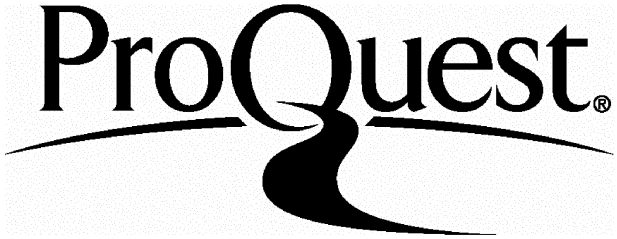
COMPETING RISKS METHODOLOGY APPLIED TO THE ANALYSIS OF CANCER THERAPY OUTCOME

ÖZLEM ATAMAN

Thesis submitted for the degree of Doctor of Philosophy (PhD)

DEPARTMENT OF ONCOLOGY
UNIVERSITY COLLEGE LONDON

JANUARY 2003



ABSTRACT

The relationships between potential prognostic factors and various types of failures in different cancer sites were studied in a competing risks model. The methods that have been applied here were based on time to the first observed event. The clinical data analyzed came from the randomized trials of continuous hyperfractionated accelerated radiotherapy (CHART) vs. conventional radiotherapy in non-small cell lung cancer (NSCLC) and head and neck squamous cell carcinoma (HNSCC). Clinical outcome was analyzed in 549 NSCLC patients and in 309 HNSCC patients. Competing risks analysis was performed using the biomedical data package (BMDP) statistical software with an accelerated failure-time model and a log-logistic hazard function. The final reduced models included age, sex, clinical stage and treatment for NSCLC and proliferative pattern, bcl2, cyclin D1, TRT*CD31, Ki-67 scores and T and N stage for HNSCC. From these models, prognostic indices for local, regional and distant failure were estimated for each individual patient and prognostic groups were formed to identify patients at risk for different types of failures.

Cumulative incidence (CI) and the Kaplan-Meier estimate (KM) are two estimates for quantifying the occurrence of an endpoint over time in the competing risks framework. Properties of the two estimators for treatment failure and late morbidity over time have been explored for different prognostic groups using the CHART NSCLC and HNSCC studies. The CI estimate showed unexpected variations in estimating tumour outcome and late side effects in unfavourable prognostic groups where there was a higher incidence of competing risks. Without a comprehensive understanding of the assumptions of KM method, the clinical interpretations must be made with caution. The KM and the CI methods should be used as complementary analyses. The natural behaviour of the tumour site and the competing events under study should be clearly understood by clinicians.

ACKNOWLEDGEMENTS

This research was funded by the Scott of Yews Trust and the Gray Cancer Institute and by the Radiation Oncology Department of the Dokuz Eylul University Medical School. I thank them all for giving me this opportunity to work on this project and also for their confidence in me.

I wish to acknowledge the great efforts of all members of the CHART Steering Committee and all the contributors to CHART studies for the excellent quality of the follow-up data.

I want to thank to my principal supervisor Søren Bentzen for introducing me to the field of biostatistics and for providing stimulating working conditions where I had the chance to present and discuss many aspects of my project in group meetings every Thursday morning. Thanks to Søren for giving me the constant support and independence during these years where I learned a lot about different aspects of research. I am also grateful to him and his lovely family for the long and lasting personal friendship.

I wish to thank my supervisor MI Saunders for the beneficial discussions on the project. Thanks to George Wilson and his colleagues for providing the molecular markers data for the HNSCC study. Simon Bond has contributed to this work by writing the SPSS syntax file for the CI calculation.

Many members of the Gray Cancer Institute have helped me and being a part of GCI was a great experience with its motivating and exciting research environment. Thanks to all former and present members of the Human Cancer Biology and Informatics group for their contribution to my project.

I am very grateful to Ann Barrett who heartened me to finish my thesis during my writing-up period in Norwich. She not only read the drafts patiently and contributed to the discussions but also helped with the language of the thesis.

During my stay in London, I had the pleasure of meeting some special friends. Azza is the first one to mention whose unique smile warmed and brightened up my days. She deserves a thank you for her endless support both at work and privately. Massimo taught me how to do graphs using Origin and also together with Olga they have been two Italian friends with whom we shared the laughter and the delicious bread, pasta or Turkish dishes we cooked at the weekends! Thanks to Simon as well who introduced me to the world of ales and beers. Kamer's e-mails kept me alive and hopeful at many dull moments. Thanks to her for reaching me that efficiently across all that miles between us.

My special thanks to my parents, who loved, supported and encouraged me during this journey even if they felt that they never wanted me to be far away and we missed each other a lot.

My most sincere appreciation goes to my husband Adnan Can for all his love and understanding who made this PhD possible in the first place when he agreed to move to UK and to my daughter Öykü for her patience.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 BACKGROUND AND OBJECTIVES

Surgery, radiation therapy and chemotherapy all play a role in the current management of cancer. These modalities, alone or in combination, allow the oncologist to give varying weights to the local, regional and systemic intensity of treatment. At the same time, the overall treatment intensity is limited by early and late treatment-related morbidity. This poses the challenge of working within patient tolerance to produce an optimal combination of the three modalities. Optimisation of therapy in the individual patient would ideally require a quantitative assessment of the risks of relapsing in primary, nodal or distant sites after a specific therapy. This assessment falls into a category of statistical analysis called "time-to-failure analysis" and the "failure" endpoint is usually defined in the context of the question under study. In many randomised controlled trials in radiotherapy, failure at a specific site e.g. loco-regional failure, or composite outcome measures such as overall or disease-free survival are the endpoints under study. Disease-free survival is a composite endpoint of considerable interest to both the patient and the physician, but it does not convey any information about site of the failure and therefore does not help towards a rational change of treatment strategy. On the other hand, studying a single site of failure does not provide any information on the risk of other types of failures. Even worse, a patient with a high risk of, say, distant relapse might appear to have a low risk of loco-regional failure, simply because a distant failure would effectively prevent prolonged observation of the loco-regional tumour outcome. Loco-regional and distant relapse

are *competing risks*, both of them requiring prolonged observation of the patient, but as soon as one of them occurs, this will in many cases effectively shorten the period of being at risk of failing from the competing cause. If the patients were followed-up rigorously beyond the time of the first event to detect possible subsequent failures then we might have been in a position to see several events occurring in sequence, which is referred to as multiple failures. The term "competing risks" encompasses the study of any failure process in which there is more than one distinct cause or type of failure. In this thesis a competing risk is defined as an event, which may or may not occur during follow-up but if it does occur will prevent the observation of time to any other event of interest. The main concern is the type and time of the first failure since the patient is often not followed with the same intensity once the first failure has occurred. Also, in terms of therapeutic management, the location and time to occurrence of the first relapse is obviously of major concern. Locally advanced non-small cell lung carcinoma (NSCLC) is an example where loco-regional failure and distant metastasis are two important and very common types of failure. These are obviously competing events because the occurrence of either one of them is likely to lead to the death of the patient and thereby to termination of the time at risk for developing the other type of failure. This is shown diagrammatically in Figure 1.1.

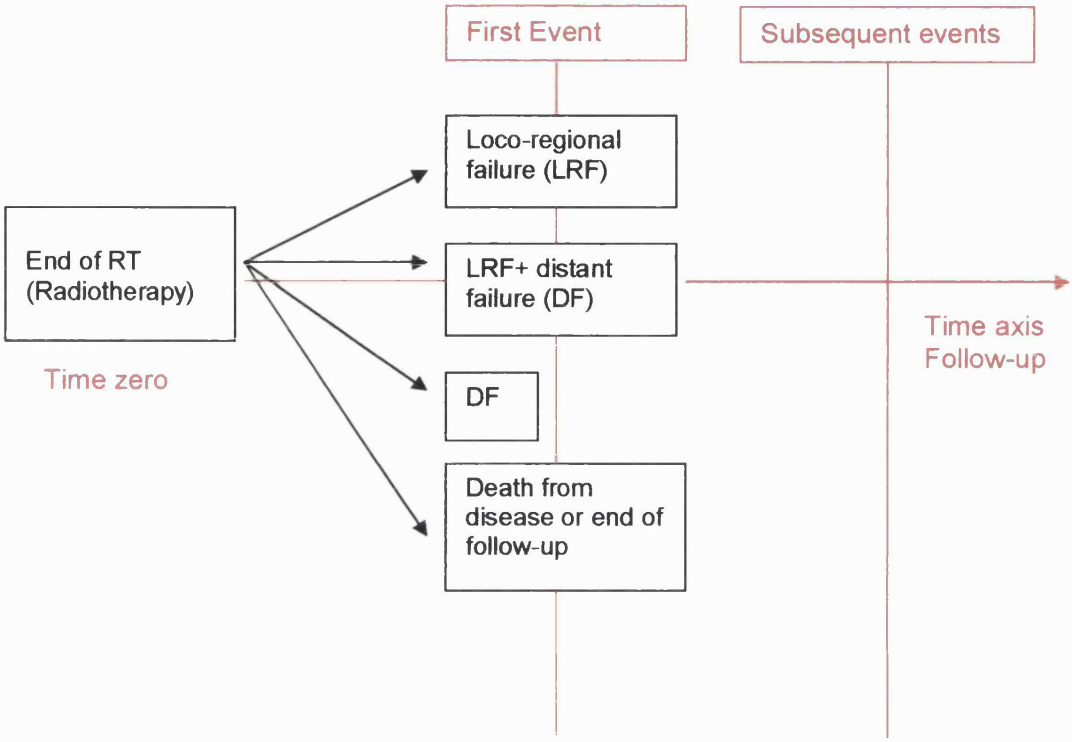


Figure 1.1 Possible outcomes as a first event after the end of radiotherapy in locally advanced NSCLC

Another example comes from allogeneic bone marrow transplantation (BMT) in high-risk (refractory or relapsed) adult acute lymphoblastic leukaemia (ALL) studies. The endpoint of interest could be disease relapse or the incidence/prevalence of a major complication of transplantation, or chronic graft versus host disease (CGVHD). Each of these endpoints is subject to the competing risks of disease relapse, CGVHD, or death due to other treatment-related complications with no evidence of the disease (Figure 1.2).

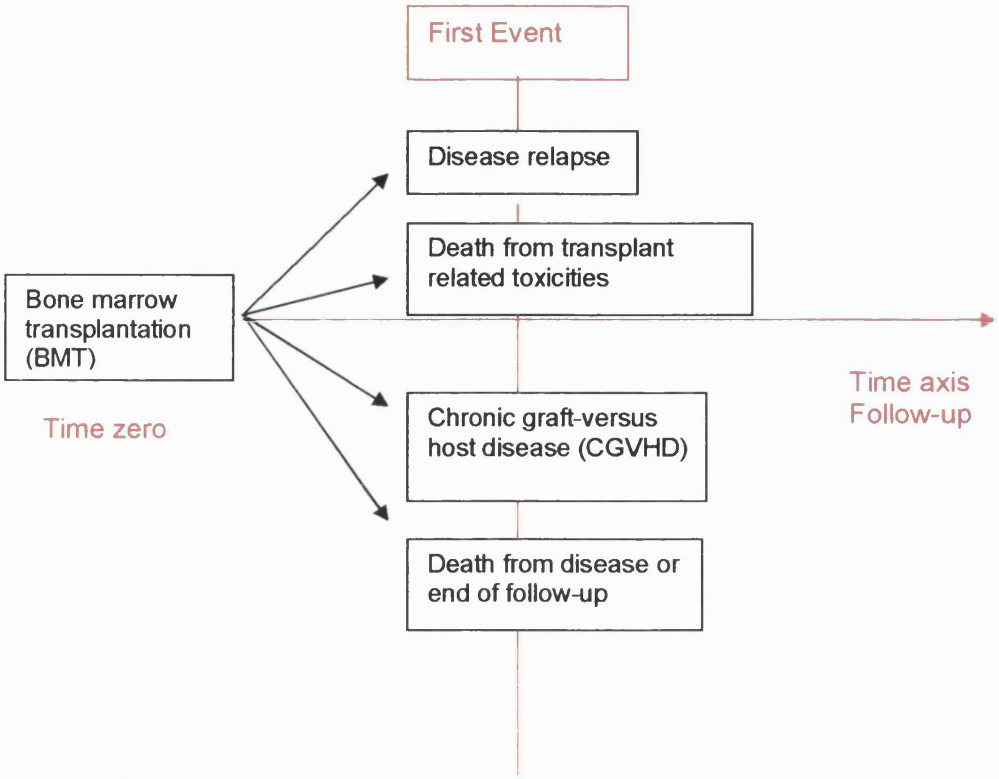


Figure 1.2 Possible outcomes as the first event after bone marrow transplantation in high-risk ALL

In these clinical situations questions such as "How effective is the loco-regional treatment?" or "How successful is the BMT?" have to be answered with the confounding influence of *competing risks*.

1.2 COMPETING RISKS METHODOLOGY

The basic estimable quantity in the competing causes (risks) situation is the cause-specific hazard. To estimate the cause-specific hazard and cause-specific failure probability, any other event is treated as censoring and then the time-to-failure analysis methods (described below in section 1.2.1) are applied. The literature

assumes the concept of latent times, where each cause-event occurs at some point and we observe the minimum time. In reality the cause-specific failure probabilities are not observable because we never will be in a position to evaluate one failure rate in isolation from others. The "net" or "true" distribution of time-to-failure (events) is impossible to estimate as this theoretical quantity depends on the joint distribution of the risk of the failure of interest and the risk of competing risks, which cannot be estimated. Interdependence of causes cannot be assessed directly and the common question of "What's the failure probability for cause A, if causes B, C... are removed?" can only be answered if the causes are assumed to be independent, from a theoretical basis. This assumption cannot be tested. On the other hand the basic estimable quantity, cause-specific hazard, is considered difficult to interpret especially for clinicians.

Statistical methods for analysing competing risks have been available since the 1970s (1,2) but have so far only found limited use in the analysis of outcome data after cancer therapy (3,4,5). The conceptual advantage of competing risks analysis is that it takes both the type of failure and the time to failure into account and allows estimation of failure-site specific prognostic factors.

1.2.1 Analysis of failure-time data

Failure is an event that occurs at a specific point in time, and the term *failure time* is used for the time-to-event. Basic convention involves the estimation of and specification of models for the underlying failure time distribution. In general, failure-time data are subject to (time-) censoring. This is the technical term for data where some individuals in the study have not experienced the event of interest at the end of

the study or time of analysis. This may be due to intercurrent death, loss to follow-up or simply because the individual was alive and without any sign of the failure at the time of the last control. The exact failure times of these individuals are unknown. These are called censored observations or censored times.

When censoring, a distinction can be made to reflect the "side" of the time dimension at which censoring occurs. In failure-time analysis there is always a certain time point where the study starts. This could be the date of start or end of the treatment or date of randomisation and the analysis take place after a certain amount of time. In this type of analysis the censoring always occurs on the right side (*right censoring*), because the researcher knows exactly when the study started. Alternatively, it is possible for the censoring to occur on the left side (*left censoring*). For example in biomedical research, one may know that a patient entered the hospital at a particular date, and that s/he survived for a certain amount of time thereafter; however, the researcher does not know when exactly the symptoms of the disease first occurred or were diagnosed. In this thesis only right censored data have been used.

Three functions are particularly useful in the analysis and discussions of failure-time problems: (1) the *survivorship function,* (2) the *probability density functio*n, and (3) *the hazard function*. These three functions are mathematically equivalent-if one of them is given, the other two can be drived.

Let T denote the failure time of an individual survival time. The probability distribution of T can be characterised by the following three equivalent functions.

1.2.1.1.Survival Function $S(t)$ The survival function is the terminology normally used, when the endpoint is death; in the present context it is the probability that an

individual is free of failure longer than time t. The probability of not having reached the endpoint by time t is given by the function $S(t)$.

$$S(t) = P \text{ (an individual is free of failure longer than } t)$$

$$= P\ (T > t\ )$$

From the definition of the cumulative distribution function $F(t)$ of T,

$$S(t) = 1\text{-}P \text{ (an individual fails before time t)}$$

$$= 1\text{-}\ F(t)$$

Here $S(t)$ is a non-increasing function of time t with the properties

$$S(t) = 1 \text{ for } t=0$$

And

$$S(t) = 0 \text{ for } t= \infty$$

That is, the probability of surviving at least at time zero is 1 and that of surviving infinite time is zero. If there are no censored observations, the $S(t)$ is estimated as the proportion of individuals free of failure longer than t:

$$S(t) = \frac{\text{number of individuals free of failure longer than } t}{\text{total number of patients}}$$

1.2.1.2 Probability Density Function $f(t)$ is defined as the limit of the probability that an individual fails in the short interval t to $t+ \Delta t$ per unit width Δt or simply the probability of failure in a small interval per unit time. It can be expressed as

$$f(t) = \lim_{\Delta t \to 0} \frac{P\{\text{an individual failing in the interval } (t, t+ \Delta t)\}}{\Delta t}$$

The proportion of individuals that fail in any time interval can be found from the density function. The density function is also known as the unconditional failure rate.

1.2.1.3 Hazard Function *h(t)* This is defined as the probability of failure during a very small time interval, assuming that the individual has survived to the beginning of the interval. The hazard function is also known as the instantaneous failure rate or conditional failure rate time t, and gives the risk of failing in the next time interval. Thus it is possible to know how the risk of a failure changes with time. The risk or the hazard rate can be estimated, within specific time intervals, by dividing the total period of failure into time segments, counting the number of events arising in the segment and dividing by the number of patients at risk during that segment. When the unit of time is a day, it is the probability of the event occurring within the (next) day, given that you have survived to that day.

The hazard function can also be defined in terms of the cumulative distribution function $F(t)$ and probability density function $f(t)$:

$$h(t) = f(t) / \{1-F(t)\} = f(t) / S(t)$$

1.2.2 Kaplan-Meier (KM) Method

A non-parametric estimate of $S(t)$ for censored data is the Kaplan-Meier estimator(6). The KM estimator of the survival function was developed to estimate the probability of death by time t from right-censored data. The time-to-recurrence distributions of competing causes of failure are commonly computed using the conventional KM procedure, although this approach is specifically developed for the cases where there

is a single, possibly censored failure time. The KM method uses the simple idea that survival to time T can be described as a sequence of survivals: First survive to time 1, then, having survived to time 1, survive to time 2, and so forth. Each time an event occurs one can estimate the probability of death as the proportion of individuals who experience the event among those "at risk" i.e. among those alive and under follow-up at that time. The probability of surviving beyond some time T is calculated as the product of the probabilities of surviving all the time points up to and including time T. The survival curve is a plot of these cumulative survival estimates over time. This method will yield consistent estimators of the true survival experience provided that a censored patient can be expected to have the same survival experience as patients whose follow-up extends beyond the time of censored patients follow-up. When the failure of an individual may be of several distinct types or causes, the most common approach is to generate a series of Kaplan-Meier (KM) curves, one for each endpoint(7) censoring all the other events other than the event of interest. In this case censoring occurs not only due to losses to follow-up but also by deaths from other causes and by other events if they preclude development of the endpoint under consideration.

One of the fundamental assumptions of the KM method is the independence between the censoring mechanism and the failure, which has also been called *noninformative censoring* (8). Without the assumption of independence the KM estimator may overestimate the $S(t)$ if the failure time and the censoring time are positively correlated, and underestimate the $S(t)$ if the times are negatively correlated(9). Any disease-related outcome treated as a censored observation might possibly have an influence on the failure time and is called dependent or *informative censoring.* In

statistics two events, A and B, are said to be statistically independent if knowing A has occurred would not provide useful information in predicting whether B will occur. If B is "having a distant failure" and A is "having a local failure" for most human malignancies it would be unreasonable to assume that these two events are independent for each patient. For example, a breast cancer patient with many positive lymph nodes may be at risk for both types of failures. Unfortunately, if one knows only time to first failure and whether it is distant or local, it is impossible to estimate the correlation of the two types of failure and it is also impossible to do a statistical test of independence (10). Estimation of a survival function and treating these competing risk events as censoring events implies that one assumes a hypothetical latent event time for the endpoint of interest. The KM curve is interpretable in the context of the latent event time model under the additional assumption that each cause-event occurs at some point and we observe the minimum time. It estimates the probability of that event occurring by time t if the competing risks could be removed without altering the risk of the event of interest. However, the various competing risks are often quite dependent. When applying this approach to the competing risks situation, a patient who experienced multiple events will 'fail' more than once and as separate KM analyses are conducted for each cause of failure, the resulting component failure probabilities may exceed the total probability of failure (overall event rate).

In competing risks situations the KM method can be applied for any event type of interest in either of the following ways: (i) by considering only the first event if of the type of interest, and censoring all other events. If the event type of interest is a

second or third event then it will not be considered. (ii) by considering any event of interest whether it is the first event or not while ignoring all other events.

Arriagada et al termed the above two approaches the 'censor' and 'ignore' methods respectively (3, 11). Koscielny et al. termed the same methods KM (1^{st}) and KM (any) respectively in their recent study. Our reference to these methods will be consistent with that defined by Koscielny et al (12). For example in locally advanced NSCLC (see Fig.1.1) the probability of loco-regional failure (LRF) estimated by the KM (1^{st}) method would censor all events other than the first LRF which means that if any LRF is preceded by any other events it has not been counted as an event. However the KM (any) method will count all the LRF's whether they occurred as the first event or not (13-15).

1.2.3 Cumulative Incidence (CI) Method

A cumulative incidence (CI) estimator has been preferred by some authors in the competing risks setting and it gives the increasing probability of cause-specific failure over time (3, 11). If there are no censored observations, the last (highest) point on the curve will be the crude rate (observed proportion) of cause-specific failures. CI estimate has been recommended for the elimination of the "independence assumption" of the KM method. With this approach, a patient will fail only once and the overall event rate at a given time would break down into a sum of the individual CI functions for each failure type (16). The sum of all estimated cause-specific failure probabilities equals one minus the event-free survival (EFS) estimate. The cumulative incidence is the cause-specific failure probability. This is the marginal probability of the event having occurred by time *t*. This function is not a true marginal probability but

rather is the chance of the event of interest occurring prior to time *t* in a system where an individual is exposed to both risks.

To illustrate how the CI estimate operates, we would like to give an example from the CHART bronchus database (details of the study and the patients are given in section 1.5 and Chapter 2). The following table provides a summary of event types occurring as the first event in 563 NSCLC patients treated in the CHART bronchus trial. We are interested to know the CI estimate for loco-regional failure (LRF). The data are grouped into time intervals in this example but the actual observed event times could be used as well.

Table 1.1 Summary of the event types and the CI estimate of LRF in NSCLC study

Time interval	N	LRF	OE	C	$P_{all\ E}$	P_{LRF}	S_{EFS}	CI of LRF
0	563						1	0
0–1	563	23	34	7	0.101	0.041	0.899	0.041
1–2	499	32	40	7	0.144	0.064	0.769	0.098
2–3	420	47	49	6	0.228	0.111	0.594	0.183
3–4	318	46	16	4	0.195	0.145	0.479	0.27
4–5	252	49	19	2	0.27	0.194	0.345	0.362
5–6	182	15	14	3	0.16	0.082	0.289	0.39
6–7	150	19	8	3	0.18	0.127	0.237	0.426
7–8	120	14	6	1	0.167	0.117	0.197	0.453
8–9	99	6	4	0	0.101	0.06	0.177	0.465
9–10	89	3	3	4	0.068	0.034	0.165	0.471
11–12	79	2	0	2	0.025	0.025	0.161	0.475
12–13	75	4	2	0	0.08	0.053	0.148	0.483
13–14	69	3	5	2	0.115	0.043	0.131	0.491
14–15	59	0	1	0	0.016	0.043	0.129	0.497
15–16	58	2	1	4	0.051	0.034	0.122	0.501
16–17	51	1	0	4	0.012	0.012	0.121	0.502
17–18	46	2	0	4	0.043	0.043	0.115	0.507
18–19	40	1	1	2	0.05	0.025	0.109	0.51
19–20	36	1	0	3	0.028	0.028	0.105	0.513

N: number of patients at risk of any event at the start of time interval T.
LRF: number of loco-regional failures.
OE: number of other (competing) events (distant metastasis) observed in time interval T.
C: number of censored observations (dead and lost to follow-up) in time interval T.
P_{allE}: conditional probability of failure from all events (LRF+OE/N)
P_{LRF}: conditional probability of loco-regional failure (LRF/N)
S_{EFS}: Event-free survival at the end of time interval T. For each time interval the S_{EFS} is calculated by multiplying the previous S_{EFS} probability by the ($1-P_{allE}$) of the current time interval.
CI for LRF: The cumulative incidence for LRF is than the cumulative sum of P_{LRF} multiplied by S_{EFS}.

CI calculation takes into account all causes of failure. Thus whenever a failure occurs, whether it is the event of interest or not, the patient with the failure is removed from the denominator, in other words is no longer "at risk" for the event of interest. While the CI gives an accurate estimate of the number of the LRF that will be seen over time, it fails to convey a total picture of risk to patients at risk because the incidence

depends upon failure due to competing risks. The CI estimates of LRF would be quite different in two groups of patients with identical treatments where one group has a higher incidence of death. The CI is not specific for the endpoint of interest and should be taken more cautiously when estimating the risk of normal tissue side effects where the results may not correlate with treatment intensity. The usefulness and interpretation of such an estimate have been the subject of some debate (17,18) especially in late normal-tissue effects reporting.

1.3 FAILURE-SPECIFIC PROGNOSTIC FACTORS

Assessment of baseline variables and their relationship to outcome with regard to cause-specific failures is of particular interest for the selection of optimal treatment for patients. Further individualisation of treatment requires determination of any influence of prognostic factors such as age, sex, etc. on the type of failure, time until failure, or both. Although many studies of prognostic factors have been conducted, they have not analysed failure-specific outcomes in the competing risks setting (19,20). Classification of patients according to a prognostic index combining multiple prognostic factors has been used extensively with the Cox model (21).

This model implicitly contains two assumptions. The first assumption is the multiplicative relationship between the underlying hazard function and the log-linear function of the covariates (the proportionality assumption). Thus the ratio of the hazard functions for two individuals with different sets of covariates does not depend upon time. The second assumption of the model is that the effect of covariates upon the hazard function is log-linear. This regression model has become popular because of its generality; i.e. no parametric model is assumed for the underlying hazard

function. Thus, the model is non-parametric with respect to time but parametric in terms of the covariates, and is often referred as a semi-parametric model.

An alternative model, the accelerated failure time model, is a class of log-linear models, which assumes that a set of covariates influence an individual's failure time distribution, by stretching, or accelerating the time scale (2). A fundamental feature distinguishes this model from the proportional hazards model. The non-parametric baseline hazard function of the proportional hazard model is replaced by a specified distribution in the accelerated time model. Parametric models such as exponential, Weibull, log-normal and log-logistic failure time distributions are commonly used in studies. Competing risks analysis allows the accelerated failure time model to be used to assess the effect of prognostic variables on time to failure for different types of failure.

1.4 ESTIMATION OF TREATMENT RELATED MORBIDITY

Another major concern in radiotherapy practice is treatment-related morbidity since any choice of treatment schedule must represent a balance between the beneficial therapeutic effect and the complications of that schedule. A specific expression of radiation injury to a tissue will occur after a characteristic time, the latent time. Biologically the latent-time distribution depends mainly on the proliferative organization and the cellular turnover in the tissue. Thus, estimating the level of late reactions requires prolonged observation of the patient. Hodgkin's disease (HD) is one of the examples where cured patients have a high incidence of a range of late side effects of treatment. Pulmonary toxicity, cardiac complications, death from infection, a high risk of leukaemia and second solid tumours are examples of late

mortality due to treatment side effects. The time-to-failure distribution of any of these late morbidities and their relation to treatment factors would be a competing risks problem. This is shown in Figure 1.3.

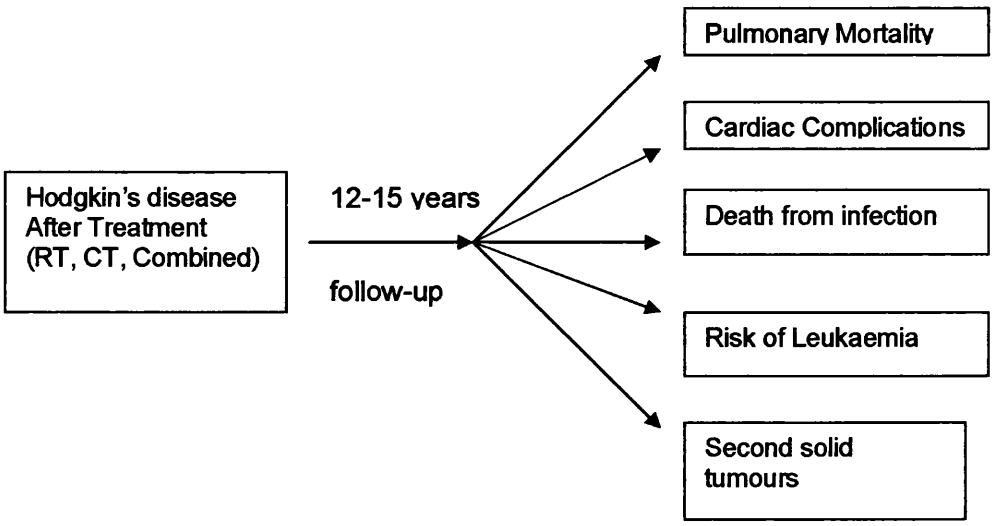


Figure 1.3 Possible late complications during prolonged follow-up in cured Hodgkin's disease

However, disease recurrence and death of the patient could interfere with the detection of late radiation morbidity in other clinical situations such as high-grade brain tumours, simply because the life expectancy of the patient might be short. Rates of toxicity will vary depending on the method of calculation thus making comparison of normal-tissue effects of various treatment schedules difficult. Crude proportion, prevalence, actuarial and cumulative incidence estimates have been discussed broadly by many authors with no agreement on the minimum requirements for reporting the outcome of radiation therapy (17,18,22,23).

The usefulness of different statistical approaches in analysing failure time data needed to be tested in data from a large clinical trial where all the relevant information had been recorded prospectively. We chose to undertake this work using data accrued from the CHART studies, which are described below.

