ri and from 35 to 91% for classification 5T (Table 5). The cutoff points on the sum score for the five groups of classification 5R are also given in Table 5. The difference in survival between the prognostic groups for each classification is illustrated in Figure 2. The c-statistic for the five groups of the IGCC classification and classifications 5R and 5Ri was slightly higher than for the three group classifications (0.739, 0.741 and 0.744, respectively) and with a small amount of optimism (0.002) for the Cox regression models. The increase of the c-statistic for the five groups of classification 5T was very limited (0.722) with an optimism of 0.011. The increase in model χ^2 was more substantial; 422 for the extended IGCC classification, 446 for classification 5R, 450 for classification 5Ri. The increase in model χ^2 for classification 5T (383) was less substantial.

Table 4 Survival of the IGCC classification, the regression-based classifications 5R and 5Ri and classification 5T based on recursive partitioning

	IGCC		5R		5Ri		5T	
Group	Surv	N	Surv	N	Surv	N	Surv	N
Good	92%	1691	92%	1691	92%	1691	91%	1865
Intermediate	81%	862	80%	872	80%	915	78%	761
Poor	50%	495	50%	485	47%	442	49%	422

Surv=5-year survival

Table 5 Survival of subgroups within the IGCC classification, the regression-based classifications 5R and 5Ri and classification 5T based on recursive partitioning

	IGCC		5R		5Ri		5T	
Group (Surv)	Surv	N	Surv	N	Surv	N	Surv	N
Good (≥ 90%)	92%	1691	92%	1691	92%	1691	91%	1865
Intermediate (75-89%)	82%	684	81%	824	82%	818	80%	619
Good-poor (60-74%)	72%	251	65%	225	63%	194	70%	142
Intermediate-poor (40-59%)	51%	321	48%	169	51%	188	51%	376
Poor-Poor (≤40%)	37%	101	34%	139	36%	157	35%	46

Surv=5-year survival
Cutoff points on sum score classification 5R: Good 0, Intermediate 2-9, Good-poor 10-16, Intermediate-poor 17-22, Poor-poor > 22

2.4 Discussion

The discriminative ability of classifications derived through Cox regression and recursive partitioning was in concordance with the IGCC classification and therefore supports the validity of the IGCC classification. We did, however, find that not all intermediate tumour markers and poor risk factors were equally important, and that taking these differences into account does affect the classification of patients. In Cox regression-based classifications, especially risk factors NPVM and AFP had less impact compared to the other risk factors. That AFP is of less importance than the other risk factors is confirmed by recursive partitioning where AFP was not selected in the final tree. Furthermore, not all risk factors had statistical interactions. In classifications 5Ri and 5T, only a limited number of interactions were included. Combining several risk factors led to differences in 5-year survival, that is, patients with one poor risk factor had a better chance of survival than patients with three risk factors. These deviations from the weights used by the IGCC classification did, however, not lead to improvements in discriminative ability, in contrast with what we expected. The use of Cox regression and recursive partitioning did allow for more flexible classifications with more subgroups, leading to a small improvement in discriminative ability and 5-year survival of 34% for the poorest risk patients.

It appears that the maximum discriminative ability might have been reached with the current IGCC risk factors and coding, making further improvement in discriminative ability difficult. The risk factors selected for the IGCC classification are in agreement with risk factors used in other studies on identifying good and poor prognosis patients with NSGCT [8,10]. Some other potentially useful risk factors include age, lung metastases and abdominal mass size. However, adding these three risk factors to the Cox model had no substantial effect on discriminative ability (c increased from 0.73 to 0.74). One could also consider using continuous codings of tumour markers, but this would lead to an undesirable increase in complexity and decrease in applicability.

The division into more prognostic groups is similar to another division by recursive partitioning of poor prognosis patients[13]. Kollmannsberger et al identified three prognosis groups: a good-poor, intermediate-poor and poor-poor risk group with 2-year survival rates of 84, 64 and 49%, respectively. These survival rates are higher than the survival rates of the good-poor, intermediate-poor and poor-poor risk groups identified in the IGCC dataset. This may be due to the difference in survival for the poor prognosis patients (72 vs 50%), and remains when the difference in follow up time is taken into account (2 vs 5 years). The data in Kolmannsberger et al (2000) are more recent and improvements in treatment may have led to the difference in survival.

The lack of improvement in discriminative ability in both the classifications with three and five groups might also be explained by the dominance of the good prognosis group, which has a similar survival for all classifications and contains more than half of all patients. We therefore examined whether discriminative ability increased within the poor prognosis group of each classification. Discriminative ability increased from 0.50 to 0.60, 0.63, 0.64 and 0.65 for the three poor prognosis groups of classifications 5T, IGCC, 5R and 5Ri, respectively. Hence, some improvement was noted within the IGCC poor prognosis group. Furthermore, even though the c-statistic is often used and easy to interpret, it is not suitable for detecting small differences in discriminative ability [25,26].

Although the use of Cox regression and recursive partitioning did not have a major effect on discriminative ability, they can still be useful tools in the construction of future prognostic classifications when other criteria are taken into account. One of the advantages of classifications such as the IGCC classification is its simplicity. Classification 5T is reasonably simple with only a few subgroups and the survival probability readily available. Classification 5R is slightly more complicated because of the sum score that has to be calculated. Finally, classification 5Ri is not so much complicated as visually unattractive. Furthermore, survival estimates for infrequent combinations of risk factors are not reliable and therefore provide little information on the prognosis of patients with these risk factors.

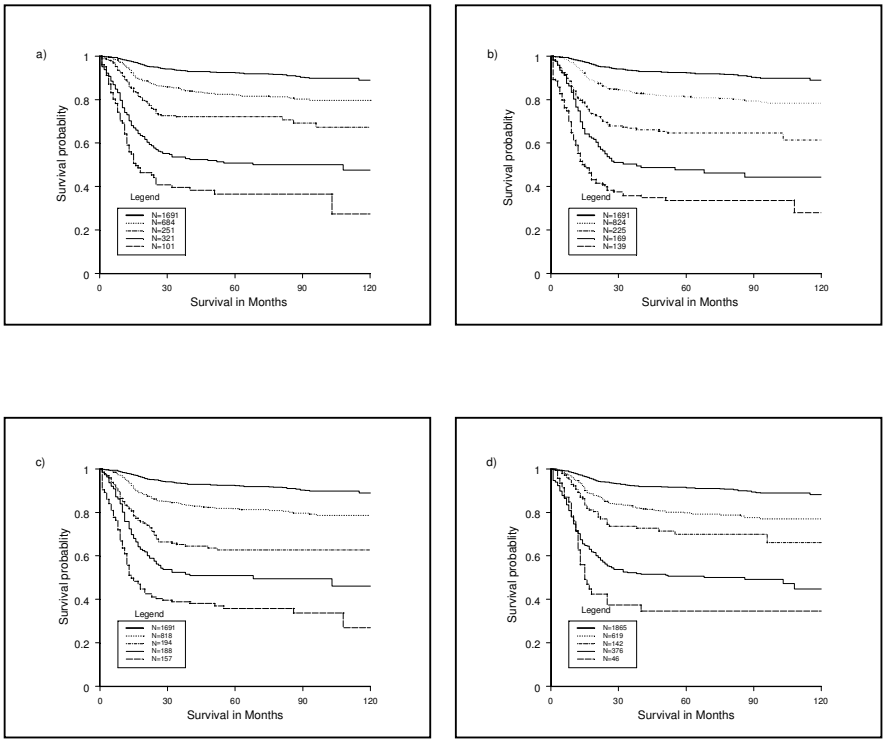


Figure 2 Survival curves for the 5 groups of the IGCC classification (a) and classifications 5R (b), 5Ri (c) and 5T (d)

A disadvantage of the IGCC classification is its inflexibility. More groups could be defined, but not in a straightforward manner. Classification 5R and classification 5Ri are very flexible with many possible cutoff points. Classification 5T is less flexible due to the limited number of subgroups, but flexibility could be increased by putting fewer restrictions on the recursive partitioning allowing for more subgroups to be identified.

The IGCC classification considered not just discrimination but also simplicity and the size of the resulting prognostic groups and was chosen by consensus from a

shortlist of possible models, which balanced these considerations. Consequently, in the IGCC classification there is a lack of transparency; it is unclear how the classification was constructed statistically because statistical considerations were not the only criteria used to derive the classification. Classification 5T shows very clearly how the subgroups were derived from the successive splits in the risk factors. Classification 5R shows the difference in importance between the risk factors and how the risk factors are combined in a sum score. Classification 5Ri could be presented in a similar way as classification 5R, but interpretation of the main and interaction effects is difficult.

The IGCC dataset suffers from a number of limitations. First, not all data were used for the multivariable regression models because of missing data. When patients with missing data differ from the other patients on prognosis, this causes a bias in the regression coefficients and the estimated 5-year survival rates [27-29]. Secondly, we could not internally validate the IGCC classification, because the exact steps taken in the modelling process (selection and categorisation of risk factors) were not defined. The IGCC classification was applied to a 30% validation set [11], and although the proportion of patients in each prognostic group was similar, the 5-year survival for poor prognosis patients was higher (57%). We did internally validate the modelling steps of the Cox regression models and found minor optimism in discriminative ability. Classification 5T, based on recursive partitioning, however, showed optimism in discriminative ability, as might be expected from a more data-driven method. This, in combination with the poorer performance, suggests that recursive partitioning is less suitable for the construction of prognostic classifications. It can be useful, however, for exploratory analyses in finding interactions between risk factors.

The survival estimates of the IGCC classification were also externally validated with more recent data from an MRC/EORTC trial (N=300). The 2-year PFS outcome largely corresponded with the IGCC estimates [11]. To gain further insight in the generalisability of the Cox regression models as well as the IGCC classification, further external validation is necessary, in larger recent datasets with longer follow up.

In conclusion, the IGCC classification appears to be a valid way to classify patients with NSGCT in three prognostic groups. Recursive partitioning is less suitable for the construction of prognostic classifications, because of its poorer performance. Although Cox regression did not lead to a clear improvement in performance, it gave a more flexible and transparent scoring system without much loss in simplicity. We therefore recommend the use of regression-based weights in the development of future prognostic classifications.

Acknowledgements

We would like to thank the members of the IGCCCG (Medical Research Council (MRC): GM Mead (Royal South Hants Hospital, Southampton, Hampshire, UK), P Cook (MRC Clinical Trials Unit, London), SD Fossa (Norwegian Radium Hospital, Montebello, Oslo, Norway), A Horwich, SB Kaye (Royal Marsden Hospital, Surrey, England), RTD Oliver (St Bartholomew's Hospital, London, England). European Organisation for the Research and Treatment of Cancer (EORTC): PHM de Mulder (Academic Hospital Nijmegen, The Netherlands), R de Wit & G Stoter (Rotterdam Cancer Institute, The Netherlands), RJ Sylvester (EORTC Data Center, Brussels Belgium). USA: DF Bajorin, GJ Bosl, M Mazumdar (Memorial Sloan-Kettering Cancer Center, New York). CR Nichols (Indiana University Hospital, Indianapolis). R Amato (University of Texas MD Anderson Cancer Center, Houston, Texas) Italy: G Pizzocaro (Istituto Nazionale per lo studio e la Cura dei Tumori, Milan). France: JP Droz (Centre Leon Benard, Lyon), A Kramar (CRLC Val d'Aurelle, Montpellier). Denmark: G Daugaard (Rigshospitalet, Copenhagen). Spain: H Cortes-Funes & L Paz-Ares (Hospital Doce de Octubre, Madrid). Australia: JA Levi (Royal North Shore Hospital, Sydney). New Zealand: BM Colls (Christchurch Hospital, Christchurch), VJ Harvey (Auckland Hospital, Auckland). Canada: C Coppin (Fraser Valley Cancer Centre, Vancouver, British Columbia) and their colleagues (see below) for kindly providing their data for the analyses in this paper.

MRC: N Aass (Norwegian Radium Hospital, Oslo, Montebello, Norway), PI Clark (Clatterbridge Hospital, Liverpool, England), MH Cullen (Queen Elizabeth Hospital, Birmingham, England), D Dearnaley (Royal Marsden Hospital, Surrey, England), SJ Harland (Middlesex Hospital, London, England), WG Jones (Yorkshire Centre for Cancer Treatment, Leeds, England (retired)), ES Newlands (Charing Cross Hospital, London, England), JT Roberts (Northern Centre for Cancer Treatment, Newcastle, England), GJS Rustin (Mt Vernon Hospital, Middlesex, England), P Wilkinson & G Read (Christie Hospital, Manchester, England), MV Williams (Addenbrookes Hospital, Cambridge, England). EORTC: D Sleijfer (Academic Hospital Groningen, The Netherlands), D WW ten Bokkel Huinink, Netherlands Cancer Institute, The Netherlands). USA: R Motzer (Memorial Sloan-Kettering Cancer Center, New York), L Finn (University of Texas MD Anderson Cancer Center). Italy: R Salvioni & L Mariani (Istituto Nazionale Tumori, Milan). Spain: P Lianes Barragan (Hospital doce de Octubre, Madrid). Denmark: S Werner Hansen (Rigshospitalet, Copenhagen). France: S Culine (Centre Paul Lamarque, Montpellier), J Bouzy (Institut Gustave Roussy, Villejuif). Australia: D Thompson (Princess Alexandra Hospital, Brisbane), T Sandeman (Peter MacCallum Institute, Melbourne), PG Gill (Royal Adelaide Hospital, Adelaide), M Tattersall, Royal Prince Alfred Hospital, Sydney), M Byrne, (Sir Charles Gairnder Hospital, Perth). New Zealand: JD Perez (Dunedin Hospital, Dunedin), P Thompson & M Bennet, (Auckland Hospital, Auckland). Canada: N Murray (Vancouver Cancer Centre, Vancouver, British Columbia)

References

1. Bosl GJ, Motzer RJ. Testicular germ-cell cancer. *N Engl J Med* 1997;337(4):242-53.

2. Hartmann JT, Kanz L, Bokemeyer C. Diagnosis and treatment of patients with testicular germ cell cancer. *Drugs* 1999;58(2):257-81.

3. Sonneveld DJ, Hoekstra HJ, van der Graaf WT, et al. Improved long term survival of patients with metastatic nonseminomatous testicular germ cell carcinoma in relation to prognostic classification systems during the cisplatin era. *Cancer* 2001;91(7):1304-15.

4. Steele GS, Richie JP, Stewart AK, Menck HR. The National Cancer Data Base report on patterns of care for testicular carcinoma, 1985-1996. *Cancer* 1999;86(10):2171-83.

5. de Wit R, Roberts JT, Wilkinson PM, et al. Equivalence of three or four cycles of bleomycin, etoposide, and cisplatin chemotherapy and of a 3- or 5-day schedule in good-prognosis germ cell cancer: a randomized study of the European Organization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group and the Medical Research Council. *J Clin Oncol* 2001;19(6):1629-40.

6. Bokemeyer C, Kollmannsberger C, Meisner C, et al. First-line high-dose chemotherapy compared with standard-dose PEB/VIP chemotherapy in patients with advanced germ cell tumors: A multivariate and matched-pair analysis. *J Clin Oncol* 1999;17(11):3450-6.

7. Bokemeyer C, Oechsle K, Hartmann JT, et al. Treatment-induced anaemia and its potential clinical impact in patients receiving sequential high dose chemotherapy for metastatic testicular cancer. *Br J Cancer* 2002;87(10):1066-71.

8. Bajorin DF, Geller NL, Bosl GJ. Assessment of risk in metastatic testis carcinoma: impact on treatment. *Urol Int* 1991;46(3):298-303.

9. Bajorin D, Katz A, Chan E, Geller N, Vogelzang N, Bosl GJ. Comparison of criteria for assigning germ cell tumor patients to "good risk" and "poor risk" studies. *J Clin Oncol* 1988;6(5):786-92.

10. Mead GM, Stenning SP, Parkinson MC, et al. The Second Medical Research Council study of prognostic factors in nonseminomatous germ cell tumors. Medical Research Council Testicular Tumour Working Party. *J Clin Oncol* 1992;10(1):85-94.

11. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor- based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *J Clin Oncol* 1997;15(2):594-603.

12. McCaffrey JA, Bajorin DF, Motzer RJ. Risk assessment for metastatic testis cancer. *Urol Clin North Am* 1998;25(3):389-95.

13. Kollmannsberger C, Nichols C, Meisner C, Mayer F, Kanz L, Bokemeyer C. Identification of prognostic subgroups among patients with metastatic 'IGCCCG poor-

prognosis' germ-cell cancer: an explorative analysis using cart modeling. *Ann Oncol* 2000;11(9):1115-20.

14. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Belmont, California: Wadsworth, 1984.

15. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *J Clin Oncol* 1997;15(2):594-603.

16. Clayton D, Hills M. Statistical Models in Epidemiology: Oxford University Press, 1993: 242-243.

17. Assmann G, Cullen P, Schulte H. Simple scoring scheme for calculating the risk of acute coronary events based on the 10-year follow-up of the prospective cardiovascular Munster (PROCAM) study. *Circulation* 2002;105(3):310-5.

18. Ahn H, Loh WY. Tree-structured proportional hazards regression modeling. *Biometrics* 1994;50(2):471-85.

19. Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 1989;8(5):539-50.

20. Therneau T, Grambsch P, Fleming T. Martingale based residuals for survival models. *Biometrika* 1990;77:147-160.

21. LeBlanc M, Crowley J. Relative risk trees for censored survival data. *Biometrics* 1992;48(2):411-25.

22. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;3(2):143-52.

23. Steyerberg EW, Harrell FE, Jr., Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54(8):774-81.

24. Efron B, Tibshirani RJ. An introduction to the bootstrap. London: Chapman & Hall Inc., 1993.

25. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15(4):361-87.

26. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 2000;19(8):1059-79.

27. Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. *J Clin Epidemiol* 2003;56(1):28-37.

28. Little R. Regression With Missing X's: A Review. *Journal of the Amserican Statistical Association* 1992;87(420):1227-1237.
29. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;18(6):681-94.

Appendix

5-year survival estimates and number of patients are given for all 108 combinations of the IGCC risk factors based on a Cox regression model of the IGCC risk factors and interactions AFP and primary site, AFP and NPVM, HCG and NPVM, and HCG and LDH

			Primary site Testis				Primary site Mediastinum			
			NPVM		NPVM		NPVM		NPVM	
AFP	HCG	LDH	No		Yes		No		Yes	
			Surv	N	Surv	N	Surv	N	Surv	N
Good	Good	Good	92%	1691	79%	27	53%	14	18%	1
		Inter	83%	459	60%	31	25%	12	2%	10
		Poor	73%	11	43%	3	10%	0	0%	1
	Inter	Good	77%	81	54%	9	15%	3	1%	0
		Inter	66%	62	38%	16	5%	1	0%	1
		Poor	60%	2	30%	0	2%	0	0%	0
	Poor	Good	64%	16	39%	8	4%	0	0%	0
		Inter	59%	56	32%	38	2%	1	0%	2
		Poor	61%	0	35%	3	3%	0	0%	0
Inter	Good	Good	88%	121	79%	5	65%	8	44%	1
		Inter	76%	104	60%	18	39%	14	17%	6
		Poor	64%	0	43%	1	21%	0	5%	0
	Inter	Good	69%	16	54%	1	28%	0	12%	0
		Inter	55%	19	37%	9	13%	0	3%	0
		Poor	48%	1	30%	3	8%	0	2%	0
	Poor	Good	52%	2	38%	1	11%	0	4%	0
		Inter	46%	13	32%	3	7%	0	2%	0
		Poor	49%	3	35%	0	9%	0	3%	0
Poor	Good	Good	81%	16	76%	5	71%	4	64%	1
		Inter	63%	43	55%	24	48%	17	38%	3
		Poor	47%	2	37%	3	30%	0	20%	0
	Inter	Good	54%	4	49%	0	37%	0	32%	0
		Inter	37%	10	31%	0	20%	0	16%	0
		Poor	29%	0	21%	0	14%	0	10%	0
	Poor	Good	33%	0	33%	1	17%	0	17%	0
		Inter	27%	1	26%	3	12%	1	12%	0
		Poor	30%	0	29%	2	15%	0	14%	0

Surv= 5-year survival
N = number of patients
Inter = Intermediate
Classification into 3 groups; Good prognosis 5-year survival >90%, Intermediate prognosis 5-year survival 65-89%, Poor prognosis 5 year survival <65%.
Classification into 5 groups; Good prognosis 5-year survival >90%, Intermediate prognosis 5-year survival 75-89%, Good-poor prognosis 5 year survival 60-74%, intermediate-poor prognosis 5 year survival 40-59%, Poor-poor prognosis 5 year survival <40%

3

Survival Estimates of a Prognostic Classification Depended more on Treatment than on Imputation of Missing Values

M. R. van Dijk, E. W. Steyerberg, S. P. Stenning, and J. D. F. Habbema
J Clin Epidemiol 2006;59(3):246-53

Abstract

Background The International Germ Cell Consensus (IGCC) classification defines good, intermediate, and poor prognosis groups among patients with nonseminomatous germ cell cancer. In the database used to develop the IGCC classification (n = 5202), >40% of patients were excluded because of missing values (n = 2154). We looked for effects of this exclusion on survival estimates in the three IGCC prognosis groups.

Methods We imputed missing values using a multiple imputation procedure. The IGCC classification was applied to patients with complete data (n = 3048) and with imputed data (n = 2154), and 5-year survival was calculated for each prognosis group.

Results Patients with missing values had a lower 5-year survival than those without missing values: 76 vs. 82%. Five-year survival in the complete and imputed data samples was 92 and 87% for the good prognosis groups and 80 and 70% for the intermediate prognosis groups, whereas 5-year survival for the poor prognosis groups in both samples was similar (50 and 47%, respectively). This difference in survival was largely explained by a higher proportion of missing values among patients treated before 1985, who had a worse survival than patients treated after 1985.

Conclusion Multiple imputation of the missing values led to lower survival estimates across the IGCC prognosis groups, compared with estimates based on the complete data. Although imputation of missing values gives statistically better survival estimates, adjustments for year of treatment are necessary to make the estimates applicable to currently diagnosed patients with testicular cancer.

3.1 Introduction

Testicular germ cell tumours (seminomatous and nonseminomatous) are the most common cancers among young adult men. Since the 1970s, long term cure rates of patients with germ cell tumours have increased to >80%, because of success of cisplatin-based chemotherapy in curing advanced disease [1-4]. Toward defining prognosis groups, the International Germ Cell Cancer Collaborative Group (IGCCCG) combined data from 5202 patients with nonseminomatous germ cell tumours (NSGCT). This resulted in the International Germ Cell Consensus (IGCC) classification, which identifies three prognosis groups - good, intermediate, and poor - based on five easily measured risk factors: primary site, presence of nonpulmonary visceral metastases (NPVM), and levels of the tumour markers alpha-fetoprotein (AFP), human chorionic gonadotropin (HCG), and lactate dehydrogenase (LDH) [5,6]. In the published report, 56, 28, and 16% of patients were allocated to the good, intermediate, and poor prognosis groups, with a 5-year survival of 92, 80, and 48%, respectively [5]. The IGCC classification is widely accepted and easy to use and remember.

Because only patients with complete data were considered, however, the data for 2154 patients were excluded, and the IGCC classification was based on roughly 3000 patients, not the full dataset of ~5000 patients. Wherever the patients with missing values differ substantially from patients without missing values, this could have led to bias in the survival estimates of the prognostic groups in the IGCC classification. Furthermore, exclusion of patients because of missing data is statistically inefficient [7]. It would therefore be preferable to estimate the missing values and use all available data instead of excluding patients with missing values.

The IGCC classification was published in 1997 [5]. Since then, application of methods for handling missing data is becoming more standard and software is more readily available. Multiple imputation (MI) is considered a sound statistical methodology for handling complex missing data problems [7-9]; its use is not widespread, however, and the implications of using MI are still unclear. We studied the effects of MI on survival estimates in the three IGCC prognosis groups. We also tried to explain possible differences between patients with and without missing values, and the implications for the clinical applicability of the IGCC classification.

3.2 Patients and methods

Patients

Centres participating in the International Germ Cell Collaborative Group provided retrospective data of 5202 adult male patients with NSGCT. All patients were treated between 1975 and 1990 with cisplatin-based chemotherapy. Data were collected on age, histology, primary site, date of diagnosis, treatment centre, levels of tumour markers AFP, HCG, and LDH, the presence and the size of nodal disease in the abdomen, mediastinum, and neck, the presence, number and size of lung metastases, spread to other visceral sites such as liver, bone and brain, previous radiotherapy, and prior node dissection [5].

Outcome and risk factors

The outcome measures were overall survival and progression-free survival from the start of chemotherapy, and response to treatment (complete, incomplete, or not assessable). We focus on overall survival in our analysis. The risk factors in the IGCC classification were primary site (testis or retroperitoneal vs. mediastinum), presence of NPVM (yes or no), and tumour markers AFP, HCG, and LDH. Each tumour marker was categorized as good, intermediate, or poor, with specific cutoff points for each marker based on a previous study on the prognostic value of tumour markers (see footnotes to Table 2) [5,10].

Some variables considered potentially related to the presence of missing values were not included in the IGCC classification: age, histology, abdominal mass, mediastinal mass, and presence of metastases in neck and lung. Furthermore, treatment variables not considered risk factors could also be related to missing values: year of treatment, region of treatment centre, previous radiotherapy (yes or no), and prior node dissection (yes or no) [5,7]. Centres were classified into four regions; North America (the United States and Canada), the United Kingdom, Europe, and Oceania (Australia and New Zealand).

Multiple imputation

Imputation is the replacement of missing data with statistical estimates. The goal of imputing missing values is to produce a complete dataset that can then be analysed using statistical methods for complete data. Examples of imputation techniques include mean imputation, regression imputation, and hot deck [11,12].

A disadvantage of these imputation methods is that they do not take the uncertainty of the imputation process into account. Imputation methods should include a random component to reflect the fact that imputed values are estimated. This will be especially relevant when the amount of missing data is substantial.

In multiple imputation (MI), each missing value is imputed several (M) times, with the imputed values drawn from their predictive distribution. The variation among the M imputations reflects the uncertainty with which the missing values can be predicted from the observed ones. MI results in M complete datasets, which can be analysed with standard complete data methods. The results are then combined to produce overall estimates and standard errors that reflect missing data uncertainty [7,9,13,14]. MI assumes that the missing data are missing at random (MAR). This means that the probability of a missing value depends only on other observed variables. This assumption implies that the missingness of a variable does not depend on unobserved variables, nor on the true unobserved value of the variable itself—in which case the missing values would be missing not at random (MNAR) [15].

