Sirpa Heinävaara

Master of Social Sciences

MODELLING SURVIVAL OF PATIENTS WITH MULTIPLE CANCERS

Academic dissertation to be presented, with the permission of the Faculty of Social Sciences of the University of Helsinki, for public criticism in Auditorium XIV, Fabianinkatu 33, on June 6th, 2003, at 12 noon.

Helsinki 2003
The author’s address:
Finnish Cancer Registry
Liisankatu 21 B
FIN-00170 Helsinki

ISBN 952-10-1187-4
Yliopistopaino
Helsinki 2003
MODELLING SURVIVAL OF PATIENTS
WITH MULTIPLE CANCERS

by

Sirpa Heinävaara

Finnish Cancer Registry, Helsinki, Finland
&
Department of Statistics, University of Helsinki, Finland

Helsinki 2003
ABSTRACT

Sirpa Heinävaara: Modelling survival of patients with multiple cancers

With increasing number of subsequent primary cancers there is a growing concern to know how cancer patients survive with their subsequent cancer compared to those with their respective first cancer. Results of earlier studies have been conflicting and have not lead to firm conclusions. One reason for conflicting results might be a lack appropriate methodology as survival from subsequent cancer has usually not been adjusted for an extra hazard due to an underlying first cancer.

This study presents four alternative models for estimating survival of patients with multiple cancers. Models are extensions and modifications to those proposed earlier for estimating relative and cause-specific survival of patients with a single cancer. The assessment of survival from subsequent cancer raised a need for introducing new concepts, especially when survival of patients with their multiple cancers of the same site is concerned. Survival estimates from cancer are compared between the models, and between a first and subsequent tumour of the same site. The importance of adjusting survival from subsequent cancer to that from a underlying first cancer is also highlighted.

The results show that survival from cancer as a first and subsequent tumour can be reliably assessed with the newly introduced models based either on the relative and cause-specific survival. The results also show that survival from cancer as a first and subsequent tumour may be dependent on the site of cancer and whether patients’ cancers are of the same site or not. Nevertheless, survival from a subsequent cancer is not usually different from that from a respective first cancer. However, even with large population-based data, a lack of power often prevents the detection of modest differences in survival.

Keywords: multiple cancers, subsequent cancer, relative survival, cause-specific survival
ACKNOWLEDGEMENTS

I thank my supervisor Professor Timo Hakulinen for his enormous support of my work, time for discussing and reading the drafts, and most of all, for his everlasting optimism. His support and encouragement was essential for finalizing this study.

I’m indebted to Professor Lyly Teppo and Docent Risto Sankila for contributing with their knowledge and understanding of the clinical background. Lyly Teppo, with his long experience in cancer registration in Finland, is a marvelous source of information which he is willing to share with those interested. Thank you both for your time!

I thank my reviewers Professors Esa Lääärä and Juni Palmgren for their valuable comments which lead to the improvement of this work.

I also thank my collaborator Hans Storm, MD, for his valuable comments on the manuscript.

I’m grateful for being able to use some of the best data on cancer patients in the world. I’m indebted to Bengt Söderman for supplying Finnish data, Niels Sønderby Christensen for Danish data, and Aage Johansen & Svein Erling Tysvær for Norwegian data from the respective cancer registries. I also thank Bengt Söderman and Martti Ojama at the Finnish Cancer Registry for technical support.

My warmest thanks are due to my parents and friends for their love, understanding and letting me work in peace.

And last, but definitely not the least, the support of my family has been essential. I dedicate my warmest thanks, hugs and love to my husband Mikko and our children Otte and Teo for their love, sense of humour and for showing the importance of all aspects of life. I couldn’t have done this without you.