1.5 CHART STUDIES

In the nineteen eighties, evidence accumulating from clinical and laboratory studies of radiotherapy fractionation suggested that alteration of conventional daily regimens might lead to considerable advantage in terms of tumour control and reduction in normal tissue morbidity. Studies investigating the cell kinetics of human tumours with bromodeooxyuridine to determine the growth fraction and the duration of the S phase showed that nearly half had the capacity to double their cell number in 5 or fewer days (24,25). This potential for very rapid repopulation suggested there could be benefit from using the shortest overall treatment time possible. It might also then be possible to complete the treatment before the peak of the acute reactions. The CHART schedule was introduced at Mount Vernon Hospital, UK in 1985 aiming to improve local tumour control by increased dose intensity using many small fractions and a reduced overall treatment time (26,27). After initial pilot studies, two multicentre randomised trials were designed in 1989 to run in parallel, for the treatment of non small cell lung cancer (NSCLC) and head and neck squamous cell carcinoma (HNSCC), both of them comparing continuous hyperfractionated accelerated radiotherapy (CHART) with conventional radiotherapy. During 5 years of accrual, 13 centres entered a large number of patients, 918 with HNSCC and 563 with NSCLC.

All patients over the age of 18 years judged suitable for a radical course of external beam radiotherapy as definitive treatment were considered for inclusion. In the head and neck cancer study, those with early tumours (T1 N0) of the oral cavity, oropharynx, hypopharynx, and larynx were excluded but all stages of nasal sinus and nasopharyngeal carcinoma were included: histological proof of squamous cell carcinoma was essential. In the NSCLC study, all patients with tumours confined to the thorax and proven by histology or unequivocal cytology were eligible for inclusion. In both studies, any suspicion of distant metastasis excluded the patient and the WHO performance status was required to be 0 or 1. Co-existing disease prejudicing survival excluded the patient, as did any possibility that follow-up study might not be completed.

The randomisation procedure was designed to produce a 60% chance of a patient receiving CHART and a 40% chance of receiving conventional radiotherapy. In HNSCC, stratification was by centre, age, tumour site and nodal status. Assuming a 2-year local control rate of 45% in the conventional arm, it was calculated that an entry of 500 patients (230 events) was needed to detect an improvement of 15%, i.e. from 45% to 60% with a power approaching 90% at the 5% level of significance. However the HNSCC study recruited a total of 918 patients when it was decided by CHART Steering Committee to continue randomisation in order to run in parallel with the NSCLC trial.

In NSCLC, stratification was by centre, nodal status and WHO performance status. Assuming a 2-year survival rate of 15% in the conventional arm, 600 patients (475 events) were needed to detect an improvement of 10% in survival, i.e. from 15% to 25% with a power approaching 90% and a 5% level of significance. Randomisation

was performed by the method of minimisation by a telephone call to the Medical Research Council (MRC) Cancer Trials Office.

The radiotherapy planning was identical for all patients regardless of the treatment allocated. Radiation doses were prescribed to the intersection point as defined by international recommendations. The variation in radiation dose through the tumour volume in the central plane, from maximum to minimum, was normally limited to 10% of the prescribed dose. The treatment technique had to be identical for each arm of the trial. During the main part of the course of radiotherapy, the volume irradiated included the primary tumour, any involved lymph nodes and the area of lymphatic drainage. The final small volume included only the primary tumour and known nodal involvement together with a margin.

In NSCLC, the large volume irradiated included the mediastinum and the primary tumour together with a 1 cm margin. The mediastinum was defined as extending from the suprasternal notch to 3 cm below the carina. The ipsilateral hilar nodes and the paratracheal nodes but not the contralateral hilar nodes were included in the field. The area of the field at the isocentre should not exceed 240 sq cms. Areas of lead shielding could be subtracted from the total. In the second phase, the small volume treated included the primary tumour and known nodal involvement with a 1 cm margin. The area of the field at the isocentre should not exceed 140 sq cms. Lung correction factors were used.

In the head and neck study any arrangement of two or more intersecting beams was permitted. The target-absorbed dose was defined as the dose at the intersection of the central axes of the beams. For electrons, the dose was prescribed at the mid-

point of the target volume. Guidelines for the inclusion of the lymph node areas in the neck in the initial volume were drawn up for each tumour site and stage.

In the conventional treatment arm, all patients received a daily treatment dose of 2 Gy, which was given 5 days per week. In both trials, the large volume received 44 Gy in 22 fractions. The small volume in the head and neck patients then received 22 Gy in 11 fractions while in the patients with lung cancer 16 Gy were given in 8 fractions. Total doses in the two trials were therefore 66 Gy given in over 45 days and 60 Gy given over 42 days. These different doses were set recognising the larger areas normally treated in the chest and the greater sensitivity of some intrathoracic tissues. Both doses accorded with existing protocols for curative radiotherapy employed in a majority of the contributing centres. In the CHART arm, a dose of 1.5 Gy was given 3 times per day on each of 12 consecutive days including the weekend with an interval of at least 6 hours between treatments. The large volume received 37.5 Gy in 25 fractions followed by 16.5 Gy in 11 fractions to the small volume, giving a total of 54 Gy in 36 fractions.

A quality assurance team comprising a physicist, radiographer and bioengineer, drawn from the staff at the Cancer Treatment Centre at Mount Vernon, visited all centres to monitor the delivery of radiotherapy. Included among the checks was the use of 'phantom' patients so that the quality assurance of all aspects of treatment delivery could be made. A data manager and radiotherapist also took part in a quality assurance survey of the data at each UK centre, when completed proformas were compared with records of ten randomly selected patients.

All patients were seen weekly for the first 6 weeks from the start of treatment. Further follow-up attendances were at 8 and 12 weeks after the first day of treatment. Subsequently, patients attended 3 monthly up to 2 years, 6 monthly to 5 years and annually thereafter. At every visit a full clinical assessment and any appropriate radiological investigations were carried out. A blood count and chest X-ray were performed routinely at 6 monthly intervals in the head and neck study. A chest X-ray was taken at each follow-up visit and a CT scan was performed at 6 months in the NSCLC study. At each follow-up, treatment related side effects were recorded carefully.

Three major tumour endpoints for NSCLC were local tumour control, disease-free interval and overall survival. Local tumour control was defined as being achieved when there was either complete disappearance of all abnormalities from a chest X-ray or CT scan, or any residual abnormality observed at 6 months remained stable for a further 6 months or more.

In the HNSCC study, tumour endpoints were loco-regional control, primary recurrence, nodal recurrence, disease-free interval, freedom from metastases and overall survival. Local tumour control was defined as being achieved when there was complete disappearance of tumour within 6 months of commencing treatment. In both trials all endpoints were from the date of randomisation, and the disease-free interval was defined as the time from randomisation to progression of loco-regional disease or appearance of metastatic disease. Patients dying without progression of local/ loco-regional disease or appearance of metastatic disease were censored at the time of death. All other patients were censored at the time of their last follow-up. For both trials, overall survival was defined as the time from randomisation to death; patients

still alive were censored at the time last seen alive. Kaplan-Meier (life-table) curves were performed for each endpoint and compared across treatments by the logrank test.

Results of mature data in NSCLC (28) showed a 22% (P=0.008) reduction in the relative risk of death and a 21% (P=0.033) reduction in the relative risk of local progression. In the large subgroup of patients with squamous cell cancer, which accounted for 81% of the cases, there was a 30% (P=0.0007) reduction in the relative risk of death and 27% (P=0.012) reduction in the relative risk of local progression. Furthermore, in squamous carcinoma there was a 24% reduction in the relative risk of metastasis (P=0.043). The morbidity was essentially dysphagia during treatment and was more troublesome in the CHART treated cases, but settled satisfactorily in both arms of the study. Late radiation morbidity was defined as occurring from 6 months and was analysed with life-tables and prevalence methods. In all the assessments of late morbidity, no clear evidence of a difference in frequency or severity has emerged between treatment arms.

In the HNSCC study (29), life table analysis of loco-regional control, primary tumour control, nodal control, disease-free interval, freedom from metastasis and survival showed no evidence of differences between the arms. From subgroup analyses there was evidence of a greater response to CHART in younger patients (P=0.041) and poorly differentiated tumours appeared to fare better with conventional radiotherapy (P=0.030). In the larynx there was evidence of a trend towards increasing benefit with more advanced T stage (P=0.002). Acute radiation mucositis was more severe with CHART, occurred earlier but settled sooner and was healed by 8 weeks in nearly all cases in both arms. Skin reactions were less severe and settled more quickly in the

CHART treated patients. Osteoradionecrosis occurred in 0.4% of patients after CHART and 1.4% of patients after conventional radiotherapy. The incidence of chondritis or cartilage necrosis was similar in both arms. Life table analysis showed evidence of reduced severity in a number of late morbidities in favour of CHART. This was most striking for skin telangiectasia, superficial and deep ulceration of the mucosa and laryngeal oedema.

1.6 AIMS OF THE THESIS

Time-to-event endpoints or survival time are widely used in clinical cancer research. Understanding the principles of event definition, censoring and possible biases by competing risks is critical to obtain the unbiased estimates of time-to-event endpoints. One of the aims of this thesis was to apply a multivariate analysis of competing risks model to identify prognostic groups with varying risks of failure types. The other aim was to explore and clarify the properties of two estimators (KM and CI) in the presence of competing risks in two large randomised controlled trials where the CHART schedule was compared with conventional radiotherapy in two tumour sites with different natural histories. This was performed both for tumour outcome and late morbidity endpoints.

A summary is provided below:

- Apply a competing risks model to identify factors associated with local, nodal and/or distant failure using all available prognostic factors in CHART NSCLC and HNSCC studies.
- Form individual prognostic indices for failure in primary (T), nodal (N) and distant (M) positions; derive individual *risk profiles* with the main aim of developing a

rational basis for individualization of treatment prescription (that would enable us to predict the risk of specific types of failure with improved precision in the presence of competing risks).

- Estimate and compare time-to-failure rates using KM (1^{st}, any) and CI methodology in clinical groups of patients with different prognoses.
- Estimate and compare late normal tissue morbidity using KM and CI estimates in clinical groups of patients with different prognoses.

1.7 THESIS OUTLINE

Summary of the analyses is described in Table 1.2.

Table 1.2 Thesis Outline

Type of analysis	Population	Endpoints for analysis Time to first event	Co-variates	Methods of analysis
Modelling Failure-Specific Prognostic Factors	Non-small cell lung cancer (NSCLC) n=549	TN (loco-regional)-, M (distant) - position failures	Clinical and pathological prognostic factors	Competing risks analysis (BMDP) Log-logistic failure time distribution Accelerated failure time model
	Head and Neck Squamous Cell Cancer (HNSCC) n=309*	T (local)-,N(regional)-,M(distant)- position failures	Clinical and pathological prognostic factors and molecular biomarkers	
Estimation of Tumour Outcome	NSCLC n=549	Loco-regional Failure (LRF) Distant Failure (DF) rates	None	KM (1^{st}) KM (any) CI
	HNSCC n=918	Loco-regional Failure (LRF) Distant Failure (DF) Rates	None	KM (1^{st}) KM (any) CI
Estimation of Late Normal Tissue Morbidity	NSCLC n=560	Dysphagia Clinical pneumonitis Radiological pneumonitis rates	None	KM (1^{st}) CI
	HNSCC N=918	Dryness of the mouth Fibrosis Laryngeal oedema rates	None	KM (1^{st}) CI

* These are the patients where biopsy material could be retrieved and who had an assessment of all the molecular biomarkers

CHAPTER 2

PATIENT DATA AND VALIDATION AGAINST PUBLISHED RESULTS

2.1 NSCLC STUDY

A total of 563 patients were entered into the trial between April 1990 and April 1995 and the mature data updated to 1 April 1998 were published in 1999 (28). This report is based upon the data updated to 1 November 1999. The endpoint variables were defined and extracted from the follow-up case report forms (CRF) of each patient using the updated database. Data management included editing data values defining and recoding variables, and computing their values using syntax commands. Statistical Package for Social Sciences (SPSS) software for Windows version 9 and 10 was used for data management. SPSS syntax commands were written using the Syntax Reference Guide.

Overall survival was defined as the time from randomisation to death; patients still alive were censored at the time last seen alive.

Loco-regional progression-free rate was defined as time to progression of the primary tumour within the irradiated volume detected on a chest X-ray or CT scan. This information was gathered together from the radiological investigations performed at each follow-up and from evidence of death records. Patients dying without progression of loco-regional disease were censored at the time of their last follow-up.

Metastases-free rate was defined as the time from randomisation to appearance of distant metastatic disease. Patients dying without appearance of metastatic disease were censored at the time of death.

Kaplan-Meier curves were formed for each endpoint and compared across treatment arms by the log-rank test using SPSS survival analysis. These results were then correlated with the published results from 1999 (28).

2.1.1 Prognostic Variables

Sex, age, WHO performance score, T stage, N stage, clinical stage and histology were the prognostic variables in the NSCLC database. All variables were considered and entered into the analysis to start the initial model. Optimisation of variables with an ordered nature such as WHO status, T stage, N stage, clinical stage was carried out by creating dummy variables with incremental coding. Histology and sex were binary variables and age of the patient was a continuous variable. Patients with full information on clinical and pathological characteristics were entered into the competing risks analysis.

2.2 HNSCC STUDY

The definitive results of the HNSCC study were published in 1997 (29). This report is based upon the data updated to 1 March 1997. Clinicians taking part in the HNSCC study were also invited to take part in the cell kinetics study. CR analysis with the covariates was performed using data from patients with the information on the molecular markers and at the time of this analysis there were a total of 309 specimens assessable for all molecular markers.

Tumour outcome analysis was performed using 309 patients who had biological specimens for marker studies. Tumour outcome endpoints in this subset of patients were compared with the whole patient population.

Overall survival was defined as the time from randomisation to the time of death and if alive, patients were censored at the time of their last follow-up.

Local and nodal failures were recorded at the time when clinically definitive tumor growth was detected in the primary site or in the nodes. Distant failure was defined as appearance of metastatic disease outside the irradiated volume. All times were calculated from the date of randomisation.

Kaplan-Meier curves were formed for each endpoint and compared across treatment arms by the log-rank test.

2.2.1 Prognostic variables

All available prognostic factors, age, sex, WHO performance status, treatment, T stage, N stage, histological grade, site, Ki-67, p53, CD31, bcl-2, cyclin D1 scores and proliferative pattern, were assigned numerical scores and entered into the model. Optimisation of the variables was carried out by creating dummy variables and incremental coding where necessary. Cyclin D1 and age were the two continuous variables.

2.2.1.1 Molecular markers

Histologic material

Histologic material was obtained, retrospectively, from the referring hospitals. Each Pathology Department was requested either to provide the original blocks for processing at Mount Vernon Hospital or to cut up to twelve 4μm sections mounted onto poly-l-lysine coated slides. Each specimen was examined by Paul I. Richman at the pathology department to confirm the presence of SCC tumour and that the specimen was assessable. George D. Wilson and his group at the Gray Cancer Institute carried out the molecular analyses. As previously mentioned a total of 309 specimens were assessable for all molecular markers.

Immunohistochemical (IHC) staining

Sections of 4 μm were dried overnight at 37°C. Before antibody staining for each marker, the slides were blocked for endogenous peroxidase activity with a 3% solution of hydrogen peroxide in methanol for 30 minutes. Microwave irradiation was used to unmask the binding epitopes using either 10mM citric acid (ph 6.0) for Ki-67, p53, bcl-2 and CD31 or 1 mM EDTA (pH 8.0) in the case of cyclin D1. A cycle of 3 five-minute irradiations was used for all antibodies except bcl-2, which required only two cycles. The slides were then left to stand for 10 minutes in buffer at room temperature before being washed thoroughly in tap water. After three washes in Triethanolamine Buffered Saline (TBS), the slides were incubated in the different antibody in TBS containing 1 drop per ml of Dako Serum Free Protein block (Dako Ltd, High Wycombe, X0909) for 1hr at room temperature. The monoclonal antibodies to p53 (clone DO-7, Dako Ltd.) were used at a 1:75 dilution, CD31 (PECAM-1, Dako

Ltd.) at 1:30, bcl-2 (clone 124, Dako Ltd.) at 1:40 and cyclin D1 (Novacastra, Peterburgh, clone NCL-Cyclin D1-GM) at a dilution of 1:75. The antibody against Ki-67 was a rabbit polyclonal (clone MIB-1, Dako Ltd.) and used at 1:75 dilution. After three further washes in TBS, biotinylated rabbit anti-mouse antibody (Dako Ltd E0354) or, in the case of Ki-67, biotinylated swine anti-rabbit antibody (Dako Ltd.) diluted 1:400 in TBS was applied for 1 hour at room temperature. After three further washes, ABC complex (Dako Ltd K0355) was added for 1 hour at room temperature. The staining was visualised by adding diaminobenzidine (Vector Labs. DAB kit SK 4100) for 5 minutes at room temperature. Cyclin D1 was stained at a later date and Dako Envision polymer kit (Dako Ltd., K4006) was added for 30mins. Following three further washes in TBS, Envision DAB (Dako Lt.d) was added for five minutes. All slides were washed well in tap water and counterstained with Mayers Haematoxylin for 10secs-1min and then dehydrated, cleared and mounted in Dextropropoxyphene (DPX).

Assessment of staining

All slides were inspected for the presence of assessable tumour and semi-quantitatively analysed by a consultant pathologist (P.I.R) except in the case of cyclin D1.

For Ki-67 the slides were visually scanned and assigned to one of three scores; 1 = less than 20% positive cells, 2 = 20 to 40% positive and 3 = greater than 40% positivity. In addition, the proliferation pattern was assessed as previously described (30) where 1 = marginal (most organised), 2 = intermediate (mainly organised), 3 =

mixed (more than one pattern usually including random) and 4 = random (diffuse, disorganised staining).

Slides stained for bcl-2 were scored as negative if less than 5% of cells were stained and positive otherwise.

p53 protein was assessed in two ways 1) pattern of expression (whole specimen looked for staining) and p53 was classified as 1=negative (less than 5% positive cells), 2=sporadic (5-75% cells positive) and 3=all (>75% positive cells); 2) variation in intensity was semiquantitated as strong, moderate or weak in those specimens showing positivity.

The vessel count identified by CD31 staining was classified in 10 high power fields as 1=up to 35 vessels, 2=35-55 vessels and 3=greater than 55 vessels.

Cyclin D1 was assessed by manual counting aided by an in-house image acquisition system incorporating a CCIR-format 3-CCD chip colour camera coupled to a PC-based video-rate frame grabber. Software routines for image acquisition, normalisation and storage have been developed as have grids and manual counting data recording and export to spreadsheets. At least 10 high power fields (X40 objective) were counted for each specimen.

As an informal validation of the extraction of endpoints from the CHART databases, we reproduced the published results for tumour outcome endpoints by performing a conventional KM analysis and the log-rank test to compare the outcome across the treatment arms.

CHAPTER 3

COMPETING RISKS MODELLING OF FAILURE-SPECIFIC PROGNOSTIC FACTORS

The influence of the prognostic variables in relation to types and times of failures (failure-specific prognostic factors) was analysed using competing risks analysis as implemented in (Biomedical Data Package) BMDP statistical software (31). BMDP CR analysis utilizes the model proposed and demonstrated by Lagakos (32) for lung cancer patients who could have recurred locally or distantly, or be free of disease (censored) at the time of investigation. It provides a systematic framework for examining the effects of covariates. Time to first failure and the type of first failure (competing risks endpoints) were defined for both databases before running the CR analysis.

3.1 COMPETING RISKS (CR) ENDPOINTS

In the competing risks method used here, the failure types considered need to be mutually exclusive events. Thus, the endpoints have to be defined in such a way that the first failure in each patient can be assigned to a unique type of failure. In each patient, the time to first failure and its site (local, regional or distant) were recorded, or, in patients without clinical progression, the time of the last follow up was used as input data (censored cases) for the analysis. All times were calculated from the date of randomisation.

3.1.1 NSCLC Study

For each patient two types of first failures was defined (1) loco-regional failure (TN-position) and (2) distant failure (M-position). Patients with synchronous distant and loco-regional failure were classified as failing distantly. Patients without clinical progression were censored at the time of the last follow up. Two columns of data were obtained after these calculations, one with the *time* variable, which was the time from randomisation to the first failure (a TN- or a M-) or if no events were recorded, time from randomisation to the last follow-up. The other column showed the *status* variable, which was coded as 1 if the first event was distant metastases, as 2 if the first event was a loco-regional failure and 3 if the case was censored. We could have chosen any number to assign codes.

3.1.2 HNSCC study

In each patient three types of clinical outcome were defined and recorded as the time to first failure (1) isolated local (T-position), (2) nodal with or without local relapse (N-position) or distant (M-position) with or without concurrent T or N failure. Patients without clinical progression were censored at the time of the last follow up. Patients with isolated distant failures as the first failure and patients failing distantly with synchronous loco-regional failures were classified as having failed in the M-position. Patients with isolated nodal failure and patients with synchronous nodal and local failures were classified as having failed in the N-position. Isolated failures in the primary site were classified as having failed in the T-position. All event times were

calculated from the date of randomisation. Two variables were obtained after these calculations as in the NSCLC database but this time the *status* variable had four code labels which were 1 if the first event was a distant failure, 2 if the first event was a nodal failure, 3 if the first event was a local failure and 4 if the patient was a censored case. The other variable included the *time* to these events.

3.2 COMPETING RISKS (CR) ANALYSIS-BMDP

BMDP software provides a simple way of performing CR analyses (module 2L.11), with the insertion of the statement COMP=CRSK in the regression paragraph of the Survival Analysis program, 2L. After having identified time and type of first failure in each database using SPSS as described above, these files were transferred to BMDP by saving files as simple text files (standard ASCII files).

CR analysis was performed using an accelerated failure-time model, which is in many ways similar to the Cox proportional hazards model (CPHM) except that the underlying hazard function has a specified mathematical form and it is possible to incorporate a specified distribution. Here, we used a log-logistic failure time distribution, which is similar to the log-normal distribution but has the computational advantage that the cumulative distribution function has a simple analytic form (33). The log-logistic distribution is a two-parameter model with a non-monotonic hazard function that increases initially to a maximum and then decreases to zero as time tends to infinity. The log-logistic distribution is specified by the following survival function;

$$S(t) = 1/(1+(\lambda t)^{\gamma})$$

Where λ has the dimension of time (i.e. median failure time) and is a position or scale parameter for the p.d.f. and τ is the shape parameter. Here t is the time after randomisation.

The probability density and hazard functions for log-logistic distribution are given below.

$$f(t) = \frac{\lambda t(\lambda t)^{\tau-1}}{(1+(\lambda t)^{\tau})^2}$$

$$h(t) = \frac{\lambda t(\lambda t)^{\tau-1}}{1+(\lambda t)^{\tau}}$$

BMDP competing risks analysis requires Input 2L.11 to be defined. The first paragraph is called the INPUT paragraph, where the relevant file name, format, and the number of variables are defined. In the VARIABLE paragraph the variable names are defined using a comma after each one. Any new variable can be added to the model by entering VARIABLE ADD=1,2...in the VARIABLE paragraph. In the GROUP paragraph, the codes of the variables are defined. In the HNSCC study the interaction between the treatment arm and the other variables were defined in a paragraph called TRANSFORM where the variables added in the VARIABLE paragraph were introduced (see input HNSCC below). In the FORM paragraph, the *time* and the *response* (status) variables are defined. Response codes are the ones that include the competing events. Any codes not included in this list will be considered censored. In the REGRESS paragraph COMP=CRSK statement invokes the competing risks analysis. The accelerated failure time model and the selection of the failure-time distribution (log-normal, log-logistic, exponential, Weibull) are also defined in the

REGRESS paragraph. In this paragraph the covariates to be tested are also selected. The BMDP input data (2L.11) for NSCLC and HNSCC for the reduced models (explained at page 51) are given below:

Input 2L.11 for NSCLC (for the reduced model)

```
BMDP Instruction File: C:\DYNAMIC\BMDPRUN
BMDP2L--SURVIVAL ANALYSIS WITH COVARIATES

/ INPUT
        FILE = 'C:\DYNAMIC\NSCLC.DAT'.
        FORMAT = FREE.
        VARIABLES = 6.
/ VARIABLE
        NAMES = AGE, SEX, TREATMENT, CLINICAL STAGE, COMPETING RISKS (CRSK),
        FAILURE TIME.
/ GROUP
        CODES
        CODES (SEX) = 1,2.
        NAMES (SEX) = MALE, FEMALE.
        CODES (TREATMENT) =1,2.
        NAMES (TREATMENT) = CHART, CONVENTIONAL.
        CODES (CRSK) = 1, 2, 3.
        NAMES (CRSK) = METASTASIS, LOCO-REGIONAL, CENSORED.
/ FORM
        TIME = FAILURE TIME.
        RESPONSE = 1, 2.
/ REGRESS
        ACCEL = LLOGISTIC.
        COVARIATES = AGE, SEX, TREATMENT, CLINICAL STAGE.
        COMP = CRSK.
/ END
```

Input 2L.11 for HNSCC (for the reduced model)

```
BMDP Instruction File: C:\DYNAMIC\BMDPRUN
BMDP2L--SURVIVAL ANALYSIS WITH COVARIATES

/ INPUT
        FILE = 'C:\DYNAMIC\HNSCC.DAT'.
        FORMAT = FREE.
        VARIABLES = 11.
/ VARIABLE  ADD =1.
        NAMES = PROLIFERATIVE PATTERN, TREATMENT, CYCLIN-D1, Ki-67, N STAGE, Bcl-2,
        CD31, TRTCD31, T STAGE, COMPETING RISKS (CRSK), FAILURE TIME.
/ GROUP
        CODES
        CODES (PROLIFERATIVE PATTERN) = 1,2,3,4.
        NAMES (PROLIFERATIVE PATTERN) = MARGINAL, INTERMEDIATE, MIXED, RANDOM.
        CODES (TREATMENT) =0,1.
        NAMES (TREATMENT) = CHART, CONVENTIONAL.
        CODES (Ki-67) = 1, 2.
        NAMES (KI-67) = 0-20%, >20%.
        CODES (N STAGE) = 1, 2.
        NAMES (N STAGE) = N0, N (+).
        CODES (Bcl-2) = 1, 2.
        NAMES (Bcl-2) = NEGATIVE, POSITIVE.
        CODES (CD31) = 1, 2.
        NAMES (CD31) = LOW, INTERMEDIATE+STRONG.
        CODES (T STAGE) = 1, 2,3,4.
        NAMES (T STAGE) = T1, T2, T3, T4.
        CODES (CRSK) = 1, 2,3,4.
        NAMES (CRSK) = METASTASIS, NODAL, LOCAL, CENSOR.
/TRANSFORM
        TRTCD31=TREATMENT*CD31.
/ FORM
        TIME = FAILURE TIME.
        RESPONSE = 1, 2, 3.
/ REGRESS
        ACCEL = LLOGISTIC.
        COVARIATES = PROLIFERATIVE PATTERN, TREATMENT, Ki-67, N STAGE, Bcl-2, CD31,
        CYCLIND1, T STAGE, TRTCD31.
        COMP = CRSK.
/ END
```

BMDP competing risks analysis output includes failure-specific regression coefficients and standard errors (SE) for each covariate and from this summary it is possible to understand the relative importance of any covariate for the given failure type. A positive regression coefficient means that increasing values of the covariate will *increase* time to failure and a negative regression coefficient means that increasing values of the covariate will *decrease* the time to failure. This convention for the signs is in contrast to the convention used with the Cox Proportional Hazards Model (CPHM), where a positive regression coefficient indicates a positive effect on hazard function and a negative effect on failure times with increasing values of the corresponding covariate. This is because 2L takes the log of the times for each patient and then reparameterizes the model based on log times.

BMDP competing risk analysis provides an opportunity to test the influence of the covariates on type of failure, time until failure, or both.

H_{comb}: this is the *combined* hypothesis that the covariate has no influence on either type of or time until failure.

H_{FT}: the covariate has no influence on *failure type*.

H_{cond}; the *conditional* hypothesis that the covariate does not influence time to failure given that it does not influence type of failure.

The three hypotheses H_{comb}, H_{FT}, H_{cond} are interdependent. In particular the latter 2 hypotheses are contained within the H_{comb} hypothesis. The H_{FT} and H_{cond} hypotheses have to be considered only if H_{comb} is significant. This ensures there is consistency across the results for the three hypotheses and also provides considerable protection of the type I error for the testing of the H_{FT}, H_{cond} hypotheses. A result was considered

significant if $p \leq 0.05$. Non-significant variables ($p>0.05$), i.e. variables with no significant influence on time to failure and/or site of failure, were excluded from the model in a stepwise manner.

In HNSCC, in the reduced model, interaction terms between trial arm and each molecular marker were included one by one. If the interaction term was more statistically significant than the term representing the marker itself, the interaction term was retained in the model.

3.3 Derivation of the prognostic index (PI)

After identification of variables which may be of prognostic importance, knowledge of these may be combined and used to define a prognostic index, which in turn defines groups of individuals at different risk. Once the estimates of the coefficients had been calculated, a prognostic index (PI) for each individual patient for a specific failure type (FT) was calculated using the equations below:

For NSCLC:

$$PI_{FT} = \beta_{FT,sex} I(sex) + \beta_{FT,treatment} I(treatment) + \beta_{FT,age} \cdot age + \beta_{FT,clinicalstage} I(clinicalstage)$$

For HNSCC:

$$PI_{FT} = \beta_{FT,pattern}(pattern) + \beta_{FT,bcl2} I(bcl2) + \beta_{FT,cyclinD} \cdot cyclinD1 + \beta_{FT,Ki-67} I(Ki-67) + \beta_{FT,TRT*CD31} I(TRT * CD31) + \beta_{FT,tstage}(tstage) + \beta_{FT,nstage} I(nstage)$$

where the β's are the regression coefficients for the various prognostic factors for the specific failure type and I is the indicator function, which can have the values of 1

and 0, depending on whether a given characteristic is present or not. Cyclin D1 and age were the continuous variables.

3.4 Modelled failure-rate estimates

The prognostic indices for each individual patient for each failure type (T-, TN-, M-) were used to estimate the log-logistic (LL) failure time distribution for the patient using the formula:

$$F_{FT}(t) = 1 - S_{FT}(t) = 1 - \frac{1}{1 + (\exp(-\alpha_{FT} - PI_{FT}) \cdot t)^{\psi}}$$

where α is a constant and ψ is a shape parameter for the specific failure type (ψ is $1/\sigma$ where σ is the scale parameter as derived from the BMDP Manual Appendix B5 (31). This formula was calculated for each patient for many time points to form the 2-year modelled failure rates for each type of failure.

For the NSCLC study, the median values of 2-year modelled failure rates for TN- and M-failures were used to divide the whole patient population into four prognostic groups with different risks of failure.

For the HNSCC study, the 33 percentiles of the 2-year modelled failure rates were used to define 3 prognostic groups for each type of failure with different risks of T-, N- and M-failure.