We applied the multiple imputation procedure according to Van Buuren et al [7]. This is a semiparametric approach in which each variable has a separate imputation model, with a set of predictors explaining missingness and the form (e.g., predictive mean matching, logistic) of the imputation model depending on the type of variable (e.g., continuous, binary). To estimate values for the missing data, the Van Buuren approach does not explicitly assume a particular form for the multivariate distribution. It does assume that a multivariate distribution exists and that draws can be generated from it by Gibbs sampling of the conditional distributions, based on the imputation models. Gibbs sampling is a Monte Carlo technique to simulate drawings from a multivariate probability density distribution by repeatedly drawing from conditional probability density distributions [16]. This is an iterative process in which the missing values of a variable are estimated using its imputation model, and the completed variables are used to estimate the missing values for the other variables - also called regression switching. Each iteration ends when all variables have been updated [7,13].

We followed three main steps to investigate the missing data and apply the multiple imputation procedure: (a) investigate missing data; (b) specify the imputation models by risk factor; and (c) generate sets of imputed values to obtain the desired number of completed datasets.

The first step, investigation of missing data, was conducted as follows:
1. Quantify the multivariate patterns of the missing data - how many patients are missing in each IGCC risk factor.
2. Check whether there is a difference in survival between patients with and without missing values.
3. Investigate for differences in measuring practice over time and region - the proportion of missing data for each risk factor against year of treatment for each region.
4. Explore the relationship between missing values of the IGCC risk factors and other risk factors and treatment variables using correlation coefficients.

Correlation coefficients were obtained by univariate logistic regression taking the square root of the explained variance (R^2).
In the second step, we specified an imputation model for each IGCC risk factor, according to the guidelines in Van Buuren et al [7]. All five risk factors from the IGCC classification and the outcome variables were included. We included those treatment variables related to missingness. We also included risk factors related to the value of the variable to be imputed. The form of the imputation model was chosen depending on the type of variable. For the continuous variables (age, AFP, HCG, LDH), a predictive mean matching model was used as the imputation model. This imputes a missing value by selecting at random, with replacement, a value from those individuals who have matching observed values for other variables. We used a logistic regression model for the binary variables (primary site, NPVM, abdominal mass, mediastinal mass, presence of metastases in neck and lung, prior radiotherapy, prior node dissection) and a polytomous regression model for the categorical variables (histology and response to treatment). The log-transformed tumour markers AFP, HCG, and LDH were imputed as continuous variables, and then categorized according to the cut-points specified in the IGCC classification.
Finally, we used the imputation model to generate 10 sets of imputed values for the missing data, which resulted in 10 versions of completed datasets. The Gibbs sampling algorithm was run for 20 iterations, with updating for each of the 10 sets of imputed values. Simulation studies have shown that as few as 10 iterations are usually sufficient to obtain convergence [7].
When 50% of the data are missing, an estimate based on M = 5 imputations has a standard deviation that is only 5% wider than an estimate based on an infinite number of imputations. Therefore, 5–10 imputations are usually sufficient [14].

Statistical Analyses

Patient characteristics on the IGCC risk factors and the IGCC classification are given for the samples with complete data (n = 3048) imputed data (n = 2154), and all data (n = 5202). We report the imputed data separately, to evaluate the effects of imputation without dilution in all data. The frequencies of the IGCC risk factors in imputed and all-data samples were obtained by averaging the frequencies over the 10 datasets [14]. In each dataset, the difference in frequencies between the complete and the imputed data was tested using a chi-square statistic. We averaged the 10 chi-square statistics to determine the overall difference in frequencies. The number of patients and 5-year overall survival was studied for each prognosis group and compared in the complete, imputed, and all-data samples. Five-year overall survival was calculated by the Kaplan–Meier method in each dataset and then averaged over the 10 datasets. Differences in survival between the complete and imputed data in each dataset were tested using a log-rank test. We averaged the log-rank statistics of each of the 10 datasets to determine the overall difference in survival between complete and imputed data.

The IGCC classification was evaluated in the complete, imputed, and all-data samples on its ability to distinguish between patients differing in survival. Discriminative ability was indicated by a c-statistic. For binary outcomes, the c-statistic is similar to the area under the receiver operating characteristic (ROC) curve. The c-statistic for survival data indicates the probability that, for a randomly chosen pair of patients, the one having the higher predicted survival is the one who survives longer [17]. The c-statistic was determined for the complete data and in each imputed and all-data dataset. We took the average of the c-statistics in the 10 datasets for the imputed data and for all data. The variance of this overall c-statistic is the average variance over the 10 datasets plus the variance between datasets. We used this overall variance to determine the 95% confidence interval CI of the c-statistics of the imputed and all-data samples.

To better compare the value of the c-statistic in the imputed data with the c-statistic in the complete data, the expected value of the c-statistic in the imputed data was calculated. This c-statistic was obtained by combining the average proportion of patients in the prognosis groups in the 10 imputed datasets with the survival estimates of the prognosis groups in the complete data in a simulated sample of 10000 patients drawn with replacement.

Finally, we did a pooled Cox regression analysis on the 10 datasets to explain possible differences between patients with and without missing values [13]. Hazard

ratios were determined for each dataset and averaged over the 10 datasets. The variance of the overall hazard ratio is the average variance over the 10 datasets plus the variance between datasets [13].
All statistical analyses were done in S-Plus 2000 (Mathsoft, Inc, Seattle, WA) using the Hmisc, Design, and MICE libraries. MICE is available from www.multiple-imputation.com.

3.3 Results

Follow up information was available on all 5202 patients. The median follow up time of surviving patients was 5 years, and 90% had been followed for at least 2 years from the start of chemotherapy. Disease progression occurred in 1313 patients, and 1056 patients died. Five-year progression-free survival was 75% (95% CI = 74–76%) and 5-year overall survival was 80% (95% CI = 79–81%).

Missing data

Of the IGCC risk factors, LDH had the most missing values: 1945 (37%). The number of missing values in the other IGCC risk factors was limited; 53 (1%) in primary site, 185 (4%) in NPVM, 114 (2%) in AFP, and 91 (2%) in HCG (Table 1). These 2388 missing values represent only 9% of all 26010 possible data values (5 risk factors × 5202 patients), but resulted in the exclusion of 2154 (41%) of 5202 patients in the development of the IGCC classification. Of the 2154 patients, 1979 (92%) had only one missing value, 124 (6%) had two missing values, 44 (2%) had three missing values, 6 (<1%) had four missing values, and 1 (<1%) patient had all five IGCC risk factors missing. Patients with missing values (n = 2154) had a lower 5-year overall survival than patients without missing values (n = 3048); 76% (95% CI = 74–78%) vs. 82% (95% CI = 81–84%) (Figure 1, P < .001).
There were differences in measuring practice of tumour marker LDH over time and across treatment centre regions (Figure 2). The proportion of missing LDH values in North American centres was small and constant over time (average 8%).

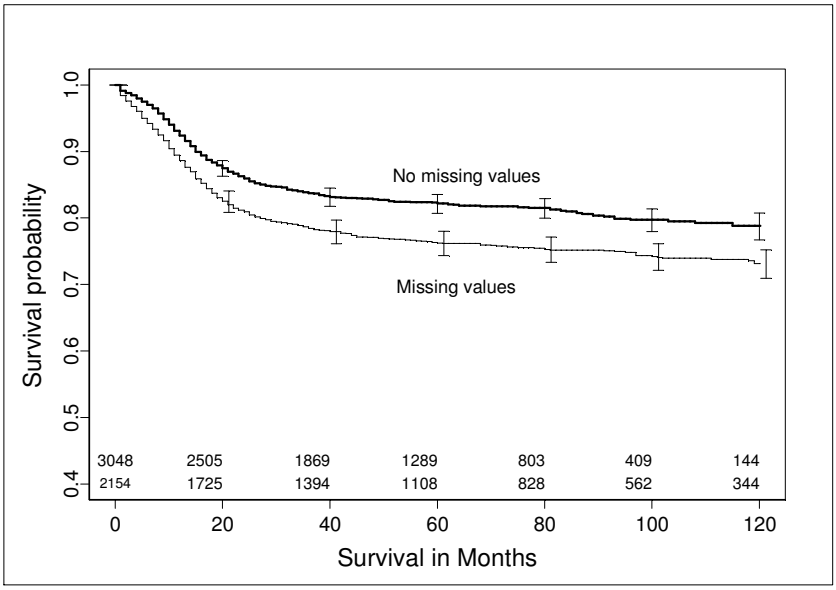


Figure 1 Survival and number of patients at risk for patients with and without missing values for the risk factors of the IGCC classification. Error bars indicate 95% confidence interval

In European centres, initially half of the LDH values were missing, but after 1982 the proportion of missing values decreased to 10%; on average, 14% of LDH values in European centres were missing. Most missing LDH values were in centres from the United Kingdom (73%) and Oceania (87%). The proportion of missing LDH values decreased over time in the United Kingdom, but it stayed more or less constant in Oceania. The other IGCC risk factors showed no differences in measuring practice over time or across regions.

In Table 1, correlation coefficients are shown between missingness and the value of LDH on the one hand and several risk factors, outcome, and treatment variables on the other hand.

Year of treatment and region of treatment centre were most strongly related to the presence of missing LDH values, with correlation coefficients of 0.29 and 0.61, respectively. Except for the risk factors age and prior radiotherapy, all variables were significantly related to the value of LDH. All 18 variables were included in the imputation model. This imputation model was used for all variables with missing values.

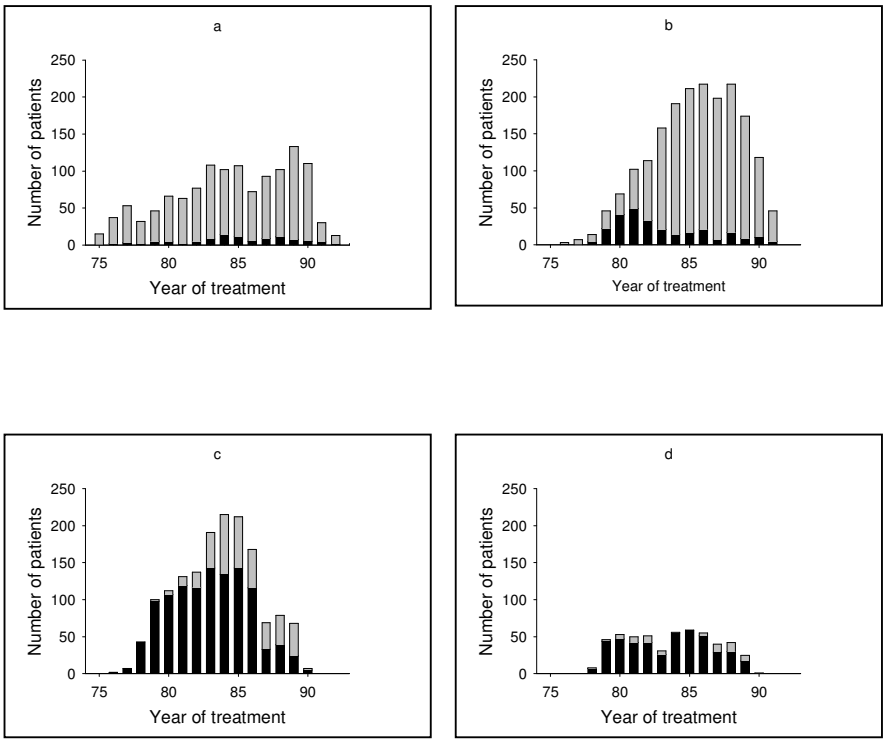


Figure 2 Number of missing and measured values of tumour marker LDH per year per region. Black bars represent the number of patients with LDH missing; grey bars, the number of patients with LDH not missing. By region, the number of patients with (without) LDH values are: (a) North America, n = 1163 (n = 96); (b) Europe, n = 1615 (n = 270); (c) United Kingdom, n = 411 (n = 1130); and (d) Oceania, n = 68 (n = 449)

Table 1 Variables considered for imputation with number of missing values and their correlation with log of the LDH value and with missingness of LDH

Variables	Missing values, no. (%)		r LDH_{value}	r $LDH_{missing}$
IGCC risk factors				
Primary site	53	(1)	.08**	.06**
NPVM	185	(4)	.29**	.03
AFP	114	(2)	.19**	.04
HCG	91	(2)	.25**	.04
LDH	1945	(37)	-	-
Outcome variables				
Survival time in months	0	(0)	.13**	.14**
Dead	0	(0)	.25**	.05**
Response to treatment	392	(8)	.23**	.08**
Other risk factors				
Age	62	(1)	.00	.04**
Histology	54	(1)	.05*	.07**
Abdominal mass	142	(3)	.18**	.02
Mediastinal mass	111	(2)	.21**	.06**
Neck mass	180	(4)	.25**	.02
Lung metastases	111	(2)	.19**	.00
Treatment variables				
Year of treatment	0	(0)	.05*	.29**
Prior radiotherapy	1097	(21)	.01	.13**
Prior node dissection	1740	(33)	.13**	.12**
Region[1]	0	(0)	.10**	.61**

AFP, alpha-fetoprotein; HCG, human chorionic gonadotropin; LDH, lactate dehydrogenase; NPVM, nonpulmonary visceral metastases. Tumour markers AFP, HCG, LDH according to IGCC classification.

* P < .05;

** P < .01.

[a] Regions were North America (United States and Canada), Europe, United Kingdom, and Oceania (Australia and New Zealand)

Distribution of IGCC risk factors and the IGCC classification

We completed 10 datasets, imputing 2388 values in each dataset. As a result, 8382 real data values could be added into each dataset (2154 patients with missing values × 5 risk factors - 2388 = 8382). In Table 2, the distribution of patients over the IGCC risk factors in the complete data (n = 3048), imputed data (n = 2154) and all-data (n = 5202) samples is given. There were significant differences between complete data and imputed data for the risk factors primary site, NPVM, HCG, and LDH — although apart from LDH the differences were relatively small. More patients in the imputed than the complete data samples had good LDH values (80 and 67%,

respectively), fewer patients had intermediate LDH values (20 and 32%, respectively), and the proportion of patients in the poor category (1%) was similar (P < .001). As a consequence, a higher proportion of patients were allocated to the good prognosis group according to the IGCC classification in the imputed data sample, compared with complete data (59 and 55%, respectively). The number of patients allocated to each category varied little over the 10 imputed datasets for the variables primary site, NPVM, AFP, and HCG because these variables had a limited number of missing values.

Table 2 IGCC risk factors and IGCC risk groups in the complete, imputed, and in all data

	Patients, no. (%)			
IGCC risk factors	Complete data	Imputed data	P-value[1]	Patients, all data, no. (%)
Sample size	n = 3048	n = 2154		n = 5202
Primary site			p<.001	
Testis	2947 (97)	2117 (98)		5064 (97)
Mediastinum	101 (3)	37 (2)		138 (3)
NPVM			p<.01	
No	2808 (92)	1934 (90)		4742 (91)
Yes	240 (8)	220 (10)		460 (9)
AFP[2]			p=.10	
Good	2559 (84)	1802 (84)		4361 (84)
Intermediate	349 (11)	269 (12)		618 (12)
Poor	140 (5)	83 (4)		223 (4)
HCG[3]			p<.001	
Good	2660 (87)	1850 (86)		4510 (87)
Intermediate	238 (8)	144 (7)		382 (7)
Poor	150 (5)	160 (7)		310 (6)
LDH[4]			p<.001	
Good	2036 (67)	1816 (84)		3852 (74)
Intermediate	977 (32)	324 (15)		1301 (25)
Poor	35 (1)	14 (1)		49 (1)
IGCC			p<.001	
Good	1691 (55)	1392 (65)		3083 (59)
Intermediate	863 (28)	365 (17)		1228 (24)
Poor	494 (16)	398 (18)		892 (17)

AFP, alpha-fetoprotein; HCG, human chorionic gonadotropin; LDH, lactate dehydrogenase; NPVM, nonpulmonary visceral metastases.

[1] P-value is based on average chi-square statistic over 10 datasets.

[2] AFP: good, <1000 ng/mL; intermediate, 1000–10000 ng/mL; poor, >10000 ng/mL.

[3] HCG: good, <5000 IU/L; intermediate, 5000–50000 IU/L; poor, >50000 IU/L.

[4] LDH: good, <1.5 × N; intermediate, 1.5–10 × N; poor, >10 × N (where N is the upper limit of normal)

By contrast, because a large number of values had to be imputed for LDH, the number of patients varied widely over LDH categories over the 10 imputed datasets: In the imputed data sample, minimum and maximum number of patients were 1683 and 1794, 348 and 457, and 10 and 19 for good, intermediate, and poor LDH, respectively.

Five-year survival was significantly lower in the good, intermediate, and poor prognosis group in the imputed (87, 70, and 47%, respectively) data sample, compared with complete data (92, 80, and 50%, respectively) (Table 3). Because multiple imputation allowed for a more efficient use of the available data, the confidence intervals of the survival estimates in the all-data sample were smaller, compared with complete data. Discriminative ability was lower in the imputed data than in the complete data sample, with c-statistics of 0.68 and 0.73, respectively.

This difference was not attributable to a difference in distribution of patients across the prognosis groups, because, given the distribution of IGCC prognosis groups in the imputed data, the expected c-statistic was 0.74.

Table 3 Five-year survival by IGCC prognosis group and performance for complete, imputed, and all data

	5-year survival, % (95% CI)				Performance, c-statistic (95% CI)
	Good	Intermediate	Poor	Total	
Complete data	92 (91-94)	80 (77-83)	50 (45-55)	82 (81-84)	0.73 (0.69-0.77)
Imputed data	87 (85-89)	70 (66-75)	47 (42-52)	76 (74-78)	0.68 (0.64-0.73)
Log-rank[1]	p<.025	p<.001	p<.025	p<.001	
All data	90 (89-91)	77 (74-79)	49 (46-52)	80 (79-81)	0.71 (0.67-0.74)

[1] Significance of log-rank for comparison of complete with imputed data based on average log-rank statistic over 10 datasets

The difference in survival between patients with and without missing values could in part be explained by differences in year of treatment. Survival increased over time, and the differences in year of treatment therefore resulted in differences in survival. With adjustment for year of treatment and IGCC prognosis group, the hazard ratio of missing IGCC (yes or no) decreased from 1.4 to 1.2 ($P < .01$). Further adjustment with other variables, such as region of treatment centre, did not further decrease the hazard ratio.

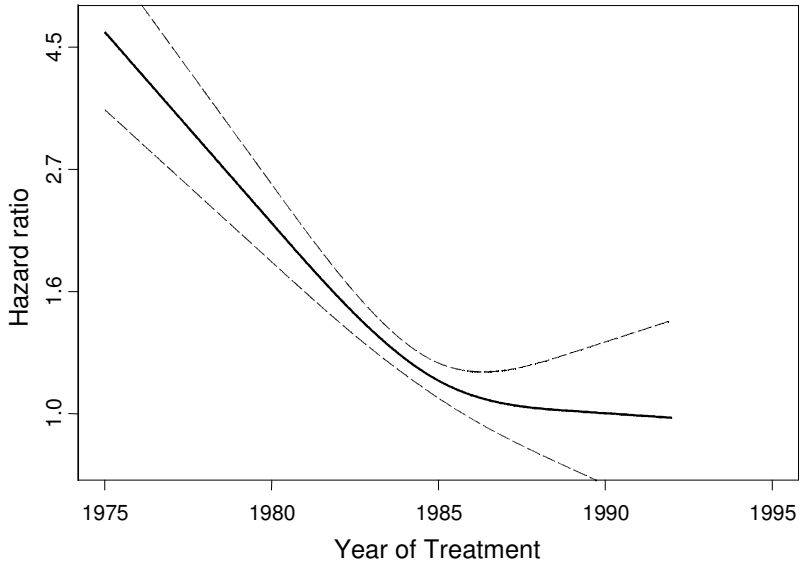


Figure 3 Hazard ratio for patients with nonseminomatous germ cell cancer by year of treatment. Dashed lines indicate the 95% confidence interval

In Figure 3, the hazard ratio of year of treatment is plotted using a restricted cubic spline function with three knots against year of treatment for all patients [18]. The relative hazard decreased strongly until 1985; thereafter, survival remained relatively constant. This relation also held when the complete data were analysed. No significant differences were seen according to region of treatment centre. The dependency on year of treatment limits the applicability of the obtained survival estimates for the good, intermediate, and poor prognosis group, especially after imputation, to more recently diagnosed patients. We therefore estimated 5-year survival for good, intermediate, and poor prognosis patients adjusted to the year 1990 for all data and the complete data. For the all-data sample, the adjusted 5-year survival estimates for the good, intermediate, and poor prognosis groups were 93, 85, and 62%, respectively. For complete data, 5-year survival estimates were very similar: 94, 86, and 62%, respectively.

When we selected only those patients treated after 1985, 5-year survival estimates for the good, intermediate, and poor prognosis groups for the all-data sample were 93, 83, and 57%, respectively, and 95, 84, and 59% for patients with complete data.

3.4 Discussion

We found lower survival estimates for patients with nonseminomatous germ cell cancer after multiple imputation of the missing values in the IGCC database, compared with patients with complete data, especially for those in the intermediate prognosis group. The differences in survival between patients with and without missing values were, however, due mainly to differences in year of treatment, because year of treatment was related to both survival and missingness.

In the analysis of the IGCC data, 2154 of 5202 patients (41%) were excluded because of one or more missing values. Missing values were due mainly to the absence of the tumour marker LDH, which occurred 1945 times. We used multiple imputation to create 10 complete datasets; by imputing the 2388 missing values in each dataset, 8382 real data values could be added. We used an extensive imputation model, including 18 variables significantly related to either the missingness or the value of LDH, to impute the missing values of the IGCC risk factors. We included the observed outcome (survival status and survival time), in line with the approach described by van Buuren et al [7].

In general, using all available information results in multiple imputations that have minimal bias [13]. The multiple imputation procedure assumes that values are missing at random: that is, that missing values are related only to observed variables and not to unobserved variables or the true unobserved value of the risk factor itself (which is missing not at random, MNAR). This seems a plausible assumption for the missing values of tumour marker LDH. LDH was established as an important risk factor only in the late 1980s, and was therefore not collected systematically by all centres. It remains possible, however, that missing values are related to unobserved values. Including many predictors in the multiple imputation model reduces the need to make special adjustments for missing-not-at-random problems [13]. The multiple imputation of missing values leads to a more efficient use of the available data. If the imputation models specified and the assumption regarding the nonresponse mechanism are valid, multiple imputation leads to statistically better estimates.

Did the multiple imputation procedure also improve the clinical applicability of the survival estimates to currently diagnosed patients? In the IGCC database, survival increased over time, probably due to improvements in treatment regimens, but year of treatment was neither included in the IGCC classification nor otherwise accounted for. Therefore, the survival estimates of the IGCC classification are too low for currently diagnosed patients, even in the complete data sample of 3048 patients. Because the probability of missingness was related to survival and year of treatment, MI led to the inclusion of even more historical patients and hence to a further underestimation of the survival of the IGCC classification. This underestimation limits the possibility to evaluate the results of currently conducted observational studies reporting on survival of patients with nonseminomatous germ cell cancer. For randomised controlled trials, this is not very problematic, because the effect of a treatment is directly compared with patients receiving another treatment; however, the results of nonrandomised phase II trials on dose-intensive or high-dose chemotherapy for poor prognosis patients might be interpreted too optimistically when compared with the 5-year survival estimate of 50% for poor-prognosis patients in the IGCC classification [19-23]. The same holds for outcome research, in which case series are reported to evaluate the treatment and outcome in one centre [24,25].

To increase the clinical applicability, we adjusted survival for year of treatment, resulting in 5-year survival estimates of 93, 85, and 62% for good, intermediate, and poor prognosis patients, respectively. The adjusted estimates were similar for the complete and for all data and in line with survival estimates from more recent studies [24,25]. This further demonstrates that, irrespective of the occurrence of missing values, year of treatment should have been taken into account in the development of the IGCC classification in order to obtain survival estimates applicable for currently treated patients. The validity of these adjusted estimates for currently diagnosed patients depends on the assumption that survival has not increased significantly over the last 15 years. Otherwise, survival estimates of the IGCC classification should be further updated.

In conclusion, our results show that when missing values are related to time of treatment and survival has increased over time, multiple imputation of missing values does not result in survival estimates that are better applicable to currently diagnosed patients. The omission of important predictive variables in the development of a prognostic classification cannot be compensated by multiple imputation of missing values.

Acknowledgements

We would like to thank the members of the International Germ Cell Cancer Collaborative Group (IGCCCG), for kindly providing their data for the analyses in this paper. Members of the IGCCCG and participating centres were listed previously [5]. This work was supported by the Netherlands Organization for Scientific Research (M.v.D.) and by the Royal Netherlands Academy of Arts and Sciences (E.S.). We also would like to thank the anonymous referees for their helpful suggestions.

References

1. Bosl GJ, Motzer RJ. Testicular germ-cell cancer. N Engl J Med 1997;337(4):242-53.
2. Hartmann JT, Kanz L, Bokemeyer C. Diagnosis and treatment of patients with testicular germ cell cancer. Drugs 1999;58(2):257-81.
3. Steele GS, Richie JP, Stewart AK, Menck HR. The National Cancer Data Base report on patterns of care for testicular carcinoma, 1985-1996. Cancer 1999;86(10):2171-83.
4. Sonneveld DJ, Hoekstra HJ, van der Graaf WT, et al. Improved long term survival of patients with metastatic nonseminomatous testicular germ cell carcinoma in relation to prognostic classification systems during the cisplatin era. Cancer 2001;91(7):1304-15.
5. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor- based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. J Clin Oncol 1997;15(2):594-603.
6. van Dijk MR, Steyerberg EW, Stenning SP, Dusseldorp E, Habbema JD. Survival of patients with nonseminomatous germ cell cancer: a review of the IGCC classification by Cox regression and recursive partitioning. Br J Cancer 2004;90(6):1176-83.
7. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. Stat Med 1999;18(6):681-94.
8. Rubin DB, Schenker N. Multiple imputation in health-care databases: an overview and some applications. Stat Med 1991;10(4):585-98.
9. Schafer JL. Multiple imputation: a primer. Stat Methods Med Res 1999;8(1):3-15.
10. McCaffrey JA, Bajorin DF, Motzer RJ. Risk assessment for metastatic testis cancer. Urol Clin North Am 1998;25(3):389-95.
11. Briggs A, Clark T, Wolstenholme J, Clarke P. Missing... presumed at random: cost-analysis of incomplete data. Health Econ 2003;12(5):377-92.
12. Perez A, Dennis RJ, Gil JF, Rondon MA, Lopez A. Use of the mean, hot deck and multiple imputation techniques to predict outcome in intensive care unit patients in Colombia. Stat Med 2002;21(24):3885-96.
13. Clark TG, Altman DG. Developing a prognostic model in the presence of missing data: an ovarian cancer case study. J Clin Epidemiol 2003;56(1):28-37.
14. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley, 1987.
15. Little R, Rubin DB. Statistical Analysis with Missing Data. Second Edition. New York: John Wiley and Sons, 2002.
16. Gilks WR, Richardson S, Spiegelhalter DJ. Markov chain Monte Carlo in practice. London: Chapman & Hall, 1996.

17. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. Stat Med 1984;3(2):143-52.
18. Harrell FE, Jr., Lee KL, Matchar DB, Reichert TA. Regression models for prognostic prediction: advantages, problems, and suggested solutions. Cancer Treat Rep 1985;69(10):1071-77.
19. Germa-Lluch JR, del Muro XG, Tabernero JM, et al. BOMP EPI intensive alternating chemotherapy for IGCCC poor-prognosis germ-cell tumors: The Spanish Germ-Cell Cancer Group experience (GG). Annals of Oncology 1999;10(3):289-293.
20. Decatris MP, Wilkinson PM, Welch RS, Metzner M, Morgenstern GR, Dougall M. High-dose chemotherapy and autologous haematopoietic support in poor risk non-seminomatous germ-cell tumours: An effective first-line therapy with minimal toxicity. Annals of Oncology 2000;11(4):427-434.
21. Tscherry G, Jacky E, Jost LM, Stahel RA. Risk-adapted chemotherapy of germ cell tumors with carboplatin, etoposide and bleomycin for low-risk and cisplatin, etoposide and ifosfamide for high-risk patients - A single-center study. Oncology 2000;59(2):110-117.
22. Bokemeyer C, Schleucher N, Metzner B, et al. First-line sequential high-dose VIP chemotherapy with autologous transplantation for patients with primary mediastinal nonseminomatous germ cell tumours: a prospective trial. British Journal of Cancer 2003;89(1):29-35.
23. Anthoney DA, McKean MJ, Roberts JT, et al. Bleomycin, vincristine, cisplatin/bleomycin, etoposide, cisplatin chemotherapy: an alternating, dose intense regimen producing promising results in untreated patients with intermediate or poor prognosis malignant germ-cell tumours. British Journal of Cancer 2004;90(3):601-606.
24. Sonneveld DJA, Hoekstra HJ, van der Graaf WT, et al. Improved long term survival of patients with metastatic nonseminomatous testicular germ cell carcinoma in relation to prognostic classification systems during the cisplatin era. Cancer 2001;91(7):1304-1315.
25. Germa-Lluch JR, Garcia del Muro X, Maroto P, et al. Clinical pattern and therapeutic results achieved in 1490 patients with germ-cell tumours of the testis: the experience of the Spanish Germ-Cell Cancer Group (GG). Eur Urol 2002;42(6):553-62; discussion 562-3.

4

Survival of Nonseminomatous Germ Cell Cancer Patients according to the IGCC Classification: An update based on Meta-analysis

M.R. van Dijk, E. W. Steyerberg, and J. D. F. Habbema

Eur J Cancer 2006;42(7):820-6

Abstract

Background The International Germ Cell Consensus (IGCC) Classification distinguishes patients with nonseminomatous germ cell tumours (NSGCT) with a good, intermediate or poor prognosis, with a reported 5-year overall survival of 92, 80 and 48%, respectively. Since the IGCC classification was based on patients treated between 1975 and 1990, we aimed to investigate whether survival has improved for more recently treated patients.

Methods We did a systematic search of the literature and included studies on survival of patients with NSGCT, treated after 1989 and classified according to the IGCC classification. Survival estimates of selected studies were pooled using meta-analytic techniques.

Results We included 10 papers, describing 1775 patients with NSGCT with good (n = 1087), intermediate (n = 232), or poor (n = 456) prognosis. Pooled 5-year survival estimates were 94, 83 and 71%, respectively.

Conclusion Since the publication of the IGCC classification, there was a small increase in survival for good and intermediate prognosis patients, and a large increase in survival for patients with a poor prognosis. This increase is most likely due to both more effective treatment strategies and more experience in treating NSGCT patients.

4.1 Introduction

Testicular germ cell tumours (seminomatous and nonseminomatous) are the most common cancers among young adult men. Since the 1970s the overall long term cure rate of patients with germ cell tumours has increased to over 80%, mainly due to the use of cisplatin-based chemotherapy that can cure advanced disease [1-4].

Due to the high overall cure rate, interest has shifted to reducing treatment related toxicity for patients with a good prognosis [5]. However, a small group of patients with a poor prognosis remains who might profit from alternative, more intensive treatment strategies [6-8].

The coexistence of classifications differing in type, complexity and ability to separate good from poor prognosis, has complicated international collaboration in randomised trials and making comparison of nonrandomised studies impossible. Therefore, the International Germ Cell Consensus (IGCC) Classification was developed, based on individual patient data from previously conducted studies, which has now been widely adopted [9]. The IGCC classification identifies good, intermediate, and poor prognosis groups using the following variables: primary site, presence of nonpulmonary visceral metastases (NPVM) and levels of the tumour markers alpha-fetoprotein (AFP), human chorionic gonadotrophin (HCG), and lactate dehydrogenase (LDH) (Table 1). According to the IGCC classification 56, 28 and 16% of patients with non-seminomatous germ cell cancer were allocated to the good, intermediate and poor prognosis groups, with a 5-year overall survival of 92, 80 and 48%, respectively [9]. The IGCC classification is currently the standard for selecting patients for randomised controlled trials (RCTs) [5,10,11] and for the analysis of observational studies, such as Phase I/II trials and hospital series. In observational studies, direct comparison with patients receiving standard treatment is not possible and, therefore, results are usually compared with the survival estimates reported by the IGCC classification.

The IGCC classification is based on patients treated between 1975 and 1990. Survival might be better for more recently treated patients with non-seminomatous germ cell tumours (NSGCT). There are some indications that indeed survival has improved. Sonneveld and colleagues reported a 10-year disease specific survival of 94, 87 and 66% for good, intermediate and poor prognosis patients treated between 1987 and 1996 [3].

Table 1 The International Germ Cell Consensus (IGCC) Classification [9]

GOOD PROGNOSIS
NONSEMINOMA
Testis/retroperitoneal primary site And No nonpulmonary visceral metastases And AFP good and HCG good and LDH good 56% of nonseminomas **5 year PFS 89%** **5 year OS 92%**
INTERMEDIATE PROGNOSIS
NONSEMINOMA
Testis/retroperitoneal primary site And No nonpulmonary visceral metastases And AFP intermediate or HCG intermediate or LDH intermediate 28% of nonseminomas **5 year PFS 75%** **5 year OS 80%**
POOR PROGNOSIS
NONSEMINOMA
Mediastinal primary site Or Nonpulmonary visceral metastases Or AFP poor or HCG poor or LDH poor 16% of nonseminomas **5 year PFS 41%** **5 year OS 48%**

Tumour markers alpha-fetoprotein (AFP)/ human chorionic gonadotrophin (HCG)/lactate dehydrogenase (LDH): Good - AFP < 1000 ng/ml, HCG < 5000 iu/l, LDH < 1.5 x upper limit of normal; Intermediate - AFP 1000 – 10000 ng/ml, HCG 5000 - 50000 ng/ml, LDH 1.5 - 10 x N; Poor - AFP > 10000 ng/ml, HCG > 50000 iu/l, LDH > 10 x N

Germa-Llurch and colleagues reported a 3 year overall survival of 97, 89 and 72% for patients treated between 1994 and 2001 [12].
However, all relevant studies should be considered for a valid update of the survival estimates of the IGCC classification. We therefore conducted a meta-analysis to update the survival estimates of the prognosis groups in the IGCC classification, using the most recent studies reporting on the survival of non-seminomatous germ cell cancer patients.

4.2 Material and Methods

Literature search

We conducted a systematic literature search of the PubMed database and the Cochrane Central library for the period January 1997 to December 2004. The terms used for the search were the text words "igcc" (OR "igccc", "igcccg", "IGCC", "IGCCC", "IGCCCG"), "germ cell" (OR "germ-cell"), "cancer" or "tumor" (OR "tumour ", "tumors", tumours") and the MeSH terms "neoplasms, germ cell and embryonal" and "testicular neoplasms". The search was limited to "human", and the English language. Furthermore, we used the ISI Web of Science to screen all papers referring to the original report by the IGCCCG and we screened all papers of the EORTC Genito-Urinary Tract Group on testicular cancer published since 1997. This resulted in 350 papers.
Figure 1 shows the selection of papers for the meta-analysis. Of the 350 papers, 284 papers reporting on children, other cancers, testicular cancer other than NSGCT (seminoma, relapse, residual mass, early disease), and the histology, biology, diagnosis, or treatment of testicular cancer, as well as review articles and double papers were excluded. The remaining 66 papers were retrieved for further evaluation. Forty-five papers were excluded because no survival estimates were given, patients were not classified according to the IGCC classification, overlapping datasets were described, no distinction between seminoma and non-seminoma was made, less than 20 patients were included, or papers were published in a journal without an impact factor. When studies reported on the same patient population, the most recent study or the one with the longest follow up time was chosen. The 21 remaining papers mostly reported on previously published trials, in which patients were retrospectively allocated to the IGCC prognosis groups. Papers describing patients treated before 1989 (n = 11) were excluded due to possible overlap with the IGCCCG database resulting in 10

studies eligible for inclusion in the meta-analysis. One of these studies made no distinction in survival between seminomatous and nonseminomatous patients [5]. Since this was the largest study on good prognosis patients, we requested the original data so that the study could be included in the meta-analysis.

Data extraction and Analysis

The following characteristics of the included 10 reports were extracted: treatment period, number of patients, median age, median follow up and overall or disease-free survival. Furthermore, we noted the design of the study (Phase I, II or III trial, or hospital registry), IGCC prognosis group (good, intermediate, poor), number of patients per IGCC prognosis group and survival per IGCC prognosis group. Survival per prognosis group was obtained from tables or Kaplan - Meier plots.

Since studies differed in years of follow up for which survival was given, survival estimates could not be obtained directly [13]. We therefore pooled the survival curves of the 10 selected studies. For each curve the survival probability at each year of follow up was calculated by measuring the height of the curve at each year of follow up. Since information on censoring was not given in all studies, survival estimates could not be pooled using the exact number of patients at risk and the exact number of deaths per year of follow up. We therefore used a life table approach in which for each study the number of patients at risk and the number of deaths per year were based on the total number of patients in the study and the survival probability per year of follow up [13]. Pooled survival estimates for each year were obtained by pooling the estimated number of patients at risk and the estimated number of deaths per year of follow up from each study. In two studies no survival curves were available [5,6]. In these cases the reported 2-year overall survival was used to estimate the number of patients at risk and the number of events at one and two years of follow up.

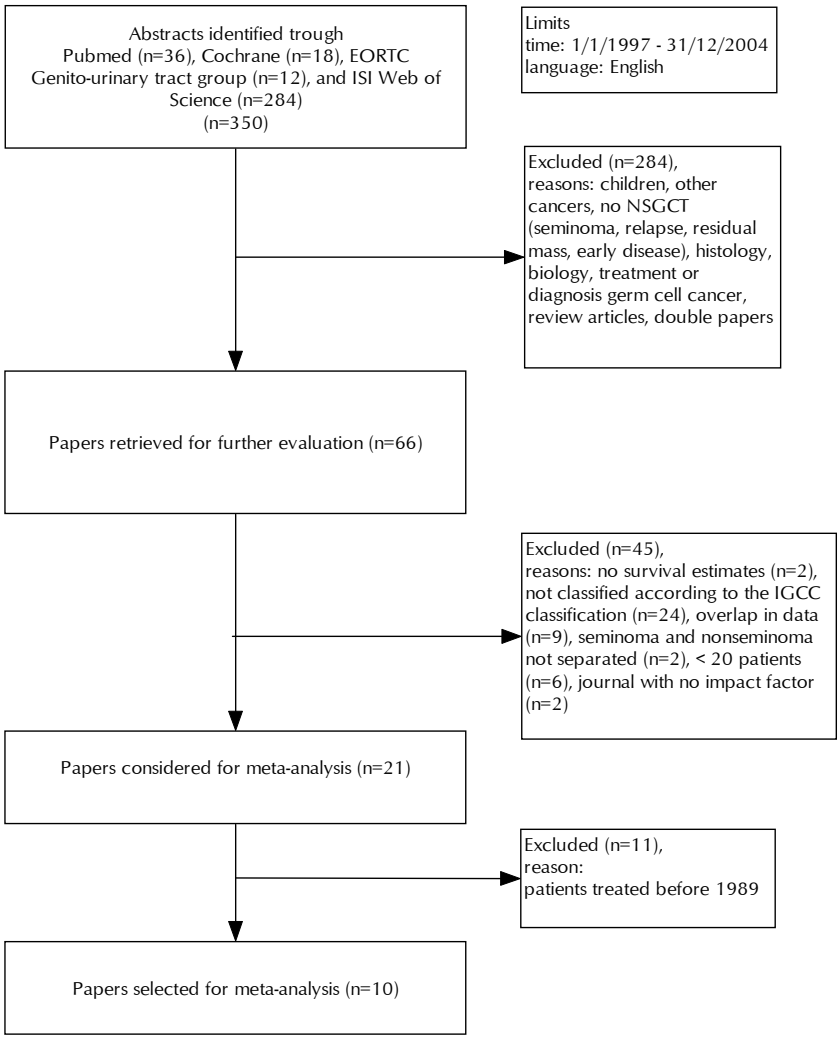


Figure 1 Flow chart showing the selection process of publications for meta-analysis. Note: NSGCT, nonseminomatous germ cell tumours

4.3 Results

The 10 included studies reported data on 1775 patients with nonseminomatous germ cell cancer, treated between 1989 and 2001 (Table 2). The median age varied from 26 to 32 years, and median follow up time from 22 to 63 months. Eight studies reported overall survival as primary outcome, and two studies reported disease-specific survival [14,15].

Table 3 shows study design, treatment regimen, number of patients per IGCC prognosis group and survival per IGCC prognosis group per study. Of the 1775 patients available for analysis, 1087 (61%) were classified as good prognosis, 232 (13%) as intermediate prognosis and 456 (26%) as poor prognosis patients. Three studies were based on hospital registries in which patients were treated with either standard-dose treatment cisplatin-etoposide-bleomycin (BEP), dose intensive regimens [12,16] or high-dose chemotherapy with stem cell support [15]. Muramaki and colleagues describe the treatment results of 8 good prognosis, 24 intermediate prognosis and 14 poor prognosis patients treated with either standard-dose chemotherapy (n = 29), or high-dose chemotherapy (n = 17) [15]. No details were given on treatment received within the IGCC prognosis groups. Five-year disease-specific survival was 88% for good prognosis, 77% for intermediate prognosis, and 71% for poor prognosis patients. Of the 178 patients described by Bhala and colleagues, 125 received standard-dose chemotherapy of whom 108 were good prognosis, 9 intermediate prognosis and 8 poor prognosis patients. Forty-eight patients received dose intense alternating chemotherapy, POMB-ACE, of which 10 were good prognosis, 15 intermediate prognosis and 23 poor prognosis patients [16]. Five-year survival was 95% for good prognosis patients, 82% for intermediate prognosis, and 57% for poor prognosis patients. Germa-Llurch and colleagues reported on 523 patients, of whom 485 patients were treated with standard-dose chemotherapy, and 38 of 96 poor prognosis patients with modified standard treatment BOMP-EPI [12,17]. Three-year overall survival was 97% for good prognosis, 89% for intermediate prognosis, and 72% for poor prognosis patients. Six Phase I or II trials evaluated new treatment regimens against standard treatment.

Table 2 Characteristics of 10 studies included in meta-analysis

No. (Refs)	Author (Year publication)	Study period	N	Median age (yr)	Median Follow up (months)	Primary outcome
1 [15]	Muramaki et al (2004)	1990 - 2001	46	31	63	5-yr DSS
2 [18]	Anthoney et al (2004)	1995 - 1999	43	27	59	3-yr OS
3 [16]	Bhala et al (2004)	1989 - 2001	178	29	22	5-yr OS
4 [14]	Schmoll et al (2003)	1993 - 1999	182	29	47	5-yr DSS
5 [19]	Christian et al (2003)	1989 - 2000	54	27	48	3-yr OS
6 [20]	Fizazi et al (2002)	1993 - 1998	57	28	31	3-yr OS
7 [12]	Germa-Lluch et al (2002)	1994 - 2001	523	26	33	3-yr OS
8 [5]	De Wit et al[1] (2001)	1995 - 1998	630	32	25	2-yr OS
9 [21]	Decatris et al (2000)	1994 - 1999	20	28	27	4-yr OS
10 [6]	Bokemeyer et al[2] (1998)	1990 - 1995	42	-	31	2-yr OS
		Total	1775			

OS, overall survival; DSS, disease specific survival
N = Number of patients
[1] only nonseminoma included
[2] intermediate prognosis patients only, poor prognosis patients included in study 4

Three Phase I/II trials evaluated the effect of dose intensive alternating chemotherapy (BOP + BEP), intensive induction chemotherapy (CBOP + BEP), or dose dense alternating chemotherapy (BOP + CISCA + BOMP + ACE) [18-20]. These studies give survival estimates for 154 intermediate and poor prognosis patients ranging from 67 to 88%.

Three studies evaluated the effect of high-dose chemotherapy with peripheral blood stem cell support [6,14,21]. Survival for 202 poor prognosis patients was 66 and 73%, respectively [14,21]. The 42 intermediate prognosis patients receiving high-dose chemotherapy had a 2-year overall survival of 89% [6]. Only one study was a RCT [5]. It showed the equivalence of 3 versus 4 cycles of bleomycin, etoposide and cisplatin for good prognosis patients, resulting in a 2-year overall survival of 97%.

In Figure 2 survival estimates up to 7 years follow up are given for patients with good, intermediate, and poor prognosis separately. At 5 years follow up, pooled survival estimates were 94, 83 and 71% respectively for patients with good, intermediate, and poor prognosis.

Table 3 Design, treatment, IGCC prognosis group, number of patients and survival per IGCC prognosis group of 10 included studies

No. (Refs)	Design	Treatment	IGCC	Number of patients	OS (95% CI)
1 [15]	Hospital registry	BEP or HD-CT	Good	8	88%
			Intermediate	24	77%
			Poor	14	71%
2 [18]	Phase II	BOP+BEP	Intermediate	24	79% (57-91%)
			Poor	19	84% (59-95%)
3 [16]	Hospital Registry	BEP or POMB+ACE	Good	120	95% (91-100%)
			Intermediate	25	82% (65-98%)
			Poor	33	57% (36-79%)
4 [14]	Phase I/II	HD-VIP + PBSC	Poor	182	73%
5 [19]	Phase II	CBOP+BEP	Poor	54	88% (71-95%)
6 [20]	Phase II	BOP+CISCA + BOMP+ACE	Intermediate	19	83% (67-100%)
			Poor	38	67% (53-84%)
7 [12]	Hospital registry	BEP or BOMP+EPI	Good	329	97% (95-99%)
			Intermediate	98	89% (81-96%)
			Poor	96	72% (62-82%)
8 [5]	Phase III	BEP vs BEP + EP	Good	630	97%
9 [21]	Phase II	BEP+HD-CEC+ PBSC	Poor	20	66%
10 [6]	Phase II	HD-VIP + PBSC	Intermediate	42	89%
		Total	Good	1087	94%
			Intermediate	232	83%
			Poor	456	71%

Treatment details: BOP+BEP=bleomycin, vincristine, cisplatin + bleomycin, etoposide, cisplatin; BEP=bleomycin, etoposide, cisplatin; POMB+ACE=cisplatin, vincristine, methotrexate, bleomycin + actinomycin D, cyclophosphamide, etoposide; HD-VIP + PBSC= high-dose cisplatin, etoposide, ifosfamide + peripheral blood stem cell support; CBOP+BEP= carboplatin, bleomycin, vincristine, cisplatin + bleomycin, etoposide, cisplatin; BOP+CISCA+BOMP+ACE= bleomycin, vincristine, cisplatin + cisplatin, cyclophosphamide, doxorubicin + cisplatin, vincristine, methorexate + bleomycin, etoposide, dactinomycin, cyclophosphamide; BOMP+EPI=bleomycin, vincristine, methotrexate, cisplatin + etoposide, ifosfamide, cisplatin; BEP + EP=bleomycin, etoposide, cisplatin + etoposide, cisplatin; HD-CEC = high-dose carboplatin, etoposide, cyclophosphamide

4.4 Discussion

This meta-analysis demonstrated that survival of recently treated poor prognosis patients with nonseminomatous germ cell tumours was better than the survival reported by the IGCC classification. There was only a small increase in survival for patients with a good or intermediate prognosis.

Based on a systematic review, we included 10 studies reporting on 1775 patients of which 1087 had a good prognosis, 232 intermediate and 456 poor.

Five-year survival of poor prognosis patients has increased strongly (71 vs. 48%). This may partly be due to promising results obtained in Phase I/II trials with new treatments, although these results still need to be confirmed in RCTs. However, survival for poor prognosis patients may have increased over time irrespective of these more recent treatment regimens.

First, because standard treatment improved when etoposide replaced vinblastine, after a RCT demonstrated that etoposide was more active and less toxic than vinblastine [3,22]. Second, because research centres gained experience with the treatment of advanced testicular cancer [23]. Furthermore improvements in second-line treatment may also have contributed to an increase in survival, especially for poor prognosis patients [24]. This is supported by two studies, excluded from the meta-analysis because of possible overlap with the IGCCG data. These two studies included patients who were treated with standard-dose chemotherapy and who were selected from hospital registries [3,8]. Hinton and colleagues [8] reported 5-year overall survival of 60% for poor prognosis patients treated with either standard-dose BEP or VIP, and Sonneveld and colleagues reported a 10-year disease-specific survival of 66% for poor prognosis patients treated with standard-dose BEP [3,8]. Furthermore, survival for poor prognosis patients treated after 1985 was 59% in a retrospective analysis of the IGCCCG data [25].

We found only a small increase in 5-year survival for good and intermediate prognosis patients compared to survival estimates reported by the IGCC classification (94 vs. 92% and 83 vs. 80%, respectively). While for good prognosis patients it has been established that 3 cycles and 4 cycles of BEP are equivalent [5], not much research has been done regarding intermediate prognosis patients, which explains the limited number of intermediate prognosis patients in this review compared to the original IGCC data (13 vs. 28%). Intermediate prognosis patients described in trials were initially selected for trials as having 'advanced' testicular cancer according to the Indiana classification, and were retrospectively

classified as intermediate prognosis according to the IGCC classification [26]. One RCT by the EORTC compared BEP and VIP in intermediate prognosis patients and found no difference [27]. However the use of different criteria than the IGCC classification to define intermediate prognosis patients limits the interpretation of the results of this RCT [9,28]. Alternatives to standard-dose BEP could further improve survival for intermediate prognosis patients.
Currently, intermediate prognosis patients are also being considered for high-dose chemotherapy with stem cell support in a RCT including both intermediate prognosis and poor prognosis patients [10]. Furthermore the EORTC is currently conducting a large RCT comparing BEP with T-BEP (BEP + paclitaxel) for intermediate prognosis patients [29].

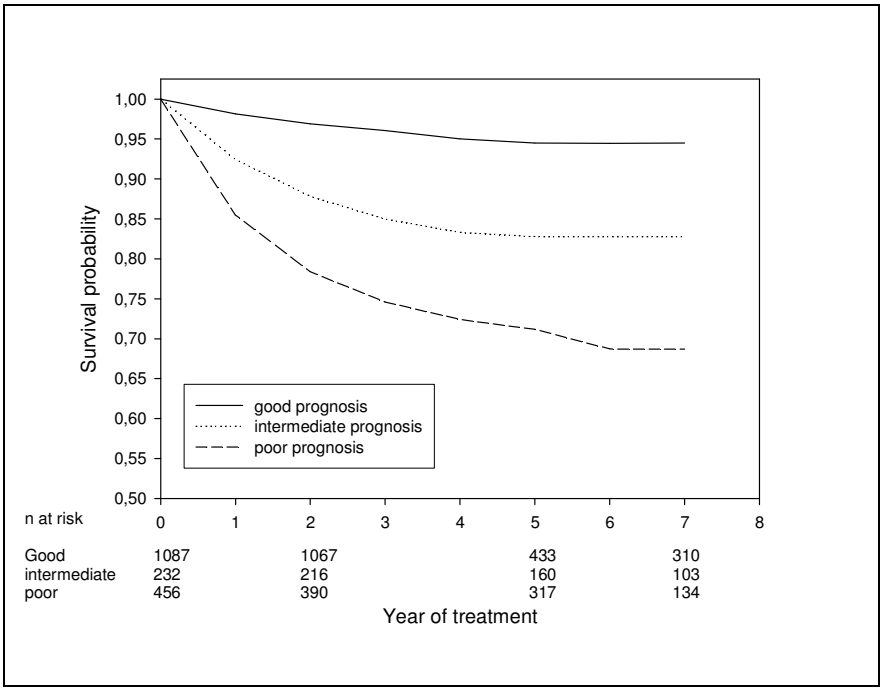


Figure 2 Pooled survival data for patients with nonseminomatous germ cell tumours treated after 1989, with either good, intermediate or poor prognosis according to the IGCC classification

Our study has some limitations. The survival estimates in our study might be overestimated due to publication bias. Trials showing good results for a certain treatment are more likely to be reported on. This effect could be even stronger in our study, since most data were published previously and later re-analysed according to the IGCC classification. This may have been done selectively, only considering those studies favouring an alternative treatment such as high-dose chemotherapy with stem cell support.