3.5 Model fit

Modelled failure time distributions were compared with Kaplan-Meier estimates over time to examine visually the goodness of fit of the model for both studies. KM curves

of failure-free rates (TN-, M-) were plotted and compared using the log-rank test in SPSS version 9.0 between four groups.

For NSCLC the CI estimates were also correlated with the KM estimates and correlated graphically for each group for TN- and M-failure rates over time.

CHAPTER 4

ESTIMATION OF TREATMENT OUTCOME

CUMULATIVE INCIDENCE (CI) AND KM ANALYSIS

Treatment outcome (tumour outcome and late morbidity) rates were estimated by two methods, the cumulative incidence (CI) and the Kaplan-Meier (KM).

4.1 CUMULATIVE INCIDENCE (CI) METHOD

CI estimates with their standard errors (SE) were calculated using the syntax file written by S. Bond given in Appendix A. A *status* variable with the coding: 1= event, 2= competing, 3= censor was required. For example in the analysis of NSCLC, when the CI estimate for loco-regional failure was to be calculated it was coded as 1 (event) and distant failure as 2 (competing event). The CI calculation syntax cannot cope with missing values. Before running the syntax the memory allocation has to be increased. This can be done in SPSS by following the algorithm; edit>option>general>special workspace memory limit. This procedure results in two variables saved for each patient in the SPSS data file: the *cuminc* variable which is the CI estimate and the *SE* variable which is the standard error of the estimate. This syntax file also creates the CI curve over time.

4.2 KAPLAN-MEIER (KM) METHOD

KM analysis was performed in two ways.

4.2.1 KM (1st) estimate

For each patient the type of first event and time to first event were used as input data. This is the same *status* variable, which was used for the CI calculation. For example if the KM (1st) estimate for distant failure (DF) was to be analysed then DF should have happened as the first event to be counted as an event.

4.2.2 KM (any) estimate

For each patient the event of interest whether it was the first event or not and the time from randomisation to this event was used as input data. If LRF rate was to be estimated then a LRF occurring after a DF would still be counted as an event with KM (any) estimate.

The difference between the KM (1st) and KM (any) estimates arises from the definition of the other event. With the KM (any) option, the only other events are death or loss to follow-up. With KM (1st) estimate, an LRF occurring first is considered as another event when studying distant failure (DF) and conversely. Hence fewer events are taken into account with the KM (1st) method as compared with the KM (any) method.

KM (1st) and KM (any) estimates were analysed in SPSS survival analysis and the complement of KM estimate (1-KM) was saved in the SPSS file with its SE.

After calculations of the estimates using SPSS, the CI and KM estimates and their SE's were transferred to Microcal Origin version 6.0 where the layered curves were plotted and compared graphically.

4.3 ESTIMATION OF TUMOUR OUTCOME

Loco-regional failure (LRF) and distant failure (DF) rates were estimated and compared between different prognostic groups. In each patient, the time to first failure and its site (LRF or DF) were recorded, or, in patients without clinical progression, the time of the last follow up was used as input data (censored cases) for the analysis. All times were calculated from the date of randomisation. Patients with synchronous distant and loco-regional failure were classified as failing distantly.

4.4. ESTIMATION OF LATE RADIATION MORBIDITY

4.4.1 NSCLC Study

4.4.1.1 Definition of late morbidity endpoints and conventional analysis

Late radiation morbidity was defined as occurring from 6 months, the start being the beginning of the treatment. Three endpoints were considered in this analysis. These were; dysphagia, clinical pneumonitis and radiological pneumonitis. These endpoints were coded as described below in Table 4.1.

Table 4.1 Grades and definitions of late radiation morbidity endpoints in NSCLC used in this analysis

GRADE	Dysphagia	Clinical pneumonitis	Radiological pneumonitis
1	None	None	None
2	Some discomfort on swallowing- no disturbance of diet	Symptomatic, not interfering with lifestyle	Slight, alterations in radiotherapy (RT) field visible
3	Difficulty with swallowing -soft diet required	Symptomatic, requiring treatment	Moderate, patchy dense abnormalities in RT field
4	Considerable difficulty with swallowing-fluids only	Symptomatic, requiring hospitalisation	Severe, dense confluent radiographic changes in RT field
5	Severe difficulty with swallowing fluids		

If radiological pneumonitis was detected after one year from the commencement of treatment it was accepted as pulmonary fibrosis and the scoring was the same as for radiological pneumonitis.

As an informal validation of the extraction of endpoints from the updated database we reproduced the published results for late morbidity endpoints (dysphagia, clinical and radiological pneumonitis) by performing a conventional KM analysis and the log-rank test to compare the outcome across treatment arms. Any grade of dysphagia was considered as an event. All patients with moderate to severe clinical pneumonitis and radiological pneumonitis (X-ray/CT scan assessment) were considered. Here all late side effects, whether they were the first event or not were considered as an event.

4.4.1.2 Definition of CR endpoints

Late morbidity could only be detected if the patient lived long enough to experience it; late morbidity in the updated database has been analyzed using CI and KM estimates where death of the patient was considered as a competing risk.

For each patient, the time to first failure (late morbidity or death) was recorded, or, in patients without any of the events, the time of the last follow up was used as input data (censored cases) for the analysis. Time to death was calculated from the date of randomisation and time to late morbidity was calculated from the beginning of the treatment. These calculations again produced two columns of data, one with the *time* variable, which was the time from randomisation to the first event or if no events were recorded, time from randomisation to the last follow-up. The other column had the *status* variable, which was coded as 1 if the first event was late morbidity, as 2 if the patient experienced death before the late side effect and 3 if the case was censored.

4.4.1.3 Comparison of the CI and KM estimates

KM (1^{st}) and CI methods were performed and compared for late dysphagia, clinical pneumonitis and radiological pneumonitis rates between the treatment arms and in different irradiated volumes (small volume $<160cm^2$ vs. large volume $>200cm^2$). Dysphagia, grade 2 or more was considered as an event and estimates were correlated at the 2^{nd} and 4^{th} year. Clinical and radiological pneumonitis, grade 3 or more was considered as an event and estimates were correlated at the 4^{th} and 8^{th} year since pneumonitis was a later event.

Pulmonary fibrosis rates (grade 3 or more) after one year were estimated using CI and KM (1st) methods and correlated for two prognostic groups of patients who received the CHART schedule i.e. a uniform treatment. Cox Proportional Hazards analysis (backward, LR) was used to select patients with different prognoses from the CHART arm using overall survival as an endpoint. Patients who received the CHART schedule were divided into two prognostic groups; (1) N stage 0 and 1; (2) N stage 2 and 3.

4.4.2 HNSCC Study

4.4.2.1 Definition of late morbidity endpoints and conventional analysis

Late radiation morbidity was defined as occurring from 6 months, the start being the beginning of the treatment. Late radiation changes in the skin, subcutaneous tissue, mucosa, bone, cartilage and the central nervous system were routinely sought at each follow-up visit and each item was separately recorded. Twelve items recorded at follow-up were condensed to nine indices of radiation change by incorporating superficial and deep mucosal ulceration, subcutaneous fibrosis with oedema and thinning of mucosa with pallor.

In this study we have chosen three late morbidity endpoints because of the higher event rate, which were dryness of the mouth, laryngeal oedema and subcutaneous fibrosis, and oedema. The scores of these three endpoints are given below in Table 4.2.

Table 4.2 Grades and definitions of late radiation morbidity endpoints in HNSCC used in this analysis

GRADE	Subcutaneous fibrosis and oedema	Dryness of mouth	Laryngeal oedema
0	None	None	No
1	Slight	Slight	Yes
2	Moderate	Moderate	
3	Severe	Severe	

Moderate and severe subcutaneous fibrosis and oedema, and moderate and severe dryness of mouth were considered as an event. All grades of laryngeal oedema were counted as events. As an informal validation of the extraction of endpoints from the CHART database we reproduced the published results for dryness of mouth, fibrosis and laryngeal oedema using life-table (KM) methods in SPSS software for the two treatment arms. All patients (918) were included in this analysis.

4.4.2.2 Definition of the competing risks endpoints

In each patient, the time to first failure (late morbidity or LRF) was recorded, or, in patients without any of the events, the time of the last follow up was used as input data (censored cases) for the analysis. Time to LRF was calculated from the date of randomisation and time to late morbidity was calculated from the beginning of the treatment. These calculations again produced two columns of data, one with the *time* variable, which was the time from randomisation to the first event or if no events were recorded, time from randomisation to the last follow-up. The other column had the *status* variable, which was coded as 1 if the first event was late morbidity, or as 2 if

the patient experienced an LRF before the late side effect and 3 if the case was censored.

4.4.2.3 Comparison of the CI and KM estimates

The CI and the KM (1^{st}) and KM (any) methods were used to estimate the dryness of mouth, subcutaneous fibrosis and oedema and laryngeal oedema rates and were compared in early versus advanced T stage disease among N_0 CHART patients where LRF was the competing event.

CHAPTER 5

RESULTS-NSCLC STUDY

5.1 Validation against published results

Patient characteristics of 563 patients in the NSCLC database are shown in Table 5.1. The distribution of patients by sex, age, performance status, T stage, N stage, and clinical stage was similar in both arms of the trial.

Tumour outcome analysis included all 563 patients. The overall survival probability was 31% for the CHART arm and 21% for the conventional arm at 2 years. The 4-year figures were 13% for the CHART arm and 7% for the conventional arm. There were more events in both treatment arms than given in the published paper (28) since an updated version was utilised. Log-rank p value was 0.007. The overall survival curves are shown in Figure 5.1.

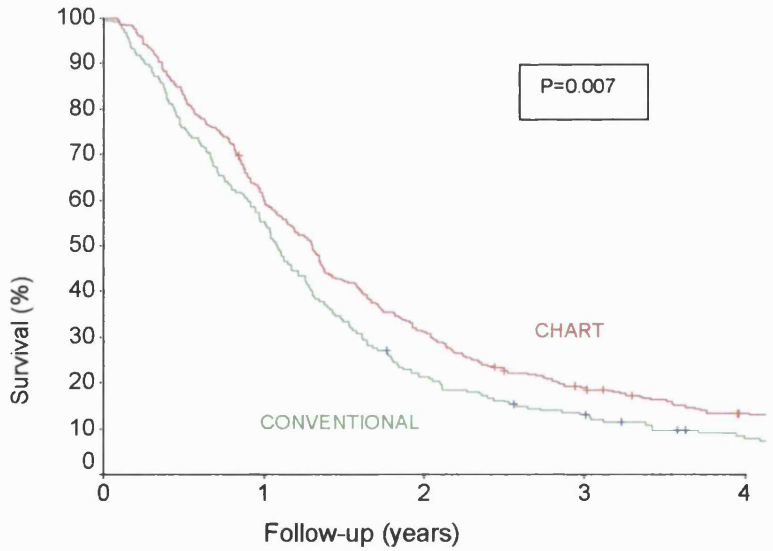


Figure 5.1 Kaplan-Meier Curves -Overall Survival by treatment allocated for all 563 patients

Table 5.1 Patient characteristics in NSCLC study (n=563)

Total patients	**CHART (n=338)**		**Conventional (n=225)**	
	No.	%	No.	%
Sex				
Male	267	79	166	74
Female	71	21	59	26
Age				
< 61	108	32	72	32
61-70	133	40	91	40
> 70	97	28	62	28
WHO				
(0) no restriction	135	40	95	42
(1) restriction	202	60	130	58
(2) ambulatory	1	0	0	0
T stage				
T1	32	10	15	7
T2	143	44	105	47
T3	84	25	57	25
T4	72	21	46	20
Unknown	2	1	2	1
Nodal stage				
NX	7	2	3	1
N0	163	48	107	48
N1	42	12	32	14
N2	117	35	77	34
N3	9	3	6	3
Clinical stage				
IA	22	7	11	5
IB	74	22	56	25
II	25	7	15	7
IIIA	129	38	86	38
IIIB	79	23	52	23
Unknown	9	3	5	2
Histology				
Squamous cell				
Well differentiated	39	12	26	12
Moderately diff	71	21	45	20
Poorly differentiated	34	25	62	27
Not specified	78	23	56	25
Other	66	19	36	16

In the published study, local tumour control was defined as either a) complete disappearance of all abnormalities in a chest X-ray or CT scan, or b) when any residual abnormality observed at 6 months remained stable for a further 6 months or more. Patients in either arm of the trial who did not achieve local control were defined as never being free of disease and thus were scored as having an event at time zero. Achievement of local control/relapse was not recorded as a separate variable for each patient and retrieving this information from the follow-up variables was not straightforward. The reason was that although all patients had CT scans for the 6^{th} month follow-up, not all of them had a CT scan for comparison at the end of the 12^{th} month. Some of them had only X-rays. The decision as to whether local control had been achieved could have been made by the clinician at the time of the follow-up physical examination and the comparison between the 6^{th} and 12^{th} month examinations and available radio diagnostic evidence.

In the present study, time to loco-regional progression was used as an endpoint and was defined as progression of the primary tumour within the irradiated volume detected on a chest X-ray or CT scan. The local control rate and local progression-free rate would be quite similar asymptotically. This information was gathered together from the radiological investigations performed at each follow-up and from evidence of death records.

There were 252 (74%) loco-regional failures (LRF) in the CHART arm and 176 (78%) LRF in the conventional arm. Loco-regional progression-free survival at 4 years was 14% for the CHART arm and 11% for the conventional arm, which was statistically significant (log-rank test p=0.015). The Kaplan-Meier curves for loco-regional

progression-free rates calculated in this study and the local tumour control from the published study are shown in Figure 5.2 (a) and (b). The 4-year estimates of these two different definitions for loco-regional effectiveness endpoint were quite similar around 10-15%.

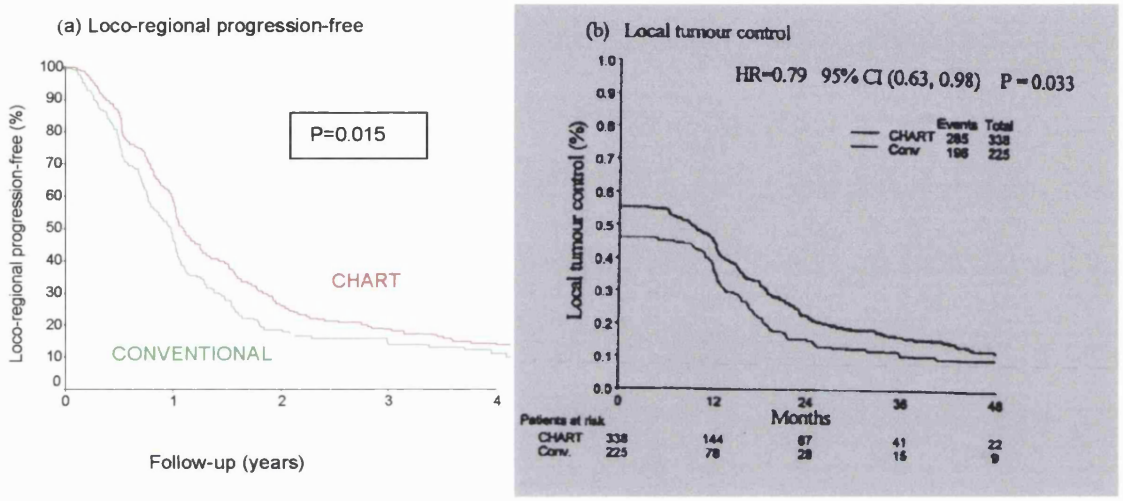


Figure 5.2 Kaplan-Meier curves (a) loco-regional progression-free rates and (b) local tumour control as published by treatment allocated for all 563 patients (28)

There were 276 patients with distant metastasis. Metastasis-free estimates were not statistically significant between treatment groups and the log-rank p value was 0.37. The metastasis-free rates are shown in Figure 5.3.

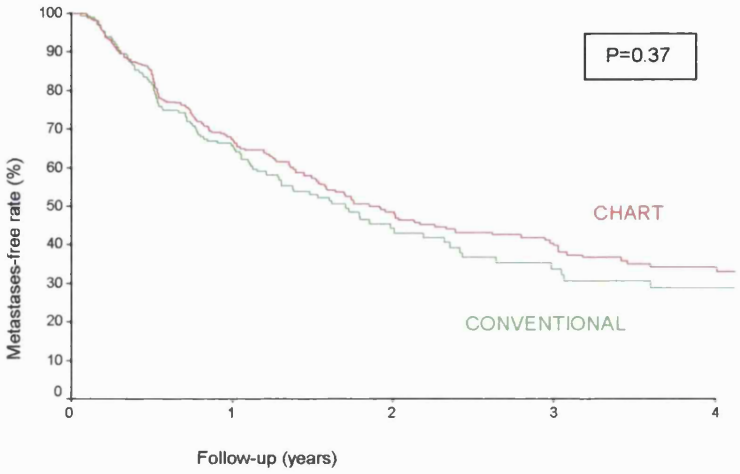


Figure 5.3 Kaplan-Meier curves-metastases-free rate by treatment allocated for all 563 patients

The results obtained from the updated database on the overall survival, loco-regional-progression free and metastasis-free rates were comparable to that of the published paper (28).

5.2 MODELLING FAILURE-SPECIFIC PROGNOSTIC FACTORS

Fourteen patients were excluded from the competing risks analysis because they did not have assessable T or N stages; thus information on all prognostic variables was available in 549 patients. Isolated loco-regional failure (TN-) was the first failure in 264 patients and distant failure with or without local failure was the first failure type in 201 patients (M-). The number of censored patients was 84. The variables (covariates) considered in the initial model after optimisation are given with the categories and numerical scorings in Table 5.2.

Table 5.2. Prognostic factors and their categories with numerical scores in the initial competing risks analysis group (n=549)

Covariate	Category	Score	No.	%
Sex	Male	1	423	77
	Female	2	126	23
Age	Continuous		549	100
WHO status				
	(0) No restriction	1	224	41
	(1) Restriction	2	325	59
T stage				
	T1	1	47	9
	T2	2	245	45
	T3	3	139	25
	T4	4	118	21
Nodal stage				
	N0	1	268	49
	N1	2	74	13
	N2	3	192	35
	N3	4	15	3
Clinical stage				
	I and II	1	203	37
	IIIA and IIIB	2	346	63
Histology				
	Squamous	1	316	58
	Other	2	102	20
Treatment				
	CHART	1	329	60
	Conventional	2	220	40

After excluding non-significant variables ($p > 0.05$) from the analysis in a stepwise manner, the final reduced model included only four variables- age, sex, clinical stage and treatment (see Appendix B).

Table 5.3 shows the regression coefficients for TN- and M-failures together with their standard errors. The magnitude of these regression coefficients should be seen in relation to their standard errors.

Table 5.3 Failure-specific regression coefficients of prognostic variables in the final model in NSCLC study

Type of Failure	Loco-regional failure (TN)		Distant failure (M)	
Covariate	Coefficient	Standard Error (SE)	Coefficient	Standard Error (SE)
Sex	0.2056	0.1241	0.1098	0.1799
Age	0.0076	0.0060	0.0168	0.0090
Treatment	-0.3379	0.1049	-0.0448	0.1545
Clinical stage	-0.0316	0.1080	-0.5455	0.1656

Information provided by Table 5.3 is shown graphically in Figure 5.4, which shows these estimates with 95% confidence limits for the 4 covariates included in the final model. If, for a given type of failure, the 95% confidence limits for the regression coefficient do not overlap zero, the coefficient has a statistically significant influence on the time to that type of failure. A positive regression coefficient corresponds to an increasing time to failure with increasing scores of the covariate and vice versa for a negative coefficient.

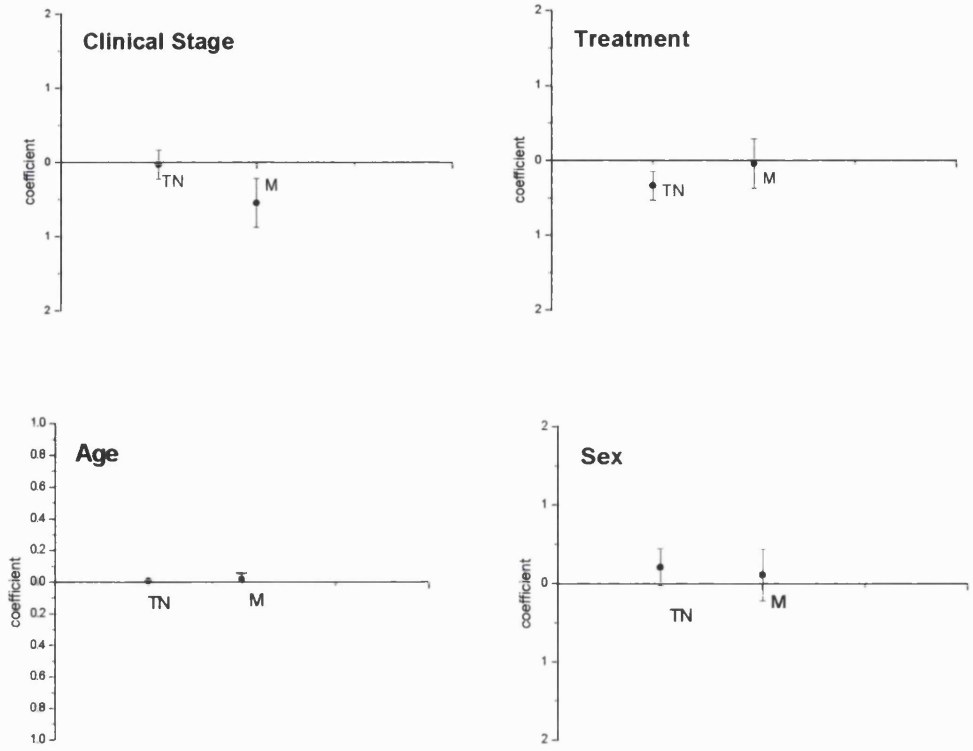


Figure 5.4 Regression coefficients with their 95% confidence limits for loco-regional (TN) and distant (M) failure for 4 covariates in the reduced model.

As shown in Table 5.2 clinical stage had two categories, stage I-II and stage IIIA- IIIB, the scores were 1 and 2 respectively. Using these scores Figure 5.4 shows us that patients with advanced clinical stage (IIIA and IIIB) had significantly decreased time to M-failures compared to patients with early stage (I and II). Clinical stage had no significant effect on TN-failures. Again from Table 5.2 the treatment arm was coded as 1 if the treatment assigned was the CHART arm, or as 2 if the treatment assigned was the conventional arm. The negative coefficient shown in the treatment panel of Figure 5.4 means that patients who received conventional radiotherapy had decreased time to TN-failure compared with the patients who received the CHART

schedule. Female gender was associated with a prolonged time to TN-failure compared to male gender.

The results of hypothesis testing are shown in Table 5.4 where the H_{comb} was significant for both the treatment arm (2P=0.005) and the clinical stage (2P=0.004), whereas a borderline significance was observed for age (2P=0.08). The analysis of H_{FT} and H_{cond} reveals that treatment arm had a significant influence on time to failure but was not selective for the type of failure. However advanced clinical stage was associated with a decreased interval to failure (2P=0.004) and a significantly increased risk (2P=0.009) of failing in distant rather than in local position.

Table 5.4 Results of hypotheses testing in the final model in NSCLC study

	H_{comb}	H_{FT}	H_{cond}
Covariate	**P-value**	**P-value**	**P-value**
Sex	0.211	0.661	0.088
Age	0.078	0.398	0.036
Treatment	0.005	0.116	0.005
Clinical stage	0.004	0.009	0.041

5.2.1 Prognostic Groups

The final model was used to construct prognostic indices and specific failure rates at 2 years were used to identify four prognostic groups. The scatter plot of the local and distant failure rates of individual patients is shown in Figure 5.5.

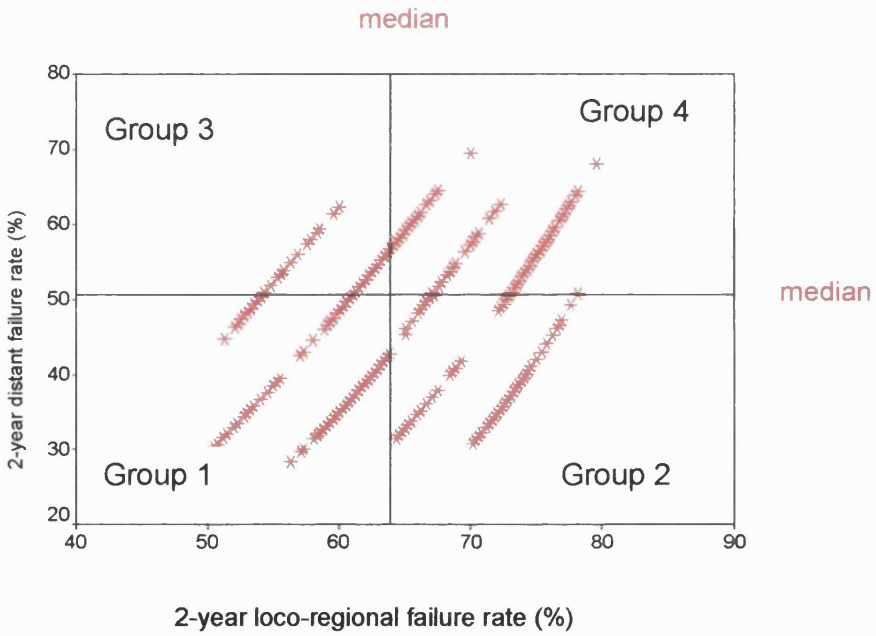


Figure 5.5 Scatter plot of the estimated local and distant failure rates at 2 years for each individual patient. The median values of failure rates were used to define four prognostic groups with different failure patterns. Each star represents a single patient

The median values of the estimated failure rates at 2 years were used to define four prognostic groups with different risks for local and distant failure.

Groups with a differing failure profile:

Group 1 (n=173): Low risk of both types of failures

Group 2 (n=103): Low risk of distant metastasis, high risk of loco-regional failure

Group 3 (n=100): High risk of distant failure, low risk of loco-regional failure

Group 4 (n=173): High risk of both types of failures

The performance of the model was first tested by standard methods and failure specific-free rates were compared using the log rank test. The results are shown in Figures 5.6 and 5.7. The local failure-free rates were 0.45 for group 2 and 0.44 for group 4 whereas they were only 0.32 and 0.28 for groups 1 and 3 at 2 years respectively. The log rank p value was 0.05 for local failure-free rate between 4 groups. The model was much more powerful for the distant metastasis-free rates, where the p value was 0.0008 and groups 3 and 4 had distant failure-free rates of 0.58 and 0.60 at 2 years respectively.

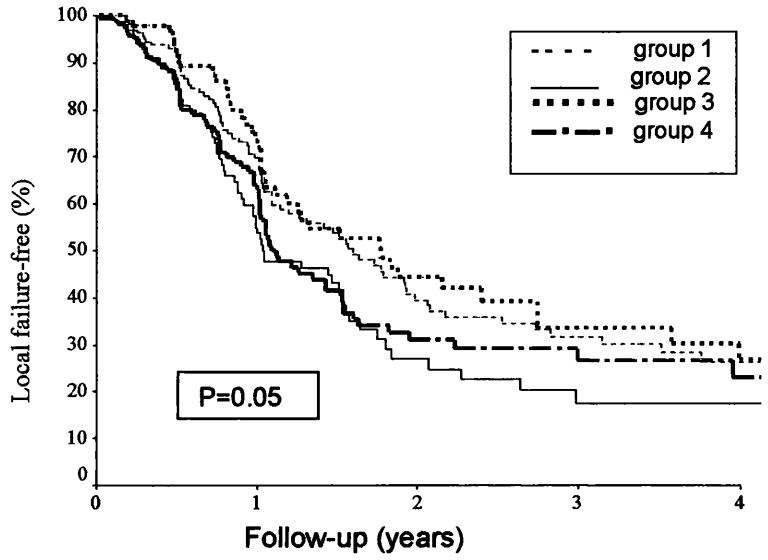


Figure 5.6 Local failure-free rates compared between the four prognostic groups using KM estimates and log-rank test. Group 1 and 3 have higher local failure-free rates than the other two groups as expected from the modelling.

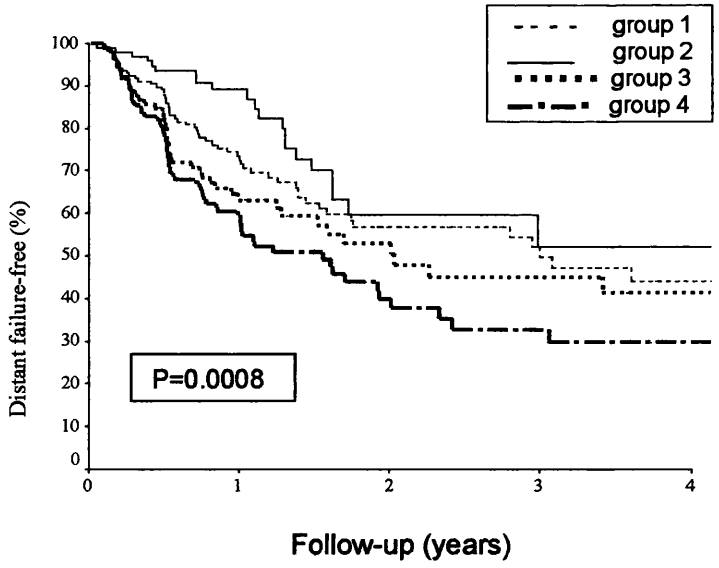


Figure 5.7 Distant failure-free rates compared between the four prognostic groups using KM estimates and log-rank test. Group 1 and 2 have higher distant failure-free rates than the other two groups as expected from the modelling.

5.2.2 Modelled Failure-Rate Estimates and Model Fit

Modelled failure time distributions were compared graphically with KM (1st) estimates over time in order to validate the model. There was close agreement between the modelled failure time distributions and KM (1st) estimates in all groups and these graphs are shown in Figures 5.8- 5.11.

(Group 1)

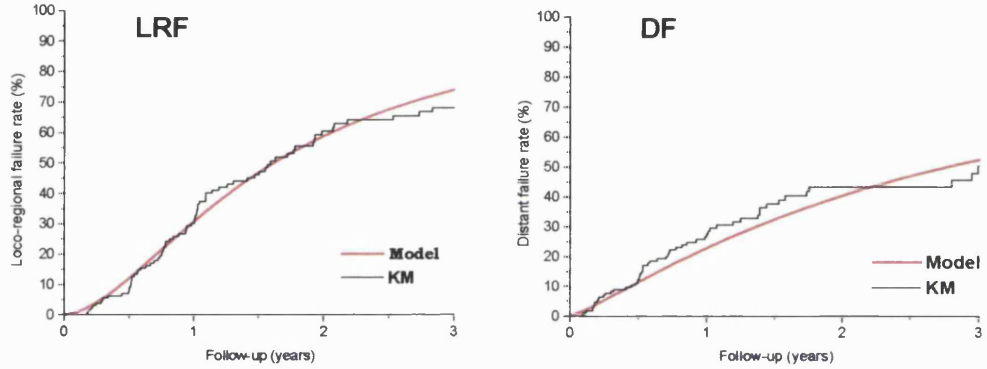


Figure 5.8 Modelled local and distant failure rate distributions in group 1 (red line) compared graphically with the KM estimate over time.