Survival estimates might also have been overestimated due to selection bias within the studies. Patients participating in clinical trials are usually selected because of their relatively 'good' health (e.g. adequate renal function, adequate bone marrow function, no other major organ dysfunction) compared to patients not treated in a clinical trial. Since follow up differed between studies, we could not directly combine 5-year overall survival. Since not all studies provided information on number of patients at risk or 95% confidence intervals of survival estimates, the 95% confidence intervals of pooled estimates could not be provided. This illustrates the importance of reporting adequate information on survival, e.g., 95% confidence intervals and number of patients at risk.

In studies describing poor prognosis patients all or almost all patients were treated with alternative treatment regimens, even in studies based on hospital registries [12,15,16]. Since no RCTs have been published demonstrating the effect of these treatment regimens, it cannot fully be determined what causes the increase in survival for poor prognosis patients.

Patients in the selected studies were mostly treated in clinical trials and centres specialized in the treatment of testicular cancer. This may limit the generalisability of the survival estimates to patients treated in other hospitals. However, Collette and colleagues concluded that poor prognosis patients should be treated in specialized treatment centres since larger centres with more experience had a better survival than smaller centres with less experience [23].

We will have to wait for the completion of ongoing RCTs for poor prognosis patients to know what survival of patients receiving standard treatment is nowadays, and if new treatment regimens are really more effective. Currently two RCTs are investigating the effect of high-dose chemotherapy, one in Europe which is still including poor prognosis patients and one in the US for which accrual of patients has closed [10,11]. Unfortunately, it may take years before the final results of

these RCTs are published, although insight into the 2-year survival would already be helpful.

Meanwhile we will have to use the information available in the literature. Based on our meta-analysis we conclude that there was a large increase in survival for poor prognosis patients, while there was only a small increase in survival for good prognosis and intermediate prognosis patients. Although this could be explained by alternative treatment regimens investigated since the introduction of the IGCC classification, there is also evidence that survival increased over time irrespective of treatment received. We therefore recommend that results of Phase I/II trials for poor prognosis patients should no longer be compared with the survival estimates for poor prognosis patients reported by the IGCC classification as this may overestimate the effect of new treatment regimens.

Acknowledgements

The authors thank the European Organization for Research and Treatment of Cancer and the Medical Research Council for permission to use the data from the MRC-TE20/EORTC 30941 trials for this research.

This work was supported by the Netherlands Organisation for Scientific Research and by the Royal Netherlands Academy of Arts and Sciences.

References

1. Bosl GJ, Motzer RJ. Testicular germ-cell cancer. *N Engl J Med* 1997;337(4):242-53.
2. Hartmann JT, Kanz L, Bokemeyer C. Diagnosis and treatment of patients with testicular germ cell cancer. *Drugs* 1999;58(2):257-81.
3. Sonneveld DJ, Hoekstra HJ, van der Graaf WT, et al. Improved long term survival of patients with metastatic nonseminomatous testicular germ cell carcinoma in relation to prognostic classification systems during the cisplatin era. *Cancer* 2001;91(7):1304-15.
4. Steele GS, Richie JP, Stewart AK, Menck HR. The National Cancer Data Base report on patterns of care for testicular carcinoma, 1985-1996. *Cancer* 1999;86(10):2171-83.
5. de Wit R, Roberts JT, Wilkinson PM, et al. Equivalence of three or four cycles of bleomycin, etoposide, and cisplatin chemotherapy and of a 3- or 5-day schedule in good-prognosis germ cell cancer: A randomized study of the European organization for research and treatment of cancer genitourinary tract cancer cooperative group and the medical research council. *Journal of Clinical Oncology* 2001;19(6):1629-1640.
6. Bokemeyer C, Harstrick A, Beyer J, et al. First-line high-dose chemotherapy for 'poor risk' metastatic non-seminomatous testicular germ cell tumors. *Onkologie* 1998;21:23-25.
7. Bokemeyer C, Kollmannsberger C, Meisner C, et al. First-line high-dose chemotherapy compared with standard-dose PEB/VIP chemotherapy in patients with advanced germ cell tumors: A multivariate and matched-pair analysis. *Journal of Clinical Oncology* 1999;17(11):3450-3456.
8. Hinton S, Catalano PJ, Einhorn LH, et al. Cisplatin, etoposide and either bleomycin or ifosfamide in the treatment of disseminated germ cell tumors - Final analysis of an intergroup trial. *Cancer* 2003;97(8):1869-1875.
9. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *J Clin Oncol* 1997;15(2):594-603.
10. EORTC. Phase III Randomized Study of Standard Cisplatin, Etoposide, and Ifosfamide (VIP) Followed By Sequential High-Dose VIP and Stem Cell Rescue Versus Bleomycin, Etoposide, and Cisplatin (BEP) in Chemotherapy-Naive Men With Poor-Prognosis Germ Cell Cancer, Accessed June 1, 2006.
11. US intergroup. Phase III Randomized Study of Bleomycin, Etoposide, and Cisplatin (BEP) With or Without High-Dose Carboplatin, Etoposide, and Cyclophosphamide Plus Autologous Bone Marrow or Peripheral Blood Stem Cell Transplantation in Male Patients With Previously Untreated Poor- or Intermediate-Risk Germ Cell Tumors. Accessed June 1, 2006.

12. Germa-Lluch JR, Garcia del Muro X, Maroto P, et al. Clinical pattern and therapeutic results achieved in 1490 patients with germ-cell tumours of the testis: the experience of the Spanish Germ-Cell Cancer Group (GG). *Eur Urol* 2002;42(6):553-62; discussion 562-3.
13. Earle CC, Pham B, Wells GA. An assessment of methods to combine published survival curves. *Med Decis Making* 2000;20(1):104-11.
14. Schmoll HJ, Kollmannsberger C, Metzner B, et al. Long-term results of first-line sequential high-dose etoposide, ifosfamide, and cisplatin chemotherapy plus autologous stem cell support for patients with advanced metastatic germ cell cancer: An extended phase I/II study of the German testicular cancer study group. *Journal of Clinical Oncology* 2003;21(22):4083-4091.
15. Muramaki M, Hara I, Miyake H, Yamada Y, Kawabata G, Kamidono S. Advances in the management of non-seminomatous germ cell tumors during the cisplatin era: a single-institution experience. *Int J Urol* 2004;11(9):768-73.
16. Bhala N, Coleman JM, Radstone CR, et al. The management and survival of patients with advanced germ-cell tumours: Improving outcome in intermediate and poor prognosis patients. *Clinical Oncology* 2004;16(1):40-47.
17. Germa-Lluch JR, Garcia del Muro X, Tabernero JM, et al. BOMP/EPI intensive alternating chemotherapy for IGCCC poor-prognosis germ-cell tumors: the Spanish Germ-Cell Cancer Group experience (GG). *Ann Oncol* 1999;10(3):289-93.
18. Anthoney DA, McKean MJ, Roberts JT, et al. Bleomycin, vincristine, cisplatin/bleomycin, etoposide, cisplatin chemotherapy: an alternating, dose intense regimen producing promising results in untreated patients with intermediate or poor prognosis malignant germ-cell tumours. *British Journal of Cancer* 2004;90(3):601-606.
19. Christian JA, Huddart RA, Norman A, et al. Intensive induction chemotherapy with CBOP/BEP in patients with poor prognosis germ cell tumors. *Journal of Clinical Oncology* 2003;21(5):871-877.
20. Fizazi K, Prow DM, Do KA, et al. Alternating dose-dense chemotherapy in patients with high volume disseminated non-seminomatous germ cell tumours. *British Journal of Cancer* 2002;86(10):1555-1560.
21. Decatris MP, Wilkinson PM, Welch RS, Metzner M, Morgenstern GR, Dougall M. High-dose chemotherapy and autologous haematopoietic support in poor risk non-seminomatous germ-cell tumours: An effective first-line therapy with minimal toxicity. *Annals of Oncology* 2000;11(4):427-434.
22. Williams SD, Birch R, Einhorn LH, Irwin L, Greco FA, Loehrer PJ. Treatment of disseminated germ-cell tumors with cisplatin, bleomycin, and either vinblastine or etoposide. *N Engl J Med* 1987;316(23):1435-40.

23. Collette L, Sylvester RJ, Stenning SP, et al. Impact of the treating institution on survival of patients with "poor-prognosis" metastatic nonseminoma. European Organization for Research and Treatment of Cancer Genito-Urinary Tract Cancer Collaborative Group and the Medical Research Council Testicular Cancer Working Party. *J Natl Cancer Inst* 1999;91(10):839-46.

24. El-Helw L, Coleman RE. Salvage, dose intense and high-dose chemotherapy for the treatment of poor prognosis or recurrent germ cell tumours. *Cancer Treat Rev* 2005;31(3):197-209.

25. van Dijk MR, Steyerberg EW, Habbema JD. Survival of non-seminomatous germ cell cancer patients according to the IGCC classification: An update based on meta-analysis. *Eur J Cancer* 2006;42(7):820-6.

26. Birch R, Williams S, Cone A, et al. Prognostic factors for favorable outcome in disseminated germ cell tumors. *J Clin Oncol* 1986;4(3):400-7.

27. de Wit R, Stoter G, Sleijfer DT, et al. Four cycles of BEP vs four cycles of VIP in patients with intermediate-prognosis metastatic testicular non-seminoma: a randomized study of the EORTC Genitourinary Tract Cancer Cooperative Group. European Organization for Research and Treatment of Cancer. *Br J Cancer* 1998;78(6):828-32.

28. Stoter G, Sylvester R, Sleijfer DT, et al. Multivariate analysis of prognostic factors in patients with disseminated nonseminomatous testicular cancer: results from a European Organization for Research on Treatment of Cancer Multiinstitutional Phase III Study. *Cancer Res* 1987;47(10):2714-8.

29. EORTC. Phase II/III Randomized Study of Bleomycin, Cisplatin, and Etoposide (BEP) Versus Bleomycin, Cisplatin, Etoposide, and Paclitaxel (T-BEP) in Men With Intermediate Prognosis Germ Cell Cancer, Accessed June 1, 2006.

5

Identifying Subgroups among Poor Prognosis Patients with Nonseminomatous Germ Cell Cancer: A Validation study

M.R. van Dijk, E.W. Steyerberg, S.P. Stenning, and J.D.F. Habbema
Ann Oncol 2004;15(9):1400-5

Abstract

Background In order to target intensive treatment strategies for poor prognosis patients with nonseminomatous germ cell cancer, those with the poorest prognosis should be identified. These patients might profit most from more intensive treatment strategies. For this purpose, a regression tree was previously developed on 332 patients. We aimed to evaluate the performance and structure of this tree.

Patients and Methods The previously developed tree was applied to 456 patients with a poor prognosis as defined by the International Germ Cell Cancer Collaborative Group (IGCCCG). Next, we developed a new tree to evaluate whether a similar structure to the previous tree was found. We assessed the internal validity of the new tree, and compared the 2-year survival estimates of each subgroup together with the discriminative ability for both the previously developed and the new tree. Discriminative ability was measured by a concordance (c) statistic, which varies between 0.5 (no discrimination) and 1.0 (perfect discrimination).

Results The 2-year survival estimates in the IGCCCG data ranged from 33 to 63%. The ordering of the subgroups was different and discriminative ability was lower than originally found (c = 0.56 in the IGCCCG data vs. 0.63 originally). The new tree differed considerably from the original tree, and identified poor prognosis subgroups with 2-year survival estimates from 38 to 73%. Internal validation showed similar discriminative ability for the new tree and the original tree (c = 0.59 vs. 0.56).

Conclusion The previously developed tree showed poor validity with respect to discriminative ability and the stability of its structure. The performance of the new tree was also unsatisfactory. Given the low proportion of patients categorised as poor prognosis, it seems that the potential to identify further subgroups with the currently available patient characteristics is limited.

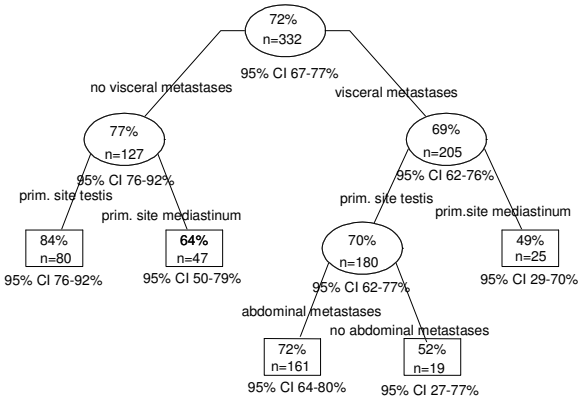
5.1 Introduction

Patients with metastatic nonseminomatous germ cell tumours nowadays have a long term cure rate of >80%, because of the ability of cisplatin-based chemotherapy to cure advanced disease [1-4]. Because of the high overall cure rate, interest has shifted to reducing treatment related toxicity for patients with a good prognosis [5]. On the other hand poor prognosis patients should be considered for more intensive treatment strategies [6-8]. The International Germ Cell Cancer Collaborative Group (IGCCCG) developed the International Germ Cell Consensus (IGCC) classification to distinguish patients according to prognosis [9]. The IGCC classification is currently widely applied. It distinguishes patients with good, intermediate and poor prognosis. The poor prognosis group consists of ~15% of all patients and is characterised by the presence of mediastinal primary site, non-pulmonary visceral metastases or poor tumour marker levels. Long term survival following standard treatment may be ~50% [9]. To further improve the long term survival in poor prognosis patients, those who are most likely to fail standard treatment should be identified. These patients are most likely to profit from novel chemotherapy approaches, such as dose intensification and high-dose chemotherapy with stem cell support [10]. Dose intensification with either sequential or alternating non-cross-resistant chemotherapy has shown promising results in nonrandomised trials [11], but this has not been confirmed in randomised clinical trials [10,11].

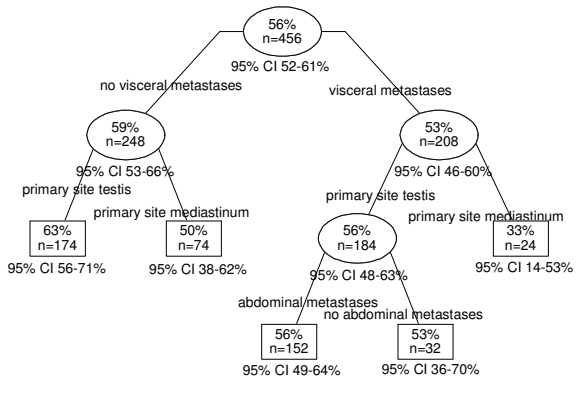
Studies conducted by the Memorial Sloan Kettering Cancer Center [12,13] and the German Testicular Cancer Group [7,14] showed beneficial effects for high-dose chemotherapy with stem cell support and high-dose chemotherapy, respectively. However these results were based on nonrandomised trials and comparisons with historical controls. To confirm these results, two randomised, multicentre trials are currently being conducted in the USA and Europe [15]. Furthermore, the identification of subgroups among poor prognosis patients would allow for a more accurate estimate of the individual patients' chances of survival, and increase the comparability of results from clinical trials [16].

Kollmannsberger et al [16] used tree modelling as an explorative method to identify important risk factors within a group of poor prognosis patients, as defined by the IGCC classification, and to find subsets of patients differing in prognosis. They developed a regression tree based on data of 332 poor prognosis patients as defined by the IGCC classification (Kollmannsberger tree). The risk factors visceral metastases, primary site and abdominal mass were used.

A


72%
n=332
95% CI 67-77%
no visceral metastases
visceral metastases
77%
n=127
95% CI 76-92%
69%
n=205
95% CI 62-76%
prim. site testis
prim. site mediastinum
prim. site testis
prim.site mediastinum
84%
n=80
95% CI 76-92%
64%
n=47
95% CI 50-79%
70%
n=180
95% CI 62-77%
49%
n=25
95% CI 29-70%
abdominal metastases
no abdominal metastases
72%
n=161
95% CI 64-80%
52%
n=19
95% CI 27-77%

B


56%
n=456
95% CI 52-61%
no visceral metastases
visceral metastases
59%
n=248
95% CI 53-66%
53%
n=208
95% CI 46-60%
primary site testis
primary site mediastinum
primary site testis
primary site mediastinum
63%
n=174
95% CI 56-71%
50%
n=74
95% CI 38-62%
56%
n=184
95% CI 48-63%
33%
n=24
95% CI 14-53%
abdominal metastases
no abdominal metastases
56%
n=152
95% CI 49-64%
53%
n=32
95% CI 36-70%

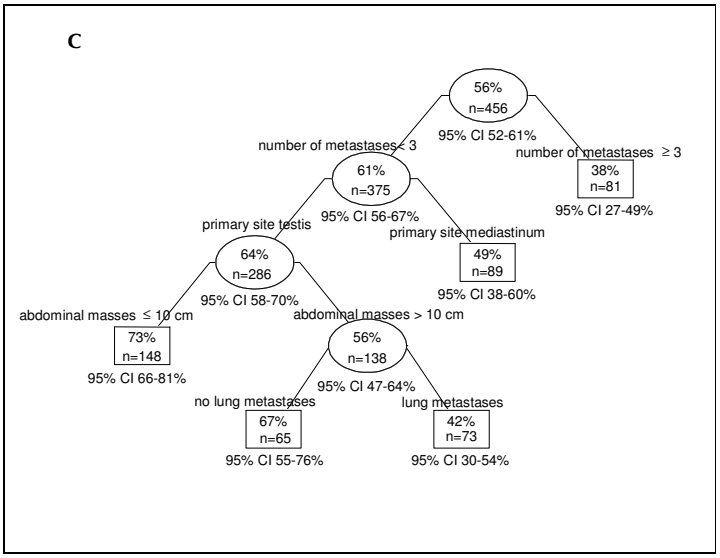


Figure 1 Trees for poor prognosis patients with nonseminomatous germ cell cancer with 2-year survival, 95% confidence interval and number of patients (n) (A) Kollmannsberger tree applied to the Kollmannsberger data (B) Kollmannsberger tree applied to the International Germ Cell Cancer Collaborative Group (IGCCCG) poor prognosis data (C) International Germ Cell Consensus (IGCC) tree applied to the IGCCCG poor prognosis data

This resulted in a tree with five poor prognosis subgroups (Figure 1a). The subgroups differed in 2-year survival, ranging from 49 to 84% [16].

Tree models are attractive ways to identify subsets of patients because they are easy to apply and interpret. They have few restrictions, which makes them suitable for finding interactions between risk factors [17-20]. Furthermore, trees have more resemblance with the way clinicians make decisions than linear models [17]. On the other hand, this flexibility makes trees 'data hungry'. Use of relatively small datasets will lead to unstable tree models, and optimism in the performance of the model due to overfitting [17,21,22].

Kollmannsberger et al [16] recognise these problems and the limitations of their tree. Some subgroups only had a small number of patients, and their identification may be the result of pure chance. Such subgroups may not be present when new data are considered. Furthermore, survival estimates of small groups are often unreliable. This was illustrated by the group of patients with visceral metastases

and primary site testis, in which patients with an abdominal mass had a higher 2-year survival (72%; 95% confidence interval (CI) 64–80%) than patients without (52%; 95% CI 27–77%). Kollmannsberger et al [16] therefore propose that further confirmatory studies in other poor prognosis patients cohorts are needed, before the tree can be used in practice.

The aim of the present study was to evaluate the validity of the Kollmannsberger tree. We consider two aspects of validity: performance and the structure of the tree model. We evaluated the performance of the Kollmannsberger tree by applying the tree to the poor prognosis patients in the IGCCCG database [9]. Furthermore, we developed a new tree for the poor prognosis patients in the IGCCCG database to study whether its structure, that is the selection and hierarchy of risk factors, was similar to the Kollmannsberger tree.

5.2 Patients and Methods

Patients

The Kollmansberger tree was based on data from 332 patients with metastatic nonseminomatous germ cell tumours, from prospective clinical trials conducted between 1984 and 1997 [23,24]. All patients were treated for poor prognosis disease, as defined by the IGCC classification [9], with either cisplatin–etoposide–ifosfamide (PEI) or cisplatin–etoposide–bleocymin (PEB).

To validate the Kollmannsberger tree, we used the 495 poor prognosis patients with nonseminomatous germ cell cancer from the IGCCCG database, which consists of 5202 adult male patients. Poor prognosis was defined by the presence of any of the poor risk factors mediastinal primary site, (nonpulmonary) visceral metastases, alpha-fetoprotein (AFP) poor (>10000 ng/ml), human chorionic gonadotrophin (HCG) poor (>10000 ng/ml) and lactate dehydrogenase (LDH) poor (>10 times the upper limit of normal) [9].

Patients were treated between 1975 and 1990 with cisplatin-based chemotherapy. Data were collected on age, primary site, date of diagnosis, levels of serum AFP, HCG and LDH, nodal disease in the abdomen, mediastinum, neck, lung metastases, spread to other visceral sites such as liver, bone and brain, and on treatment details such as previous therapy [9]. Data for 39 patients were excluded because of missing values on the risk factors age, and lung, liver, bone and brain metastases. The endpoint was overall survival, calculated from the beginning of chemotherapy.

Statistical analyses

We assessed the performance of the Kollmannsberger tree by applying it to the 456 poor prognosis patients in the IGCCCG database. Two-year overall survival was calculated using the Kaplan–Meier method.

Furthermore, a new tree was developed in the poor prognosis patients in the IGCCCG database and the result compared with the Kollmannsberger tree to evaluate its structure, that is the selection and hierarchy of the risk factors. We will refer to this new tree as the IGCC tree. For the development of this tree we used the same candidate risk factors and coding as Kollmannsberger et al [16]. We first determined the risk factor that best split the data into two subgroups, leading to the largest decrease in prediction error. Splitting continued until a group reached a minimum size of five, or until no further improvement in discrimination could be made, based on the loss in exponential log-likelihood. The full tree, which might be too complex and overfit, was pruned using 10-fold crossvalidation. All trees within one standard error of the lowest crossvalidated prediction error were considered as equivalent. A final tree was selected from these equivalent trees[19,20,25].

Modelling was performed with S-plus version 2000 using the RPART library, which contains a recursive partitioning method for survival data. The RPART library (rpart2.zip) and manual (rpart2doc.zip) can be found at http://www.stats.ox.ac.uk/pub/SWin.

For comparison we fitted a Cox regression model using the same risk factors in the poor prognosis patients in the IGCCCG database. All risk factors were entered in the model and the final model was obtained with a backward stepwise selection procedure using a P-value of 0.05.

Predictive performance

We determined the discriminative ability, to indicate the predictive performance of the Kollmannsberger tree, the IGCC tree and the Cox regression model. The discriminative ability indicates how well a model can distinguish between patients with different survival expectations and was measured by a concordance statistic (c-statistic). For binary outcomes, the c-statistic is identical to the area under the ROC curve [26,27]. The c-statistic for survival data estimates the probability that for a randomly chosen pair of patients, the one having the higher predicted survival is the one who survives longer [26]. The c-statistic varies between 0.5 and 1.0 for

sensible models. A predictive model with a c of 0.5 has no predictive value, while a model with a c of 1.0 discriminates perfectly between patients differing in survival. C-statistics were computed for the Kollmannsberger tree, when applied to the Kollmannsberger patients or the IGCCCG patients, for the IGCC tree and for the Cox regression model. The steps taken in the development of the IGCC tree were internally validated by taking 100 random bootstrap samples. The development process of the tree was repeated for every bootstrap sample and the resulting tree tested on the original sample, to estimate and correct for the optimism in discriminative ability [28,29]. The original sample hence served as the test sample for models developed in the bootstrap samples. The Cox regression model was validated according to the same procedure. The standard error of the corrected c-statistic was taken from the empirical distribution of the c-statistics in the test sample. This standard error was used to calculate the 95% CI of the optimism-corrected c-statistic.

5.3 Results

Patient characteristics are given in Table 1, both for the 456 poor prognosis patients of the IGCCCG study and the 332 patients of the Kollmannsberger study. More than half of the IGCCCG poor prognosis patients had primary site testis (67%), lung metastases (62%) or abdominal masses (70%). Sixty-three per cent of the patients had poor AFP, HCG or LDH levels. The presence of liver metastases was common (34%). The distribution of the patient characteristics age, lung metastases, visceral metastases, abdominal masses, number of metastatic sites and tumour markers, combined as well as separate, and follow up time was largely similar for the IGCCCG and the Kollmannsberger studies. Disease progression occurred in 252 of the IGCCCG poor prognosis patients. Of the 223 patients who died, 213 were categorised as disease-related.

The corresponding 2-year survival was 56% (95% CI 52–61%) for the IGCCCG poor prognosis patients. This differs from the Kollmannsberger data, where 2-year survival was 72% (95% CI 67–77%).

The Kollmannsberger tree was applied to the poor prognosis patients in the IGCCCG database, as presented in Figure 1b. The split according to the presence of visceral metastases resulted in two subgroups with only slightly different 2-year survival (59 and 53%).

Table 1 Patient characteristics, Kolmannsberger and IGCCCG poor prognosis data

	Kolmannsberger		IGCCCG	
Patient characteristics	Number of patients (%)		Number of patients (%)	
Age				
Median	28		27	
Range	15-62		14-67	
Primary site				
Mediastinum	72	(22)	98	(22)
Retroperitoneal	31	(9)	51	(11)
Testis	229	(69)	307	(67)
Lung metastases	247	(74)	283	(62)
Visceral metastases	205	(62)	232	(51)
Liver	131	(39)	153	(34)
Bone	35	(11)	31	(7)
CNS/Brain	33	(10)	36	(8)
Other	6	(2)	12	(3)
Abdominal mass	205	(62)	318	(70)
Abdominal mass >10 cm	120	(36)	179	(39)
Marker combined				
Good	18	(6)	37	(8)
Intermediate	104	(31)	133	(29)
Poor	210	(63)	286	(63)
AFP				
Good	189	(57)	247	(54)
Intermediate	68	(20)	80	(18)
Poor	72	(22)	129	(28)
Missing	3	(1)		
HCG				
Good	180	(54)	260	(57)
Intermediate	34	(10)	54	(12)
Poor	117	(35)	142	(31)
Missing	1	(1)		
LDH				
Good	79	(24)	104	(23)
Intermediate	197	(59)	318	(70)
Poor	38	(12)	34	(7)
Missing	18	(5)		
Number of metastatic sites				
Median	2		2	
Range	0-5		1-5	
Number of patients ≥ 3 metastatic sites	159	(48)	216	(47)
Status at last follow up				
CR/PR	215	(65)	204	(45)
AWD	14	(5)	29	(6)
Dead by disease	95	(28)	213	(47)
Dead other cause	8	(2)	10	(2)
Follow up in months				
Median	23+		23	
Range	0-99		1-170	
Time of treatment	1984-1997		1975-1990	
Total number of patients	332		456	

Good/intermediate/poor tumour markers according to the IGCC classification.
IGCCCG, International Germ Cell Cancer Collaborative Group; AFP, alfa-fetoprotein; HCG, human chorionic gonadotrophin; LDH, lactate dehydrogenase; CR, complete remission; PR, partial remission; AWD, alive with disease; IGCC, International Germ Cell Consensus

Both branches were split further according to primary site, resulting in two final groups in the no visceral metastases branch, with 2-year survival of 63 and 50%, and in the visceral metastases branch in one final group with 2-year survival of 33% and one further subgroup with 2-year survival of 56%; the last final groups from this branch were defined by the presence of abdominal mass, with similar 2-year survival (56 and 53%). As can be seen by comparing figure 1b and 1a, respectively, the 2-year survival estimates for the five final groups identified in the poor prognosis patients in the IGCCCG database by the Kollmannsberger tree were less extreme than in the original Kollmannsberger tree. This was reflected by the lower discriminative ability in the IGCCCG poor prognosis data (c = 0.56; 95% CI 0.49–0.64) compared with the original data (c = 0.63; 95% CI 0.56–0.70).
The newly developed IGCC tree is presented in Figure 1c. Trees with two to six groups gave equivalent performance based on the loss in exponantial log-likelihood. However, a tree with five final groups was chosen for fair comparability with the Kollmannsberger tree. The 2-year survival ranged from 38 to 73%. The principal determinant of survival was the total number of metastases, where three or fewer metastases resulted in a subgroup with a 2-year survival of 61% and the presence of more than three metastases in a final group with a 2-year survival of 38%. The next split was made by primary site, resulting in a subgroup of patients with testis as primary site and a 2-year survival of 64% and a final group of patients with mediastinal primary site and a 2-year survival of 49%. A further distinction was made by the size of an abdominal mass, resulting in a 2-year survival of 73% for patients with a mass ≤10 cm, and 56% for a mass >10 cm. Finally, the presence of lung metastases resulted in two final groups with a 2-year survival of 67% for patients without lung metastases and 42% for patients with lung metastases.
Thus, both the Kollmannsberger tree and the new IGCC tree selected primary site and abdominal mass as important risk factors, although the Kollmannsberger tree selected the presence of abdominal mass rather than size. Furthermore, the presence of lung metastases was used in the IGCC tree, but not in the Kollmannsberger tree. The apparent discriminative ability of the new IGCC tree was similar to the discriminative ability of the Kollmannsberger tree, with a c-statistic of 0.63 (95% CI 0.62–0.72). Internal validation revealed an optimism in c-statistic of 0.04, leading to an optimism-corrected estimate of c = 0.59 (95% CI 0.54–0.63) for the new IGCC tree when applied to future patients similar to those included in the IGCCCG database.

The Cox regression model selected the risk factors primary site, presence of abdominal mass, size of abdominal mass, total number of metastases, AFP and tumour markers combined. The discriminative ability of the Cox regression model was slightly higher, with a c-statistic of 0.64, but decreased to 0.61 after correction for optimism.

5.4 Discussion

The previously developed tree to identify subgroups among poor prognosis patients with nonseminomatous germ cell cancer showed poor validity in the poor prognosis patients in the IGCCCG database. First, the discriminative ability of the Kollmannsberger tree was substantially lower (c = 0.56) than in the original data (c = 0.63). Secondly, a new tree was developed in the poor prognosis patients in the IGCCCG database, which used the number of metastases, primary site, size of abdominal mass and the presence of lung metastases to identify subgroups, whereas the Kollmannsberger tree used the presence of visceral metastases, primary site and abdominal mass as risk factors. The discriminative ability of the new tree, c = 0.59 at internal validation, was similar to the Kollmannsberger tree (c = 0.56).

In our case, the selected risk factors were rather similar in the trees developed with the Kollmannsberger data and the IGCCCG poor prognosis data (presence versus size of abdominal mass and presence versus number of metastases). The structure of the trees, however, was very different. Since primary site was an important risk factor in patients with and without visceral metastases in the Kollmannsberger tree, both risk factors can be interpreted statistically as main effects. In the IGCC tree, primary site occurred only once, and no main effects were present except for the number of metastases. Furthermore, the trees fitted in the bootstrap samples varied in size, the smallest tree having only two groups and the largest tree 20. A tree size of four was most prevalent (14% of cases). The flexibility in structure and tree size led to optimism in the performance of the tree developed on the IGCCCG poor prognosis data, where the c-statistic was expected to decrease from 0.63 to 0.59 according to a bootstrap validation technique. Hence, the flexibility of tree modelling may have more cons than pros in small datasets, that is datasets with relatively few deaths. Larger datasets are required for reliable application of tree modelling.

The discriminative ability of a Cox regression model was slightly higher than the regression tree models with a c-statistic of 0.61 at internal validation. A Cox

regression model was not available for the Kollmannsberger data. Should such a model be applied to the Kollmannsberger data, the selection and the number of risk factors might differ. However, the structure is likely to be similar for a given selection of risk factors, since we usually fit main effects only in a regression model. In relatively small datasets, this may be the most sensible, since we have insufficient statistical power to identify important interaction terms [30].

To assess whether the Kollmannsberger tree can be generalised to other patients, we applied it to poor prognosis patients from the IGCCCG dataset. The comparison of the Kollmannsberger and the IGCC tree was, however, limited by the differences in the two datasets. Although the distribution of risk factors in the Kollmannsberger and IGCCCG poor prognosis data was largely similar, there was a difference in 2-year survival (72 and 56%, respectively). These differences may reflect the different time periods in which the data were collected (Kollmannsberger 1984–1997, IGCCCG 1975–1990). Owing to improved treatment strategies, survival has increased over time [4,31]. Patients in the Kollmannsberger study were treated with regimens of either PEB or PEI. Although all poor prognosis patients in the IGCCCG database were treated with cisplatin-based chemotherapy, patients from the late 1970s and early 1980s were probably treated with the cisplatin–vinblastine–bleocymin regimen rather than PEB. The differences between the populations suggest that a more recent population of poor prognosis patients might be more suitable to assess the generalisability of the Kollmannsberger tree. Ideally, to assess the differences between the Kollmannsberger and the IGCCCG poor prognosis data and to make an honest comparison between both trees, the new IGCC tree should be applied to the Kollmannsberger data.

Better performance might be achieved by adding stronger risk factors that have not been used before in the classification of patients with germ cell cancer. Besides pre-treatment characteristics, the rate of tumour marker decline during the first two cycles of chemotherapy has been identified as an important risk factor. Rate of tumour marker decline predicted outcome in 189 patients, independent of risk status as defined by the IGCC classification, especially in poor prognosis patients [32]. Similar results were found in 139 poor prognosis patients in a previously conducted study [33]. These results will be validated with data from a multicentre, randomised clinical trial [32]. Furthermore, promising research is being carried out on the prognostic value of molecular and genetic markers. Knowledge on the role of such markers will not only allow for a better understanding of the

development and progression of testicular gem cell cancer, but may also lead to a more refined assessment of prognosis and better management of germ cell tumours [34,35].

In conclusion, survival of IGCC poor risk patients in the present day may have improved compared with the historical IGCCCG data. This justifies the investigation of poorer risk subgroups, although the difficulties in evaluating new treatment approaches through randomised trials in these small groups must be acknowledged. The performance of the current regression trees was unsatisfactory. The currently available risk factors are not strong enough to clearly identify subgroups among poor prognosis patients with nonseminomatous germ cell cancer, who comprise a small subgroup (~15%) of metastatic germ cell tumour patients. A new model might incorporate molecular and genetic markers in addition to the risk factors currently incorporated in the IGCC classification. We suggest the use of Cox regression for the construction of such a new model, rather than tree modelling. This method is proven to be more stable and gives less optimistic results, especially in smaller datasets. Tree modelling can give insight into possible interactions between risk factors provided sufficient data is available, but should be restricted to exploratory analyses.

Acknowledgements

We would like to thank the members of the IGCCCG for kindly providing their data for the analyses in this paper. Members of the IGCCCG and participating centres have been listed previously [9]. This work was supported by the Netherlands Organisation for Scientific Research and by the Royal Netherlands Academy of Arts and Sciences.

References

1. Bosl GJ, Motzer RJ. Testicular germ-cell cancer. N Engl J Med 1997;337(4):242-53.
2. Hartmann JT, Kanz L, Bokemeyer C. Diagnosis and treatment of patients with testicular germ cell cancer. Drugs 1999;58(2):257-81.
3. Sonneveld DJ, Hoekstra HJ, van der Graaf WT, et al. Improved long term survival of patients with metastatic nonseminomatous testicular germ cell carcinoma in relation to prognostic classification systems during the cisplatin era. Cancer 2001;91(7):1304-15.
4. Steele GS, Richie JP, Stewart AK, Menck HR. The National Cancer Data Base report on patterns of care for testicular carcinoma, 1985-1996. Cancer 1999;86(10):2171-83.
5. de Wit R, Roberts JT, Wilkinson PM, et al. Equivalence of three or four cycles of bleomycin, etoposide, and cisplatin chemotherapy and of a 3- or 5-day schedule in good-prognosis germ cell cancer: a randomized study of the European Organization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group and the Medical Research Council. J Clin Oncol 2001;19(6):1629-40.
6. Hinton S, Catalano PJ, Einhorn LH, et al. Cisplatin, etoposide and either bleomycin or ifosfamide in the treatment of disseminated germ cell tumors: final analysis of an intergroup trial. Cancer 2003;97(8):1869-75.
7. Bokemeyer C, Kollmannsberger C, Meisner C, et al. First-line high-dose chemotherapy compared with standard-dose PEB/VIP chemotherapy in patients with advanced germ cell tumors: A multivariate and matched-pair analysis. J Clin Oncol 1999;17(11):3450-6.
8. Bokemeyer C, Harstrick A, Beyer J, et al. The use of dose-intensified chemotherapy in the treatment of metastatic nonseminomatous testicular germ cell tumors. German Testicular Cancer Study Group. Semin Oncol 1998;25(2 Suppl 4):24-32; discussion 45-8.
9. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor- based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. J Clin Oncol 1997;15(2):594-603.
10. Kaye SB, Mead GM, Fossa S, et al. Intensive induction-sequential chemotherapy with BOP/VIP-B compared with treatment with BEP/EP for poor-prognosis metastatic nonseminomatous germ cell tumor: a Randomized Medical Research Council/European Organization for Research and Treatment of Cancer study. J Clin Oncol 1998;16(2):692-701.
11. Fizazi K, Zelek L. Is one cycle every three or four weeks' obsolete? A critical review of dose-dense chemotherapy in solid neoplasms. Ann Oncol 2000;11(2):133-49.

12. Motzer RJ, Mazumdar M, Gulati SC, et al. Phase II trial of high-dose carboplatin and etoposide with autologous bone marrow transplantation in first-line therapy for patients with poor-risk germ cell tumors. J Natl Cancer Inst 1993;85(22):1828-35.
13. Motzer RJ, Mazumdar M, Bajorin DF, Bosl GJ, Lyn P, Vlamis V. High-dose carboplatin, etoposide, and cyclophosphamide with autologous bone marrow transplantation in first-line therapy for patients with poor-risk germ cell tumors. J Clin Oncol 1997;15(7):2546-52.
14. Schmoll HJ, Kollmannsberger C, Metzner B, et al. Long-term results of first-line sequential high-dose etoposide, ifosfamide, and cisplatin chemotherapy plus autologous stem cell support for patients with advanced metastatic germ cell cancer: an extended phase I/II study of the German Testicular Cancer Study Group. J Clin Oncol 2003;21(22):4083-91.
15. Motzer RJ, Sheinfeld S. Poor-risk testicular cancer and high-dose chemotherapy. J Clin Oncol 2003;21(22):4073-4.
16. Kollmannsberger C, Nichols C, Meisner C, Mayer F, Kanz L, Bokemeyer C. Identification of prognostic subgroups among patients with metastatic 'IGCCCG poor-prognosis' germ-cell cancer: an explorative analysis using cart modeling. Ann Oncol 2000;11(9):1115-20.
17. Marshall RJ. The use of classification and regression trees in clinical epidemiology. J Clin Epidemiol 2001;54(6):603-9.
18. Radespiel-Tröger M, Rabenstein T, Schneider HT, Lausen B. Comparison of tree-based methods for prognostic stratification of survival data. Artificical Intelligence in Medicine 2003;28(3):323-341.
19. LeBlanc M, Crowley J. Relative risk trees for censored survival data. Biometrics 1992;48(2):411-25.
20. Davis RB, Anderson JR. Exponential survival trees. Stat Med 1989;8(8):947-61.
21. Crichton NJ, Hinde JP, Marchini J. Models for diagnosing chest pain: is CART helpful? Stat Med 1997;16(7):717-27.
22. Harrell FE, Jr. Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis. New York: Springer, 2001.
23. Nichols CR, Catalano PJ, Crawford ED, Vogelzang NJ, Einhorn LH, Loehrer PJ. Randomized comparison of cisplatin and etoposide and either bleomycin or ifosfamide in treatment of advanced disseminated germ cell tumors: an Eastern Cooperative Oncology Group, Southwest Oncology Group, and Cancer and Leukemia Group B Study. J Clin Oncol 1998;16(4):1287-93.

24. Nichols CR, Williams SD, Loehrer PJ, et al. Randomized study of cisplatin dose intensity in poor-risk germ cell tumors: a Southeastern Cancer Study Group and Southwest Oncology Group protocol. J Clin Oncol 1991;9(7):1163-72.
25. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Belmont, California: Wadsworth, 1984.
26. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. Stat Med 1984;3(2):143-52.
27. Grunkemeier GL, Jin R. Receiver operating characteristic curve analysis of clinical risk models. Ann Thorac Surg 2001;72(2):323-6.
28. Efron B, Tibshirani RJ. An introduction to the bootstrap. London: Chapman & Hall Inc., 1993.
29. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996;15(4):361-87.
30. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD. Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Med Decis Making 2001;21(1):45-56.
31. Bokemeyer C. Current trends in chemotherapy for metastatic nonseminomatous testicular germ cell tumors. Oncology 1998;55(3):177-88.
32. Mazumdar M, Bajorin DF, Bacik J, Higgins G, Motzer RJ, Bosl GJ. Predicting outcome to chemotherapy in patients with germ cell tumors: the value of the rate of decline of human chorionic gonadotrophin and alpha-fetoprotein during therapy. J Clin Oncol 2001;19(9):2534-41.
33. Fizazi K, Kramar A, Culine IC, Droz J, Logothetis CJ. The early decline of tumor markers predicts outcome in poor prognosis non-seminomatous germ-cell tumors (NSGCT): results of a multi-institution study. Proc Am Soc Clin 2002.
34. Heidenreich A, Srivastava S, Moul JW, Hofmann R. Molecular genetic parameters in pathogenesis and prognosis of testicular germ cell tumors. Eur Urol 2000;37(2):121-35.
35. Donadio AC, Bosl GJ. The future of therapy for nonseminomatous germ cell tumors. Chest Surg Clin N Am 2002;12(4):769-89.

6

A Decision-analytic Approach to Define Poor Prognosis Patients: A Case Study for Nonseminomatous Germ Cell Cancer Patients

M.R. van Dijk, E.W. Steyerberg, and J. D. F. Habbema

Submitted (BMC Medical Informatics and Decision Making)

Abstract

Introduction Classification systems may be useful to direct more intensive treatment to cancer patients with a relatively poor prognosis. The definition of 'poor prognosis' often lacks a formal basis. We propose a decision analytic approach to weigh benefits and harms explicitly to define the treatment threshold for more intensive treatment. This approach is illustrated by a case study in advanced testicular cancer, where patients with a high risk of mortality under standard treatment may be eligible for high-dose chemotherapy with stem cell support.

Materials and methods We used published literature to estimate the benefit and harm of high-dose chemotherapy (HD-CT) versus standard-dose chemotherapy (SD-CT) for patients with advanced nonseminomatous germ cell cancer. Benefit and harm were defined as the reduction and increase in absolute risk of mortality due to HD-CT respectively. Harm included early and late treatment related death, and treatment related morbidity (weighted by 'utility').

Results We considered a conservative and an optimistic benefit of 30 and 40% risk reduction respectively. We estimated the excess treatment related mortality at 2%. When treatment related morbidity was taken into account, the harm of HD-CT increased to 5%. With a relative benefit of 30% and harm of 2 or 5%, HD-CT might be beneficial for patients with over 7 or 17% risk of cancer specific mortality with SD chemotherapy, while with a relative benefit of 40% HD-CT was beneficial over 5 and 12.5% risk respectively.

Conclusion Benefit and harm can be used to define 'poor prognosis' explicitly for nonseminomatous germ cell cancer patients who are considered for high-dose chemotherapy. This approach can readily be adapted to new results and extended to other cancers to define candidates for more intensive treatments.

6.1 Introduction

The prognosis of a cancer patient is of key importance in the choice of more or less intensive treatment. Prognostic estimates can be based on extent of disease, as for example reflected in TNM stage, on age and comorbidity, and on specific characteristics, such as values of tumour markers [1]. Prognostic classifications can facilitate decision-making by grouping patients with a similar prognosis. Poor prognosis patients may be considered candidates for more intensive treatment strategies, while good prognosis patients may be treated with less burdensome interventions, for example by less toxic chemotherapy regimens [2,3]. Prognostic classifications use estimated survival to identify poor prognosis patients eligible for alternative treatments. However this approach only implicitly takes the possible side effects of an alternative treatment into account. Ideally both the expected gain in survival (benefit) and the toxic side effects or burden due to treatment (harm) are considered [4].

We propose a decision analytic approach in which both benefit and harm of an alternative treatment are explicitly specified and weighed to determine which patients could profit from this alternative treatment strategy.

The decision analytic approach is illustrated in Figure 1. Benefit of treatment is the reduction in absolute risk of cancer mortality due to treatment. Benefit increases linearly with risk of cancer mortality assuming that patients with the highest risk have most to gain. Harm is the increase in absolute risk of treatment mortality (e.g. related to toxicity) due to treatment. The level of harm is the same for all patients, assuming that for example the toxicity of treatment is independent of prognosis. Patients are candidates for more intensive treatment when their risk of cancer mortality is above the threshold, i.e. when benefit is higher than harm.

As an example we consider high-dose chemotherapy (HD-CT) as first line treatment to improve survival of patients with nonseminomatous germ cell cancer. Several nonrandomised trials reported a higher survival for poor prognosis patients treated with HD-CT as first line treatment (including etoposide, ifosfamide, cisplatin) with autologous stem cell support, compared to standard-dose chemotherapy (SD-CT) (including bleomycin, etoposide, cisplatin) [5-7]. Furthermore, HD-CT is currently considered in two RCTs by the European Organisation for Research and Treatment of Cancer (EORTC) and by the US intergroup [8,9]. However, HD-CT is related to a higher toxicity, both during treatment (e.g. granulocytopenia, anaemia, nausea/vomiting, diarrhoea), shortly

after treatment (e.g. pulmonary toxicity) and long after treatment (e.g. leukaemia, cardiovascular disease) [5,10].

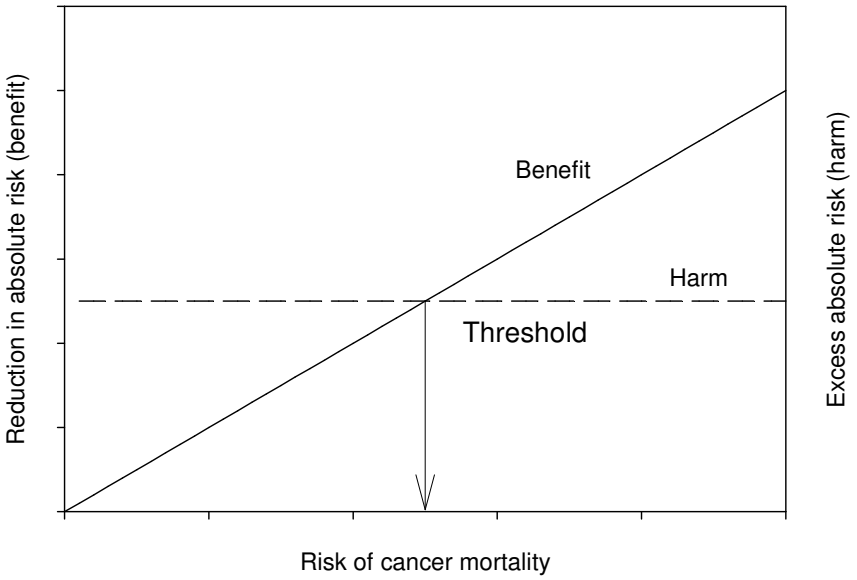


Figure 1 Benefit and harm of treatment, expressed on the same scale. Benefit of treatment (reduction in absolute risk) increases with risk, while harm of treatment (excess absolute risk, e.g. due to toxicity of treatment) is constant. Net benefit occurs only when risk is above the threshold [4]

So far studies on HD-CT focus on patients with a poor prognosis according to the International Germ Cell Consensus (IGCC) Classification [11]. The IGCC classification combined 5 risk factors to define a good, intermediate and poor prognosis group based on survival. Good prognosis patients are considered eligible for less intensive treatment reduce treatment related toxicity [12], intermediate prognosis patients usually receive standard treatment, and poor prognosis patients are considered candidates for more intensive treatment.
The aim of this study is to use a decision-analytic approach to determine how high the risk of patients with nonseminomatous germ cell cancer should be in order to profit from high-dose chemotherapy with stem cell support. Estimates of benefit and harm of high-dose chemotherapy were based on currently available literature.

6.2 Materials and methods

Of the different high-dose chemotherapy (HD-CT) treatment strategies currently investigated we considered the benefit and harm of the HD-CT approach by the German testicular cancer group [5].
We considered benefit and harm till 10 years after treatment, since longer term evidence is scarce.

Benefit
Benefit is based on the reduction in relative risk due to HD-CT compared to standard chemotherapy.

Benefit is expressed as:

(1) $1 - (R_{\text{C-MORT HD-CT}}/R_{\text{C-MORT SD-CT}})$

where $R_{\text{C-MORT HD-CT}}$ is the risk of cancer mortality with HD-CT and $R_{\text{C-MORT SD-CT}}$ the risk of cancer mortality with standard chemotherapy. This relative risk reduction translates into a decrease in absolute risk of cancer mortality at the patient level. When HD-CT results in a relative risk reduction of 25%, absolute risk decreases 10% for patients with a risk of cancer mortality of 40% (0.25 x 40%), whereas for a patient with a risk of cancer mortality of 80% the absolute risk reduction is 20% (0.25 x 80%).
Although benefit should preferably be based on results of RCTs it will take several more years before the results of RCTs comparing HD-CT to SD-CT become available. Therefore we estimated risk of cancer mortality due to HD-CT and SD-CT from three observational studies; two on patients treated with SD-CT and one on patients treated with HD-CT [5,13,14]. These observational studies reported on either 5-year or 10-year survival. To estimate benefit we need the risk of cancer mortality due to SD-CT and HC-CT 10 years after treatment.
We therefore translated survival into risk of cancer mortality at 10 years.
Firstly, overall survival (S_{OVERALL}) in each study was translated to risk of overall mortality due to treatment (R_{OVERALL}).

(2) $R_{\text{OVERALL}} = 1 - S_{\text{OVERALL}}$

From the overall risk of mortality we determine the risk of cancer mortality ($R_{C\text{-}MORT}$) by subtracting risk of treatment mortality ($R_{T\text{-}MORT}$). We ignore mortality due to other causes since testicular cancer patients are relatively young.

(3) $R_{C\text{-}MORT} = R_{OVERALL} - R_{T\text{-}MORT}$

Finally, we assumed that the relative increase in risk between 5 years and 10 years after treatment was 20% and increased the risk of cancer mortality accordingly [11]. The resulting estimates of cancer mortality 10 years after treatment of the two studies on SD-CT were combined in a weighted average by study size.

Harm

Harm is the excess risk of mortality due to HD-CT and is assumed to remain comparatively constant. We considered the excess risk of mortality and morbidity using published literature.

Treatment mortality consisted of early treatment mortality (<6 months) and late treatment mortality (>6 months). We based late treatment mortality ($R_{LATE\ T\text{-}MORT}$) on the incidence and fatality of long term complications. Fatality was assumed to be identical for patients treated with HD-CT or SD-CT once a complication occurred, although no information was available on similarity of fatality between patients treated with either SD-CT or HD-CT.

The excess risk of late treatment mortality is the difference in incidence multiplied by the estimated fatality:

(4) $\Delta R_{LATE\ T\text{-}MORT} = (\text{incidence}_{HD\text{-}CT} - \text{incidence}_{SD\text{-}CT}) \times \text{fatality}$.

Late treatment morbidity ($R_{LATE\ T\text{-}MORB}$) was made comparable to mortality by weighing complications by their utility value. Utility (U) is a measure of health related quality of life, ranging from 0 to 1, where a weight of 1 corresponds to perfect health and a weight of 0 corresponds to a health state judged equivalent to death [15]. By expressing long term complications in utilities, treatment related morbidity could be directly compared with treatment related mortality.

We estimated late treatment morbidity for SD-CT and HD-CT by combining the incidence and utilities of long term complications up to 10 years after treatment. We obtained utilities for long term complications from available literature [15].

The risk of excess late treatment morbidity for surviving patients is:

(5) $\Delta R_{\text{LATE T-MORB}}$ = (incidence $_{\text{HD-CT}}$ – incidence $_{\text{SD-CT}}$) x (1-U) x (1-fatality).

Sensitivity analysis

We considered a conservative and an optimistic scenario for benefit, since only observational data were available. Further, a constant relative risk reduction assumes a linear relationship between benefit and risk, where benefit is absent for patients with no risk, and maximal for patients with 100% risk of cancer mortality. Alternatively we considered a nonlinear relationship between benefit and risk, in which benefit is absent for patients with no risk or 100% risk and maximal for patients with a 50% risk of cancer mortality. We determined the threshold for such a parabolic relation between benefit and risk, for both the optimistic and conservative scenario. Finally, we calculated treatment thresholds for more intensive therapy when benefit and harm were varied over wide ranges.