(Group2)

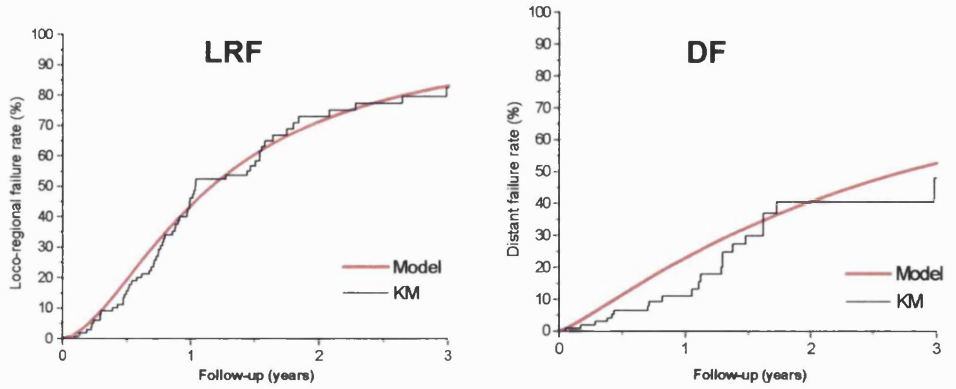


Figure 5.9 Modelled local and distant failure rate distributions in group 2 (red line) compared graphically with the KM estimate over time.

(Group 3)

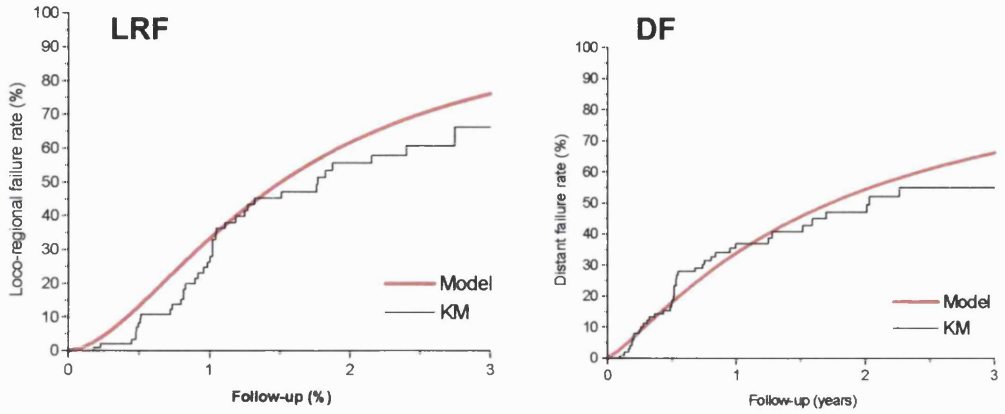


Figure 5.10 Modelled local and distant failure rate distributions in group 3 (red line) compared graphically with the KM estimate over time.

(Group 4)

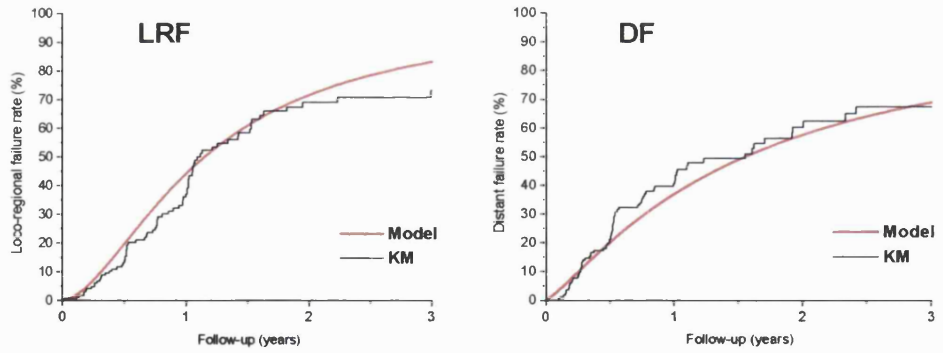


Figure 5.11 Modelled local and distant failure rate distributions in group 4 (red line) compared graphically with the KM estimate over time.

5.3 ESTIMATION OF TREATMENT OUTCOME -COMPARISON OF THE CI AND THE KM ESTIMATES

5.3.1 The CI and KM Estimates of the Prognostic Groups

The CI estimates and KM (1^{st}) estimates were calculated and correlated for two different types of failures for 4 prognostic groups at 4 years. These results are shown in Table 5.5.

Table 5.5 CI and KM estimates of failure specific prognostic groups at 4 years

	CI Estimates (SE)		**KM (1^{st}) Estimates (SE)**	
Prognostic Groups	**LRF**	**DF**	**LRF**	**DF**
Group 1 DF↓ LRF↓	.52 (.04)	.37 (.04)	.74 (.05)	.56 (.06)
Group 2 DF↓ LRF↑	.68 (.05)	.23 (.04)	.83 (.05)	.48 (.09)
Group 3 DF↑ LRF↓	.45 (.05)	.44 (.05)	.73 (.07)	.59 (.07)
Group 4 DF↑ LRF↑	.47 (.04)	.46 (.04)	.77 (.06)	.70 (.05)

DF↓ LRF↓: Low risk of both distant and loco-regional failures
DF↓ LRF↑: Low risk of distant, high risk of loco-regional failure
DF↑ LRF↓: High risk of distant, low risk of loco-regional failure
DF↑ LRF↑: High risk of both distant and loco-regional failure

The CI estimate for LRF rate in group 4 was only .47 when it was estimated to be .77 by the KM estimate. Groups 2 and 4 were predicted to have higher LRF rates but differing DF rates. The CI estimate of LRF rate in group 2 was .68 where there were fewer competing events i.e. DF. However the CI estimate showed a much lower figure (.47) compared to the KM estimate (.77) in group 4 where patients had a higher risk of failing distantly. KM and CI curves for groups 2 and 4 are shown in Figures 5.12 and 5.13.

(Group 2)

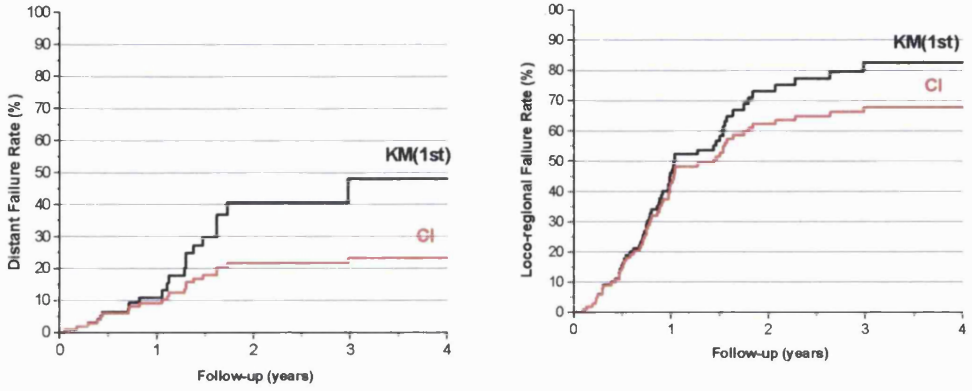


Figure 5.12 CI and KM (1st) estimates of distant and loco-regional failure rates in group 2

(Group 4)

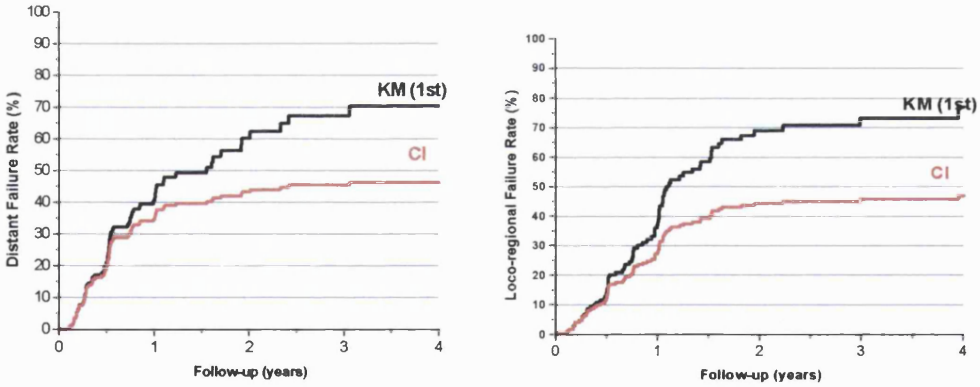


Figure 5.13 CI and KM (1st) estimates of distant and loco-regional failure rates in group 4

5.3.2 Estimation of tumour outcome

LRF and DF rates were correlated using KM (1st), KM (any) and CI estimates in early (stage I-II) and advanced (stage III) stages using 549 patients with complete clinical stage information. The characteristics of the patients in the early and advanced stage groups are shown in Table 5.6.

Table 5.6 Patient characteristics of early and advanced clinical stage groups (n=549)

Patient Characteristic		Early stage (n=203) No.	%	Advanced stage (n=346) No.	%
Sex					
	Male	148	73	275	80
	Female	55	27	71	20
Age					
	< 61	51	25	161	47
	61-70	83	41	143	41
	> 70	69	34	42	12
WHO					
	(0) no restriction	70	35	154	45
	(1) restriction	132	65	192	55
	(2) ambulatory	1	.5	0	0
T stage					
	T1	39	19	8	2
	T2	164	91	81	23
	T3	0	0	139	40
	T4	0	0	118	35
Nodal stage					
	N0	163	80	105	30
	N1	40	20	34	10
	N2	0	0	192	56
	N3	0	0	15	4
Clinical stage					
	IA	33	16	0	0
	IB	130	64	0	0
	II	40	20	0	0
	IIIA	0	0	215	62
	IIIB	0	0	131	38
Treatment					
	CHART	121	60	208	60
	Conventional	82	40	138	40
Competing risk endpoints					
	LRF	103	51	161	47
	DF	58	29	143	41
	Censor	42	20	42	12
LRF status					
	Event	147	72	273	79
	Censor	56	28	73	21
DF status					
	Event	85	42	185	53
	Censor	118	58	161	47

The KM (1st) estimates were almost identical around .72 for both clinical stages at 3 years for LRF rates. For early clinical stage, KM (any) estimate was .80 while it was .85 for advanced clinical stage for LRF, again at 3 years. The CI estimates for LRF rates were .56 for early and .47 for advanced clinical stage at 3 years. Distant failure rates were higher for advanced clinical stage compared to the early clinical stage for all three estimates. These rates are shown in Table 5.7.

Table 5.7 KM and the CI estimates of LRF and DF for early versus advanced clinical stage at 3 years

Type of failure at 3 years	Early Stage (I-II)			Advanced Stage (III)		
	KM (1st) (SE)	KM (any) (SE)	CI (SE)	KM (1st) (SE)	KM (any) (SE)	CI (SE)
LRF	.73 (.04)	.80 (.03)	.56 (.04)	.72 (.03)	.85 (.02)	.47 (.03)
DF	.45 (.05)	.54 (.04)	.30 (.03)	.60 (.04)	.69 (.03)	.42 (.03)

LRF and DF curves are shown in Figures 5.14 and 5.15 respectively.

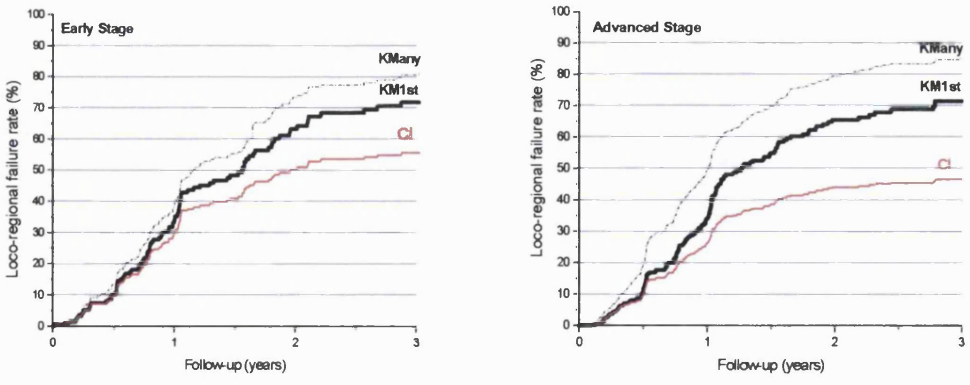


Figure 5.14 Loco-regional failure rate estimates using KM (1st), KM (any) and CI estimates in early and advanced clinical stage

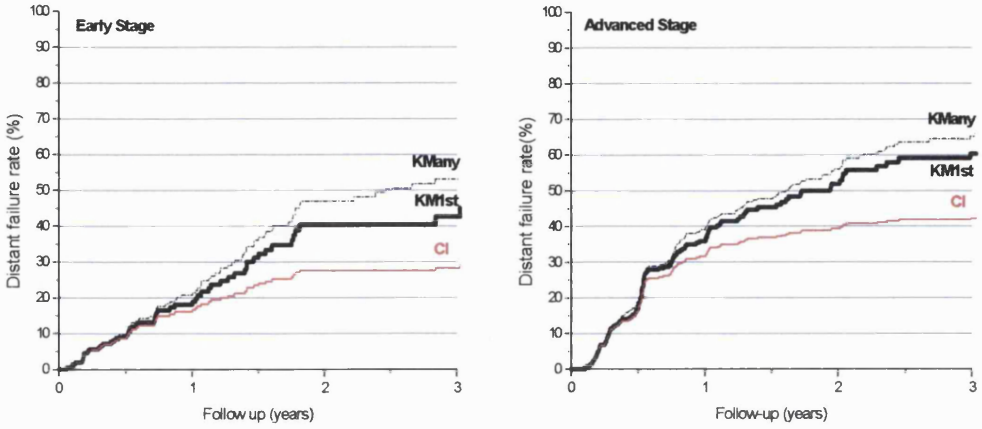


Figure 5.15 Distant failure rate estimates using KM (1st), KM (any) and CI estimates in early and advanced clinical stage

5.3.3 ESTIMATION OF LATE RADIATION MORBIDITY

5.3.3.1 Validation against published results

Late morbidity has been confined to pneumonitis and dysphagia. No case of radiation myelitis has been recorded in follow-up. Late radiation morbidity was defined as occurring from 6 months, the start being the beginning of the treatment. In the published study (28) the late morbidity was analysed using the life-table method and the prevalence at each follow-up.

For this informal validation of the extraction of endpoints from the CHART database we reproduced the published results for three endpoints using the life-table (KM) method. Three patients were excluded (date of treatment unknown) and 560 patients were in the analysis. Any grade of dysphagia was considered as an event and 85 patients were identified with events (440 dead, 35 censored). All patients with moderate to severe clinical pneumonitis were considered and 82 patients had this

event (445 dead, 33 censored). Moderate to severe radiological pneumonitis (X-ray/CT scan assessment) was recorded in 250 patients (298 dead, 12 censored).

The KM estimates comparing two treatment arms were almost identical with the published data and there was no statistically significant difference between the arms of the treatment. These reproduced graphs are shown in Figure 5.16.

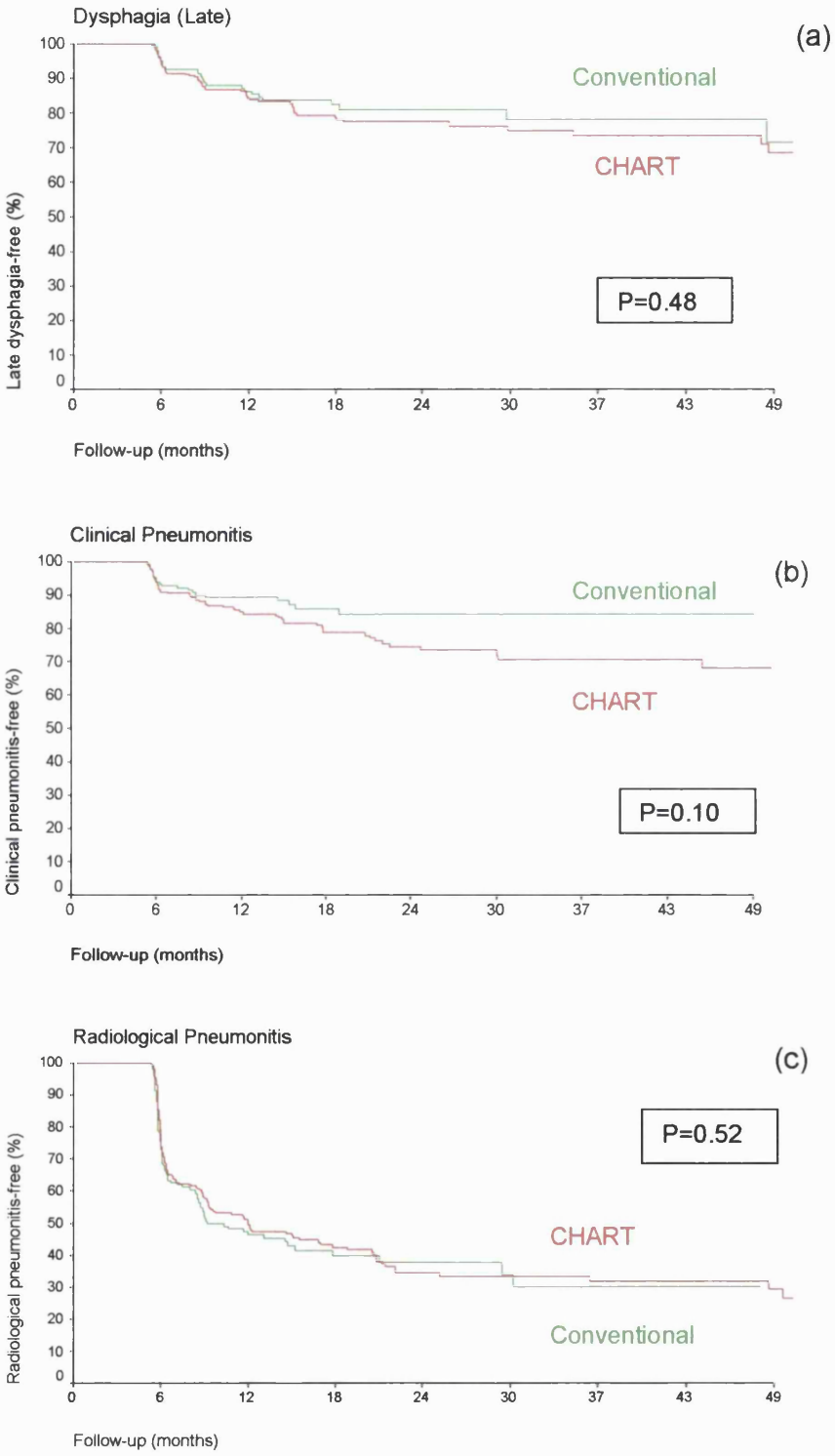


Figure 5.16 KM curves initiated at 6 months from commencement of treatment for all cases showing: (a) freedom from any dysphagia related to RT, (b) freedom from moderate or severe radiation pneumonitis as assessed clinically, (c) freedom from moderate or severe radiation pneumonitis as assessed radiologically

5.3.3.2 Comparison of the CI and KM estimates

The CI and the KM (1^{st}) estimates were studied between the treatment arms for dysphagia and clinical and radiological pneumonitis in 560 patients. In NSCLC study the death of the patient was considered as a competing risk.

Dysphagia

There were 60 patients who experienced late mild to severe dysphagia as the first failure and the number of competing events was 464. Thirty-six patients were censored at the time of the analysis. Table 5.8 shows CI and KM (1^{st}) estimates for grade 2 or more dysphagia with their standard errors at 2 and 4 years after treatment for all patients and for both treatment groups. These figures are also shown graphically in Figure 5.17.

Table 5.8 CI and KM estimates of grade 2 or more dysphagia at 2 and 4 years

Time	**2^{nd} year**		**4^{th} year**	
Dysphagia	**CI (SE)**	**KM (1^{st}) (SE)**	**CI (SE)**	**KM (1^{st}) (SE)**
All patients	.08 (.01)	.14 (.02)	.10 (.01)	.27 (.04)
CHART	.10 (.02)	.17 (.03)	.12 (.02)	.29 (.05)
Conventional	.06 (.02)	.12 (.03)	.08 (.02)	.22 (.06)

Morbidity related to dysphagia appeared similar in both arms of the study and it was estimated as .10 in the CHART arm and .06 in the conventional arm at 2 years by CI estimates. The KM estimates at 2 years were .17 and .12 for the CHART and

conventional arms respectively. The difference between the two estimates was larger (around 17%) at 4 years compared to the figures at 2 years (around 7%).

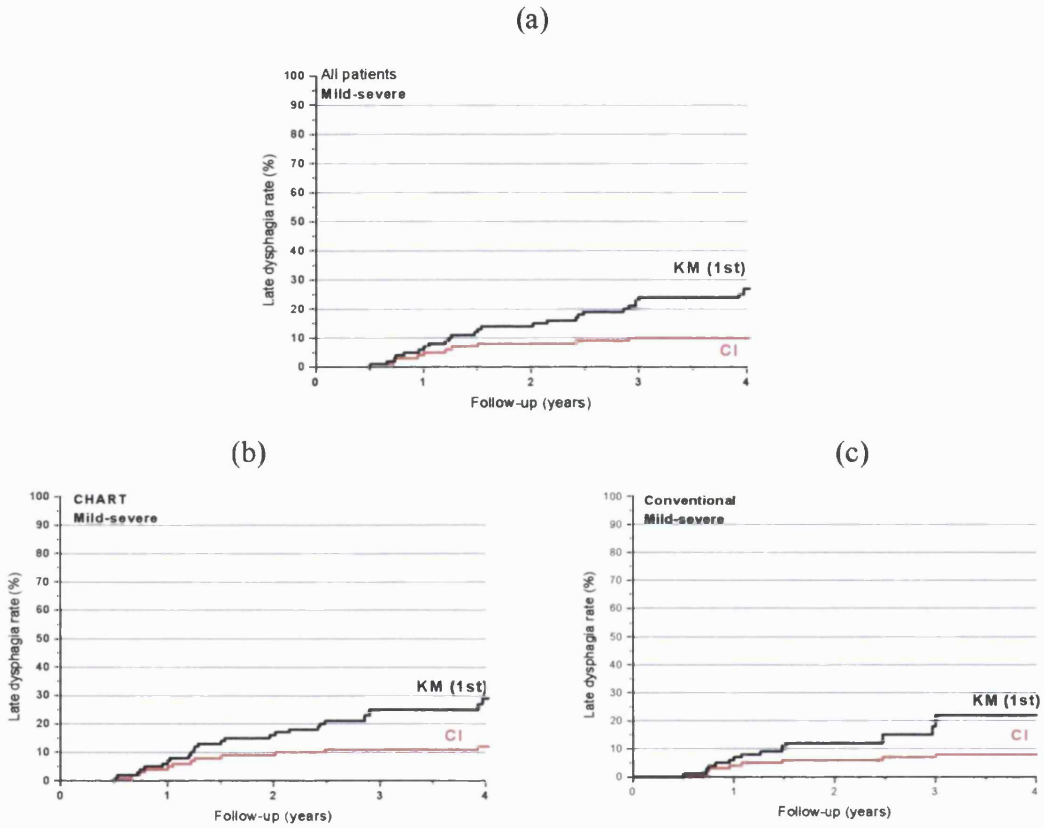


Figure 5.17 Late mild to severe dysphagia rates for (a) all patients (b) CHART arm and (c) conventional arm by KM and CI estimates

Clinical Radiation Pneumonitis

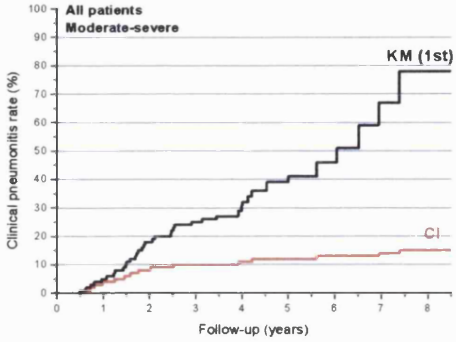
Moderate to severe late pneumonitis (grade 3, 4) was the first type of failure in 70 cases and there were 457 competing events. Thirty-three cases were censored. The CI and KM (1^{st}) estimates for late moderate or severe radiation pneumonitis rates with their standard errors at 4 and 8 years after treatment for all patients and for treatment groups are shown in Table 5.9. These figures are also shown graphically in Figure 5.18.

Table 5.9 CI and KM estimates of grade 3 or more clinical radiation pneumonitis at 4 and 8 years

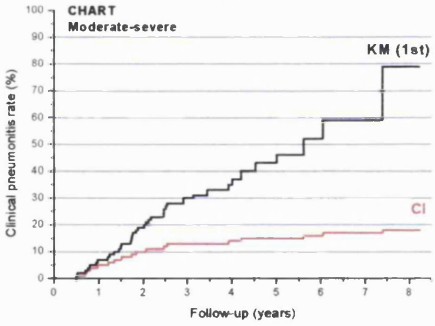
Time	**4^{th} year**		**8^{th} year**	
Clinical Pneumonitis	**CI (SE)**	**KM (1^{st}) (SE)**	**CI (SE)**	**KM(1^{st}) (SE)**
All patients	.11 (.01)	.32 (.04)	.15 (.02)	.78 (.11)
CHART	.14 (.02)	.37 (.05)	.18 (.02)	.79 (.15)
Conventional	.06 (.02)	.22 (.07)	.09 (.02)	.77 (.19)

Patients with clinical pneumonitis requiring treatment or hospitalization were estimated to be .14 in the CHART arm and .06 in the conventional arm at 4 years by CI estimate. However corresponding KM (1^{st}) estimates were higher (.37 in the CHART arm and .22 in the conventional arm) and the difference between the two estimates was even more pronounced at 8 years. The KM (1^{st}) method estimated that 78% of all patients would experience this morbidity regardless of the treatment arm. However CI estimates that this figure would only be 15%.

(a)



(b)



(c)

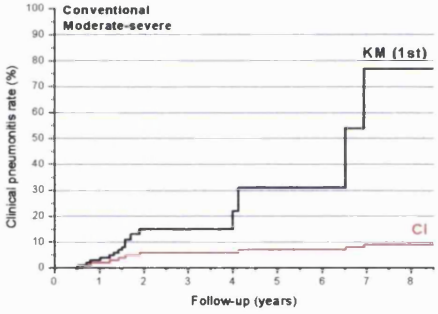


Figure 5.18 Late moderate or severe clinical pneumonitis rates for (a) all patients (b) CHART arm and (c) conventional arm by KM and CI estimates

Radiological Radiation Pneumonitis

Moderate to severe radiological pneumonitis was the first event in 217 patients. There were 329 competing events (death) and 14 cases were censored. The CI and KM (1^{st}) estimates for late moderate or severe radiation pneumonitis rates with their standard errors at 4 and 8 years after treatment for all patients and for treatment

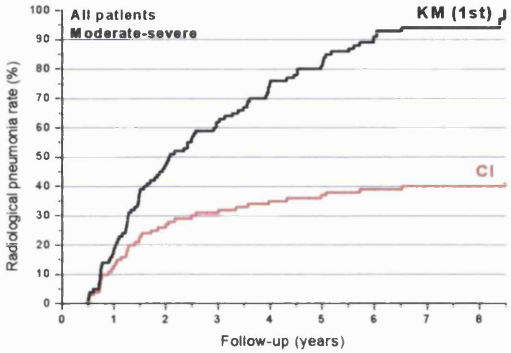
groups are shown in Table 5.10. These figures are also shown graphically in Figure 5.19.

Table 5.10 CI and KM estimates of grade 3 or more radiological radiation pneumonitis at 4 and 8 years

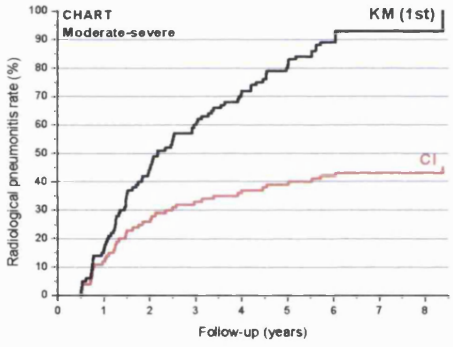
Time	4th year		8th year	
Radiological Pneumonitis	CI (SE)	KM (1st) (SE)	CI (SE)	KM (1st) (SE)
All patients	.35 (.02)	.76 (.03)	.40 (.02)	.94 (.02)
CHART	.37 (.03)	.72 (.04)	.43 (.03)	.93 (.03)
Conventional	.33 (.03)	.84 (.06)	.36 (.03)	.97 (.03)

With the previous morbidity data, the two estimates both showed increased morbidity estimates for the CHART treatment arm even if the figures were not correlated most of the time. The KM estimates of the radiological assessment of pneumonitis showed that the patients receiving conventional treatment were more likely to experience this morbidity compared to patients receiving CHART treatment. However the CI estimate showed a reverse trend. The KM method again estimated radiologically assessed radiation pneumonitis to be as high as 94% for all patients at 8 years whereas the CI estimate was only 40%.

(a)



(b)



(c)

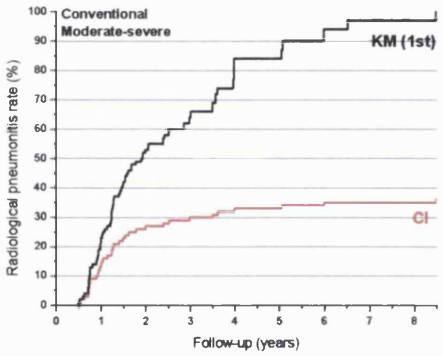


Figure 5.19 Late moderate or severe radiological radiation pneumonitis rates for (a) all patients (b) CHART arm and (c) conventional arm by KM and CI estimates

5.3.3.3 The effect of the irradiated volume

The dysphagia, clinical and radiological pneumonitis rates after 6 months and radiological pulmonary fibrosis rate after the first year have been estimated in two groups of patients (small volume <160cm^2 vs. large volume >200cm^2) using CI and KM methods. Characteristics of these 355 patients are shown in Table 5.11.