6.3 Results

Benefit

The three observational studies on which our estimate of benefit of HD-CT was based are presented in Table 1. Sonneveld et al reported 10-year disease specific survival of 66% for 22 patients treated with SD-CT in their hospital between 1987 and 1996 [16]. A RCT comparing standard-dose bleomycin-etoposide-cisplatin (BEP) with standard-dose etoposide-ifosfamide-cisplatin (VIP) reported a 5-year overall survival of 60% for 181 poor prognosis patients [13]. Schmoll et al reported 5-year survival of 73% for 182 patients treated with HD-CT between 1993 and 1999 [5]. Combined, the 203 patients treated with SD-CT had an estimated 10-year risk of cancer mortality of 43%, which was substantially higher than that for the 182 patients treated with HD-CT chemotherapy (10-year risk of cancer mortality 28%). The pooled estimate of benefit is 35% (RRR=1 – (28%/43%)). For our conservative scenario we assume a benefit of 30% and for our optimistic scenario a benefit of 40%.

Table 1 Survival and early treatment related death in nonseminomatous germ cell cancer patients treated with high-dose (HD) or standard-dose (SD) chemotherapy

Reference	Tx	Year Tx	N	$S_{OVERALL}$	F-up	$R_{OV\text{-}MORT}$	Early toxic death[2]	$R_{C\text{-}MORT}$	$R_{C\text{-}MORT\ 10\ yrs}$
Hinton et al [13]	SD	1987-1992	181	60%	5	40%	3%	37%	44%
Sonneveld et al [16]	SD	1987-1996	22	66%[1]	10	34%	NA	31%	31%
Schmoll et al [5]	HD	1993-1999	182	73%[1]	5	27%	4%	23%	28%

Tx = treatment
$S_{OVERALL}$ = Overall survival at year of follow up
F-up = follow up in years
$R_{OV\text{-}MORT}$ = Risk of overall mortality at year of follow up
$R_{C\text{-}MORT}$ = Risk of cancer mortality at year of follow up
$R_{C\text{-}MORT\ 10\ yrs}$ = Risk of cancer mortality 10 year after treatment
1 disease specific survival
2 early toxic death [5]: neutropenic infections (decreased white blood cells) and septic multi-organ failure. HD toxic death: any death occurring within 100 days from grafting and not directly related to the disease itself

Harm

Early treatment related mortality was 3% for patients treated with SD-CT in a RCT [17]. This is concordant with an early treatment related mortality of 3% reported in other series [18,19]. HD-CT early treatment related mortality was 4%. The European Group for Blood and Marrow Transplantation (EBMT) Solid Tumours registry has recently reported an update of the mortality rate of germ cell tumour patients treated in Europe between 1990 and 1999. The rate of toxic death, defined as any death occurring within 100 days from grafting and not related to the disease itself, declined from 8% in 1990 to 3% in 1999 (overall 5%) [19]. We estimate the excess early treatment mortality as 1% (4-3%).

Table 2 lists the most common complications due to treatment of non-seminomatous germ cell cancer [10,20]. For each complication the incidence for SD-CT and HD-CT is given and the suspected agent. Leukaemia is the main cause of late treatment mortality in patients treated for NSGCT. More patients are expected to develop leukaemia after HD-CT than SD-CT (1.5 vs. 0.5%). With a mortality of 70% for leukaemia, this results in a difference in late treatment mortality of 0.7% [10,21,22].

Table 2 Incidence, mortality and utility of long term complications due to high-dose (HD) or standard-dose (SD) chemotherapy for nonseminomatous germ cell cancer

Morbidity (references)	Incidence		Suspected agent	Mort	Δ Mort[3]	Utility[4]	Δ Morb[5]
	SD	HD					
Therapy related leukaemia (10,21,22)	0.5%	1.5%	Etoposide (< 2 g/m^2 > 2 g/m^2)	70%	0.7%	0.90	0.03%
Vascular toxicity (10,15,23,24)							
Raynaud's phenomenon	25%	>25%	Bleomycin			-	-
Cardiovascular disease	7%	10%	Cisplatin	10%	0.3%	0.7	0.81%
Neurotoxicity (5,10,20)							
Peripheral neuropathy	4%	5%	Cisplatin			-	-
Ototoxicity	5%	65%	Cisplatin (<400 mg/m^2, > 400 mg/m^2)			-	-
Nephrotoxicity (5,10,15)							
Renal failure	1%	4%	Cisplatin			0.6	1.2%
Hypertension	10%	24%	Cisplatin (<400 mg/m^2, > 400 mg/m^2)			0.99	0.14%
Gonadal toxicity (10,15,25,26)							
Infertility[1]	50%	>50%	Cisplatin				
Sexual functioning[2]	15%	27%				0.92	0.96%
Total					1%		3.14%

[1] oligospermia/azoospermia
[2] sexual dissatisfaction
[3] Δ Mortality calculated as (incidence $_{HD\text{-}CT}$ – incidence$_{SD\text{-}CT}$) x fatality
[4] Utility ranges from 0-1and is a measure of health related quality of life
[5] Δ Morbidity calculated as (incidence $_{HD\text{-}CT}$ – incidence $_{SD\text{-}CT}$) x (1-U) x (1-fatality)

Cardiovascular disease further contributes to treatment mortality of patients treated for NSGCT [10,23,24]. The incidence of cardiovascular disease is estimated as 7% for SD-CT patients. We estimated the incidence of cardiovascular disease at, 10% for HD-CT patients, although no firm empirical estimates were available for HD-CT. With a fatality of 10% this results in 0.3% excess mortality. The combination of early and late treatment related mortality resulted in an estimated harm of 2%.

Other long term complications vary from relatively mild (Raynaud's phenomenon, ototoxicity) to severe (renal failure) [5,10,26]. In estimating the difference in long term morbidity between SD-CT and HD–CT we only took the more severe complications into account. No utility was known for acute myeloid leukaemia. Although physical and emotional functioning of long term leukaemia survivors is near normal, sexual functioning and fertility is often affected [27]. We therefore estimated a utility of 0.9 for treatment related leukaemia.

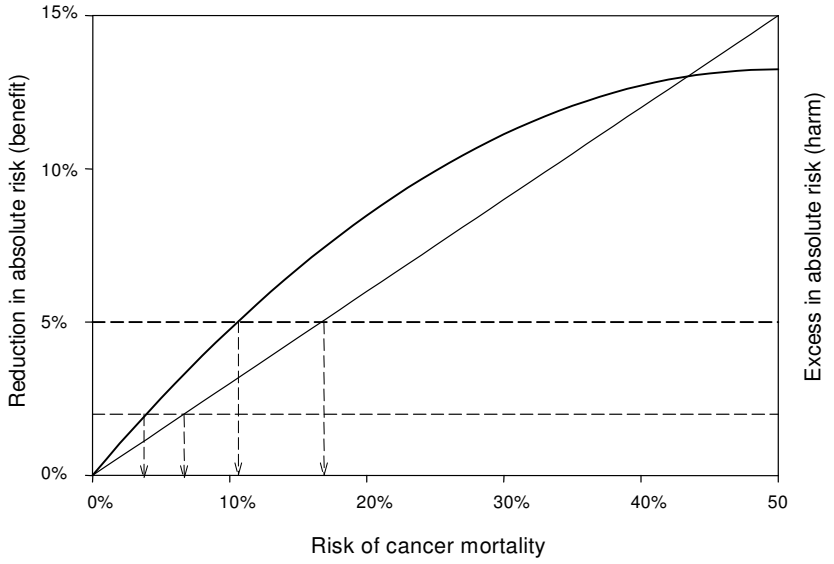


Figure 2 Linear benefit (30%) and nonlinear benefit (30%) vs. harm of high-dose chemotherapy, with harm defined as 10-year treatment related mortality (2%) or mortality plus morbidity (5%). The arrows indicate the thresholds to define poor prognosis (7% and 17% respectively for linear benefit, 4% and 11% respectively for nonlinear benefit)

The overall difference in utility weighted long term morbidity was 3.1%. The total harm due to HD-CT was approximately 5% (excess mortality 2% + excess morbidity 3.1%).

Treatment thresholds for HD-CT

At a benefit of 30% and only treatment related mortality included in our estimate of harm (2%), patients with only 7% risk of cancer mortality or higher should be treated with HD-CT (Figure 2). With a benefit of 40% the treatment threshold was as low as 5%.

When we also take treatment related morbidity into account in our estimate of harm (5%) and benefit is 30%, patients with a 17% risk of cancer mortality or higher should be treated with HD-CT (Figure 2). With a benefit of 40% the treatment threshold was 12.5%.

When we assumed a nonlinear benefit of 30% and a harm of either 2 or 5% treatment thresholds were 4 and 11% respectively (Figure 2). With a nonlinear benefit of 40% threshold values were below 10% (3 and 8% respectively).

The estimates of benefit and harm determine the treatment thresholds as shown in Figure 3 for treatment benefits from 0 to 50% and harms from 0% to 40%. For nonseminomatous germ cell cancer patients an estimated benefit of 30% and harm of 5% resulted in a threshold of 17% (block 1). When we assumed a benefit of 40%, with the same harm of 5%, the threshold decreased to 12.5% (block 2). The same threshold could be obtained with a smaller benefit, and a much smaller harm, for example 10% and 1% (block 3). We could also consider more harmful therapies, which would naturally only be considered for types of cancer with a very poor prognosis. With harm as high as 20% and a benefit of 50%, the treatment threshold for such patients is a 40% risk of cancer mortality (block 4).

6.4 Discussion

We illustrated how decision analysis can explicitly assist in defining poor prognosis testicular cancer patients who have a net benefit of high-dose chemotherapy (HD-CT) with stem cell support. Based on the currently available literature we considered a conservative estimate of 30% for the benefit and an optimistic estimate of 40%. We estimated a harm of 5%, based on both treatment mortality (2%) and treatment morbidity expressed in utilities (3%). Even with a

conservative estimate of 30% for the benefit of treatment, and taking both treatment related mortality and morbidity into account, patients with a risk of cancer mortality of 17% or higher might already benefit from HD-CT. With a benefit of 40% this threshold was reduced to 12.5%. When we assumed benefit to be nonlinear, treatment thresholds were 11 and 8% for benefit of 30 and 40% respectively.

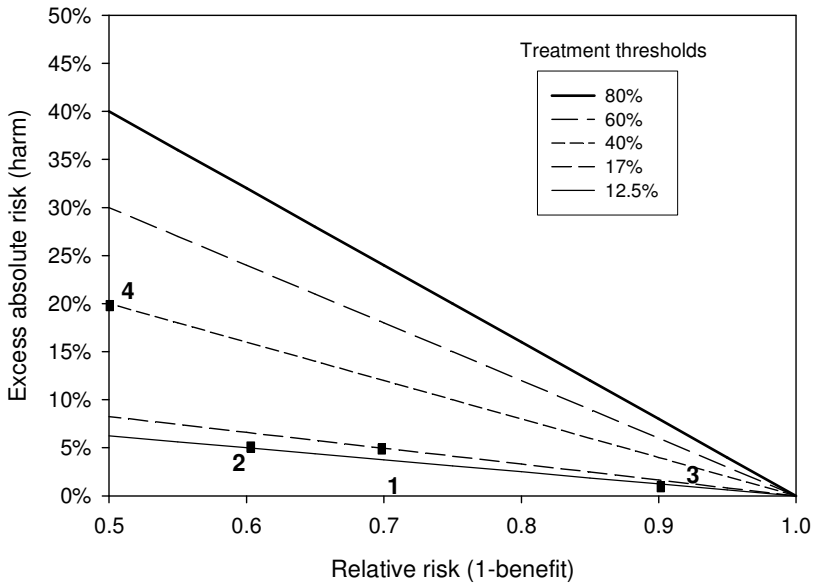


Figure 3 Thresholds according to risk with standard treatment for a range of hypothetical benefits (reduction in relative risk, RR) and harms associated with a more intensive treatment.
1. benefit 30%, harm=5%, threshold=17% (– – –) , 2. benefit 40%, harm=5%, threshold = 12.5% (—), 3. benefit 10%, harm=1%, threshold = 12.5% (—), 4. benefit 50%, harm=20%, threshold = 40% (- - -).

To what extent is the group of patients above the threshold comparable to the poor prognosis patients as defined by the IGCC classification? The 5-year survival for the good, intermediate and poor prognosis patients was originally reported as 92, 80 and 48% respectively [11]. However, year of treatment was ignored, and 2154 out of 5202 patients with missing values were discarded from the analysis. We have recalculated the expected survival for the IGCC prognosis groups with

statistical adjustment to the year 1990 and with consideration of all 5202 patients by imputation of missing values, and found 10-year estimates of 93, 84 and 57% respectively [28]. These numbers are confirmed in more recently reported series [13,16]. Also with these revised estimates, risk of death from cancer in the poor prognosis group is clearly above the threshold, confirming that these patients are likely candidates for HD-CT.

The patients in the intermediate prognosis group have mortality risks around the threshold. We previously modelled the 5 risk factors of the IGCC classification in more detail in a Cox regression model [29]. The risk estimates from this model show some spread for the 862 patients classified as intermediate prognosis: 580 (67%) have risks lower than 17%, and 282 (33%) have risks higher than 17%. However, only 97 (11%) had 10-year mortality risks higher than 20%, which would more strongly support considering them for HD-CT. On the other hand a minor fraction of the IGCC category 'poor prognosis' had modelled risks below 17% (43/495, 9%).

The IGCC classification and our decision analysis hence largely agree on which patients are candidates for HD-CT. In the future, a more refined prognostic classification is however desirable, with prognostic groups defined in more detail and with more powerful predictors, e.g. new biomarkers [30,31].

Although we considered a conservative and optimistic estimate of the benefit of HD-CT our estimate may still be too optimistic. Differences in treatment other than HD-CT may have affected the difference in survival between patients treated with SD-CT and patients treated with HD-CT. Firstly the patients treated with SD-CT were mainly treated in the US whereas patients treated with HD-CT were treated in Germany. However, the estimated risk of cancer mortality for SD-CT is in line with the IGCC survival estimate for poor prognosis patients adjusted for year of treatment, which is based on patients treated in both Europe and the US [28]. Secondly, patients treated with SD-CT were treated earlier than patients treated with HD-CT. Improvements over time in second line treatment may have effected the difference in survival [17].

Our estimate of harm may be too low. We estimated harm due to treatment related mortality and morbidity at 10 years after treatment. Direct estimates of early treatment mortality were available for both SD-CT and HD-CT. However information on long term complications is merely available for SD-CT, and limited for HD-CT. As a consequence our estimate of etoposide induced leukaemia, which is very difficult to cure, may be too low.

Similarly, the harm due to complications such as cardiovascular disease and hypertension may be higher since they pose a lifetime risk. Finally, little is known about the harm due to chronic fatigue and neuropsychological sequelae [17]. Figure 3 helps to directly calculate the risk threshold if more conservative assumptions are made. For example, when the relative risk reduction due to HD-CT is only 20% and the harm 8%, only patients with at least a 40% risk will benefit from more intensive treatment.

Our analysis has some other limitations. To compare harm and benefit of HD-CT we expressed both in 10-year risks, without considering the time of the event since treatment (early or late). This is a simplification. An alternative would be a more extensive decision analysis, in which expected life years and the probability of complications are modelled, e.g. with a Markov model with yearly cycles [32]. However given the uncertainty in the estimates of harm and benefit such a more complicated model was not considered desirable.

We also did not consider costs of HD-CT or SD-CT. There are currently no data available on the difference in costs between HD-CT and SD-CT for testicular cancer patients but in other diseases, such as non-Hodgkin's lymphoma, multiple myeloma and breast cancer, the costs of HD-CT have been reported to be one to four times higher than SD-CT [33]. Hence, HD-CT needs to have a substantial net benefit to be relevant from a societal perspective.

Evidence of the benefit of HD-CT as first line treatment in the literature has not been conclusive, and the results of two ongoing RCTs have to be awaited for more reliable decision making. One RCT by the EORTC (BEP vs. high-dose VIP) is still including poor prognosis patients [8]. The inclusion of intermediate and poor prognosis patients for an RCT by the US intergroup (BEP vs. high-dose CEC) has closed and preliminary results have been presented [9,34]. There was no significant difference in complete response after 1 year between standard and high-dose chemotherapy (48 vs. 52%). We will have to await the publication of the final results of these RCTs before a more precise estimate of the benefit of HD-CT can be made.

Based on the number of patients enrolled in these trials, a relative risk reduction over approximately 50% can be detected with sufficient statistical power. This may be an optimistic estimate, and results of the trials may be inconclusive when HD-CT in fact has a smaller effect. Our analysis suggests that HD-CT may not be beneficial for the full group of intermediate prognosis patients, especially because of excess long term mortality and morbidity. Special attention should be given to

the intermediate prognosis patients in the analysis of the RCT that includes these patients [9]. Further, it is important that more precise information becomes available on the long term complications of HD-CT by longer follow up, since testicular cancer occurs mostly at a young age.

Besides HD-CT other approaches are being investigated to improve survival of NSGCT patients, such as dose intensification and the introduction of new agents [35-37].

A recently published phase II trial investigating the intensive induction chemotherapy carboplatin, bleomycin, vincristine, cisplatin + bleomycin, etoposide, cisplatin (C-BOP/BEP) showed promising results with 2-year survival of 94 and 85% for intermediate and poor prognosis patients respectively. However, 2-year progression free survival was much lower for poor prognosis patients (56%) suggesting that the benefit will be smaller at 5 or 10-year follow up [38].

Furthermore the EORTC currently conducts a RCT targeted especially at intermediate prognosis patients which investigates the combination of paclitaxel with BEP (T-BEP) [39]. The results of these trials can be incorporated in the decision analytic approach described in this study to determine which treatment is optimal at what harm and benefit.

In conclusion, we illustrated how decision analysis can support treatment choices on more intensive therapy. From the decision analysis we learn at what risk a treatment becomes beneficial. A prognostic model or prognostic classification can then be used to estimate the risk of an individual patient or a subgroup of patients. This approach can be adapted to new results from ongoing trials and extended to many other cancers to explicitly define candidates for more intensive treatments. Hence, patients who are expected to benefit will be treated more intensively, without overtreatment of those at relatively low risk, and patients who are not expected to benefit will be treated in a more standard way, without undertreatment of those at relatively high risk.

Acknowledgements

We would like to thank the members of the International Germ Cell Cancer Collaborative Group (IGCCCG), for kindly providing their data for the analyses in this paper. Members of the IGCCCG and participating centres were listed previously [40].
This work was supported by the Netherlands Organisation for Scientific Research (MvD) and by the Royal Netherlands Academy of Arts and Sciences (ES).

References

1. Piccirillo JF, Tierney RM, Costas I, Grove L, Spitznagel EL, Jr. Prognostic importance of comorbidity in a hospital-based cancer registry. *Jama* 2004;291(20):2441-7.

2. Simon R, Altman DG. Statistical aspects of prognostic factor studies in oncology. *Br J Cancer* 1994;69(6):979-85.

3. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15(4):361-87.

4. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *Bmj* 1995;311(7016):1356-9.

5. Schmoll HJ, Kollmannsberger C, Metzner B, et al. Long-term results of first-line sequential high-dose etoposide, ifosfamide, and cisplatin chemotherapy plus autologous stem cell support for patients with advanced metastatic germ cell cancer: An extended phase I/II study of the German testicular cancer study group. *Journal of Clinical Oncology* 2003;21(22):4083-4091.

6. Bokemeyer C, Harstrick A, Beyer J, et al. First-line high-dose chemotherapy for 'poor risk' metastatic non-seminomatous testicular germ cell tumors. *Onkologie* 1998;21:23-25.

7. Bokemeyer C, Oechsle K, Hartmann JT, et al. Treatment-induced anaemia and its potential clinical impact in patients receiving sequential high dose chemotherapy for metastatic testicular cancer. *Br J Cancer* 2002;87(10):1066-71.

8. EORTC. Phase III Randomized Study of Standard Cisplatin, Etoposide, and Ifosfamide (VIP) Followed By Sequential High-Dose VIP and Stem Cell Rescue Versus Bleomycin, Etoposide, and Cisplatin (BEP) in Chemotherapy-Naive Men With Poor-Prognosis Germ Cell Cancer, Accessed June 1, 2006.

9. US intergroup. Phase III Randomized Study of Bleomycin, Etoposide, and Cisplatin (BEP) With or Without High-Dose Carboplatin, Etoposide, and Cyclophosphamide Plus Autologous Bone Marrow or Peripheral Blood Stem Cell Transplantation in Male Patients With Previously Untreated Poor- or Intermediate-Risk Germ Cell Tumors. Accessed June 1, 2006.

10. Chaudhary UB, Haldas JR. Long-term complications of chemotherapy for germ cell tumours. *Drugs* 2003;63(15):1565-77.

11. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *J Clin Oncol* 1997;15(2):594-603.

12. de Wit R, Roberts JT, Wilkinson PM, et al. Equivalence of three or four cycles of bleomycin, etoposide, and cisplatin chemotherapy and of a 3- or 5-day schedule in good-prognosis germ cell cancer: a randomized study of the European Organization for Research and Treatment of Cancer Genitourinary Tract Cancer Cooperative Group and the Medical Research Council. *J Clin Oncol* 2001;19(6):1629-40.
13. Hinton S, Catalano PJ, Einhorn LH, et al. Cisplatin, etoposide and either bleomycin or ifosfamide in the treatment of disseminated germ cell tumors: final analysis of an intergroup trial. *Cancer* 2003;97(8):1869-75.
14. Sonneveld DJ, Hoekstra HJ, Van Der Graaf WT, Sluiter WJ, Schraffordt Koops H, Sleijfer DT. The changing distribution of stage in nonseminomatous testicular germ cell tumours, from 1977 to 1996. *BJU Int* 1999;84(1):68-74.
15. Tengs TO, Wallace A. One thousand health-related quality-of-life estimates. *Med Care* 2000;38(6):583-637.
16. Sonneveld DJA, Hoekstra HJ, van der Graaf WT, et al. Improved long term survival of patients with metastatic nonseminomatous testicular germ cell carcinoma in relation to prognostic classification systems during the cisplatin era. *Cancer* 2001;91(7):1304-1315.
17. Bokemeyer C, Kollmannsberger C, Meisner C, et al. First-line high-dose chemotherapy compared with standard-dose PEB/VIP chemotherapy in patients with advanced germ cell tumors: A multivariate and matched-pair analysis. *J Clin Oncol* 1999;17(11):3450-6.
18. Bosl GJ, Motzer RJ. Testicular germ-cell cancer. *N Engl J Med* 1997;337(4):242-53.
19. Rosti G, Ferrante P, Ledermann J, et al. High-dose chemotherapy for solid tumors: results of the EBMT. *Crit Rev Oncol Hematol* 2002;41(2):129-40.
20. Bokemeyer C, Berger CC, Kuczyk MA, Schmoll HJ. Evaluation of long-term toxicity after chemotherapy for testicular cancer. *J Clin Oncol* 1996;14(11):2923-32.
21. Kollmannsberger C, Hartmann JT, Kanz L, Bokemeyer C. Therapy-related malignancies following treatment of germ cell cancer. *Int J Cancer* 1999;83(6):860-3.
22. Xie Y, Davies SM, Xiang Y, Robison LL, Ross JA. Trends in leukemia incidence and survival in the United States (1973-1998). *Cancer* 2003;97(9):2229-35.
23. Huddart RA, Norman A, Shahidi M, et al. Cardiovascular disease as a long-term complication of treatment for testicular cancer. *J Clin Oncol* 2003;21(8):1513-23.
24. Meinardi MT, Gietema JA, van der Graaf WT, et al. Cardiovascular morbidity in long-term survivors of metastatic testicular cancer. *J Clin Oncol* 2000;18(8):1725-32.
25. Hartmann JT, Albrecht C, Schmoll HJ, Kuczyk MA, Kollmannsberger C, Bokemeyer C. Long-term effects on sexual function and fertility after treatment of testicular cancer. *Br J Cancer* 1999;80(5-6):801-7.

26. Jonker-Pool G, Van de Wiel HB, Hoekstra HJ, et al. Sexual functioning after treatment for testicular cancer–review and meta-analysis of 36 empirical studies between 1975-2000. *Arch Sex Behav* 2001;30(1):55-74.

27. Redaelli A, Stephens JM, Brandt S, Botteman MF, Pashos CL. Short- and long-term effects of acute myeloid leukemia on patient health-related quality of life. *Cancer Treat Rev* 2004;30(1):103-17.

28. van Dijk MR, Steyerberg EW, Stenning SP, Habbema JD. Survival estimates of a prognostic classification depended more on year of treatment than on imputation of missing values. *J Clin Epidemiol* 2006;59(3):246-53.

29. van Dijk MR, Steyerberg EW, Stenning SP, Dusseldorp E, Habbema JD. Survival of patients with nonseminomatous germ cell cancer: a review of the IGCC classification by Cox regression and recursive partitioning. *Br J Cancer* 2004;90(6):1176-83.

30. Donadio AC, Bosl GJ. The future of therapy for nonseminomatous germ cell tumors. *Chest Surg Clin N Am* 2002;12(4):769-89.

31. Heidenreich A, Srivastava S, Moul JW, Hofmann R. Molecular genetic parameters in pathogenesis and prognosis of testicular germ cell tumors. *Eur Urol* 2000;37(2):121-35.

32. Sonnenberg FA, Beck JR. Markov models in medical decision making: a practical guide. *Med Decis Making* 1993;13(4):322-38.

33. Simnett SJ, Stewart LA, Sweetenham J, Morgan G, Johnson PW. Autologous stem cell transplantation for malignancy: a systematic review of the literature. *Clin Lab Haematol* 2000;22(2):61-72.

34. Bajorin DF, Nichols CR, Margolin KA, et al. Phase III trial of conventional-dose chemotherapy alone or with high-dose chemotherapy for metastatic germ cell tumors (GCT) patients (PTS): A cooperative group trial by Memorial Sloan-Kettering Cancer Center, ECOG, SWOG, and CALGB. *J Clin Oncol*;2006 ASCO Annual Meeting Proceedings Part I. Vol 24, No. 18S (June 20 Supplement), 2006: 4510.

35. Bower M, Newlands ES, Holden L, Rustin GJ, Begent RH. Treatment of men with metastatic non-seminomatous germ cell tumours with cyclical POMB/ACE chemotherapy. *Ann Oncol* 1997;8(5):477-83.

36. Germa-Lluch JR, Garcia del Muro X, Tabernero JM, et al. BOMP/EPI intensive alternating chemotherapy for IGCCC poor-prognosis germ-cell tumors: the Spanish Germ-Cell Cancer Group experience (GG). *Ann Oncol* 1999;10(3):289-93.

37. Kaye SB, Mead GM, Fossa S, et al. Intensive induction-sequential chemotherapy with BOP/VIP-B compared with treatment with BEP/EP for poor-prognosis metastatic nonseminomatous germ cell tumor: a Randomized Medical Research Council/European

Organization for Research and Treatment of Cancer study. *J Clin Oncol* 1998;16(2):692-701.

38. Fossa SD, Paluchowska B, Horwich A, et al. Intensive induction chemotherapy with C-BOP/BEP for intermediate- and poor-risk metastatic germ cell tumours (EORTC trial 30948). *Br J Cancer* 2005.

39. EORTC. Phase II/III Randomized Study of Bleomycin, Cisplatin, and Etoposide (BEP) Versus Bleomycin, Cisplatin, Etoposide, and Paclitaxel (T-BEP) in Men With Intermediate Prognosis Germ Cell Cancer, Accessed June 1, 2006.

40. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor- based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *J Clin Oncol* 1997;15(2):594-603.

7
Discussion

This thesis describes methodological aspects of prognostic classifications in oncology, using the IGCC classification for patients with nonseminomatous germ cell cancer for illustration.
We evaluated the validity of the IGCC classification, with respect to assumptions made in the development of the classification and the generalisability of the survival estimates for currently diagnosed patients. Furthermore we studied alternative ways of defining prognostic groups, especially for poor prognosis patients.
In this chapter the findings of our studies are discussed according to the research questions as specified in Chapter 1. We end with conclusions and recommendations for future research.

7.1 Answers research questions

Research question 1: Are the assumptions made in the development of the IGCC classification valid with regard to the inclusion of prognostic factors, or can discriminative ability be improved?
Answer: Incorporating differences in predictive strength, or specifying interaction terms did not result in an increase in discriminative ability over the original IGCC classification. Hence we support the validity of the IGCC classification to define prognosis groups.

Explanation: The IGCC classification did not consider differences in strength of prognostic factors and made no distinction between the number of prognostic factors within a prognosis group. Better discrimination might be achieved by incorporating differences in predictive strength and testing specific interaction terms.
Differences in predictive strength*:* Simplifications in the modelling process, such as categorising continuous risk factors and using strongly rounded regression weights, are usually regarded as undesirable as the associated loss in information can result in a decrease in predictive performance [1,2]. Our Cox regression analysis demonstrated differences in predictive strength between the IGCC prognostic factors. But an alternative classification that took these differences into account did not perform better than the IGCC classification in defining a poor, an intermediate and a good prognosis group.

Of the 3048 patients with complete data, 204 patients (7%) were classified differently by the Cox regression model compared to the IGCC classification; 97 intermediate prognosis patients were reclassified as poor prognosis, while 107 poor prognosis patients were reclassified as intermediate prognosis (Figure 1). However, 5-year survival of these patients was identical (69 vs. 69%), and re-classification did hence not result in a better performance. So, the differences could affect the patient at the individual level with respect to treatment choice but did not change the overall performance.

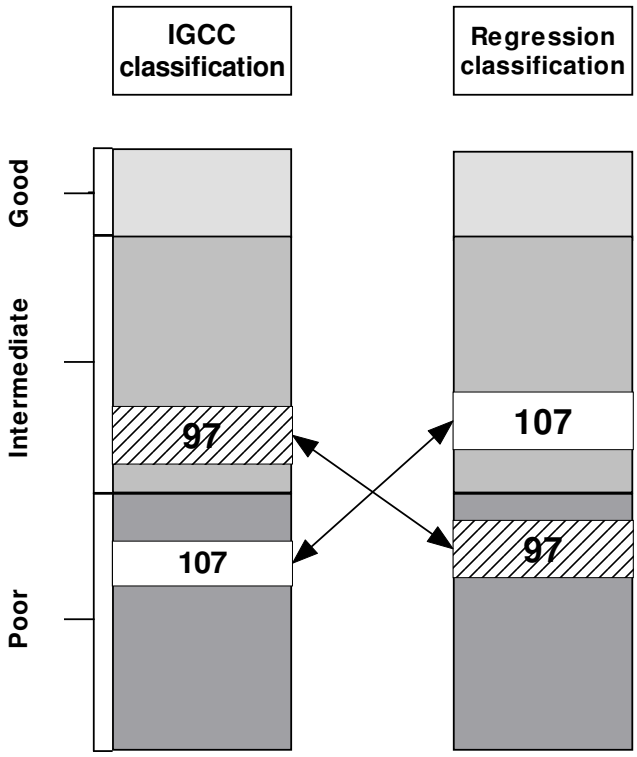


Figure 1 Classification of 3048 nonseminomatous germ cell cancer patients in the IGCC database according to the IGCC classification and a Cox regression based classification

Interactions between IGCC prognostic factors: Regression models assume additivity of the effects of their prognostic factors [2]. The effect of a prognostic factor is assumed not to modify the effect of other prognostic factors. In Cox regression, additivity refers to the additivity of regression coefficients on the log hazard scale. Non-additivity is taken into account by including interaction terms in the regression model. Usually we only consider two-way interactions, where a significant interaction term indicates that the effect of the prognostic factors together is smaller or larger than the sum of separate effects of two prognostic factors.

The IGCC classification considers the presence of more than one intermediate or poor prognostic factors as equal to having only one intermediate or poor prognostic factor respectively. This is an example of a 'max rule', which assumes a negative interaction between prognostic factors. We tested the implicit non-additivity assumption in the IGCC classification in two ways. Interaction terms were included in Cox regression models. And we applied regression tree modelling which is very suitable of capturing non-additive effects of prognostic factors [3].

Both methods showed that the tumour marker AFP was less important than the other IGCC prognostic factors, while the IGCC classification assumed that all prognostic factors were equally important. The tree model did not include AFP, while in the Cox regression analysis the effect of AFP was modified by the presence of primary site and NPVM. The limited contribution of AFP was also demonstrated by the much lower weight of AFP, indicating less predictive strength, in the regression based classification. The alternative classifications did not classify patients better than the IGCC classification in terms of discriminative ability.

There were some limitations in testing the non-additivity assumption of the IGCC classification. Tree modelling only allows for binary splits of variables. The tumour markers were split in the categories good/intermediate vs. poor, which limits comparability with the IGCC classification in which all three tumour marker categories are considered. Second, in a Cox regression analysis only two-way interactions can readily be interpreted, while the IGCC classification assumes more complicated higher order interactions between the prognostic factors.

Research question 2: What is the effect of missing values on survival estimates of the IGCC classification?
Answer: Multiple imputation of the missing values in the IGCC database led to lower survival estimates across the IGCC prognosis groups, compared with estimates based on the complete data. This was explained by a correlation between missingness and year of treatment, while year of treatment was associated with survival.

Explanation: In the development of the IGCC classification 2154 (41%) of 5202 patients were excluded. Remarkably, only 2388 (9%) of all 26010 required data values of the 5 IGCC risk factors were missing (5 x 5202). Imputation of the 9% missing values added 32% observed values to the analysis. Exclusion of patients because of missing data was statistically inefficient. Moreover, it could have led to bias in the survival estimates of the prognostic groups in the IGCC classification if missingness was not completely at random [4]. Patients with missing values had poorer survival.
Missingness was mostly caused by missing values for LDH. This tumour marker was not systematically collected by all participating centres before 1985, since its prognostic value was not yet fully recognized at that time.
Imputation of missing values resulted in lower survival estimates across the IGCC prognosis groups, compared with the analysis of the 3048 patients with complete data. This difference could largely be explained by year of treatment, since an earlier year of treatment was both related to missingness and to a lower survival. Year of treatment hence acted as a kind of confounder (Figure 2).
The imputation of missing values made us aware of the relevance of year of treatment, independent of the use of complete case analysis as the statistical method to handle missing values. The survival estimates reported by the IGCCCG are therefore not valid for current patients, and should be adjusted for year of treatment.

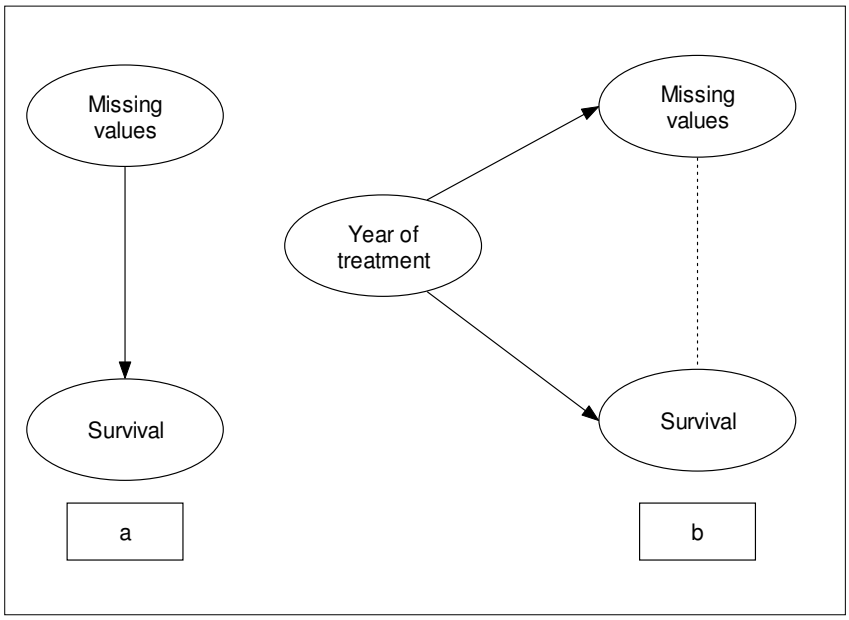


Figure 2 Relation between missing values and survival. Patients in the IGCC database with missing values have a lower survival than patients with no missing values (a). This difference is explained by year of treatment; patients with missing values were treated earlier, when survival was lower (b)

Handling of missing values in prognostic studies: A recent review demonstrated that missing values are a common problem in prognostic studies in oncology [5]. In this review 100 articles were considered, published in 2002 and selected from 7 clinical cancer journals with high impact factors

Missing values occurred in almost all studies reviewed. In 81 articles missing values were present, and another 13 articles had availability of data as an inclusion criteria but did not report how many patients were excluded. In 52 articles the number of missing values for each variable was given, while the number of complete cases used for analysis was given in only 39 studies.

When a method for handling missing data was reported this was most often complete case analysis. Three articles used single imputation, while only one article used multiple imputation. Finally, the possible reasons for missingness were

discussed in 21 articles, while only 10 studies investigated possible differences in characteristics or outcome between patients with and without missing values.
These results show that many researchers are not fully aware of how to deal with missing values in prognostic studies. Guidelines were proposed for reporting prognostic studies with missing data (Table 1) [5]. Table 1 shows that the proposed guidelines focus on 3 main issues: quantification of the completeness of predictor data, approaches to dealing with missing predictor data (including imputation methods), and exploration of the missing data (e.g. comparing for complete and incomplete cases).

Table 1 Guidelines for reporting prognostic studies with missing covariate data [5]

Quantification of completeness of covariate data
• If availability of data is an inclusion criterion, specify the number of patients excluded for this reason • Provide the total number of eligible patients and the number with complete data • Report the frequency of missing data for every variable considered
Approaches for handling missing covariate data
• Provide sufficient details of the methods adopted to handle missing covariate data for all incomplete covariates • Give appropriate references for any imputation method used • For each analysis, specify the number of patients included and the associated number of events
Exploration of the missing data
• Discuss any known reason for missing covariate data • Present the results of any comparisons of characteristics between the cases with and without missing data.

Use of imputation techniques: Multiple imputation is a state of the art method for handling missing data. It is preferred over single imputation since it takes the uncertainty of the imputed values into account.
However, the application of multiple imputation is complex and requires knowledge of advanced statistical software. Under certain conditions single imputation may therefore be a reasonable method for handling missing data. Also, a complete case analysis can sometimes be performed.

Complete case analysis could be applied when all of the following conditions are met:

- The number of missing values (per variable) is limited, and therefore the number of excluded patients is small (e.g. <10%)
- There is no difference in outcome or patient characteristics between patients with and without missing values
- Missing values can be considered missing completely at random (MCAR), i.e. they are a random sample of the whole dataset.

Multiple imputation should be applied when any of the following occurs:

- The number of missing values (per variable) is large resulting in the exclusion of many patients (e.g. >30%).
- There is a substantial difference in outcome of patients with and without missing values

Single imputation can be considered for intermediate situations, e.g. around 20% missing values, but not completely at random. Any imputation method (single or multiple) has to assume that missing values are missing at random (MAR). This means that missingness only depends on other variables in the dataset. The MAR assumption is not testable, but becomes more reasonable with imputation models that include a wide range of characteristics, including all potential prognostic factors, auxiliary variables (such as year of treatment or treatment centre), and the outcome (e.g. survival time and the censoring variable) [6].
Including the outcome may appear a bit circular, since the aim of a prognostic model is to predict outcome. However, it can easily be shown that not including the outcome in the imputation model causes bias, even in the MCAR situation.

Research question 3: Has survival of patients with advanced testicular cancer improved since the introduction of the IGCC classification?
Answer: Yes, survival estimates reported by the IGCC investigators are lower than survival of patients currently diagnosed with advanced testicular cancer.

Explanation: The survival estimates of the IGCC classification were based on patients treated between 1975 and 1990. A systematic review of the literature on studies reporting on survival of patients with nonseminomatous germ cell cancer treated after 1989 showed an increase in survival especially for poor prognosis patients. For these patients we estimated a 5-year survival of 71% instead of 48%.

Even within the IGCC data survival increased over time; when survival estimates were adjusted in a Cox model for the last year of treatment (1990) 5-year survival was 61% for poor prognosis patients (Figure 3).

The increase in survival for poor prognosis patients is most likely due to more effective treatment and to more experience in treating patients with nonseminomatous germ cell cancer [7]. Which of these factors is most responsible for the increase in survival, should become clear from randomised controlled trials. In such trials patients in both treatment arms profit from improved experience in treating patients, and hence a pure effect of treatment can be determined.

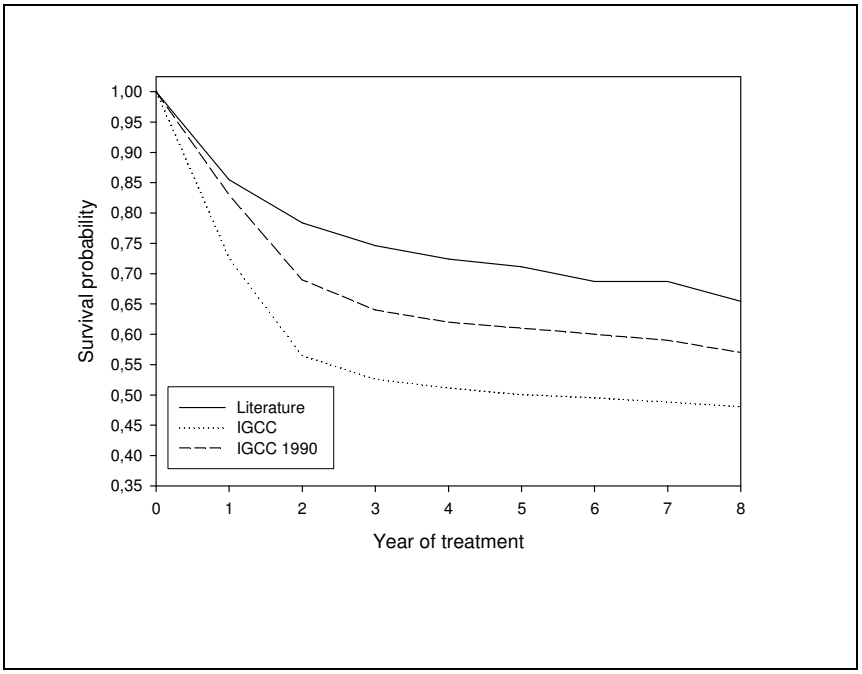


Figure 3 Survival estimates of poor prognosis nonseminomatous germ cell cancer patients according to a) a meta-analysis of available literature on patients treated since 1990 (5-year survival 71%), b) patients in the IGCC database adjusted for year of treatment 1990 (5-year survival 61%), c) patients in the IGCC database (5-year survival 48%)

Implications for interpretation of clinical trials: Survival estimates from the IGCC data [8] are currently the historical control for the comparison of results from nonrandomised Phase I/II trials in which new treatment strategies for patients with advanced testicular cancer are studied. However, this comparison leads to a too optimistic impression of a new treatment, since the survival of currently diagnosed patients is most likely better than reported by the IGCC investigators.
We can assume that 5-year survival of poor prognosis patients diagnosed nowadays will lie between 61 and 71%. However, a more reliable estimate can be obtained from patients treated with standard chemotherapy in current RCTs. Until then, results of nonrandomised trials should not be compared with the 48% reported by the IGCCCG.

Research question 4: Is regression tree analysis an appropriate method for further subgrouping within poor prognosis patients?
Answer: No, regression tree analysis leads to unstable and optimistic models, and is therefore not appropriate for identifying subgroups in prognostic classifications.

Explanation: Regression tree analysis is a nonlinear method to define groups of patients differing in prognosis. It is more flexible than standard regression approaches, especially in detecting complex interactions between prognostic factors. The German Testicular Cancer (GTC) group used this method for further subgrouping of poor prognosis patients to allow for a more refined identification of individuals patients at high risk, eligible for high-dose chemotherapy [9].
Predictive performance of this tree was lower than expected when applied to the IGCC data. A new tree model developed with similar methods as the GTC group differed in the selection and hierarchy of prognostic factors. Internal validation of this new tree showed a large degree of optimism in performance.

Linear modelling versus nonlinear modelling: In our analyses, prognostic classifications based on linear models (Cox regression) outperformed prognostic classifications based on nonlinear models (regression tree analysis) both in small (n=332) and large (n=3048) datasets (Chapters 5 and 2 respectively).
Previous studies confirm that linear models may perform just as well or better than nonlinear models in prognostic studies. Ennis et al (1998) compared a previously developed logistic regression model to predict 30-day mortality, based on 40830 patients, with nonlinear models such as tree models and neural

networks [10]. They found that the nonlinear models did not outperform the relatively simple logistic regression model.
Similar results were found in another study in which the performance of Cox regression was compared with regression tree analysis and neural network in 3 large urological datasets (prostate cancer and renal cell carcinoma) [11]. Both studies conclude that these datasets apparently did not include highly predictive nonlinear or interaction effects, and that maximum performance could be reached by using a relatively simple linear model. We may wonder how often such complex effects occur in the real world, and when they are not predefined, whether we can capture them without overfitting our models to our data.
Further advantages of the Cox regression model over tree models and neural networks include insights in the prediction model (e.g. hazard ratios, significance testing) and its reproducibility. The same results are obtained each time Cox regression is applied to a dataset, while nonlinear modelling strategies may give different results because random processes are sometimes used in determining the final model.
We therefore suggest to use tree modelling for exploratory purposes only to detect possible interaction effects, which could then be considered in a regression model [12].

Research question 5: At what risk of cancer-mortality should patients with advanced testicular cancer be treated with high-dose chemotherapy?
Answer: With current estimates on benefit and harm of high-dose chemotherapy, patients with a risk of cancer mortality of 17% or higher might profit from high-dose chemotherapy.

Explanation: Based on evidence available in the literature we estimated the benefit of high-dose chemotherapy (HD-CT) with autologous stem cell support for patients with advanced testicular cancer at 30%, i.e. HD-CT results in relative reduction in risk mortality of 30%. Harm was estimated at 5%, i.e. treatment with HD-CT leads to an increase in absolute risk of mortality of 5%. Combined in a decision analysis this resulted in a treatment threshold of 17% risk of cancer specific mortality with SD chemotherapy. This threshold leads to a similar selection of patients as defined by the IGCC classification for poor prognosis patients.

The IGCCCG investigators considered several aspects in deriving their poor prognosis group. It should be simple to define, have a relatively low 5-year survival, and contain a sufficiently large number of patients. The latter aspects reflect a research perspective and not necessarily the patient's perspective. In decision making for individual patients, the net effectiveness of a therapy is of interest. This is based on weighing benefits of treatment, such as lower cancer-specific mortality, against harm of treatment, such as toxic side effects, burden, and long term risks (e.g. cardiovascular morbidity and mortality). The point where benefits and harms are equal is the treatment threshold. Decision analysis provides an appropriate and useful framework to define such treatment thresholds in prognostic classifications. Our estimates of benefit and harm of high-dose chemotherapy in 'poor prognosis patients' are yet to be confirmed in future studies. Benefit was based on results from nonrandomised trials, while there was only limited information available on long term harm of high-dose chemotherapy.

Ongoing randomised clinical trials: New treatment strategies are under study for the treatment of patients with advanced testicular cancer. Benefits and harms of these new treatments may differ from those of high-dose chemotherapy. Therefore, the IGCC poor prognosis definition might not always be appropriate to allocate patients to such treatments. Decision analysis can be used to determine a treatment threshold for each pair of treatments.

To determine these thresholds reliable estimates of benefit and harm are necessary. Such estimates are ideally based on RCTs. Only preliminary results are available from current RCTs.

For high-dose chemotherapy, two different approaches are being considered. First results of a treatment strategy investigated by the US intergroup (standard chemotherapy vs. two cycles of BEP plus two cycles of high-dose carboplatin, etoposide, cyclophosphamide with stem cell support) were presented at the 2006 American Society of Clinical Oncology (ASCO) conference [13]. There was no significant difference in complete response after 1 year between standard and high-dose chemotherapy (48 vs. 52%). This is in contrast with results from earlier nonrandomised studies, in which patients who received this treatment compared favourably to historical controls. Furthermore, the low complete response at 1 year suggests that 5-year survival will be lower than expected from our analyses (Figure 3). This is especially remarkable, since this study also included intermediate prognosis patients.

The German approach consists of standard treatment vs. one cycle cisplatin, etoposide, ifosfamide (VIP) followed by three cycles of high-dose VIP, and was used as an example in Chapter 5. It is studied in an European trial [14]. By May 2006 135 of the required 222 patients were accrued. First results are expected in 2009.

Other approaches are the addition of active drugs to standard treatment, alternating chemotherapy to prevent drugs resistance and dose dense sequential combination chemotherapy [15-17]. RCTs studying these treatments are still including patients, and results are not expected before 2008. This overview of ongoing trials demonstrates that the relative rarity of testicular cancer makes accrual of patients difficult, even though in all trials multiple centres collaborate in all these trials.

An even bigger challenge in determining treatment thresholds is to obtain reliable estimates of long term benefits and harms for these treatments. Limited information is available on risk up to 10 years, such as leukaemia, neurotoxicity and gonadal toxicity. Even less is known, both for standard treatment and alternative treatment strategies, about mechanisms influencing life long risk of for instance cardiovascular disease through hypertension and hypercholesterolaemia.

Table 2 Ongoing randomised controlled trials in patients with nonseminomatous germ cell cancer[1]

Protocol ID	Prognosis group	Treatment	Status	Start – end	N req	N ent
EORTC-30974 [14]	Poor	High-dose VIP	Active	1999 - 2009	222	135
EORTC-30983 [16]	Interm.	T-BEP	Active	1998 - 2008	498	300
MRC-TE23 [15]	Poor	CBOP/BEP	Active	2005 - 2008	84	18
FNCLCC-13/0206 [17]	Poor	BEP+dose dense CT	Active	2005 - 2012	240	55
MSKCC-94076 [13]	Poor/ Interm.	BEP + HD-carboplatin based CT	Closed	1999 - 2001	270	270

[1] Personal communication trial leaders, May 2006.
EORTC = European Organisation for research and treatment of cancer
MRC = Medical Research Council Clinical Trial Unit
MSKCC = Memorial Sloan-Kettering Cancer Center
FNCLCC = Federation Nationale des Centres de Lutte Contre le Cancer
VIP = cisplatin, etoposide, ifosfamide
BEP = bleomycin, etopside, cisplatin
T-BEP = Taxol, bleomycin, etopside, cisplatin
CBOP/BEP = cisplatin, vincristine, bleomycin, carboplatin/bleomycin, etopside, cisplatin

7.2 Conclusions and recommendations for future research

Methodological aspects of prognostic classifications

Conclusions

- Prognostic classifications should preferably be based on regression analysis, since this method is transparent, reliable and generalisable to other patients. Tree models might be used for exploratory purposes, to investigate possible interactions between prognostic factors.

- In the development of prognostic classifications missing values should be accounted for to prevent statistical inefficiency and (possible) bias. Depending on the research question, how many values are missing and the missing data mechanism, either complete case, single or multiple imputation might be used as a method for handling missing data.

- When a prognostic classification is developed with data collected over a long calendar period, there is possibly an increase of survival over time.

- Classifications may remain valid over longer periods to discriminate between good prognosis and poor prognosis patients, but require regular updating of prognostic estimates such as predicted 5-year survival.

- Decision analysis provides an appropriate and useful framework to define treatment thresholds in prognostic classifications.

Recommendations for future research

- More research is needed on how tree models can aid in detecting complex relationships between prognostic factors and how this can be used in the development of prognostic classifications.

- Single and multiple imputation should be compared in empirical studies to determine under what conditions single imputation is sufficient for handling missing data.

Prognosis in advanced testicular cancer

Conclusions

- The relatively simple IGCC classification performed just as well as more complex alternatives. The IGCC classification is a valid method for distinguishing good and poor prognosis patients with advanced testicular cancer.

- The IGCC classification underestimates survival of currently diagnosed patients with advanced testicular cancer. The results of nonrandomised trials should not be compared with the survival estimates reported by the IGCCCG. Instead updated estimates adjusted for year of treatment or based on our literature review should be used.

- Tree modelling is not an appropriate method for subgrouping the poor prognosis patients with advanced testicular cancer.

Recommendations for future research

- Our estimates of survival of currently diagnosed advanced testicular cancer patients should be confirmed by ongoing trials.

- The prognostic value of the rate of tumour marker decline during treatment and genetic markers should be determined. These prognostic factors might allow for a more refined and dynamic assessment of prognosis of patients with advanced testicular cancer.