Table 5.11 Patient characteristics of small (<160cm2) and large (>200cm2) volume groups (n=355)

Patient Characteristic	Small Volume (n=178) No.	%	Large Volume (n=177) No.	%
Sex				
Male	112	63	155	88
Female	66	37	22	12
Age				
< 61	56	32	55	31
61-70	63	35	60	34
> 70	59	33	62	35
WHO				
(0) no restriction	74	42	63	35
(1) restriction	104	58	113	64
(2) ambulatory	0	0	1	1
Clinical stage				
I & II	78	44	56	32
IIIA &B	100	56	121	68
Treatment				
CHART	98	55	111	60
Conventional	80	45	66	40
Dysphagia				
Event	31	17	28	16
Death	133	75	143	81
Censor	14	8	6	3
Clinical pneumonitis				
Event	28	16	24	13
Death	134	75	148	84
Censor	16	9	5	3
Radiological Pneumonitis				
Event	80	45	76	43
Death	93	52	98	55
Censor	5	3	3	2
Pulmonary Fibrosis				
Event	51	29	15	8
Death	120	67	157	89
Censor	7	4	5	3

Competing risk was the death of the patient. Table 5.12 shows the two estimates for four morbidity endpoints at 2 and 4 years with their SE's for small and large irradiated areas.

Table 5.12 CI and KM estimates of 4 morbidity endpoints at 2 and 4 years for the small and large irradiated area groups

Morbidity	Follow-up (years)	Irradiated area <160 cm^2		Irradiated area > 200 cm^2	
		CI (SE)	KM (1^{st}) (SE)	CI (SE)	KM (1^{st}) (SE)
Dysphagia (Grade 2 or more)	2^{nd} year	.16 (.03)	.18 (.03)	.15 (.03)	.24 (.04)
	4^{th} year	.25 (.04)	.35 (.07)	.16 (.03)	.30 (.06)
Clinical pneumonitis (Grade 3 or more)	2^{nd} year	.14 (.03)	.24 (.04)	.13 (.03)	21 (.04)
	4^{th} year	.15 (.03)	.28 (.05)	.14 (.03)	.30 (.09)
Radiological pneumonitis (Grade 3 or more)	2^{nd} year	.44 (.03)	.61 (.05)	.42 (.04)	.69 (.06)
	4^{th} year	.45 (.04)	.66 (.05)	.43 (.04)	83 (.08)
Pulmonary fibrosis (Grade 3 or more)	2^{nd} year	.27 (.03)	.52 (.05)	.24 (.03)	.61 (.07)
	4^{th} year	.29 (.04)	.59 (.06)	.26 (.03)	.83 (.09)

The CI estimates were very close to each other for small or large irradiated areas for all morbidity endpoints. The KM estimates showed similar results for dysphagia and clinical pneumonitis for small and large volumes. However the KM estimates were quite different for radiological pneumonitis and pulmonary fibrosis rates. The KM estimate was .59 at 4 years for the small irradiated area and .83 for large irradiated area for the pulmonary fibrosis endpoint whereas the CI estimates were almost identical for both areas and the CI estimate was even lower for the large irradiated area. The KM and CI estimates of the radiological pneumonitis and pulmonary fibrosis rates are shown in Figure 5.20 and Figure 5.21 respectively.

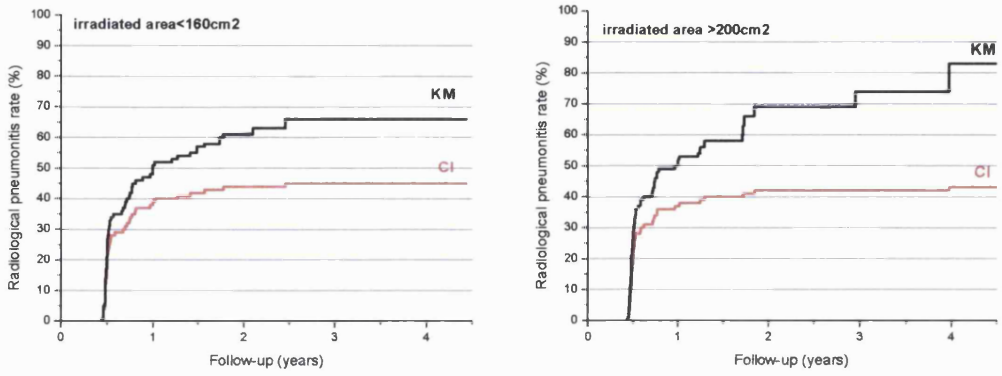


Figure 5.20 CI and KM curves for grade 3 or more radiological pneumonitis for the two different irradiated volumes

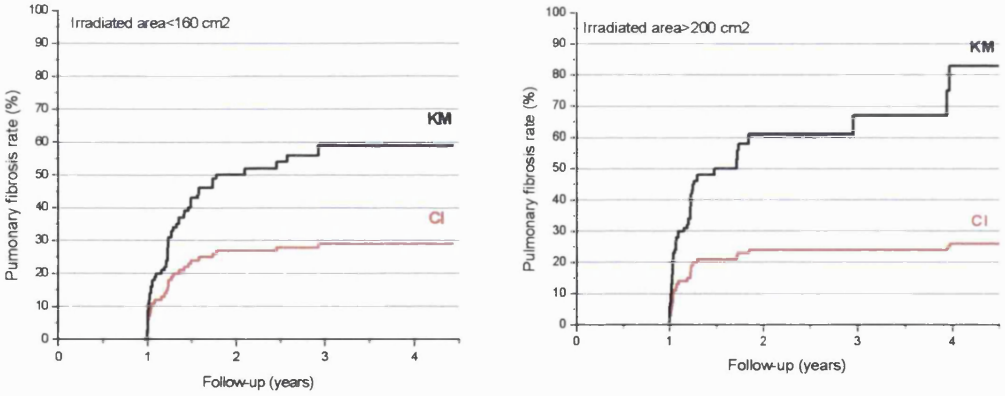


Figure 5.21 CI and KM curves for grade 3 or more pulmonary fibrosis for the two different irradiated volumes

5.3.3.4 Early versus advanced N stage-CHART Arm

Pulmonary fibrosis rates after 1st year were estimated and compared for two prognostic groups from the CHART arm. In the CHART arm (n=328) there were 203 patients with N0-1 disease and 125 patients with N2-3 disease. Patient characteristics are shown in Table 5.13.

Table 5.13 Patient characteristics in the CHART arm in NSCLC study

Characteristic	N0-1 (n=203)		N2-3 (n=125)	
	No.	%	No.	%
Sex				
Male	152	75	106	85
Female	51	25	18	15
T stage				
T1	27	13	5	4
T2	94	47	48	39
T3	45	22	37	30
T4	37	18	34	27
Nodal stage				
N0	161	49	0	0
N1	42	21	0	0
N2	0	0	116	35
N3	0	0	9	3
Treatment				
CHART	203	100	125	100
Pulmonary fibrosis				
Event	63	31	120	69
Censor	140	69	54	31
Competing risk endpoints				
Event	47	23	27	15
Death	117	58	76	44
Censor	39	19	71	41

The CI and KM estimate curves with their 95% confidence limits for pulmonary fibrosis after the 1st year in early and advanced N stage disease are shown in Figure 5.22. The analysis performed using the CI method suggested that the pulmonary fibrosis rate was lower in patients with advanced N stage disease. The CI estimate was .29 for N 0-1 disease and .25 for N 2-3 disease at 4 years. However the KM estimate showed an increase of almost 25% in the pulmonary fibrosis rate in the N2-3 patient group at 4 years.

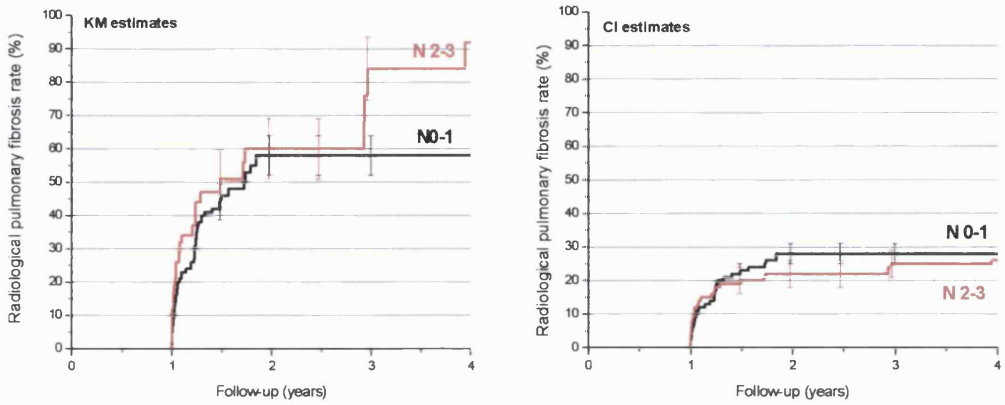


Figure 5.22 KM and CI curves of pulmonary fibrosis after 1st year for N0-1 and N2-3 disease with the 95% confidence limits

CHAPTER 6

RESULTS-HNSCC STUDY

6.1 Validation against published results

This analysis included 309 patients. The overall survival probability was 56% for the CHART arm and 50% for the conventional arm at 3 years. The 5-year figures were equal at 44% for both arms. Log-rank p value was 0.74. The overall survival curves are shown in Figure 6.1.

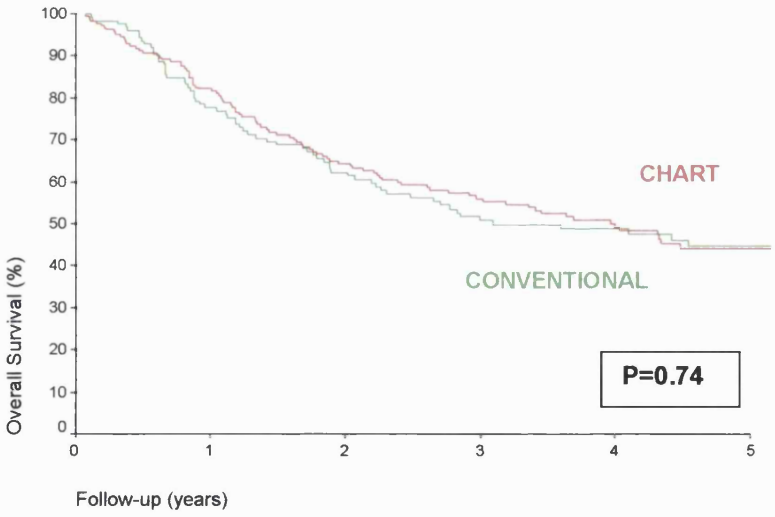


Figure 6.1 Kaplan-Meier curves -Overall survival by treatment allocated for 309 patients

Loco-regional progression free rates were 53% for the CHART arm and 46% for the conventional arm at 3 years. The 5-year rates were 50% and 46% for the CHART and conventional arms respectively (Figure 6.2). The log-rank p value was 0.28. In

the CHART arm there were 87 events (47%) and 97 censored cases whereas in the conventional arm 66 events (52%) were observed and 59 cases were censored.

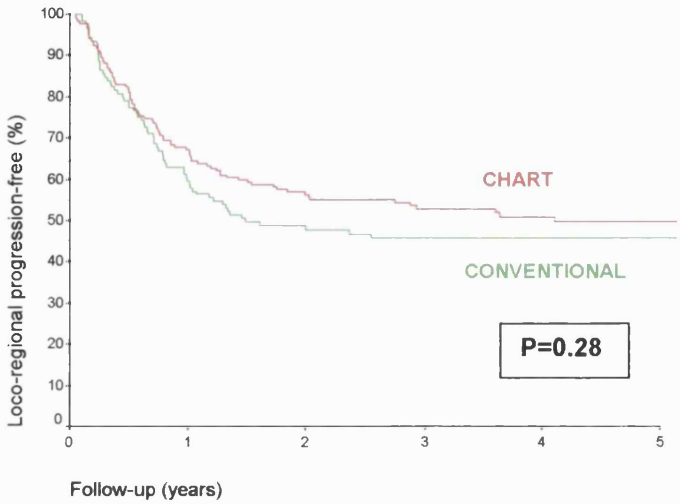


Figure 6.2 Kaplan-Meier curves of loco-regional progression-free rates by treatment allocated for 309 patients

Metastases-free rates were quite similar for both treatment arms and log-rank p value was 0.31. KM curves are shown in Figure 6.3.

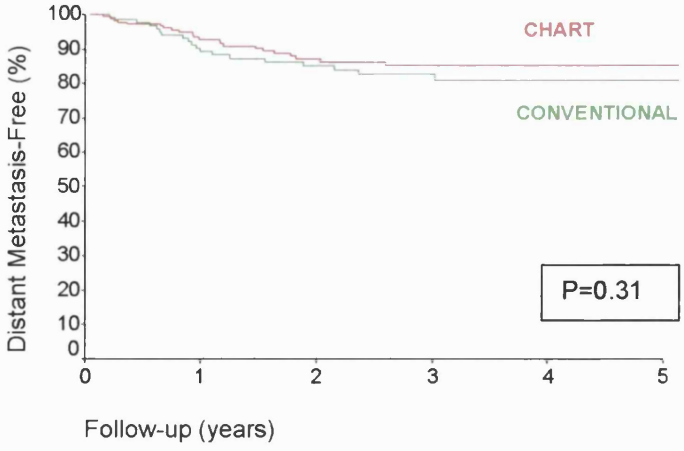


Figure 6.3 Kaplan-Meier curves of distant metastases-free rates by treatment allocated for 309 patients

In the published study no evidence of significant differences was identified by life table analysis of loco-regional control, primary nodal control, primary local control, disease-free survival, freedom from metastasis and overall survival (29). Our analysis of three endpoints showed no significant difference between the treatment arms and thus informally we confirmed that the subset of 309 patients who were analysed for the prognostic factor study had representative and comparable outcome to that of the whole group of 918 patients.

6.2 MODELLING FAILURE- SPECIFIC PROGNOSTIC FACTORS

The clinical and pathological prognostic factors and their covariate scores are given in Table 6.1 for 309 patients entered into the modelling study.

Isolated local failure, T-failure, was the first failure in 99 patients and nodal failure with or without local failure, N-failure, was the first failure in 41 patients. Twenty patients experienced an M-failure as their first failure, i.e. they had a distant failure with or without a synchronous loco-regional failure. The number of censored patients was 149.

Table 6.1 Analysed clinical and pathological prognostic factors and their covariate scores (n=309)

Characteristic	Category	Score	No.	%
Age	continuous		309	100
Sex	Male	0	225	73
	Female	1	84	27
WHO				
	No restriction	0	202	65
	Restriction	1	107	35
Treatment arm				
	CHART	0	184	60
	Conventional	1	125	40
T stage				
	T1	0	9	3
	T2	1	137	44
	T3	2	103	33
	T4	3	60	20
N stage				
	N0	0	188	61
	N+	1	121	39
Histology				
	Well diff	0	66	21
	Moderately diff	1	145	47
	Poorly diff	2	61	20
	Not specified	3	37	12
Site				
	Oropharynx	0	82	28
	Hypopharynx	1	34	11
	Larynx	2	138	46
	Oral cavity	3	42	15
Ki-67				
	0-20%	0	144	47
	> 20%	1	165	53
Proliferative pattern				
	marginal	0	52	17
	Intermediate	1	104	33
	mixed	2	42	15
	random	3	111	35
p53 score				
	negative	0	156	50
	sporadic	1	77	25
	positive	2	76	25
CD31 score				
	low	0	125	40
	Intermediate+ strong	1	184	60
Bcl-2 score				
	negative	0	269	87
	positive	1	40	13
CyclinD1	continuous		309	100

All available prognostic factors, age, sex, WHO performance status, treatment, T stage, N stage, histological grade, site, Ki-67, p53, CD31, bcl-2, cyclin D1 scores and proliferative pattern were entered into the model (Table 6.1). Stepwise regression was conducted as described in page 51 and the final reduced model included 7 variables: proliferative pattern, T and N stage, bcl2, cyclinD1, Ki-67 and the interaction term: treatment-arm x CD 31 score. Table 6.2 shows the regression coefficients for T-, N- and M-failures together with their standard errors (see Appendix C).

Table 6.2 Failure-specific covariate coefficients with standard errors in the final model

	Local failure (T)		Nodal failure (N)		Distant failure (M)	
Covariate	Coefficient	SE	Coefficient	SE	Coefficient	SE
P. pattern	-0.13	0.15	-0.23	0.24	-0.77	0.30
Cyclin-D1	-0.007	0.005	-0.02	0.008	0.001	0.01
Ki-67	0.42	0.34	0.42	0.51	1.53	0.64
N stage	0.24	0.33	-3.06	0.64	-0.86	0.58
Bcl-2	1.36	0.51	2.08	0.89	0.22	0.70
TRT*CD31	0.40	0.17	-0.14	0.27	0.40	0.32
T stage	-0.76	0.19	0.05	0.30	-1.33	0.40

Figure 6.4 shows these estimates with 95% confidence limits for the 7 covariates included in the final model. If, for a given type of failure, the 95% confidence limits for the regression coefficient do not overlap zero, the coefficient has a statistically significant influence on the time to that type of failure. A positive regression coefficient corresponds to an increasing time to failure with increasing scores of the covariate and vice versa for a negative coefficient.

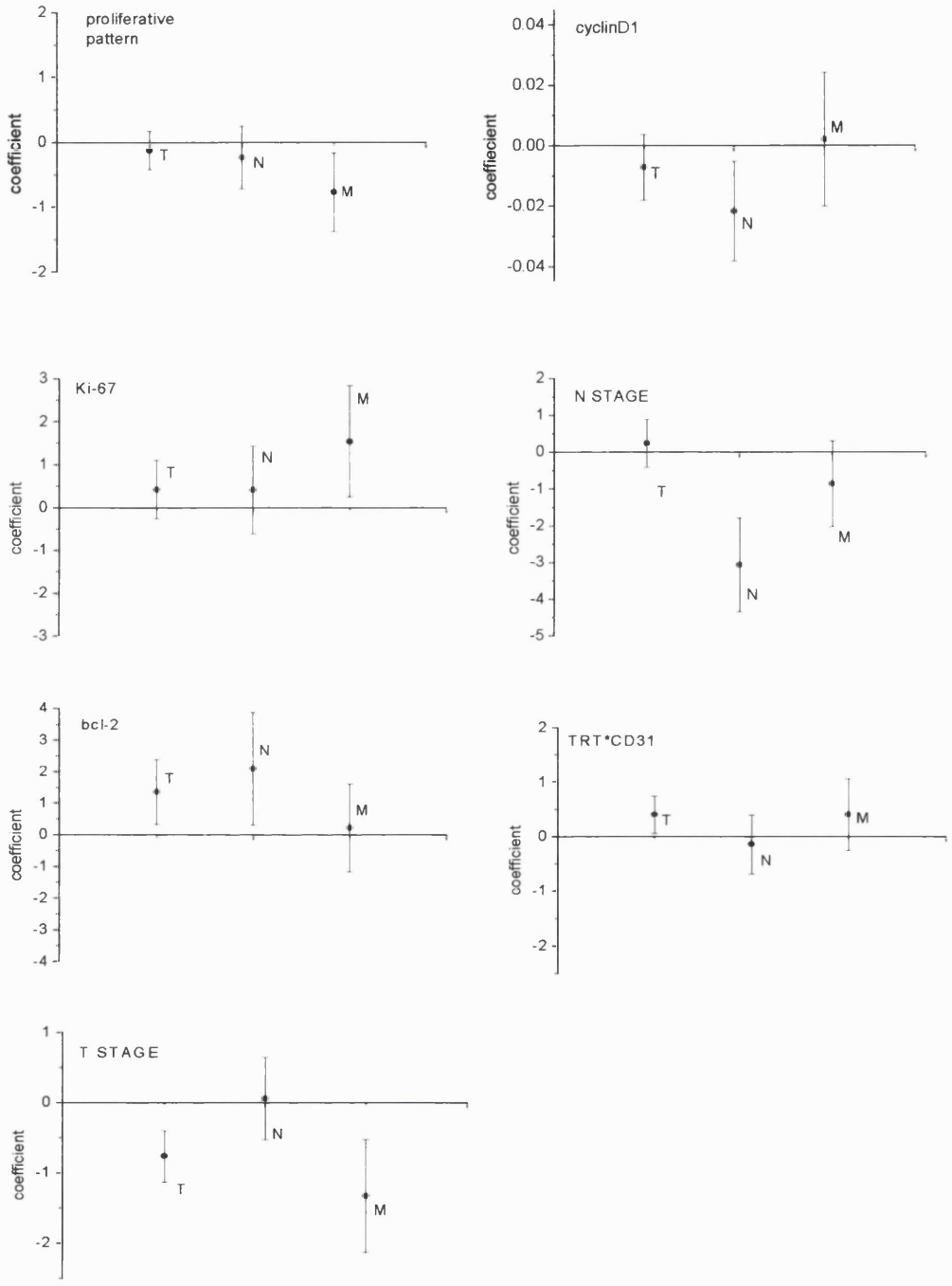


Figure 6.4 Regression coefficients with their 95% confidence limits for distant (M), nodal (N) and local (L) failure for 7 covariates in the reduced model. Positive values of a coefficient correspond to a prolonged time to the relevant type of recurrence. Negative values of a coefficient correspond to a shortened time to the relevant type of recurrence.

Take as an example the regression coefficient for M-failure for the covariate "proliferative pattern". This is -0.8 (Table 6.2) indicating that the time to M-failure is shorter in patients with the "random pattern", scored as 4, than in patients with the "marginal pattern", scored as 1. Figure 6.4 shows that proliferative pattern was not significantly associated with the time to N- or T-failures.

Patients with higher CyclinD1 scores had significantly decreased time to N-failure, whereas these covariates had no significant effect on time to T- and M-failures. (Figure 6.4) Node positive patients at the time of diagnosis had decreased time to N-failure as expected. Increasing T stage was associated with a shorter time to both T- and M-failures. Higher Ki-67 scores (>20%) were associated with increased time to M-failure without any significant effect on T- and N- failures. Bcl-2 positivity was associated with increased time to T- and N-failures, i.e. it was a favourable prognostic marker (Figure 6.4).

Among the first order interactions with therapy, only CD31 score was statistically significant (2P=0.02) with a regression coefficient showing an increased benefit of CHART relative to conventional fractionation in patients with a high CD31 score.

In order to explore this association further, we looked at the association between vascularity and histopathological grade of differentiation in the database. In 345 patients, who had a CD31 score as well as a histopathological grading, increasing values of the CD31 score were associated with more differentiated tumours (Kendall's tau, 2P=0.009).

6.2.1 Prognostic groups-modelled failure-rate estimates and model fit

From the regression coefficients of each covariate, prognostic indices for T-, N- and M-failures were calculated for each individual patient. Prognostic indices were used to estimate 2-year rates for each failure type.

The 33 percentiles of the 2-year failure estimates were used to define 3 prognostic groups for each type of failure with different risks of T-, N- and M-failure. In Figure 6.5 patients with low-, intermediate and high risk of local failure are shown in comparison to KM estimates over time. Patients in group T1 had the lowest local failure rate, which was only .23 at 3 years. However patients in group T3 had the highest local failure rate and this was as high as .51 at 3 years. A middle group was identified with a T-failure rate of .34 at 3 years. For each failure type, failure rates were highly significantly different (2P<0.00001, log rank test).

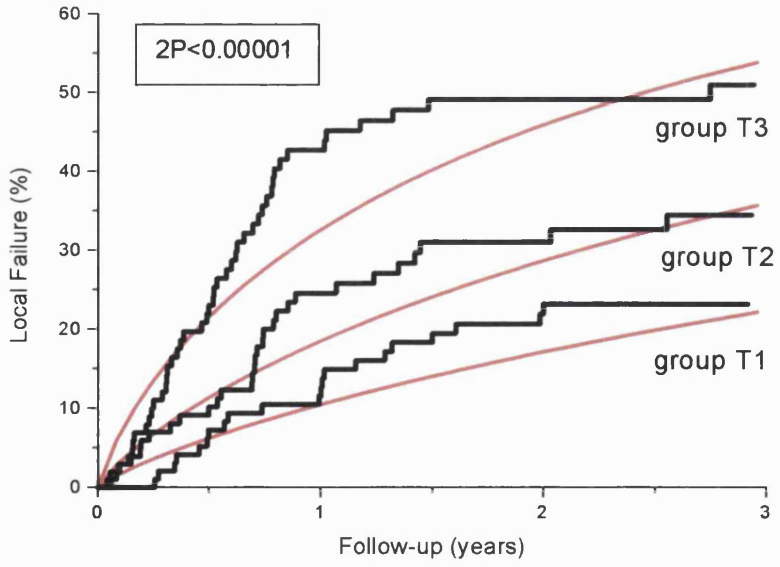


Figure 6.5 Modelled local failure rates over time in low-, intermediate and high-risk patients compared with KM estimates (black line).

In Figure 6.6 three prognostic groups defined from the 33 percentiles of the 2-year nodal failure rates are shown in comparison to KM estimates over time. While groups N1 and N2 showed nodal failure rates lower than .10 at 3 years, patients in group N3 had a N-failure rate of .42. Note that these groups are not the same groups as defined in T failure and that only the nodal failure rates were used to create N- groups.

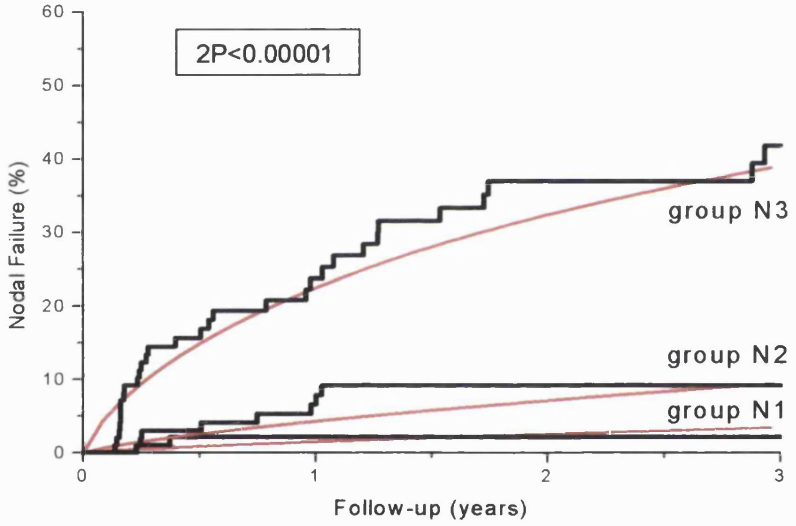


Figure 6.6 Modelled nodal-failure rates over time in low-, intermediate and high-risk patients compared with KM estimates (black line).

The 33 percentiles of the 2-year distant failure estimates were used to define 3 prognostic groups and the KM curves with the modelled failure curves of the prognostic groups are shown in Figure 6.7. While groups 1 and 2 showed M-failure rates as low as .03 at 3 years, the patients in group 3 had .26 rate of failing in M-position.

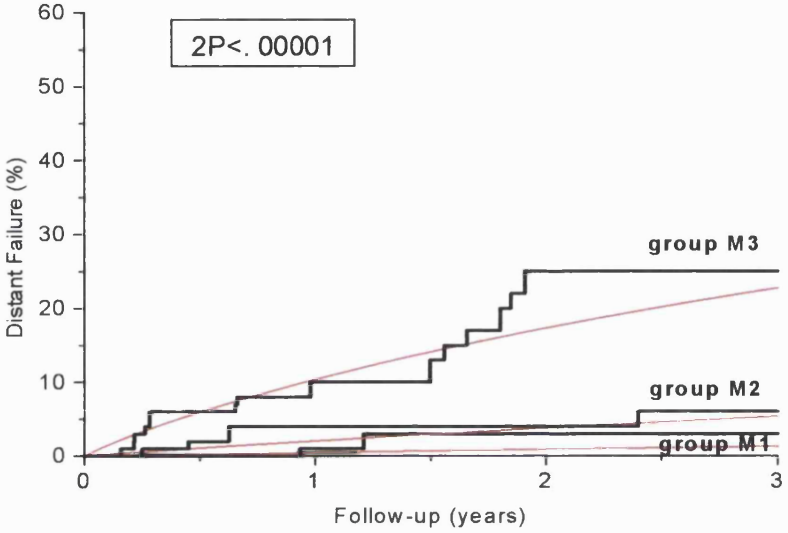


Figure 6.7 Modelled distant-failure rates over time in low-, intermediate and high-risk patients compared with KM estimates (black line).

Modelled failure rates were compared graphically with Kaplan-Meier estimates over time in order to check the model fit and for each failure type, this appeared to be very good.

6.2.2 Individual Risk Profiling

Two-year failure rates for each type of failure for each patient could be used to assess the risk profile of an individual patient for the three types of failures. Namely instead of assigning patients to prognostic groups we could profile the risk of failing at T-, N- and M-positions one by one for each single patient. This could be best shown in a three-dimensional scale where each patient stands at a particular point determined by the T, N, M co-ordinates. Figure 6.8 shows a pseudo-3D graph of the estimated 2-year failure rates in 10 selected cases. The number of cases plotted was limited for graphical clarity and the ten cases were selected to give an impression of the variation in failure rates.

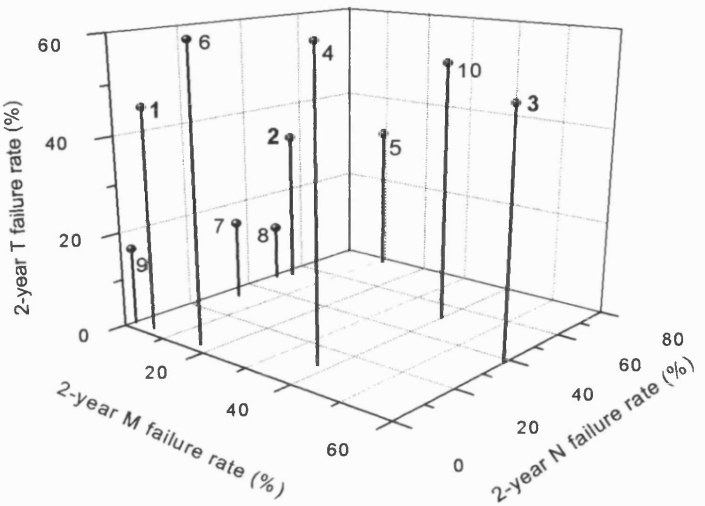


Figure 6.8 Scatter plot of the estimated T, N and M failure rates at 2 years for ten selected patients. Each little sphere represents a single patient and the drop lines show their projection on the M, N-failure plane.