- Follow up of patients with advanced testicular cancer is needed to determine long term benefit and harm of both standard treatment and alternative treatment strategies.

References

1. Harrell FE, Jr., Lee KL, Califf RM, Pryor DB, Rosati RA. Regression modelling strategies for improved prognostic prediction. *Stat Med* 1984;**3**(2)**:**143-52.
2. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;**15**(4)**:**361-87.
3. Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Belmont, California: Wadsworth, 1984.
4. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: Wiley, 1987.
5. Burton A, Altman DG. Missing covariate data within cancer prognostic studies: a review of current reporting and proposed guidelines. *Br J Cancer* 2004;**91**(1)**:**4-8.
6. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;**18**(6)**:**681-94.
7. Collette L, Sylvester RJ, Stenning SP, et al. Impact of the treating institution on survival of patients with "poor-prognosis" metastatic nonseminoma. European Organization for Research and Treatment of Cancer Genito-Urinary Tract Cancer Collaborative Group and the Medical Research Council Testicular Cancer Working Party. *J Natl Cancer Inst* 1999;**91**(10)**:**839-46.
8. IGCCCG. International Germ Cell Consensus Classification: a prognostic factor-based staging system for metastatic germ cell cancers. International Germ Cell Cancer Collaborative Group. *J Clin Oncol* 1997;**15**(2)**:**594-603.
9. Kollmannsberger C, Nichols C, Meisner C, Mayer F, Kanz L, Bokemeyer C. Identification of prognostic subgroups among patients with metastatic 'IGCCCG poor-prognosis' germ-cell cancer: an explorative analysis using cart modeling. *Ann Oncol* 2000;**11**(9)**:**1115-20.
10. Ennis M, Hinton G, Naylor D, Revow M, Tibshirani R. A comparison of statistical learning methods on the Gusto database. *Stat Med* 1998;**17**(21)**:**2501-8.
11. Kattan MW. Comparison of Cox regression with other methods for determining prediction models and nomograms. *J Urol* 2003;**170**(6 Pt 2)**:**S6-9; discussion S10.
12. Segal MR, Bloch DA. A comparison of estimated proportional hazards models and regression trees. *Stat Med* 1989;**8**(5):539-50.
13. US intergroup. Phase III Randomized Study of Bleomycin, Etoposide, and Cisplatin (BEP) With or Without High-Dose Carboplatin, Etoposide, and Cyclophosphamide Plus Autologous Bone Marrow or Peripheral Blood Stem Cell Transplantation in Male Patients With Previously Untreated Poor- or Intermediate-Risk Germ Cell Tumors. Accessed June 1, 2006.

14. EORTC. Phase III Randomized Study of Standard Cisplatin, Etoposide, and Ifosfamide (VIP) Followed By Sequential High-Dose VIP and Stem Cell Rescue Versus Bleomycin, Etoposide, and Cisplatin (BEP) in Chemotherapy-Naive Men With Poor-Prognosis Germ Cell Cancer, Accessed June 1, 2006.
15. MRC. Phase II Randomized Study of Intensive Induction Chemotherapy Comprising Cisplatin, Vincristine, Bleomycin, and Carboplatin Followed By Bleomycin, Etoposide Phosphate, and Cisplatin (BEP) Versus Standard BEP Chemotherapy in Male Patients With Poor-Prognosis Extracranial Nonseminoma Germ Cell Tumors, Accessed June 1, 2006.
16. EORTC. Phase II/III Randomized Study of Bleomycin, Cisplatin, and Etoposide (BEP) Versus Bleomycin, Cisplatin, Etoposide, and Paclitaxel (T-BEP) in Men With Intermediate Prognosis Germ Cell Cancer, Accessed June 1, 2006.
17. FRE. Phase III Randomized Study of Bleomycin, Etoposide, and Cisplatin With or Without Dose-Dense Sequential Combination Chemotherapy in Patients With Poor Prognosis Stage II or III Non-Seminomatous Germ Cell Tumors, Accessed June 1, 2006.

Summary

Patients with similar characteristics can be grouped together in a prognostic classification to estimate a patient's prognosis and guide treatment decisions. The topic of this thesis is methodological aspects of defining prognosis classifications. We specifically looked at patients with advanced testicular cancer, who are currently classified into good, intermediate and poor prognosis groups according to the International Germ Cell Consensus (IGCC) Classification. The IGCC classification aims to guide treatment decisions, it is used as a stratification method for clinical trials.
Two main topics are investigated: (1) the **validity of the IGCC classification**: are the assumptions underlying the IGCC classification method valid and can the survival estimates of the IGCC classification be generalised to currently diagnosed patients, (2) alternative methods of **defining prognostic groups**, especially for poor prognosis patients.

Validity of the IGCC classification

Issues in the development of prognostic classifications are discussed, and background on testicular cancer and the development of the IGCC classification is given in **Chapter 1**.
Chapters 2, 3 and 4 describe studies on the validity of the IGCC classification, with respect to the assumptions made in the development of the IGCC classification (**Chapters 2 and 3**) and the generalisability of the survival estimates of the IGCC classification to currently diagnosed patients (**Chapters 3 and 4**).

The IGCC classification assumes that all prognostic factors are equally important in predicting a patient's prognosis. Furthermore, within a prognosis group no distinction is made in the number of adverse prognostic factors for a patient. **Chapter 2** looks at whether incorporating differences in importance between prognostic factors and considering the number of adverse prognostic factors within a prognosis group improves the performance of the IGCC classification.
We therefore developed alternative classifications for the 3048 patients in the IGCC database based on both Cox regression analysis and tree modelling, and evaluated their performance.
Both methods demonstrated that there are differences in importance between the prognostic factors in the IGCC classification. Furthermore, within a prognosis group, prognosis differs for patients with different numbers of adverse prognostic factors.

This, however, did not result in a relevant increase in performance compared relative to the original IGCC classification. Hence we support the validity of the IGCC classification is reasonably valid to define prognosis groups.

In the development of the IGCC classification 41% of patients were excluded because of missing values. These values were mainly missing in the tumour marker LDH.
The effect of excluding these patients on the survival estimates in the IGCC classification was investigated in **Chapter 3** by filling in the missing values using a multiple imputation technique.
After multiple imputation, 5-year survival was lower for the group of patients with missing values than for the group of patients without missing values. We could explain the difference in survival between patients with and without missing values as follows.
For patients with advanced testicular cancer in general, and hence those in the IGCC database, average survival probability increased over time. However, year of treatment is not a prognostic factor used in the IGCC classification, nor is improvement in treatment accounted for in any other way. As a result, the survival estimates of the IGCC classification are too low for more recently treated patients. Because missing values were mainly found in relatively historical patients, multiple imputation led to a further underestimation of survival in the IGCC classification when no adjustment for year of treatment was made.

In **Chapter 4** we investigated whether survival estimates of patients with advanced testicular cancer increased further since the introduction of the IGCC classification in 1997.
We did a systematic search of the literature and found ten studies on survival of patients treated after 1989, with advanced testicular cancer, with outcome reported according to the IGCC classification. These ten studies describe a total of 1775 patients. We pooled the estimates of the selected studies using meta-analytic techniques.
Pooled 5-year survival estimates were 94, 83 and 71% for good, intermediate and poor prognosis patients respectively. The original IGCC classification reported lower 5-year survival estimates (92, 80 and 48% respectively). The large increase in survival for the poor prognosis patients is most likely due to more effective

treatment strategies and more experience in treating patients with advanced testicular cancer.

Chapters 3 and 4 demonstrated that survival estimates reported by the IGCC classification are not valid for currently diagnosed patients, and should not be used as comparator for results of nonrandomised trials evaluating new treatment strategies.

Defining prognostic groups

Chapters 5 and 6 describe the results of tree modelling and decision analysis as alternative methods for creating classification groups.

In **Chapter 5** we evaluated the validity of a regression tree previously developed by the German Testicular Cancer group to identify subgroups within the IGCC poor prognosis group.
Performance of this tree was substantially lower when applied to the IGCC data. We developed a new tree model, with similar methods as the tree developed by the German Testicular Cancer group, that differed in the selection and hierarchy of prognostic factors. Furthermore, internal validation of this new tree showed a large degree of optimism in performance.
We conclude that regression tree analysis leads to unstable and optimistic models and is not an appropriate method for identifying subgroups within the group of poor prognosis patients.

In **Chapter 6** a decision-analytic approach was applied to determine how high the risk of patients with advanced testicular cancer should be in order to profit from high-dose chemotherapy with stem cell support.
In the decision analysis both harm and benefit of treatment were weighed explicitly to determine a treatment threshold for the more intensive treatment.
Benefit and harm were defined as the reduction and increase in absolute risk of mortality, including treatment related death and morbidity, due to high-dose chemotherapy. Estimates of benefit and harm were based on literature data, while using data from randomised controlled trials (RCT) would have been preferred.
The decision analysis resulted in a treatment threshold of a 17% risk of cancer mortality. This threshold leads to a similar selection of patients as defined by the

IGCC classification for poor prognosis patients. Future results from RCTs should be used to update the analysis.

This thesis ends with a discussion of the study results (**Chapter 7**), and conclusions and recommendations:

- More complex classifications did not perform better than the relatively simple IGCC classification, which supports its validity as a method for distinguishing good, intermediate, and poor prognosis patients with advanced testicular cancer.

- The previously reported IGCC classification survival estimates are too low for currently diagnosed patients with advanced testicular cancer. The results of nonrandomised trials should not be compared with the survival estimates as reported in 1997 by the IGCC investigators. Instead updated estimates adjusted for year of treatment or based on more recent literature should be used.

- Decision analysis is an appropriate framework to define treatment thresholds in prognostic classifications. Tree models are recommended for exploratory purposes only.

Samenvatting

In een prognostische classificatie worden patiënten met dezelfde eigenschappen gegroepeerd met als doel patiënten die verschillen in prognose te kunnen onderscheiden. Inzicht in het verwachte ziekteverloop van een patiënt kan de arts helpen bij het maken van een keuze tussen verschillende behandelmethodes.
Het onderwerp van dit proefschrift is de evaluatie van methodologische aspecten bij de ontwikkeling van prognostische classificaties. We hebben ons daarbij gericht op patiënten met testiscarcinoom. Deze patiënten worden momenteel geclassificeerd in 3 groepen met een goede, gemiddelde of slechte prognose aan de hand van de 'International Germ Cell Consensus (IGCC)' classificatie. De IGCC classificatie wordt gebruikt ter ondersteuning van behandelingsbeslissingen en als methode om patiënten te selecteren voor klinische trials.
Dit proefschrift heeft twee onderwerpen: (1) de **validiteit van de IGCC classificatie**: zijn de aannames die ten grondslag liggen aan de IGCC classificatie valide en kunnen de overlevingskansen zoals gerapporteerd door de IGCC classificatie worden gegeneraliseerd naar patiënten die nu met testiscarcinoom worden gediagnosticeerd, (2) zijn **alternatieve methoden nuttig voor het bepalen van prognostische groepen**, met name voor patiënten met een slechte prognose.

Validiteit van de IGCC classificatie

In **Hoofdstuk 1** worden de belangrijkste methodologische aspecten bij de ontwikkeling van een prognostische classificatie besproken en wordt achtergrondinformatie over testiscarcinoom gegeven.
Hoofdstukken 2, 3, en 4 gaan over de validiteit van de IGCC classificatie, te weten de aannames die gemaakt zijn bij de ontwikkeling van de IGCC classificatie (**Hoofdstukken 2 en 3**), en de generaliseerbaarheid van de overlevingskansen van de IGCC classificatie voor patiënten die nu worden gediagnosticeerd (**Hoofdstukken 3 en 4).**

In de IGCC classificatie wordt aangenomen dat alle prognostische factoren even belangrijk zijn bij het voorspellen van de prognose van een patiënt. Verder, wordt er binnen een prognose groep geen onderscheid gemaakt tussen patiënten met een of meerdere prognostische factoren. In **Hoofdstuk 2** wordt onderzocht in hoeverre het discriminerend vermogen van de IGCC classificatie verbeterd kan worden door rekening te houden met verschillen in voorspellende waarden tussen prognostische factoren en het aantal ongunstige prognostische factoren.

Het discriminerend vermogen van een prognostische classificatie geeft aan hoe goed een prognostische classificatie in staat is patiënten met een slechte prognose te onderscheiden van patiënten met een goede prognose.
De validiteit is getoetst door alternatieve classificaties te ontwikkelen in de originele IGCC data (n=3048) op basis van Cox regressie analyses en het gebruik van regressiebomen en vervolgens het discriminerend vermogen van deze alternatieve classificaties te vergelijken met die van de IGCC classificatie.
Beide methoden lieten verschillen zien in voorspellende waarde tussen de prognostische factoren in de IGCC classificatie. Tevens vonden we dat binnen een prognose groep het aantal ongunstige prognostische factoren van invloed is op de prognose van een patiënt.
Dit resulteerde echter niet in een verbetering van het discriminerend vermogen ten opzichte van de IGCC classificatie. De IGCC classificatie is dus voldoende valide om groepen verschillend in prognose te onderscheiden.

In de ontwikkeling van de IGCC classificatie werd 41% van de patiënten niet meegenomen in de analyses omdat hun gegevens niet compleet waren. De voornaamste reden hiervoor was dat de tumormarker LDH niet was vastgesteld.
In **Hoofdstuk 3** zijn de consequenties van het uitsluiten van deze patiënten op de geschatte overlevingskansen in de IGCC classificatie onderzocht door de ontbrekende gegevens in te vullen met een multipele imputatie techniek.
Na imputatie van de ontbrekende gegevens, bleek dat de vijfjaarsoverlevingskansen lager waren voor patiënten met ontbrekende gegevens dan voor patiënten zonder ontbrekende gegevens. Dit verschil in overlevingskans kan als volgt verklaard worden.
De kans op overleving voor patiënten met testiscarcinoom, en dus ook voor de patiënten in de IGCC data, is in de loop der tijd gestegen. Het jaar van behandeling van de patiënt is echter niet als prognostische factor meegenomen in de IGCC classificatie. Ook is er geen rekening gehouden met verbeteringen in behandelmethodes.
Hierdoor zijn de door de IGCC gerapporteerde overlevingskansen te laag voor recenter behandelde patiënten. Omdat met name bij historische patiënten gegevens ontbraken, leidde multipele imputatie tot een verder onderschatting van de overlevingskansen in de IGCC classificatie, wanneer er niet werd gecorrigeerd voor jaar van behandeling.

In **Hoofdstuk 4** werd onderzocht of de overlevingskansen van patiënten met testiscarcinoom verder zijn gestegen sinds de introductie van de IGCC classificatie in 1997.
Een systematische inventarisatie van de literatuur leverde 10 studies op naar de overlevingskansen van patiënten met testiscarcinoom, geclassificeerd volgens de IGCC classificatie. Deze 10 studies beschreven 1775 patiënten, waarvan de overlevingskansen werden gecombineerd in een meta-analyse.
De gecombineerde vijfjaarsoverlevingskansen waren 94, 83 en 71% voor patiënten met een goede, gemiddelde en slechte prognose respectievelijk. De originele IGCC classificatie rapporteerde lagere overlevingskansen (92, 80 en 48%). Deze toename in overlevingskans voor patiënten met een slechte prognose kan verklaard worden door het gebruik van effectievere behandelmethodes en meer ervaring met het behandelen van patiënten met testiscarcinoom.

Hoofdstukken 3 en 4 tonen aan dat de overlevingskansen zoals gerapporteerd door de IGCC classificatie niet valide zijn voor patiënten die op dit moment met testiscarcinoom gediagnosticeerd worden. Deze schattingen kunnen daarom ook niet gebruikt worden als referentie voor resultaten van niet gerandomiseerde trials om nieuwe behandelmethoden te vergelijken.

Alternatieve methoden voor het bepalen van prognostische groepen

Hoofdstukken 5 en 6 beschrijven in hoeverre besliskundige analyse en regressiebomen geschikte methodes zijn om prognostische groepen te definiëren.

In **Hoofdstuk 5** is de validiteit onderzocht van een regressieboom, ontwikkeld door de 'German Testicular Cancer group', die subgroepen binnen de slechte prognose groep van de IGCC classificatie onderscheidt. Het discriminerend vermogen van de regressieboom, dat wil zeggen het vermogen om patiënten met een goede en slechte prognose van elkaar te onderscheiden, was lager wanneer deze werd toegepast op de IGCC data.
We ontwikkelden een nieuwe regressieboom volgens dezelfde principes als de regressieboom van de German Testicular Cancer group. In onze regressieboom werden andere prognostische factoren geselecteerd en werden de prognostische factoren anders geordend. Tenslotte, bleek uit interne validatie van de regressieboom dat het discriminerend vermogen te optimistisch is.

We concluderen dat het gebruik van regressiebomen leidt tot wisselvallige uitkomsten en een te optimistische inschatting van het discriminerend vermogen. Deze methode is daarom niet geschikt om subgroepen van patiënten binnen de slechte prognose van de IGCC classificatie te onderscheiden.

In **Hoofdstuk 6** werd een besliskundige analyse toegepast om te bepalen vanaf welk risico patiënten met testiscarcinoom zouden kunnen profiteren van hoge dosis chemotherapie met stamcelsupport in plaats van standaard dosis chemotherapie.
In een besliskundige analyse worden de negatieve en de positieve effecten van een behandeling expliciet gewogen om een behandelingsdrempel te bepalen.
Het positieve effect van een behandeling wordt gedefinieerd als de afname in het absolute risico op overlijden ten gevolgde van de behandeling; het negatieve effect als de stijging in het absolute risico op overlijden ten gevolge van de behandeling. Op basis van beschikbare literatuur is een schatting gemaakt van de positieve en negatieve effecten van hoge dosis chemotherapie. Idealiter worden hiervoor resultaten van gerandomiseerde klinische trials gebruikt.
Op basis van de besliskundige analyse zouden patiënten met een risico van 17% of hoger om te overlijden aan kanker in aanmerking komen voor een behandeling met hoge dosis chemotherapie. Deze behandeldrempel leidt tot een vergelijkbare selectie van patiënten als het gebruik van de IGCC classificatie.
Deze besliskundige analyse moet verder geactualiseerd worden aan de hand van de resultaten van gerandomiseerde klinische trials die betere schattingen van de positieve en negatieve effecten van hoge dosis chemotherapie mogelijk maken.
Dit proefschrift eindigt met een discussie van de onderzoeksresultaten (**Hoofdstuk 7**), en een aantal conclusies en aanbevelingen:

- Complexere classificaties zijn niet beter in het onderscheiden van prognostische groepen dan de relatief simpele IGCC classificatie. Dit ondersteunt de validiteit van de IGCC classificatie om testiscarcinoom patiënten met een goede, gemiddelde en slechte prognose van elkaar te onderscheiden.

- De in de IGCC classificatie gerapporteerde overlevingskansen zijn te laag voor patiënten die tegenwoordig gediagnosticeerd worden met testiscarcinoom. De resultaten van niet gerandomiseerde klinische trials

mogen niet vergeleken worden met de overlevingskansen zoals die in 1997 door de IGCC classificatie zijn gerapporteerd. In plaats daarvan moeten overlevingskansen gebruikt worden die zijn gecorrigeerd voor jaar van behandeling of die gebaseerd zijn op meer recente literatuur.

- Een besliskundige analyse is een geschikte methode om te bepalen vanaf welk risico een alternatieve behandelbeslissing te rechtvaardigen is. Regressiebomen mogen alleen gebruikt worden voor exploratief onderzoek naar interacties tussen prognostische factoren.

Dankwoord

Het is bij het afronden van mijn proefschrift geruststellend om te concluderen dat mijn geheugen selectief is: de obstakels op het proefschriftpad zijn zo goed als vergeten, wat overheerst is blijdschap dat het 'boekje' nu echt bijna af is en dankbaarheid voor alle mensen die er op hun eigen wijze een bijdrage aan hebben geleverd. Een aantal wil ik bij deze persoonlijk bedanken.

Ten eerste mijn twee promotoren, Ewout Steyerberg en Dik Habbema. Ik ben erg dankbaar voor jullie vertrouwen in mij en jullie steun in de lastige laatste fase.

Beste Ewout, dank voor jouw vriendelijke en inspirerende begeleiding de afgelopen jaren. Onze wekelijkse afspraken leidden altijd tot hernieuwde inzichten waardoor ik weer vol goede moed verder kon. Ik heb bewondering voor jouw vermogen om over elke onderwerp scherpe en zinnige vragen te stellen, mijn inziens de belangrijkste eigenschap van een wetenschapper. Waarschijnlijk mede de reden dat je nu niet als co-promotor maar als promotor in mijn commissie plaats neemt.
Beste Dik, ons contact was minder intensief maar niet minder belangrijk voor dit proefschrift. Jouw scherpe inzichten hielpen het proefschrift de juiste richting te geven en jij wist met jouw commentaar artikelen altijd naar een hoger niveau te tillen.

I would like to thank the members of the International Gern Cell Consensus Group for providing me with the IGCC data, on which almost all chapters in this thesis are based. I would especially like to thank Sally Stenning who helped me get started with the IGCC data and provided me with details on the decisions made in the development proces of the IGCC classification.

Ik wil alle (oud-) MGZ-ers, bedanken voor een gezellige en leerzame periode. Ik wil met name mijn CMB-collega's bedanken voor de fijne samenwerking. Door mijn prachtige uitzicht vanaf de 20e over de stad, alle etentjes, promotiefeesten en uitjes op bijzondere locaties in Rotterdam is deze verstokte 'Amsterdammer' ook van Rotterdam gaan houden. Dank ook voor jullie belangstelling en behulpzaamheid tijdens het afronden van mijn proefschrift.

Ik wil in het bijzonder Nino en Laetitia bedanken voor hun gezelligheid en gastvrijheid, Elsbeth voor de gedeelde proefschriftsmart - en vreugd in de laatste fase, buren René en Cecile voor hun behulpzaamheid, Resi voor de leuke treingesprekken en Marloes voor haar aanstekelijk enthousiasme.

Uiteraard mogen ook mijn kamergenoten niet onvermeld blijven. Elske, Gerrit en Iris, ik heb met veel plezier met jullie een kamer gedeeld en veel van jullie geleerd; dat of je je kerstpakket weggeeft niet afhangt van wat er in zit, de kunst van het relativeren en dat je een voetbalkenner in een 'voetbalarme' omgeving moet koesteren.

Dank aan mijn LTP (Impact) collega's voor het meeleven met mijn proefschrift perikelen. Het valt niet altijd mee samen te werken met iemand die haar dagelijkse werkzaamheden combineert met het afronden van haar proefschrift....

Lieve vrienden en (schoon)familie, bedankt voor jullie interesse in mijn vorderingen de afgelopen jaren (of het gebrek daaraan...). Vanaf nu is er gelukkig weer meer tijd om af te spreken!
Meike, ik had 15 jaar geleden niet kunnen bedenken dat jij nu als arts mijn proefschrift zou proeflezen. Dank daarvoor en dat we nog maar heel lang vriendinnen mogen blijven.
Olga, bedankt dat je me er af en toe aan herinnerde dat het schrijven van een proefschrift eigenlijk iets heel bijzonders is.

Lieve Johan en Gerda, bedankt voor jullie steun en enorme betrokkenheid bij alles wat ik doe!
Johan, op het laatst moest zelfs Feyenoord wijken voor het proefschrift; wellicht breken er betere tijden aan wanneer ik na mijn promotie weer vaker met je richting stadion ga.
Gerda, fysiek ver weg in Portugal, maar in gedachten altijd dichtbij. Wordt nu wel wel weer eens tijd voor een ticket Portugal.
Erica, Tom, Sara en Julia; bij jullie langsgaan voelt niet als op bezoek gaan maar als thuiskomen. Dank dat ik altijd bij jullie kan binnenvallen en aanschuiven. Tom, bedankt voor je hulp bij het ontwerpen van de voorkant van het proefschrift.

Tenslotte mijn paranimfen, Judith en Mart. Met hun aan mijn zijde ga ik vol vertrouwen mijn promotiedag tegemoet!

Lieve Judith, van collega tot vriendin en nu ook nog paranimf. Bedankt voor de vele gezellige momenten op werk (truckerscorner!) en daarbuiten (van filmfestival tot Chicago), de bemoedigende woorden en het geduldig beantwoorden van al mijn proefschriftvragen in de laatste fase. Volgend jaar weer filmfestival?

Lieve Mart, dank voor je steun in de afgelopen jaren; je bood me op de juiste momenten de helpende hand, wist me altijd weer op te beuren op momenten dat ik het even niet meer zag zitten en gaf me het vertrouwen dat het allemaal goed zou komen met mijn proefschrift. Maar ik wil je vooral bedanken voor de vele leuke en bijzondere dingen die we samen doen, die het leven zoveel aangenamer maken.

En nu, op naar nieuwe uitdagingen!

Curriculum Vitae

Merel van Dijk werd op 21 september 1978 geboren te Den Burg, Texel. Nadat ze in 1996 haar gymnasiumdiploma behaalde aan de OSG Willem Blaeu te Alkmaar, startte ze haar studie Psychologie aan de Universiteit van Amsterdam en specialiseerde zich in Arbeids- en Organisatiepsychologie.
In het kader van deze studie deed ze haar afstudeeronderzoek bij Human Company te Amersfoort, resulterend in de scriptie 'De constructie van een anticipatietest voor machinisten bij de Nederlandse Spoorwegen'. Voor deze scriptie ontving zij in 2002 de David van Lennepprijs voor uitzonderlijk goede doctoraalscriptie op het gebied van de psychologie van arbeid en organisatie ingesteld door de Nederlandse Stichting voor Psychotechniek. In 2001 behaalde zij cum laude haar doctoraal examen. Aansluitend bleef ze nog enkele maanden werken als R&D medewerker bij Human Company.
In juli 2001 werd ze aangesteld als Assistent in Opleiding bij het Instituut Maatschappelijke Gezondheidszorg van het Erasmus Medisch Centrum Rotterdam voor het project 'Defining poor and good prognosis groups in oncology' gefinancierd door ZonMw. Hier verrichtte zij het onderzoek beschreven in dit proefschrift. In 2003 behaalde zij haar Master of Science in Klinische Epidemiologie aan het Netherlands Institute for Health Sciences.
Sinds juni 2005 werkt zij als research psycholoog bij LTP, te Amsterdam.