6.3 ESTIMATION OF TREATMENT OUTCOME-COMPARISON OF THE CI AND KM ESTIMATES

6.3.1 Estimation of tumour outcome

This analysis included 912 patients (6 patients did not have date of last follow-up recorded) and the characteristics of early and advanced T stage patients are given in Table 6.3.

Table 6.3 Patient characteristics of early (T1-2) and advanced (T3-4) stage groups

Patient Characteristic	Early T Stage (n=438) No.	%	Advanced T Stage (n=474) No.	%
Sex				
Male	332	76	349	74
Female	106	24	125	26
Treatment				
CHART	251	57	297	63
Conventional	187	43	177	37
T Stage				
T1	28	6	0	0
T2	410	94	0	0
T3	0	0	293	62
T4	0	0	181	38
N Stage				
N0	323	74	274	58
N1	56	13	84	18
N2	45	9	78	16
N3	14	4	38	8
CR Endpoints				
LRF	164	37	264	56
DF	27	6	39	8
Censor	247	43	171	36

Table 6.4 shows the CI and KM estimates at 5 years for LRF and DF rates for early and advanced T stages.

Table 6.4 KM and the CI estimates of LRF and DF rates for early versus advanced T stage at 5 years

Type of failure at 5 years	Early Stage (T1-2)			Advanced Stage (T3-4)		
	KM (1st) (SE)	KM (any) (SE)	CI (SE)	KM (1st) (SE)	KM (any) (SE)	CI (SE)
LRF	.41 (.03)	.43 (.03)	.40 (.03)	.62 (.04)	.65 (.03)	.60 (.03)
DF	.09 (.01)	.14 (.02)	.07 (.01)	.14 (.02)	.22(.03)	.09 (.01)

The KM (1st), the KM (any) and the CI estimates of LRF rates for early T stage were almost identical around .40-.43. The advanced T stage LRF rates were higher (.60-.65) and were quite similar for the three estimates at 5 years. The KM (1st) and the CI estimates of the DF rates in the early T stage group were .09 and .07 respectively at 5 years. The KM (any) estimate showed a higher figure of .14. In the advanced T stage group the difference between the KM (1st) estimate and the CI estimate was .05 and the KM (any) estimate was as high as .22. These estimates are shown over time in Figures 6.9 and 6.10.

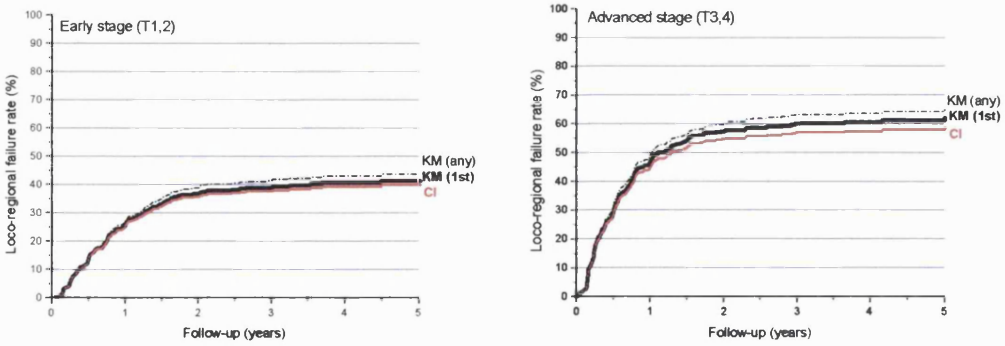


Figure 6.9 Loco-regional failure rates estimated by KM (any), KM (1st) and CI methods for early and advanced T stage groups.

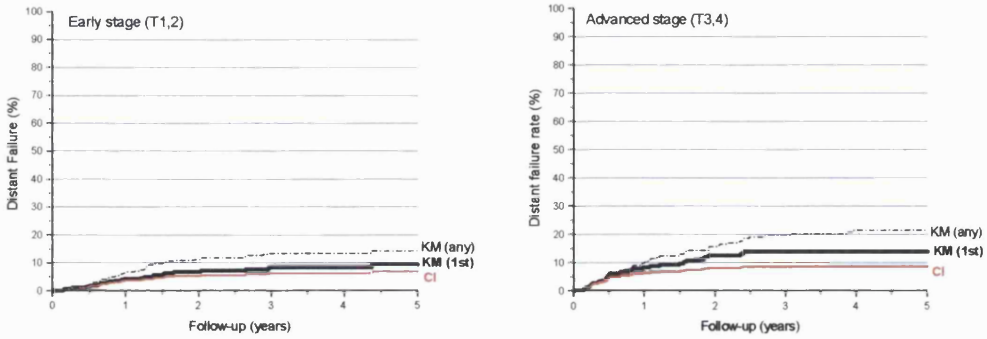


Figure 6.10 Distant failure rates estimated by KM (any), KM (1st) and CI methods for early and advanced T stage groups.

6.3.2.1 Validation against published results

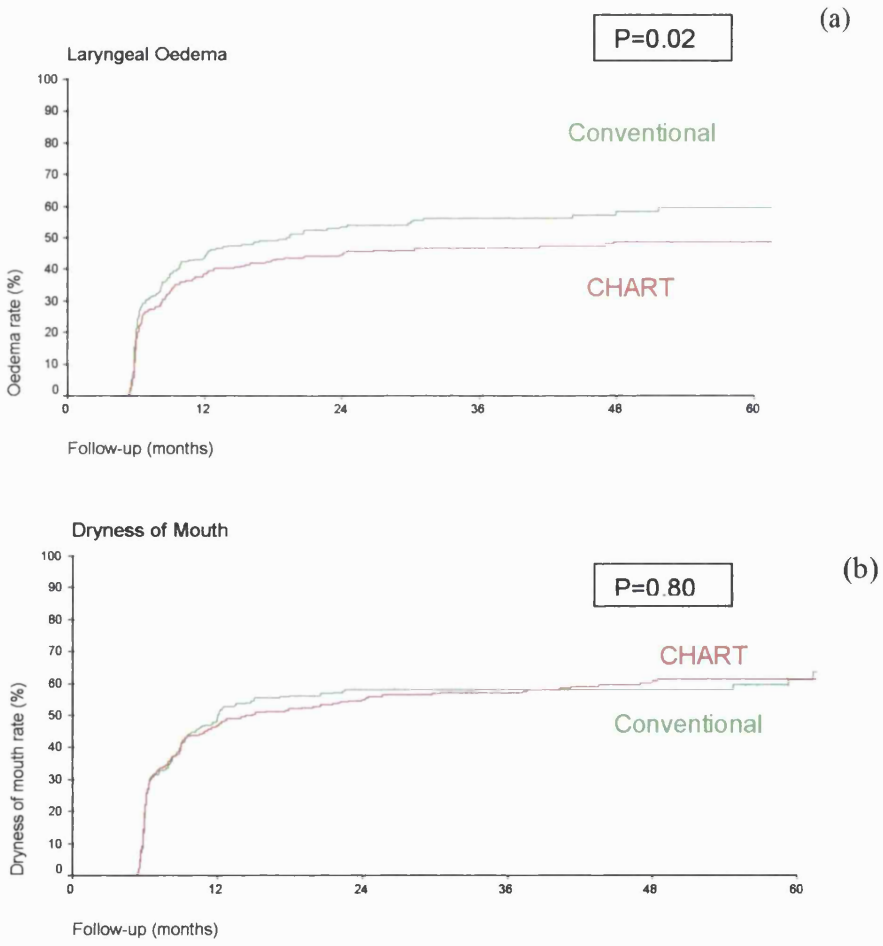
All patients (918) were included in this analysis. The frequencies of the morbidity endpoints evaluated in this study are shown in Table 6.5. For dryness of mouth and subcutaneous fibrosis and oedema, grade 2 or more were considered as events and all patients with laryngeal oedema were included.

Table 6.5 The frequency and percent of the three morbidity endpoints

Morbidity	Dryness of mouth	Laryngeal oedema	Subcutaneous fibrosis and oedema
Frequency (n)	434/918	370/918	294/918
Percent	47.3	40.3	32

The comparability of the database utilised with the published results was shown for dryness of mouth, subcutaneous fibrosis and oedema and laryngeal oedema using life-table (KM) methods for treatment arms. The log-rank test was used to compare the treatment arms.

The incidence of morbidity was significantly less in the CHART arm for laryngeal oedema and the log-rank p value was 0.02. The incidence of subcutaneous fibrosis and oedema was again less in the CHART arm but this was not significant at the 0.05 level. The KM curves with the treatment allocated comparison are shown in Figure 6.11.



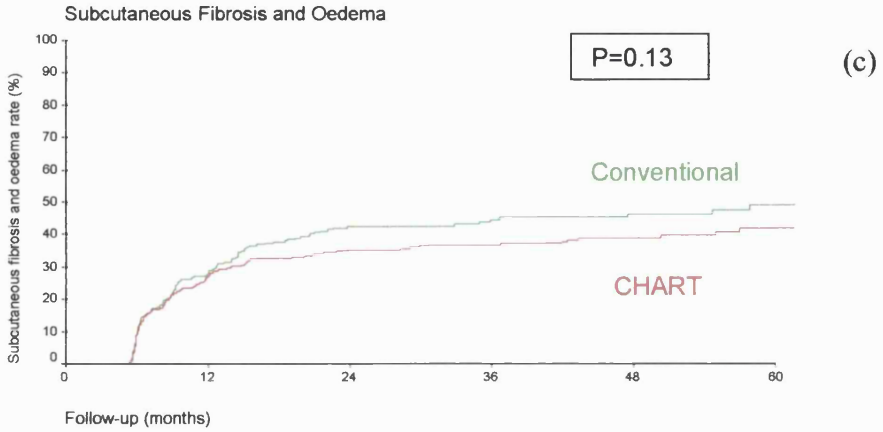


Figure 6.11 Incidence of late morbidity, KM curves: time 'zero' is set at 6 months from commencement of treatment for all cases showing: (a) laryngeal oedema, (b) moderate/severe dryness of mouth, (c) moderate/severe subcutaneous fibrosis and oedema

6.3.2.2 Comparison of the CI and the KM estimates

There were 360 N0 patients who received the CHART schedule. Two groups of patients with N0 disease were identified from the CHART arm. One group consisted of patients with T1-2 tumours and the other group with T3-4 tumours. The patient characteristics with the CR endpoints are given in Table 6.6.

Table 6.6 Patient characteristics of N_0 CHART patients (n=360)

Characteristic	T1-2 (n=186)		T3-4 (n=174)	
	No.	%	No.	%
Sex				
Male	144	77	121	70
Female	42	23	53	30
T stage				
T1	1	1	0	0
T2	185	99	0	0
T3	0	0	112	64
T4	0	0	62	36
Nodal stage				
N_0	186	100	174	100
Treatment				
CHART	186	100	174	100
Morbidity endpoints				
Dryness of mouth				
Event	100	54	106	61
Censor	86	46	68	39
Laryngeal oedema				
Event	105	57	104	60
Censor	81	43	70	40
Subcutaneous fibrosis and oedema				
Event	137	74	120	69
Censor	49	26	54	31
Competing risk endpoints				
Dryness of mouth				
Event	79	43	56	32
LRF	45	24	68	39
Censor	62	33	50	29
Laryngeal oedema				
Event	73	39	61	35
LRF	47	25	67	39
Censor	66	36	46	26
Subcutaneous fibrosis and oedema				
Event	32	17	27	15
LRF	55	30	76	44
Censor	99	53	71	41

The KM and CI estimates at 5 years for three morbidity endpoints are shown in Table 6.7. The CI estimates were lower for advanced T stage group for all three endpoints when compared to early T stage group. The most striking difference was noticed for dryness of mouth. The CI estimate indicated that there was 9% less dryness of mouth in patients with T3-4 disease. However in the advanced T stage group 39% of the cases experienced an LRF before the morbidity and this figure was only 24% in T1-2 disease. In Figure 6.12 the KM (any) LRF rate is also shown and in T1-2 disease it is .40 at 5 years, lower than the late morbidity rates. However in T3-4 disease LRF rate is not only raised to .55 at 5 years but also showed a gradual increase starting from 1.5 months to 1.5 years. KM (any) rates were very close to KM (1^{st}) rates for the dryness of mouth and the laryngeal oedema endpoints. For subcutaneous fibrosis and oedema rates KM (any) estimate was higher than the KM (1^{st}) estimate and this difference was more pronounced in T3-4 disease.

Table 6.7 KM and CI estimates of morbidity at 5 years for the early versus advanced T stage among N_0 CHART patient group

Morbidity at 5 years	CHART N_0 T1-2			CHART N_0 T3-4		
	KM (1^{st}) (SE)	KM (any) (SE)	CI (SE)	KM (1^{st}) (SE)	KM (any) (SE)	CI (SE)
Dryness of mouth	.56 (.05)	.55 (.04)	.46 (.04)	.55 (.05)	.52(.05)	.37 (.04)
Subcutaneous fibrosis and oedema	.23 (.04)	.32 (.04)	.19 (.03)	.28 (.05)	.43 (.05)	.18 (.03)
Laryngeal oedema	.49(.04)	.50 (.05)	.42 (.04)	.55 (.05)	.52 (.05)	.39 (.04)

The CI and KM estimates are shown in Figures 6.12-6.14. In Figure 6.12 the KM estimate for LRF was also plotted to show the increase in LRF rate between the two T stages with N_0 disease.

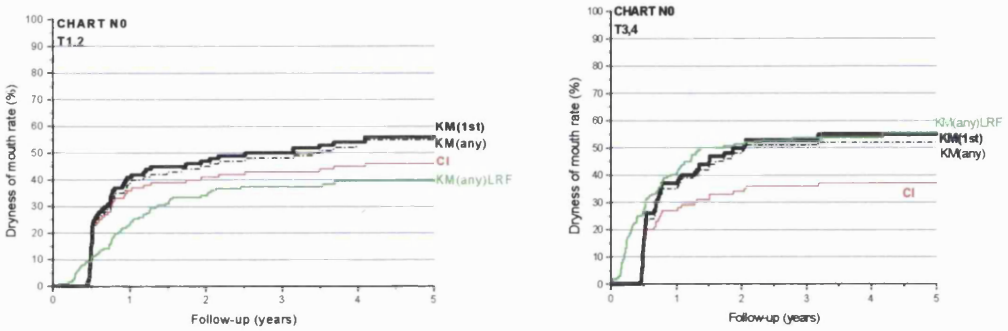


Figure 6.12 KM and CI estimates for dryness of mouth among CHART (N_0) early versus advanced T stage groups. Green curves show the KM (any) estimate for the LRF rate.

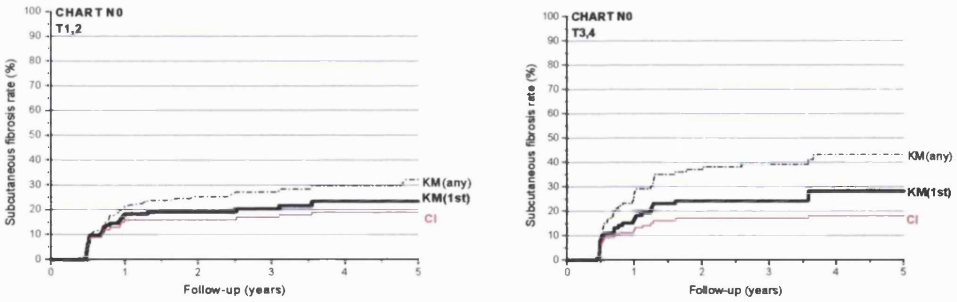


Figure 6.13 KM and CI estimates for subcutaneous fibrosis and oedema among CHART (N_0) early versus advanced T stage groups.

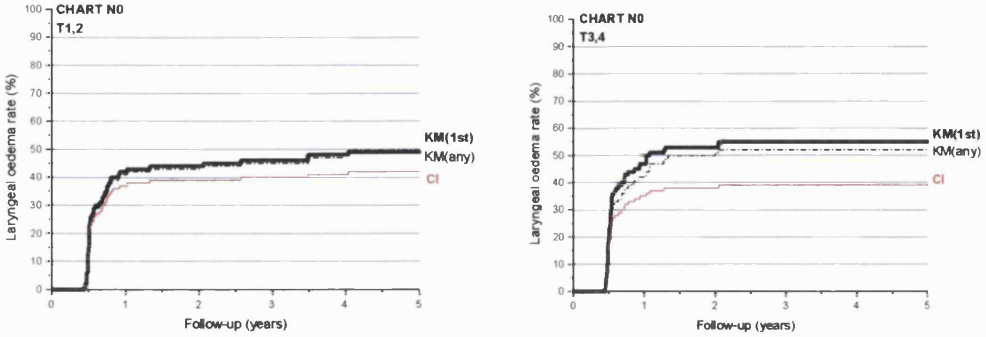


Figure 6.14 The KM and CI estimates for laryngeal oedema among CHART (N_0) early versus advanced T stage groups.

CHAPTER 7

DISCUSSION

We have presented an analysis of competing risks on cancer therapy outcome in two large RCTs comparing CHART versus conventional radiotherapy in NSCLC and HNSCC. Both studies had complete and carefully documented morbidity and follow-up data over 10 years. HNSCC and NSCLC are two clinical situations where the majority of the patients present with advanced disease (stage III-IV) at the time of diagnosis. Loco-regional failure and distant metastasis are two almost equally important failure types responsible for the poor prognosis in NSCLC and hence very interesting for the CR analysis. In HNSCC loco-regional failure is the major cause of death with less frequent distant metastasis compared to NSCLC, but still important, and can be as high as 5-30%, according to some reports(34,35). The HNSCC study was also very interesting since it gave a chance to predict the relationship of T-, N- and M failures using promising molecular markers in a competing risks framework. The assessment of the balance between the treatment effectiveness and normal tissue tolerance was very important for both studies since CHART is the most accelerated schedule which has ever been used.

Our analysis concentrated mainly on three problems arising in the analysis of data with competing risks:

1. The estimation of the relationship between covariates and the type of outcome and its timing.
2. The estimation of the rate of occurrence of failures of specific types.
3. The estimation of late radiation morbidity.

The methods that have been applied here are based on time to the first observed event. This makes it possible to separate out the treatment effects on initial events and concentrate on the major cause of failure. If the first observed event is not the death of the patient the subsequent events could also be observed although the probability of observing these events might be interfered with possible interventions (treatments for disease recurrence or late complications) or less rigorous follow-up procedures. Subsequent events are not considered and analysed in this thesis.

7.1 MODELLING FAILURE-SPECIFIC PROGNOSTIC FACTORS

7.1.1 NSCLC Study

Many clinical trials employing a combination of treatment modalities for medically or technically unresectable locally advanced non-small cell lung cancer (NSCLC) have evolved in recent years (20;36). Although modest improvements in outcome have been reported, the optimal combination of these treatment modalities and the selection of patients for the treatment options from which they are most likely to benefit remain controversial (37). Local and distant failure are two major competing risks for medically or technically unresectable locally advanced non-small cell lung cancer (NSCLC) since, despite radical treatment, it has been estimated that over 80% of patients die with disease in their chest (38). In our study 48% of the patients failed first in TN-position and 37% of the patients failed first in M-position with or without TN-failure suggesting the high risk of both failures. A recent patterns of failure study from the Radiation Therapy Oncology Group (RTOG) for locally advanced NSCLC suggested the possibility of important differences in failure-specific outcomes by pre-treatment characteristics and the treatment assigned (39). Although many studies of

prognostic factors have been conducted in NSCLC, they have not analysed failure-specific outcomes in the competing risks setting (40,41).

The purpose of this part of the study was to apply a competing risks model in order to identify factors associated with local and/or distant failure for patients treated with radiotherapy alone in medically or technically inoperabl locally advanced NSCLC.

In the present analysis, the time to failure increased with increasing age. Time to failure was also associated with treatment. Patients receiving conventional radiotherapy failed earlier than those receiving CHART, for both types of failures. This effect was more pronounced for local rather than distant failure as seen from the magnitude of the two regression coefficients (Table 5.3).

The competing risks analysis applied here also provides a framework for testing three hypotheses concerning the importance of each covariate for specific failure types. The results of these tests are shown in Table 5.4. If we combine the results of the hypothesis testing and the regression coefficients, it is possible to understand the influence of each variable on the type as well as the time of failure. H_{FT} tests whether the regression coefficients for local and distant failure are significantly different. Take the effect of patient's age as an example. Looking at the regression coefficients in Table 5.3, age is seen to prolong the time to both distant and local failure. The magnitude of these regression coefficients should be seen in relation to their standard errors. Clearly, these two regression coefficients are not statistically significantly different and this is reflected by the P-value for rejecting H_{FT}, which is 0.4 (Table 5.4). In this case, the conditional hypothesis, H_{cond}, that age has no influence on time to failure given that it has no influence on type of failure, is relevant to test, and this can

be rejected with P=0.036 in the present analysis. Thus, increasing age prolongs the time to failure but has a similar effect for the two competing failure types. H_{FT} could only be rejected for clinical stage (2P=0.009), which means that this factor had a significantly different effect on distant and local failure times. In this case, the conditional hypothesis, H_{cond}, is clearly not relevant. The combined hypothesis, H_{comb}, could be rejected (2P=0.004), i.e. clinical stage is associated with a decreased time to failure and discriminates between the two failure types. None of the other covariates were significantly associated with a specific type of failure, although from the regression coefficients it is seen in Figure 5.4 that patients receiving CHART had a numerically greater prolongation of the time to local failure than of the time to distant failure.

The final model was used to construct prognostic indices and specific failure rates at 2 years and these were used to identify four groups with different failure patterns. The scatter plot of the local and distant failure rates of individual patients are shown in Figure 5.5, where each patient is represented by a star. The 'streaky' appearance of the patients' predicted failure rates results from the continuous age parameter. This plot shows the wide variation of failure probability in distant or in loco-regional positions for the whole patient group. The performance of the model as tested using KM estimates over time showed the effective selection of patients for prognostic groups, this was especially observed for distant failure-free rates as shown in Figure 5.7.

Classification of patients according to a prognostic index combining multiple prognostic factors using a survival endpoint has been used extensively with the Cox model (21,41). The CPH model could be used to establish a prognostic index (PI),

which would allow estimation of outcome in an individual patient. This method has been used extensively where models were fitted separately for each type of failure in turn, treating other failure types as censored data. This method does not treat the different types of failures jointly, complicating the comparison of parameter estimates corresponding to different failure types. In 1995, Lunn and McNeil proposed an approach based on an extension of the Cox proportional hazards regression, which enables comparison between failure types jointly (42). This method is an adaptation of Cox regression requiring data augmentation and can be run using existing software.

A recent retrospective study by Chen et al. evaluated the local control predictors in patients with NSCLC who received definitive radiotherapy (43). They used actuarial local progression-free survival rate as an endpoint and constructed 4 prognostic groups using the Cox analysis. The local progression-free endpoint was not defined clearly and they have not considered competing events in their paper. Failure specific endpoints have not been used in any of the studies in NSCLC, which aimed to identify the relationship of various prognostic factors with the types of failures except in our recently published paper (5). Three recent papers from the RTOG group (44,45,46) used recursive-partitioning analysis (RPA) for defining prognostic subgroups of patients with locally advanced inoperable NSCLC. The RPA method creates branches of the prognostic factors from the stem of all patients (47,48). The entire patient population is partitioned into subclasses according to the variable producing the most significant survival difference to form final prognostic classes. One of these studies was concerned with patterns of failure and aimed to identify groups of patients with different failure rates (45). Four RPA classes were constructed

from Kaplan-Meier survival estimates. For each of these, the frequency distribution of the site of first failure was presented. Clearly, this is not a competing risks approach, and in fact the four RPA classes showed an almost counterintuitive trend towards more disease failures in the best prognostic class. In the worst prognostic group, 58% of the patients died without disease progression as compared with 27% in the best group. It could seem that patients with a long life expectancy survived long enough to have detectable disease progression. However, it is difficult to interpret the findings of this analysis in terms of failure patterns.

We have shown that for stage III disease, distant failure rates are relatively higher than the local failure rates. Thus, intensified local treatment may not suffice to improve therapeutic outcome in this group of patients. While the present analysis identified four subgroups with significantly different failure patterns, there is still a high failure rate at both local and distant positions in patients with inoperable NSCLC.

7.1.2 HNSCC Study

Within the classical competing risks framework, the failure of an individual may be from one of the several distinct types. However, some patients experience simultaneous events and these have to be categorised as one of the failure types depending on the biology of the tumour. The propensity to form nodal or distant metastasis is usually accepted as characterising the aggressiveness of the tumour biology. In our study, the clinical outcome was recorded as the time to first failure and the type of failure: isolated local (T), nodal with or without local relapse (N) or distant (M) with or without concurrent T or N failure. The T category represented tumours tending to remain localised while the N category represented tumours that were able

to invade lymphatic vessels independently of their T position status. The M category included all patients with distant failures. Multiple molecular markers have been assessed to derive prognostic indices for each type of failure using competing risks analysis in a large cohort of uniformly treated patients with HNSCC. Apart from p53 all other molecular biomarkers in this study were predictive of at least one of the types of failures. The regression coefficients in the final model and their relation to T, N and M position failures were plotted in Figure 6.4, which allows one to understand whether a variable has a poor or good prognostic effect and whether this influence is significant or not.

In this study we have combined the molecular markers to produce models to predict the two-year failure rates in the competing risks setting. Prognostic groups defined from the modelled 2-year failure rates differed significantly in the observed incidence of the 3 types of failure. Optimising loco-regional control, disease-free and overall survival remains a challenging goal in the management of locally advanced squamous cell carcinoma of the head and neck. Although significant therapeutic gains from altered fractionation and combined chemoradiation have been demonstrated in a number of randomised controlled trials, therapeutic outcome is far from satisfactory, with relatively high locoregional failure rates of around 50% after 3 years and an overall 5-year survival rate of 50% (49). As these intensified therapies tend to be associated with increased toxicity, selection of subgroups of patients who would benefit from a specific treatment schedule could potentially improve the overall therapeutic efficacy.

Disease-free survival (DFS) is a composite endpoint where an 'event' is any recurrence (local, regional or distant) or death from any cause whichever comes first. This endpoint does not provide information on the location of the first failures and therefore is less useful in pointing towards specific therapeutic management strategies. However, selection of primary therapeutic interventions should be based on consideration of the contribution of relative risks of failures at local (T), nodal (N) and distant (M) positions. For example the prediction of the precise local failure rate for an individual patient at the time of presentation would give us the possibility of intensifying local treatment or selecting ideal patients for organ preservation, thus avoiding the need for salvage surgery. Likewise, prediction of a patient at high risk for nodal failure with a clinically negative neck at presentation would have been useful in identifying patients requiring elective neck dissection. Administration of adjuvant chemotherapy would have had prime importance if we were able to predict the precise distant failure rate. It is also conceivable that various molecular tumour markers may have a differential effect on the risk of specific types of failure and this may lead to a loss of statistical power – as well as a loss of information – in analyses using DFS or cause-specific survival as the endpoint. Two recent meta-analyses of the management of HNSCC concluded that there was a need for specific biological markers to identify subgroups of patients who can derive the greatest benefit from a specific therapy (50,51). Markers may identify molecular and biological factors that may be used to select patients at high risk of treatment failure and who might therefore benefit from an alternative or more aggressive approach.

This is the first report of the potential of utilising multiple molecular markers to derive prognostic indices for each type of failure using competing risks analysis in a large

cohort of uniformly treated patients in HNSCC. The markers had been chosen to reflect different biological processes that might be involved in the response to treatment mainly proliferation (Ki-67), cell cycle deregulation (p53 and cyclin D1), apoptosis (bcl-2) and vascularity (CD31). Except for p53 all other molecular biomarkers in this study were predictive for at least one of the types of failures. The data presented in Table 6.2 and Figure 6.4 show both expected and unexpected associations of the factors with local, nodal and distant failure. As expected increasing T stage was associated with a decreasing time to both local and distant failure but without effect on nodal failure, whilst increasing N stage predicted strongly for both nodal and distant failure but not isolated local relapse. Amongst the biological parameters, loss of proliferative organisation as indicated by the proliferation pattern was associated with a significantly shorter time to distant failure. Paradoxically, increased proliferation rate (Ki-67 >20%) was predictive of an increased period to metastasis. This observation may seem at odds with the generally held view that increased proliferation is linked to increased biological aggressiveness but it may instead reflect an enhanced response to treatment. Increasing cyclin D1 was associated with a higher propensity for both local and, in particular, nodal failure; this is in agreement with reports in the literature (52). Bcl-2, as we have previously shown predicts for both longer local control and locoregional control times (53).

Using this approach selection of primary therapeutic interventions could be based on the consideration of the contribution of relative risks of failures at local (T), nodal (N) and distant (M) positions. Patient one in Figure 6.8 exemplifies this, where the estimated 2-year T, N and M failure rates were 46%, 0.5% and 0.3% respectively. This patient could be a candidate for intensified local treatment, small volume, high

dose, altered fractionation schedule or IMRT depending on the site of origin. Patient two (Fig. 6.8) could be considered for intensified loco-regional treatment possibly requiring elective neck dissection due to a higher risk of both local (33%) and nodal (53%) failure but very low M failure (2%) risk. However, for patient three (Fig.6.8) the estimated 2-year M failure rate was 56%, much higher than the other types of failures, and for this patient adjuvant chemotherapy would be indicated in addition to radiotherapy. In Figure 6.8, we have selected examples from the whole patient group but all patients could be considered in a similar way.

These individual estimates were then grouped together (Figs. 6.5-6.7) using the 33-percentiles of the 2-year failure rates to define different prognostic groups. Estimated local failure rates at 3 years were 23%, 35% and 51% for groups T1, T2 and T3, respectively. The differences between groups were more pronounced for N and M failures. A patient in group N3 would have a nodal failure rate of 40%, compared to a patient with 10% failure rate in group N2 and 2% in group N1. In group M3 the distant failure rate was 27%. The comparisons produced a log-rank p value of $2P < 0.00001$ in all cases.

Overall, the CHART trial found no significant difference in loco-regional tumour control between the two arms of the trial (29). This has been thought to reflect that a sub-population of tumours might have had a disadvantage from the relatively low absorbed dose whereas others had an advantage of the very short, intensive radiotherapy schedule. Another concern has been that the 12-day CHART schedule might not have given sufficient time for re-oxygenation causing potential problems with the control of hypoxic tumours. All of this has stimulated interest in the

identification of clinical or tumour characteristics that would be predictive for a benefit from CHART relative to conventional therapy. In the present analysis, a significant interaction was seen between randomization to CHART and the vascularity, with increasing CD31 scores being associated with a benefit from CHART. This raises two interesting speculations: the well-vascularised tumours may be less affected by the possible reduced re-oxygenation during the CHART regimen or it may be that these are the tumours where a proliferative response to the radiotherapy trauma can be most effectively activated. It is also interesting, that the pre-treatment Ki-67 index was not predictive for a benefit from CHART. A previous report (54) suggested that well-differentiated tumours might have a more pronounced treatment-time dependency. It is possible that differentiation is only an epi-marker for an underlying biological characteristic of the tumour and in this study we actually found a significant association between histopathological grade and CD31 score. Thus, the effect seen here is consistent with the findings from the previous report by Hansen et al. (54) but histopathological grade in itself was not of prognostic importance in our study.

Advances in the understanding of the molecular mechanisms of tumour development and disease progression have been the stimulus for many studies seeking to identify a relationship between various biomarkers and the risk for recurrence of the disease or selecting patients for more aggressive therapy or alternative treatment modalities(50). So far, none of these markers has been established in clinical practice for selection of treatment. One reason for this is probably that the literature largely reports diverging or inconclusive results due to (1) biological and technical, (2) tumour related and (3) statistical problems. An undefined biological basis, cut-off values set according to clinical outcome rather than from biological considerations and assays

often with low reproducibility may influence the reliability of the results of the marker under study. Wide variations in patient populations, different sites in the head and neck, stages of disease and treatment modalities involved should also be taken into consideration.

Amplification of the chromosome 11q13 results in overexpression of cyclin D1 (CCND1), causing growth advantage and enhancing tumorigenesis (55). There are many studies with contradictory results in the literature investigating the correlation of the cyclinD1 amplification with clinical outcome. Some studies showed a correlation to poor survival (56-58). A review of studies published after 1994 which were included in the PUBMED and studied the prognostic importance of CyclinD1 with or without other molecular markers (immunohistochemical or gene amplification) in relation to clinical outcome is shown in Table 7.1. Fourteen out of 22 studies were performed with less than 100 patients and five out of fourteen studies a multivariate analysis was also performed. Subgroup analyses were also performed in many studies even if the sample sizes were too small to ensure appropriate statistical power. Almost all studies except five studies included all tumour sites and all stages. Thirteen studies used overall survival as an endpoint and applied KM estimates, log-rank tests or the Cox proportional hazards (CPH) model in combination or alone. None of the them studied time-to failure or competing risks analysis. Only few worked on failure patterns. Comparison of these studies to be able to derive meaningful conclusions is almost impossible and very confusing for any researcher who is interested in this subject.

Table 7.1 Studies evaluating the effect of CyclinD1 on outcome in HNSCC

Author (Date)	Site, stage	N	Study Design	TRT	Marker	LF	LRF	RF	DM	DFS	Overall Survival
Muller (1994) (59)	OC, OP HP, L Stage I-IV	178	N(+) N(-) χ^2	S	11q13 amplification	NR	NR	Yes (+) corr	NR	NR	NR
Akervall (1995) (55)	All HN Stage I-IV	116	KM LR CPH	S ± RT ± CT	11q13 amplification	NR	NR	NR	NR	NR	Yes (+) corr
Michalides (1995) (52)	HP, OP, L Stage I-IV	47	KM CPH	S, S+RT	IHC	NR	NR	NR	NR	Yes (+) (DFI)	Yes (+)
Meredith (1995) (56)	All sites, all stages	56	KM	Mixed	11q13 amplification	Yes (+) corr	NR	NR	NR	NR	Yes (+)
Masuda (1996) (60)	HP, Stage I-IV	42	KM LR	S ± CT	IHC	NR	NR	NR	NR	NR	Yes (+) for S+CT group
Bellacosa (1996) (61)	L Stage I-IV	51	KM LR CPH	S S+ RT	11q13 amplification	NR	NR	NR	NR	Yes (+) trend	Yes (+) corr
Akervall (1997) (62)	All sites Stage I-IV	75	KM LR CPH	S ± RT ± CT	IHC	NR	NR	NR	NR	NR	Yes (+) corr
Takes (1997) (63)	L, Primary or Recurrent Stage I-IV	31	N(+) N(-) χ^2	S, RT	11q13 amplification	NR	NR	Yes (+) corr, No FTA	NR	NR	NR
Fortin (1997) (64)	OC, OP Stage I-IV	50	N0 N(+) Pearson χ^2	S ± RT ± CT	IHC and 11q13 amplification	NR	Yes (-) corr No FTA	NR	NR	DFS?	NR
Michalides (1997) (65)	SGL, HP, OP Stage I-IV	115	KM, CPH	S S+RT	IHC	NR	NR	NR	NR	Yes (DFI) (+) corr	NR
Muller (1997) (58)	OC, OP, HP, L Stage I-IV	282	KM LR CPH	S S+RT	11q13 amplification	Yes frequency	Yes frequency	Yes frequency		Yes (EFS) (-)	Yes (-) corr
Kyomoto (1997) (66)	All sites Stage I-IV	45	KM LR CPH	S	IHC and 11q13 amplification	NR	NR	NR	Nr	NR	Yes (+) corr

Table 7.1 Studies evaluating the effect of CyclinD1 on outcome in HNSCC-continued

Author (Date)	Site, stage	N	Study Design	TRT	Marker	LF	LRF	RF	DM	DFS	Overall Survival
Nogueria (1998) (67)	All HN Stage I-IV	56	χ^2	S, S+RT	11q13 amplification	NR	NR	NR	NR	NR	NR
Pignataro (1998) (68)	L, Stage I-IV	149	KM LR CPH	S, S+RT	IHC	NR	NR	NR	NR	Yes (+)	Yes (+,univar) (borderlin e CPH)
Matthias (1998) (69)	L, OC, OP, HP Stage I-IV	384	KM CPH	S	CCND1 Genotype (GG,AG, AA)	NR	NR	NR	NR	Yes DFI (+) in GG	NR
Bova (1999) (70)	OC (anterior tongue) Stage I-IV	148	KM LR CPH	S S+RT	IHC	NR	NR	NR	NR	Yes DFI (+) corr	Yes (+) corr
Gleisch (1999) (71)	OP, L, OC, HP Stage I-IV	43	KM LR	S S+RT	11q13 amplification	NR	NR	NR	NR	NR	Yes (-) corr
Rodrigo (2000) (72)	OC, OP, L, HP Stage I-IV	104	KM LR CPH	S	11q13 amplification	NR	NR	NR	NR	Yes (+) corr	Yes (+) corr
Capaccio (2000)(73)	OP, OC, L Stage I-IV	96	L. Regre ssion	S	IHC	NR	NR	Yes (+) corr occul t met	NR	NR	NR
Yoo (2000) (74)	L T1-T2/N0	60	Match ed case-control Pears on χ^2	RT	IHC	Yes (-) corr	NR	NR	NR	NR	NR
Smith (2001) (75)	OP, OC Stage I-IV	56	KM CPH	S+RT	IHC	NR	Yes (-) corr	NR	Yes (+) corr	NR	Yes (-) corr
Dong (2001) (76)	L Stage I-IV	102	KM LR CPH	S S+RT	IHC	NR	NR	NR	NR	Yes (+) corr	Yes (+) corr

L: any laryngeal site
SGL: supraglottic larynx
H: hypopharynx
OC: oral cavity

OP: oropharynx
NP: nasopharynx

RT: radiotherapy
S: surgery
CT: chemotherapy
P: palliative

NR: not reported
CCRT: concomitant chemoradiotherapy

7.2 ESTIMATION OF TUMOUR OUTCOME

Our analysis compared the failure rates (LRF, DF) as estimated by the actuarial method KM (1st), the CI method (both based on time to first event) and the KM (any) method (event of interest can occur 1st, 2nd, etc.) where LRF and DF were competing events. Here we used the terms "any" and "first" to describe the handling of the event of interest in two different ways in the KM analyses as proposed recently by Koscielny and Thames (12). In their study they tried to relate these estimates to the region containing the "true" risk (net risk, or risk that would be observed in the absence of competing risks) of local and distant failure in breast carcinoma. They concluded that, for breast carcinoma, the KM (any) estimate of the failure rate is higher than the KM (1st) estimate, which is higher than the CI estimate.

In our study at any follow-up time, these estimates were ordered as CI less than KM (1st) less than KM (any) for both NSCLC and HNSCC studies for the tumour outcome. That the CI estimate is less than KM (1st) or KM (any) estimate is a general result that holds whatever the group of patients under analysis simply because for the CI calculation patients are not considered at risk whenever an event of no interest or censoring occurs (see page 27). The CI estimate will always give relatively the lowest figure independent from the pattern of the events under study. However the relation between KM (1st) and KM (any) depends on the pattern of occurrence of the first event under study and other events. The difference between the KM (1st) and the KM (any) estimates become smaller and smaller when almost all events of interest occur

as the first event. Koscielny and Thames explained this relation as the effect of each type of failure on the likelihood of occurrence of the other.

The KM (any) method is widely used and applied by clinicians with user-friendly statistical software, who may not be aware of the competing risks problem. Although this estimate gives the upper limit of a failure rate by ignoring all the other events, it does not address the question of which event is likely to occur first. Thus it is inappropriate if the objective is to evaluate the effectiveness of the primary treatment and which events are contributing to the site of first failure. KM (any) estimate is also subject to two sources of bias: 1) administration of additional therapy when the other event occur 2) reduced vigilance in following up the event of interest after the other event occur. When estimating the tumour outcome these sources of bias should be kept in mind since for example administration of a systemic therapy after a LRF will interfere with the DF rate and thus the KM (any) estimate from the observed data will underestimate the true DF rate of the initial treatment. KM (any) estimate is not directly comparable to the CI estimate since these two estimates are not looking at the same quantities.

The KM (1^{st}) method considers only the first event type rather than any event type in the analysis. This estimate focuses on the first failure and is not influenced by other competing events, so it is always a higher figure compared to the CI estimate where the competing events has a direct influence on the resulting estimate.

The CI method has been promoted on the basis that lack of independence is a problem in actuarial analysis of data describing outcome with competing risks.

However the CI estimate is subject to the assumption of exclusivity of outcomes and gives the lowest possible figure and it should be interpreted with this in mind.

7.2.1 NSCLC Study

The CI and the KM (1^{st}) estimates were correlated for LRF and DF for 4 prognostic groups at 4 years (Table 5.5). Both estimators showed the expected trend in failure pattern in the four groups except the CI estimate of the LRF rate in group 4. The CI estimate showed a much lower figure (.47) in group 4 (DF↑ LRF↑) compared to group 2 (DF↓ LRF↑) figure of .68. The KM (1^{st}) estimate of the LRF rate was .83 and .77 for the two groups respectively. This may simply mean that the low figure estimated by the CI method indicates either low risk of LRF or high risk of DF, as Group 4 had patients with higher risk of failing distantly.

LRF and DF rates were correlated using KM (1^{st}), KM (any) and the CI estimates for early (I-II) and advanced (III) stage groups. The CI estimates suggested a lower LRF rate for the advanced stage group compared to the early stage group, whereas the KM (1^{st}) estimates were around .72 for both clinical stages and KM (any) estimate showed a 5% increase for the advanced stage group (Figures 5.14 and 5.15). If this data were analysed only by the CI estimator, it would not be unreasonable for anyone looking at the results to conclude that there were less local failures in the advanced stage patient group. However, the explanation might well be that patients with advanced stage experience higher DF rate and this influences the CI estimate for LRF rate, lowering it to a figure even below than the stage I-II group.

7.2.2 HNSCC Study

In the HNSCC study, the tumour outcome rates resulted in almost no difference for the LRF rate at 5 years between the three estimates for early versus advanced stage patient groups (Figure 6.9 and 6.10). This might indicate the lower incidence of competing events (DF) in this example. However even if the overall DF rates were estimated in the range of .07-.22 by all the three estimates, the CI estimate was the one that was very stable for the both stage groups. The KM (any) estimate was as high as .22 for the advanced stage group that shows the highest observable DF rate if LRF was not an operating competing (other) event.

The underlying assumptions and the difficulties in interpretation of the three estimators compared here lead us to the conclusion that no single method is appropriate on its own when estimating treatment outcome. The assumption of independence of different event types in general may not be clinically meaningful where the lowest risk of LRF might be very much related to the highest risk of DF. A second assumption independence of the censoring mechanism is statistically untestable. In clinical research censoring could indicate an unfavourable prognosis (patients too sick, or bedridden), or conversely a favourable prognosis (drop out of the study once they are cured). This problem is common to all survival analysis.

These two examples showed the dependence of these estimates on the natural behaviour of the cancer site and prognosis of the group under study. The CI estimates were very much affected by the high incidence of competing events

between different prognostic groups as shown in NSCLC example. However, the three estimates showed very close figures when the incidence of competing risks were not high as shown in the HNSCC example.

7.3 ESTIMATION OF LATE RADIATION MORBIDITY

During the course of follow-up, each patient is assumed to go through a finite number of states depending upon the occurrence of complications of varying severity, of disease recurrence, and of death. In this study, the recurrent nature of complications was not considered and the time to first normal tissue injury was considered. Crude proportions, which are simply the number of complications divided by the number of patients in the group, could give misleading results especially when comparing two treatment schedules with different death rates. Death from disease is a censoring event, which, if proper statistical methods are not applied, will obscure the level of normal tissue injury. Actuarial methods have been recommended instead of crude proportions since the time factor and censoring is taken into account.

The application of the CI estimates to report morbidity data have been discussed by various authors. Caplan et al. suggested the use of the CI and alternative calculations such as the integrated hazard, which are mathematical transforms of the product limit estimate (18). The interpretation of the hazard rates is not straightforward and could be more difficult than the KM estimate. Bentzen et al (17) suggested the use of the actuarial estimates when reporting morbidity since the CI is not specific for toxicity and that the CI estimates of the patients with identical treatments with different prognosis will be completely different. Our results confirmed the unreliability of the CI

estimate in reporting late morbidity. Another suggestion " the cumulative conditional probability" is subject to the same criticism as the cumulative incidence, namely that it is nonspecific for toxicity.

Others have suggested the use of prevalence rather than incidence estimates in describing late toxicity (22). The prevalence is an estimate of the proportion of patients still living with a specific complication at various times after treatments. While this is obviously of interest to the clinician (and the patient), we would maintain that such information should not be given as an alternative to but rather as a supplement to actuarial estimates. This is because prevalence estimates do not in themselves reflect the toxicity of the treatment but are influenced by the management of complications.

7.3.1 NSCLC Study

In NSCLC study the KM (1^{st}) estimate was compared with the CI estimate in groups of patients with different prognoses. The death of the patient was considered as a competing risk since overall survival probability was around 20-30% at 2 years. Post-radiation fibrosis and deep positioning of tumours within thorax prevents easy assessment or biopsy and thus LRF could not easily be considered as a competing event. Also the high rate of both distant and local failures in locally advanced NSCLC justifies specifying death as the main competing risk.

For each morbidity endpoint, the time to the first event whether death or the morbidity was recorded separately. Mild to severe dysphagia rates estimated by the KM (1^{st}) and CI methods were similar for the two treatment arms showing a difference of only

5-7% at 2 years. (Fig.5.17). Comparison of the estimates showed that the difference between the estimates grew larger with longer follow-up suggesting that there were less events and less patients at risk for the calculations. The situation was even more dramatic for clinical and radiological radiation pneumonitis where morbidity occurred later in follow-up. The KM estimate at 8 years was 78% for clinical radiation pneumonitis for all patients while the CI estimate was as low as 15% (Fig. 5.18). The KM method, however, censors all the events other than the event of interest and thus still incorporates information from these patients. Using censored times gives the estimate greater precision than if the curve were based only on those with sufficient time in the study to have experienced late morbidity. The KM method assumes that patients who are currently alive will eventually experience the morbidity and that the future distribution of the time span of this will follow the same pattern as in those who have already had the morbidity. This assumption is justifiable when the endpoint is survival since those who are censored are in general just the more recent entrants into the series. Censoring the patients who have died would be logically imprecise but in this way they are treated as if they are still at risk, which would result in potentially higher estimates of complications. On the other hand, the assumption of independence of different event types in general may not be clinically meaningful. In NSCLC, death from disease and morbidity could be related in a patient with a larger tumour volume due to advanced T stage. A presumption with the actuarial method is that the cause of censoring should be unrelated to the event being considered. For example, in a clinical trial comparing two treatments, those who experience success with the therapy may be more likely to drop out of the study once they are cured or conversely in a trial of glioblastome multiforme where patients who often present with

unresectable tumours may be too sick to return for follow-up. In these situations the occurrence of the endpoint will be underestimated but this is not a flaw in the actuarial methods: this represents a real lack of information. Similarly, a differential diagnostic problem of distinguishing progressive disease from a normal-tissue complication will be a real obstacle independent of the statistical methodology. The standard errors of the KM estimates are larger figures since the estimates of event-free functions, cumulative hazards and so on become unstable for longer *t* (time). This is because progressively more patients either suffer an event or are censored as *t* increases and there is little information available for longer *t*. This is inevitably true and all the plots should be interpreted with this in mind.

The effect of the irradiated area on the occurrence of late morbidity was compared using two estimates in the same competing risks frame. The patients with large irradiated volume (>200 cm2) would be expected to experience a higher morbidity rate compared to the patients treated with a small volume (<160cm2). For this analysis, the pulmonary fibrosis rate after the 1st year was also considered and was estimated to be lower for the larger irradiated volume (.24 and .26 at 2nd and 4th years respectively) by the CI estimate conversely to the expectation of higher morbidity (Table 5.12, Fig.5.21). The KM estimate at 4 years was .83 for the large volume and .59 for the small volume (Table 5.12 and Figure 5.21). Sixty eight percent of the patients with a large irradiated volume had stage III clinical disease and 89% were dead at the time of the analysis (Table 5.11).

The consistency of the CI and KM methods in estimating morbidity in different prognostic groups was analysed in patients treated with the CHART schedule i.e.

uniform treatment. The pulmonary fibrosis rates after 1st year of follow-up in two prognostic groups (N0-1 vs N2-3) were compared. The CI estimate showed that the pulmonary fibrosis rate was 4% lower in the N2-3 stage group at 4 years (Fig. 5.22). To explore this further we plotted the hazard rates of having pulmonary fibrosis and death after 1st year in CHART arm (Fig. 7.1). The hazard rate of having pulmonary fibrosis in the N2-3 group became higher after 2 years while the N0-1 stage group have a constant hazard rate after 2 years (Figure 7.1(a)). On the other hand, the hazard rate of death is almost always higher in the advanced N stage group (Figure 7.1(b)). This shows the unreliability of the CI estimate in patients with the same treatment but different prognosis.

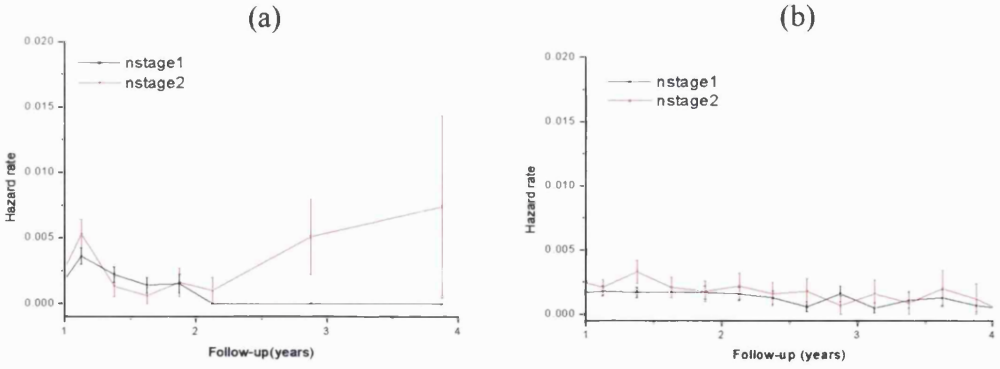


Figure 7.1 Hazard rates of having (a) radiological pulmonary fibrosis and (b) death after 1st year

7.3.2 HNSCC Study

In the HNSCC study the KM (1st), the KM (any) and the CI estimates were correlated. The LRF was considered as the main competing event due to longer overall survival and higher risk of LRF in locally advanced HNSCC. The diagnosis of LRF leads to

salvage treatments thus masking the morbidity of primary treatment completely. In some cases, LRF or, for example subcutaneous fibrosis, might also cause difficulty in distinguishing morbidity from LRF. For each morbidity endpoint, the time to the first event whether LRF or morbidity was recorded separately. Here CHART N0 patients with early or advanced T stage were compared for three morbidity endpoints. The difference between the KM and the CI estimates increases as more competing risk events precede occurrences of the event of interest namely late normal tissue morbidity in this case. When comparing the same group of patients with an identical treatment, patients with better prognosis will experience fewer failures than those with worse prognoses. The treatment complications will therefore be more pronounced for the better prognosis group. However failure for the worse prognostic group will precede morbidity and the probability of having complications will be lower even if the treatment intensity was the same. In our study, the CI estimate showed pronounced dryness of mouth rate for T1-2 stage group (.46 at 5 years) compared to T3-4 stage group (.37 at 5 years) (Fig. 6.12). The KM (any) estimates were slightly lower than the KM (1^{st}) estimates for both stage groups. Although we would expect KM (any) to be usually higher values compared to the KM (1^{st}) estimate, this might be reversed under some circumstances, most probably related to the censoring mechanism of the other events. The way each method deals with the censoring of observations that are failures from a competing risk leads to the difference between CI and KM estimates. These estimates will change over time due to their dependence on the frequency and timing of censored observations and failure from competing risks.

7.4 CONCLUSIONS

7.4.1 Modelling failure specific prognostic factors

Multivariate analysis using competing risks models as implemented in the BMDP software package allows exploring the relation between various prognostic factors and different types of first recurrence to be explored. With this approach prognostic factors may be used as predictive factors for individualising treatments based on the most likely failure types. This analysis shows the influence of a particular covariate on time and type of a specific failure by providing three hypothesis tests and has two potential fields of application.

Firstly, it enables us to predict the risk of specific types of failure with improved precision, which could be used to identify subgroups of patients with significantly different failure patterns. When this approach was used for locally advanced NSCLC the two common failure types (LRF and DF) were considered together and patients were divided into four prognostic groups with different risks for LRF and DF. These four prognostic groups defined by the model showed a close agreement when the model predicted failure rates over time were compared graphically with the KM estimates. This informal validation method used overall data and applied it to the strata. Even if the model fits the data properly in our study we are aware that the most appropriate way would be to use an independent data set to validate the method. The same approach was used in a slightly different way in HNSCC where the subgroups were formed to include patients with different risks for T, N and M failure separately. Three prognostic groups for each type of failure were defined, compared and found to be significantly different. The individual estimates could also be used for tailoring of

therapy based on individual risk profiling when they are not grouped as above. Using this approach, selection of primary therapeutic interventions could be based on the consideration of the contribution of relative risks of failures at local (T), nodal (N) and distant (M) positions. We have shown this approach for HNSCC analysis.

Secondly, the influence of each covariate on the time and the type of failure could be considered one by one in exploring the clinical significance of established clinico-pathological factors or new biological markers and stratification of patients into trials of new treatments.

In NSCLC model, the clinical stage was the only significant factor associated with both the time and the type of failure. Patients with stage III disease were prone to have a higher and earlier distant failure relative to local failure. Thus, intensified local treatment may not suffice to improve therapeutic outcome in this group of patients. The literature is rapidly expanding in the field of biological markers and their potential role as prognostic or predictive factors remains to be tested. It is possible that further biological characterisation of these tumours in a competing risk analysis would enable us to predict the risk of specific types of failure with improved precision.

In the HNSCC model, all other molecular biomarkers except for p53 were predictive for at least one of the types of failures. The predictive information of molecular markers on failure types for an individual patient at the time of presentation would provide the possibility of intensifying and thus selecting ideal patients for different treatment schedules. Moreover this analysis showed that this approach could also possibly be used to understand the relation between the CHART arm and various cell kinetic mechanisms. With the current move towards more combined loco-regional and

systemic therapies, studies of failure-type specific predictive markers are of great interest and we feel that the approach taken here warrants further study.

The BMDP competing risks analysis is based on the KM principle and assumes that the failure times for the types of failure are independent. This assumption is untestable.

Competing risks analysis of failure-specific prognostic markers is a powerful tool for selection of patients with different risk profiles for specific types of failures, which eventually may lead to tailored therapies.

7.4.2 Estimation of treatment outcome

The underlying assumptions and the difficulties in interpretation of the three estimators compared here lead us to the conclusion that no single method is appropriate on its own when estimating treatment outcome. The assumption of independence of different event types in general may not be clinically meaningful where the lowest risk of LRF might be very much related to the highest risk of DF. A second assumption independence of the censoring mechanism is statistically untestable. In clinical research censoring could indicate an unfavourable prognosis (patients too sick, or bedridden), or conversely a favourable prognosis (drop out of the study once they are cured). This problem is common to all survival analysis.

The KM (any) method gives the upper limit of an event rate under study by ignoring all the other events, it does not address the question of which event is likely to occur first. The KM (any) estimate should be interpreted with caution when estimating late

side effects because this estimate could be misleading if any other treatments have been used for any types of failures.

The KM (1^{st}) method considers only the first event type rather than any event type in the analysis. This estimate focuses on the first failure and is not influenced by other competing events, so it is always a higher figure compared to the CI estimate where the competing events has a direct influence on the resulting estimate. The KM (1^{st}) estimate should be preferred to have an idea about the upper limit of an event of interest occurring first. The biases discussed for the KM (any) estimate are not relevant for KM (1^{st}) estimate so it is a reliable estimate of the highest probability of treatment failure and late side effects.

The CI method has been promoted on the basis that lack of independence is a problem in actuarial analysis of data describing outcome with competing risks. However the CI estimate is subject to the assumption of exclusivity of outcomes and gives the lowest possible figure, where should be interpreted with this in mind. Whenever late radiation morbidity is under study, the CI estimate should be interpreted with great caution since the estimate could give misleading results in a group of patients with same treatment intensity but different prognosis.

The KM and the CI methods should be used as complementary analyses whenever possible considering the various clinical situations. The underlying mechanisms and assumptions of statistical methods used in competing risks situations should be well understood by the clinicians who will interpret them in the clinical practice.

Publications from the material presented in this thesis

1. Ö Uruk, SM Bentzen, MI Saunders, M. Parmar. Pattern of failure after CHART or conventional radiotherapy in locally advanced non-small cell lung cancer: A competing risk analysis *Radiother Oncol*, 2000: **56** Suppl: 1,S57-208, *(conference abstract).*

2. Ataman OU, Bentzen SM, Saunders MI, Dische S. Failure-specific prognostic factors after continuous hyperfractionated accelerated radiotherapy (CHART) or conventional radiotherapy in locally advanced non-small-cell lung cancer: a competing risks analysis. *Br J Cancer* 2001: **85:** 1113-8.

3. Ö. Ataman, SM. Bentzen, GD. Wilson, PI. Richman, MI. Saunders , S. Dische Prognostic value of molecular biomarkers in predicting specific types of failures after radiotherapy for head and neck cancer. *Radiother Onc*ol, 2002: **64:** Suppl: 1 S251. (*conference abstract*)

4. Özlem U. Ataman, Søren M. Bentzen, George. D. Wilson, Frances M. Daley, Paul I. Richman, Michele I. Saunders, Stanley Dische. Molecular biomarkers and site of first recurrence after radiotherapy for head and neck cancer *(Submitted to Journal of Clinical Oncology)*

Appendix A (SPSS syntax for calculating CI estimate with its SE)

```
COMMENT  These commands require that 'Status'
COMMENT   has the coding: 1=event, 2=Competing, 3=censor.
COMMENT  It cannot cope with missing values.
COMMENT  Check that the files tmp1.sav & tmp2.sav can be created or altered on
your drive.
COMMENT You may have to increase the memory allocation
COMMENT through edit>option>general>special workspace memory limit.


SET MXLOOP=1000.
SORT CASES BY time(A), status(A) .

save outfile=tmp1.sav .
COMPUTE event=0.
COMPUTE compete=0.
COMPUTE censor=0.

IF (status=1) event=1 .
IF (status=2) compete=1.
IF (status=3) censor=1 .
execute.
matrix.
get time /var=time.
get event /var=event.
get compete /var=compete.
get censor /var=censor.
get status /var=status.
compute N=nrow(time).
compute natrisk=make(N,1,N).
compute vchaz=make(N,1,0).
loop i = 2 to N.
        compute natrisk(i)=natrisk(i-1)-event(i-1)-censor(i-1)-compete(i-1).
        compute vchaz(i)=(compete(i-1)+event(i-1))/natrisk(i-1)/(natrisk(i-1)-
compete(i-1)-event(i-1) ).
        compute vchaz(i)=vchaz(i)+vchaz(i-1).
end loop.
compute localhaz=event&/natrisk.
compute evehaz=(event+compete) &/natrisk.
release censor, compete.
compute efs=make(N+1,1,1).
compute fail=make(N,1,0).
compute CI=make(N+1,1,0).
loop i = 1 to N.
        compute efs(i+1)=efs(i)*(1-evehaz(i)).
        compute fail(i)=localhaz(i)*efs(i).
        compute CI(i+1)=CI(i)+fail(i).
end loop .
release efs, evehaz, localhaz.
compute omega=make(N,N,0).
```

```
compute trifail=make(N,N,0).
loop i = 1 to N.
        compute trifail( : ,i)=fail .

        loop j = i to N.
        do if (j>i).
                compute trifail( j,i)=0.
        end if.
        do if (event(i)=1).
                do if (i=j).
                        compute omega(i,j)=-1/natrisk(i)+ vchaz(i)+1.
                else.
                        compute omega(i,j)=-1/natrisk(i)+ vchaz(i).
                end if.
        end if.
        compute omega(j,i)=omega(i,j).
        end loop.
end loop.
compute sd=sqrt(abs(DIAG(T(trifail)*omega*trifail ))).
compute CI1=CI(1:N+1,1).
release trifail, omega, natrisk, vchaz,CI, fail,event.
save {time,status,CI1,sd}
/outfile=tmp2.sav
/variables=time,status,cuminc,sd.

end matrix.

get file=tmp1.sav.
MATCH FILES /FILE=*
 /FILE=tmp2.sav
 /BY time status.
execute.
GRAPH  /SCATTERPLOT=time  WITH cuminc  /MISSING=LISTWISE .
```

APPENDIX B (BMDP OUTPUT FOR NSCLC STUDY)

```
BMDP Instruction File   : C:\DYNAMIC\BMDPRUN&.TMP
BMDP Program Output File: C:\DYNAMIC\BMDPOUT&.OUT

BMDP2L--SURVIVAL ANALYSIS WITH COVARIATES


/ INPUT FILE   = 'C:\DYNAMIC\NSCLC.DAT'.
   FORMAT      = FREE.
   VARIABLES   = 7.

/ VARIABLE NAMES = sex, age, failuretime, treatment, competing (CR), clinical
stage.

/GROUP     CODES(CR) = 1, 2, 3.
          NAMES(CR) = metastasis, loco-regional, censor.

/ FORM     TIME = failuretime.
   RESPONSE = 1, 2.

/ REGRESS  ACCEL = LLOGISTIC.
   COVARIATES = age,sex, trt, clstage.
   COMP = CR.
/ END
```

CASE	1 sex	2 age	3 ftime	4 TRT	5 CR	6 clstage
1	male	75.00	507.00	CHART	Metastasis	I-II
2	male	74.00	164.00	CHART	Metastasis	IIIA-IIIB
3	male	74.00	69.00	CONV	Metastasis	IIIA-IIIB
4	male	66.00	634.00	CHART	Metastasis	I-II
5	male	67.00	440.00	CONV	Censor	IIIA-IIIB
6	male	76.00	29.00	CHART	Censor	IIIA-IIIB
7	female	61.00	178.00	CHART	Metastasis	IIIA-IIIB
8	female	72.00	34.00	CHART	Metastasis	I-II
9	male	73.00	67.00	CONV	Metastasis	IIIA-IIIB
10	female	53.00	40.00	CONV	Metastasis	IIIA-IIIB

```
NUMBER OF CASES READ. . . . . .     549

*** N O T E *** TYPE OF FAILURE: CENSORED CONTAINS NO RESPONSES.
                      THE CASES WILL BE USED, BUT NO ANALYSIS FOR
               THIS TYPE OF FAILURE WILL BE PERFORMED.


  ACCELERATED FAILURE TIME MODEL
  ------------------------------
  DISTRIBUTION IS  LLOGIST


THE NATURAL LOGARITHM OF SURVIVAL TIME IS USED IN THE ANALYSIS

      TYPE OF FAILURE: Metastasis
       --------------------------

  INITIAL VALUES OF PARAMETERS
  CALCULATED BY LEAST SQUARES

CONSTANT          sex             age          trt         clstage
 5.279058      0.149837      0.010279     -0.197075    -0.109923

SCALE
 0.517436


LOG LIKELIHOOD =     -526.8738
GLOBAL CHI-SQUARE =  18.11  D.F.=   4  P-VALUE =0.0012
FOR ACCELERATED FAILURE TIME MODEL, GLOBAL CHI-SQUARE IS BASED ON
LIKELIHOOD RATIO TEST
NORM OF THE SCORE VECTOR=   0.134E-05#
```

VARIABLE	COEFFICIENT	STANDARD ERROR	COEFF./S.E.	EXP(COEFF.)
-1 CONSTANT	6.3748	0.7778	8.1960	586.8797
1 sex	0.1096	0.1799	0.6091	1.1158
3 age	0.0168	0.0090	1.8696	1.0169
5 trt	-0.0448	0.1545	-0.2899	0.9562
7 clstage	-0.5455	0.1656	-3.2943	0.5795
-2 SCALE	0.8217	0.0472		2.2744

TYPE OF FAILURE: Loco-regional

INITIAL VALUES OF PARAMETERS
CALCULATED BY LEAST SQUARES

CONSTANT	sex	age	trt	clstage
5.279058	0.149837	0.010279	-0.197075	-0.109923

SCALE
0.517436

LOG LIKELIHOOD = -521.3292
GLOBAL CHI-SQUARE = 13.75 D.F.= 4 P-VALUE =0.0082
FOR ACCELERATED FAILURE TIME MODEL, GLOBAL CHI-SQUARE IS BASED ON
LIKELIHOOD RATIO TEST
NORM OF THE SCORE VECTOR= 0.617E-01

VARIABLE	COEFFICIENT	STANDARD ERROR	COEFF./S.E.	EXP(COEFF.)
-1 CONSTANT	6.0777	0.5181	11.7316	436.0087
1 sex	0.2056	0.1241	1.6558	1.2282
3 age	0.0076	0.0060	1.2640	1.0077
5 trt	-0.3379	0.1049	-3.2217	0.7133
7 clstage	-0.0316	0.1080	-0.2929	0.9689
-2 SCALE	0.5903	0.0292		1.8045

TEST STATISTICS

H1: COVARIATE HAS NO INFLUENCE ON TYPE OF FAILURE AND TIME TO FAILURE.
H2: COVARIATE HAS NO INFLUENCE ON TYPE OF FAILURE.
H3: COVARIATE HAS NO INFLUENCE ON TIME TO FAILURE, GIVEN IT HAS NO INFLUENCE ON TYPE OF FAILURE.

REFERENCE: LAGAKOS (1978), APPLIED STATISTICS, VOL. 27, PP. 235-241.

	H1 (D.F.= 3)		H2 (D.F.= 2)		H3 (D.F.= 1)	
COVARIATE	STAT.	P-VALUE	STAT.	P-VALUE	STAT.	P-VALUE
sex	3.113	0.211	0.193	0.661	2.920	0.088
age	5.093	0.078	0.714	0.398	4.379	0.036
trt	10.463	0.005	2.464	0.116	7.999	0.005
clstage	10.938	0.004	6.757	0.009	4.181	0.041

*** N O T E *** COMPETING RISKS ANALYSIS ASSUMES THAT THE SURVIVAL TIMES FOR THE TYPES OF FAILURE ARE INDEPENDENT.

REFERENCE: KLEIN AND MOESCHBERGER (1988), BIOMETRICS, VOL. 44, PP. 529-538.

NUMBER OF INTEGER WORDS USED IN PRECEDING PROBLEM 7060

BMDP2L--SURVIVAL ANALYSIS WITH COVARIATES

END OF INSTRUCTIONS

PROGRAM TERMINATED

APPENDIX C (BMDP OUTPUT FOR HNSCC STUDY)

```
BMDP Instruction File   : C:\DYNAMIC\BMDPRUN&.TMP
BMDP Program Output File: C:\DYNAMIC\BMDPOUT&.OUT


BMDP2L--SURVIVAL ANALYSIS WITH COVARIATES


/ INPUT    FILE = 'C:\DYNAMIC\HNSCC.DAT'.
   FORMAT = FREE.
   VARIABLES = 10.
   RECLENGHT= 140.
/ VARIABLE ADD = 1.
   NAMES = pattern, bcl2, treatment, cyclinD1, competing (CR), failuretime,
           Ki-67, CD31, T stage, N stage, CDTRT.

/TRANSFORM CDTRT=CD31*treatment.

/GROUP     CODES(CR) = 1, 2, 3, 4.
   NAMES(CR) = metastasis, nodal, local, censor.

/ FORM     TIME = failuretime.
   RESPONSE = 1, 2, 3.

/ REGRESS  ACCEL = LLOGISTIC.
   COVARIATES = pattern, Bcl-2, cyclinD1, Ki-67, T stage, N stage, CDTRT.
   COMP = CR.
/ END
```

CASE NO.	1 pattern	2 bcl2	3 cyclinD1	4 TRT	5 CR	6 Ftime	7 Ki-67	8 CD-31	9 T stage	10 N stage	11 CDTRT
1	2.00	1.00	52.00	1.00	2.00	184.00	2.00	2.00	2.00	2.00	2.00
2	2.00	1.00	5.00	0.00	4.00	2141.00	2.00	1.00	3.00	2.00	0.00
3	4.00	1.00	19.00	1.00	2.00	197.00	2.00	1.00	3.00	2.00	1.00
4	1.00	1.00	54.00	1.00	3.00	196.00	1.00	1.00	2.00	2.00	1.00
5	2.00	2.00	8.00	1.00	3.00	366.00	2.00	2.00	3.00	1.00	2.00
6	1.00	1.00	42.00	1.00	4.00	583.00	1.00	2.00	2.00	1.00	2.00
7	2.00	1.00	33.00	0.00	4.00	2196.00	1.00	1.00	3.00	1.00	0.00
8	3.00	1.00	60.00	0.00	3.00	88.00	2.00	2.00	4.00	2.00	0.00
9	4.00	1.00	25.00	1.00	3.00	371.00	2.00	2.00	2.00	1.00	2.00
10	1.00	1.00	14.00	0.00	3.00	492.00	1.00	1.00	2.00	1.00	0.00

```
NUMBER OF CASES READ. . . . . .     309

*** N O T E *** TYPE OF FAILURE: CENSORED CONTAINS NO RESPONSES.
                    THE CASES WILL BE USED, BUT NO ANALYSIS FOR
                THIS TYPE OF FAILURE WILL BE PERFORMED.


  ACCELERATED FAILURE TIME MODEL
  ------------------------------
  DISTRIBUTION IS  LLOGIST

THE NATURAL LOGARITHM OF SURVIVAL TIME IS USED IN THE ANALYSIS

       TYPE OF FAILURE: Metastasis
       ---------------------------
```

```
  INITIAL VALUES OF PARAMETERS
  CALCULATED BY LEAST SQUARES

CONSTANT      pattern       cyclinD1      Ki-67         T-stage
 6.877478    -0.105186     -0.005482     0.222822     -0.334627

N-stage         CDTRT         Bcl-2         SCALE
-0.392986       0.097804     0.678445      0.605148
LOG LIKELIHOOD =     -82.1457
GLOBAL CHI-SQUARE =  28.02  D.F.=   7  P-VALUE =0.0002
FOR ACCELERATED FAILURE TIME MODEL, GLOBAL CHI-SQUARE IS BASED ON
LIKELIHOOD RATIO TEST
NORM OF THE SCORE VECTOR=   0.131E-02

                              STANDARD
VARIABLE      COEFFICIENT      ERROR          COEFF./S.E.    EXP(COEFF.)
--------      -----------      --------    -----------    -----------
-1 CONSTANT     13.9960        2.0860        6.7093 1197757.4200
1 pattern       -0.7716        0.3043       -2.5360         0.4623
2 Bcl2           0.2202       0.6947        0.3169         1.2463
3 cyclinD1       0.0019        0.0114        0.1679         1.0019
7 Ki-67          1.5355        0.6472        2.3728         4.6436
9 T-stage       -1.3300        0.4099       -3.2447         0.2645
10 N-stge       -0.8610        0.5837       -1.4751         0.4227
11 CDTRT         0.3992        0.3255        1.2267         1.4907
-2 SCALE         1.1010        0.2013                       3.0072


      TYPE OF FAILURE: nodal
      ----------------------

  INITIAL VALUES OF PARAMETERS
  CALCULATED BY LEAST SQUARES

CONSTANT      pattern       cyclinD1      Ki-67         T-stage
 6.877478    -0.105186     -0.005482     0.222822     -0.334627

N-stage         CDTRT         Bcl-2         SCALE
-0.392986       0.097804     0.678445      0.605148


LOG LIKELIHOOD =    -154.0201
GLOBAL CHI-SQUARE = 46.82  D.F.=   7  P-VALUE =0.0000
FOR ACCELERATED FAILURE TIME MODEL, GLOBAL CHI-SQUARE IS BASED ON
LIKELIHOOD RATIO TEST
NORM OF THE SCORE VECTOR=   0.373E-05

VARIABLE      COEFFICIENT      ERROR          COEFF./S.E.    EXP(COEFF.)
--------      -----------      --------    -----------    -----------
-1 CONSTANT     12.2635        1.7497        7.0089 211818.6671
1 pattern       -0.2344        0.2410       -0.9724         0.7911
2 Bcl2           2.0874       0.8906        2.3439         8.0638
3 cyclinD1      -0.0230        0.0086       -2.6688         0.9773
7 Ki-67          0.4195        0.5141        0.8160         1.5212
9 T-stage        0.0562        0.2967        0.1896         1.0578
10 N-stge       -3.0687        0.6419       -4.7805         0.0465
11 CDTRT        -0.1434        0.2706       -0.5298         0.8664
-2 SCALE         1.3122        0.1735                       3.7145


      TYPE OF FAILURE: Local
      ----------------------

INITIAL VALUES OF PARAMETERS
 CALCULATED BY LEAST SQUARES

CONSTANT      pattern       cyclinD1      Ki-67         T-stage
 6.877478    -0.105186     -0.005482     0.222822     -0.334627

N-stage         CDTRT         Bcl-2         SCALE
-0.392986       0.097804     0.678445      0.605148


LOG LIKELIHOOD =    -314.0638
GLOBAL CHI-SQUARE = 30.37  D.F.=   7  P-VALUE =0.0001
```

FOR ACCELERATED FAILURE TIME MODEL, GLOBAL CHI-SQUARE IS BASED ON
LIKELIHOOD RATIO TEST
NORM OF THE SCORE VECTOR= 0.148E-02

VARIABLE	COEFFICIENT	STANDARD ERROR	COEFF./S.E.	EXP(COEFF.)
-1 CONSTANT	7.3777	0.9160	8.0543	1599.9780
1 pattern	-0.1280	0.1496	-0.8554	0.8799
2 Bcl2	1.3548	0.5153	2.6291	3.8760
3 cyclinD1	-0.0068	0.0056	-1.2231	0.9932
7 Ki-67	0.4190	0.3389	1.2363	1.5204
9 T-stage	-0.7621	0.1889	-4.0339	0.4667
10 N-stge	0.2384	0.3251	0.7331	1.2692
11 CDTRT	0.4009	0.1710	2.3446	1.4932
-2 SCALE	1.1924	0.1021		3.2950

TEST STATISTICS

H1: COVARIATE HAS NO INFLUENCE ON TYPE OF FAILURE AND TIME TO FAILURE.
H2: COVARIATE HAS NO INFLUENCE ON TYPE OF FAILURE.
H3: COVARIATE HAS NO INFLUENCE ON TIME TO FAILURE, GIVEN IT HAS NO INFLUENCE ON TYPE OF FAILURE.

REFERENCE: LAGAKOS (1978), APPLIED STATISTICS, VOL. 27, PP. 235-241.

	H1 (D.F.= 3)		H2 (D.F.= 2)		H3 (D.F.= 1)	
COVARIATE	STAT.	P-VALUE	STAT.	P-VALUE	STAT.	P-VALUE
pattern	8.108	0.044	3.608	0.165	4.500	0.034
Bcl-2	12.506	0.006	3.057	0.217	9.449	0.002
CyclinD1	8.647	0.034	3.681	0.159	4.966	0.026
Ki-67	7.824	0.050	2.498	0.287	5.326	0.021
T-stage	26.836	0.000	8.756	0.013	18.081	0.000
N-stage	25.567	0.000	21.546	0.000	4.021	0.045
CDTRT	7.282	0.063	3.077	0.215	4.206	0.040

*** N O T E *** COMPETING RISKS ANALYSIS ASSUMES THAT THE SURVIVAL TIMES FOR THE TYPES OF FAILURE ARE INDEPENDENT.

REFERENCE: KLEIN AND MOESCHBERGER (1988), BIOMETRICS, VOL. 44, PP. 529-538.

NUMBER OF INTEGER WORDS USED IN PRECEDING PROBLEM 5620

BMDP2L--SURVIVAL ANALYSIS WITH COVARIATES

END OF INSTRUCTIONS

PROGRAM TERMINATED

References

1. Gail M. A review and critique of some models used in competing risk analysis. *Biometrics* 1975: **31:** 209-22.

2. Prentice RL, Kalbfleisch JD, Peterson AV Jr et al. The analysis of failure times in the presence of competing risks. *Biometrics* 1978: **34:** 541-54.

3. Arriagada R, Rutqvist LE, Kramar A, Johansson H. Competing risks determining event-free survival in early breast cancer. *Br J Cancer* 1992: **66:** 951-7.

4. Chapman JW, Fish EB, Link MA. Competing risks analyses for recurrence from primary breast cancer. *Br J Cancer* 1999: **79:** 1508-13.

5. Ataman OU, Bentzen SM, Saunders MI, Dische S. Failure-specific prognostic factors after continuous hyperfractionated accelerated radiotherapy (CHART) or conventional radiotherapy in locally advanced non-small-cell lung cancer: a competing risks analysis. *Br J Cancer* 2001: **85:** 1113-8.

6. Kaplan E.L. and Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Ass* 1958: **53**: 457-81.

7. Margaret Sullivan Pepe. Inference for events with dependent risks in multiple endpoint studies . *J Am. Stat Ass* 1991: **86** [415]: 771-8.

8. Lagakos SW. General right censoring and its impact on the analysis of survival data. *Biometrics* 1979: **35:** 139-56.

9. Kwan-Moon Leung, Robert M Elashoff and Abdelmonem A Afifi. Censoring Issues in survival analysis. *Annu. Rev. Public Health* 1997:**18**, 83-104.

10. Gelman R, Gelber R, Henderson IC, Coleman CN, Harris JR. Improved methodology for analyzing local and distant recurrence. *J Clin Oncol* 1990: **8:** 548-55.

11. Arriagada R, Kramar A, Le Chevalier T, De Cremoux H. Competing events determining relapse-free survival in limited small- cell lung carcinoma. The French Cancer Centers' Lung Group. *J Clin Oncol* 1992: **10:** 447-51.

12. Koscielny S, Thames HD. Biased methods for estimating local and distant failure rates in breast carcinoma and a "commonsense" approach. *Cancer* 2001: **92:** 2220-7.

13. Gooley TA, Leisenring W, Crowley J, Storer BE. Estimation of failure probabilities in the presence of competing risks: new representations of old estimators. *Stat Med* 1999: **18:** 695-706.

14. Pepe MS, Longton G, Pettinger M et al. Summarizing data on survival, relapse, and chronic graft-versus-host disease after bone marrow transplantation: motivation for and description of new methods. *Br J Haematol* 1993: **83:** 602-7.

15. Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med* 1993: **12:** 737-51.

16. Kalbfleisch J.D and Prentice R.L. The statistical analysis of failure time data. New York: John Wiley, 1980.

17. Bentzen SM, Vaeth M, Pedersen DE, Overgaard J. Why actuarial estimates should be used in reporting late normal-tissue effects of cancer treatment ... now! *Int J Radiat Oncol Biol Phys* 1995: **32:** 1531-4.

18. Caplan RJ, Pajak TF, Cox JD. Analysis of the probability and risk of cause-specific failure. *Int J Radiat Oncol Biol Phys* 1994: **29:** 1183-6.

19. Albain KS, Crowley JJ, LeBlanc M, Livingston RB. Survival determinants in extensive-stage non-small-cell lung cancer: the Southwest Oncology Group experience. *J Clin Oncol* 1991: **9:** 1618-26.

20. Takigawa N, Segawa Y, Okahara M et al. Prognostic factors for patients with advanced non-small cell lung cancer: univariate and multivariate analyses including recursive partitioning and amalgamation. *Lung Cancer* 1996: **15:** 67-77.

21. Bentzen SM, Poulsen HS, Kaae S et al. Prognostic factors in osteosarcomas. A regression analysis. *Cancer* 1988: **62:** 194-202.

22. Haie-Meder C, Kramar A, Lambin P et al. Analysis of complications in a prospective randomized trial comparing two brachytherapy low dose rates in cervical carcinoma. *Int J Radiat Oncol Biol Phys* 1994: **29:** 953-60.

23. O'Brien PC. Radiation injury of the rectum. *Radiother Oncol* 2001: **60:** 1-14.

24. Wilson GD, McNally NJ, Dische S et al. Measurement of cell kinetics in human tumours in vivo using bromodeoxyuridine incorporation and flow cytometry. *Br J Cancer* 1988: **58:** 423-31.

25. Denekamp J. Cell kinetics and radiation biology. *Int J Radiat Biol Relat Stud Phys Chem Med* 1986: **49:** 357-80.

26. Dische S, Saunders MI. Fractionation--a review of the clinical data. *Br J Radiol Suppl* 1988: **22:** 84-7.

27. Dische S, Saunders MI. The rationale for continuous, hyperfractionated, accelerated radiotherapy (CHART). *Int J Radiat Oncol Biol Phys* 1990: **19:** 1317-20.

28. Saunders M, Dische S, Barrett A et al. Continuous, hyperfractionated, accelerated radiotherapy (CHART) versus conventional radiotherapy in non-small cell lung cancer: mature data from the randomised multicentre trial. CHART Steering committee. *Radiother Oncol* 1999: **52:** 137-48.

29. Dische S, Saunders M, Barrett A et al. A randomised multicentre trial of CHART versus conventional radiotherapy in head and neck cancer. *Radiother Oncol* 1997: **44:** 123-36.

30. Bennett MH, Wilson GD, Dische S et al. Tumour proliferation assessed by combined histological and flow cytometric analysis: implications for therapy in squamous cell carcinoma in the head and neck. *Br J Cancer* 1992: **65:** 870-8.

31. *BMDP User's Guide* , Version 7, ed. Dixon WJ, Wiley 1992: 825-864.

32. Lagakos SW. A covariate model for partially censored data subject to competing causes of failure. *Appl Statist* 1978: **27**[3]: 235-41.

33. Bentzen SM, Thames HD, Travis EL et al. Direct estimation of latent time for radiation injury in late- responding normal tissues: gut, lung, and spinal cord. *Int J Radiat Biol* 1989: **55:** 27-43.

34. Hong WK, Bromer RH, Amato DA et al. Patterns of relapse in locally advanced head and neck cancer patients who achieved complete remission after combined modality therapy. *Cancer* 1985: **56:** 1242-5.

35. Vikram B. Changing patterns of failure in advanced head and neck cancer. *Arch Otolaryngol* 1984: **110:** 564-5.

36. Ardizzoni A, Grossi F, Scolaro T et al. Induction chemotherapy followed by concurrent standard radiotherapy and daily low-dose cisplatin in locally advanced non-small-cell lung cancer. *Br J Cancer* 1999: **81:** 310-5.

37. Stevens CW, Lee JS, Cox J, Komaki R. Novel approaches to locally advanced unresectable non-small cell lung cancer. *Radiother Oncol* 2000: **55:** 11-8.

38. Cox JD. Induction chemotherapy for non-small cell carcinoma of the lung: limitations and lessons. *Int J Radiat Oncol Biol Phys* 1991: **20:** 1375-6.

39. Cox JD, Scott CB, Byhardt RW et al. Addition of chemotherapy to radiation therapy alters failure patterns by cell type within non-small cell carcinoma of lung (NSCCL): analysis of radiation therapy oncology group (RTOG) trials. *Int J Radiat Oncol Biol Phys* 1999: **43:** 505-9.

40. Wigren T. Confirmation of a prognostic index for patients with inoperable non-small cell lung cancer. *Radiother Oncol* 1997: **44:** 9-15.

41. Wigren T, Oksanen H, Kellokumpu-Lehtinen P. A practical prognostic index for inoperable non-small-cell lung cancer. *J Cancer Res Clin Oncol* 1997: **123:** 259-66.

42. Lunn M, McNeil D. Applying Cox regression to competing risks. *Biometrics* 1995: **51:** 524-32.

43. Chen M, Jiang GL, Fu XL et al. Prognostic factors for local control in non-small-cell lung cancer treated with definitive radiation therapy. *Am J Clin Oncol* 2002: **25:** 76-80.

44. Scott C, Sause WT, Byhardt R et al. Recursive partitioning analysis of 1592 patients on four Radiation Therapy Oncology Group studies in inoperable non-small cell lung cancer. *Lung Cancer* 1997: **17 Suppl 1:** S59-74.

45. Komaki R, Scott CB, Byhardt R et al. Failure patterns by prognostic group determined by recursive partitioning analysis (RPA) of 1547 patients on four radiation therapy oncology group (RTOG) studies in inoperable nonsmall-cell lung cancer (NSCLC). *Int J Radiat Oncol Biol Phys* 1998: **42:** 263-7.

46. Werner-Wasik M, Scott C, Cox JD et al. Recursive partitioning analysis of 1999 Radiation Therapy Oncology Group (RTOG) patients with locally-advanced non-small-cell lung cancer (LA-NSCLC): identification of five groups with different survival. *Int J Radiat Oncol Biol Phys* 2000: **48:** 1475-82.

47. Ciampi A, Hogg SA, McKinney S, Thiffault J. RECPAM: a computer program for recursive partition and amalgamation for censored survival data and other situations frequently occurring in biostatistics. I. Methods and program features. *Comput Methods Programs Biomed* 1988: **26:** 239-56.

48. Ciampi A, Lawless JF, McKinney SM, Singhal K. Regression and recursive partition strategies in the analysis of medical survival data. *J Clin Epidemiol* 1988: **41:** 737-48.

49. Smith BD, Haffty BG. Molecular markers as prognostic factors for local recurrence and radioresistance in head and neck squamous cell carcinoma. *Radiat Oncol Investig* 1999: **7:** 125-44.

50. Salesiotis AN, Cullen KJ. Molecular markers predictive of response and prognosis in the patient with advanced squamous cell carcinoma of the head and neck: evolution of a model beyond TNM staging. *Curr Opin Oncol* 2000: **12:** 229-39.

51. Bourhis J, Pignon JP. Meta-analyses in head and neck squamous cell carcinoma. What is the role of chemotherapy? *Hematol Oncol Clin North Am* 1999: **13:** 769-75.

52. Michalides R, van Veelen N, Hart A et al. Overexpression of cyclin D1 correlates with recurrence in a group of forty-seven operable squamous cell carcinomas of the head and neck. *Cancer Res* 1995: **55:** 975-8.

53. Wilson GD, Saunders MI, Dische S et al. bcl-2 expression in head and neck cancer: an enigmatic prognostic marker. *Int J Radiat Oncol Biol Phys* 2001: **49:** 435-41.

54. Hansen O, Overgaard J, Hansen HS et al. Importance of overall treatment time for the outcome of radiotherapy of advanced head and neck carcinoma: dependency on tumor differentiation. *Radiother Oncol* 1997: **43:** 47-51.

55. Akervall JA, Jin Y, Wennerberg JP et al. Chromosomal abnormalities involving 11q13 are associated with poor prognosis in patients with squamous cell carcinoma of the head and neck. *Cancer* 1995: **76:** 853-9.

56. Meredith SD, Levine PA, Burns JA et al. Chromosome 11q13 amplification in head and neck squamous cell carcinoma. Association with poor prognosis. *Arch Otolaryngol Head Neck Surg* 1995: **121:** 790-4.

57. Bova RJ, Quinn DI, Nankervis JS et al. Cyclin D1 and p16INK4A expression predict reduced survival in carcinoma of the anterior tongue. *Clin Cancer Res* 1999: **5:** 2810-9.

58. Muller D, Millon R, Velten M et al. Amplification of 11q13 DNA markers in head and neck squamous cell carcinomas: correlation with clinical outcome. *Eur J Cancer* 1997: **33:** 2203-10.

59. Muller D, Millon R, Lidereau R et al. Frequent amplification of 11q13 DNA markers is associated with lymph node involvement in human head and neck squamous cell carcinomas. *Eur J Cancer B Oral Oncol* 1994: **30B:** 113-20.

60. Masuda M, Hirakawa N, Nakashima T, Kuratomi Y, Komiyama S. Cyclin D1 overexpression in primary hypopharyngeal carcinomas. *Cancer* 1996: **78:** 390-5.

61. Bellacosa A, Almadori G, Cavallo S et al. Cyclin D1 gene amplification in human laryngeal squamous cell carcinomas: prognostic significance and clinical implications. *Clin Cancer Res* 1996: **2:** 175-80.

62. Akervall JA, Michalides RJ, Mineta H et al. Amplification of cyclin D1 in squamous cell carcinoma of the head and neck and the prognostic value of chromosomal abnormalities and cyclin D1 overexpression. *Cancer* 1997: **79:** 380-9.

63. Takes RP, Baatenburg de Jong RJ, Schuuring E et al. Markers for assessment of nodal metastasis in laryngeal carcinoma. *Arch Otolaryngol Head Neck Surg* 1997: **123:** 412-9.

64. Fortin A, Guerry M, Guerry R et al. Chromosome 11q13 gene amplifications in oral and oropharyngeal carcinomas: no correlation with subclinical lymph node invasion and disease recurrence. *Clin Cancer Res* 1997: **3:** 1609-14.

65. Michalides RJ, van Veelen NM, Kristel PM et al. Overexpression of cyclin D1 indicates a poor prognosis in squamous cell carcinoma of the head and neck. *Arch Otolaryngol Head Neck Surg* 1997: **123:** 497-502.

66. Kyomoto R, Kumazawa H, Toda Y et al. Cyclin-D1-gene amplification is a more potent prognostic factor than its protein over-expression in human head-and-neck squamous-cell carcinoma. *Int J Cancer* 1997: **74:** 576-81.

67. Nogueira CP, Dolan RW, Gooey J et al. Inactivation of p53 and amplification of cyclin D1 correlate with clinical outcome in head and neck cancer. *Laryngoscope* 1998: **108:** 345-50.

68. Pignataro L, Capaccio P, Carboni N et al. p53 and cyclinX D1 protein expression in carcinomas of the parotid gland. *Anticancer Res* 1998: **18:** 1287-90.

69. Matthias C, Branigan K, Jahnke V et al. Polymorphism within the cyclin D1 gene is associated with prognosis in patients with squamous cell carcinoma of the head and neck. *Clin Cancer Res* 1998: **4:** 2411-8.

70. Bova RJ, Quinn DI, Nankervis JS, et al. Cyclin D1 and p16INK4A expression predict reduced survival in carcinoma of the anterior tongue. *Clin Cancer Res* 1999: **5**(10):2810-9.

71. Gleich LL, Li YQ, Wang X, Stambrook PJ, Gluckman JL. Variable genetic alterations and survival in head and neck cancer. *Arch Otolaryngol Head Neck Surg* 1999: **125:** 949-52.

72. Rodrigo JP, Garcia LA, Ramos S, Lazo PS, Suarez C. EMS1 gene amplification correlates with poor prognosis in squamous cell carcinomas of the head and neck. *Clin Cancer Res* 2000: **6:** 3177-82.

73. Capaccio P, Pruneri G, Carboni N et al. Cyclin D1 expression is predictive of occult metastases in head and neck cancer patients with clinically negative cervical lymph nodes. *Head Neck* 2000: **22:** 234-40.

74. Yoo SS, Carter D, Turner BC et al. Prognostic significance of cyclin D1 protein levels in early-stage larynx cancer treated with primary radiation. *Int J Cancer* 2000: **90:** 22-8.

75. Smith BD, Smith GL, Carter D et al. Molecular marker expression in oral and oropharyngeal squamous cell carcinoma. *Arch Otolaryngol Head Neck Surg* 2001: **127:** 780-5.

76. Dong Y, Sui L, Sugimoto K, Tai Y, Tokuda M. Cyclin D1-CDK4 complex, a possible critical factor for cell proliferation and prognosis in laryngeal squamous cell carcinomas. *Int J Cancer* 2001: **95:** 209-15.