This study was carried out at the Finnish Cancer Registry. I’m grateful for the financial support of the Cancer Society of Finland and MaDaMe research project of the Academy of Finland.
ABBREVIATIONS

The following abbreviations are used throughout this thesis:

c site of cancer of primary interest
c1 a first primary cancer of site c
c0 a first primary cancer of site non-c
c2 a subsequent primary cancer of site c diagnosed after a first primary cancer

cv, v=1,2 cancers c1 and c2 of site c, or
cv, v=0,1,2 cancers c1 and c2 of site c and cancer c0 of site non-c

cf, f=0,1 a first cancer c0 or c1

λv, v=0,1,2 a hazard of dying from cancer cv

λe a hazard of dying in a general population group

λT a total mortality hazard of dying from any cause

TO SUMMARIZE:

• If patients have their multiple cancers at different sites, cancers are called c0 and c2.
• If patients have their multiple cancers of the same site, cancers are called c1 and c2.

ERRATUM

The sentence on the right-hand column on page 147 of article III, 9th line from the bottom, starting with

Their model was based on an assumption that the follow-up times of those non-cured follow a Weibull distribution, ... should be

Their model was based on an assumption that the survival of those non-cured follow a Weibull distribution, ....

In addition, equation for $S_v(t, z_v)$ on page 148 of article III should be

$S_v(t, z_v) = [P_v + (1 − P_v) \exp(−(α_v t)^γ_v)]^{\exp(β_v^T z_v)}.$
8 REFERENCES  
9 APPENDIX
1 LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following articles which are referred to in the text by their Roman numerals:

I. Heinävaara S and Hakulinen, T.
Relative survival of patients with multiple primary breast cancers.
*Journal of Cancer Epidemiology and Prevention. In press*

II. Heinävaara S, Teppo L and Hakulinen, T.
Cancer-specific survival of patients with multiple cancers.
An application to patients with multiple breast cancers.

III. Heinävaara S and Hakulinen, T.
Parametric mixture model for analysing relative survival of patients with multiple cancers.
*Journal of Cancer Epidemiology and Prevention* 2002;7(3):147-53

IV. Heinävaara S, Sankila R, Storm H, Langmark F and Hakulinen T.
Relative survival of patients with prostate cancer as a first or subsequent tumour - A Nordic collaborative study.
*Cancer Causes and Control* 2002;13:797-806.
2 INTRODUCTION

At an individual level, a diagnosis of cancer can be regarded as a human tragedy, and at the level of society, cancers are the major diseases causing a notable amount of health administrative costs. Prognosis and possible cure from cancer are thus important measures. Both of them can be assessed by analysing the survival of cancer patients.

Survival of cancer patients is known to be dependent on prognostic factors such as age at diagnosis, gender of the patient and site of the cancer. Since the site of the cancer often has a significant effect on survival, one may be especially interested in survival from cancer at a specific site while simultaneously accounting for the effects of prognostic factors. Survival from cancer can be assessed by estimating relative or cause-specific survival.

With improved survival from cancer and ageing of populations, the numbers of subsequent primary cancers are increasing. In Finland, the number of registered subsequent primary cancers increased from 7,400 in calendar period 1970-79 to almost 32,000 in 1990-99. The increase is evident although that 7,400 can be an underestimation since the information on the first primary cancers diagnosed before 1953 is not available. As a consequence, estimates of survival from subsequent cancer are of growing interest, and there is an interest in comparing survival from a certain cancer as a first and subsequent tumour. For such comparisons, however, survival from subsequent cancer should be properly estimated and adjusted for available prognostic factors.

New concepts are needed in assessing survival from subsequent cancer. The need is evident especially when survival of patients with two primary cancers of the same site are studied. One should, for example, be able to distinguish between survival from breast cancer as a first tumour and breast cancer as a subsequent tumour. Moreover, there should be a concept for assessing survival from a first or subsequent cancer as well as survival from both cancers together.

This study aims at developing concepts and models for assessing survival from cancer as a first and subsequent tumour while simultaneously controlling for the effects of other prognostic factors. Special attention is paid to survival of patients with two cancers of the same site. Survival from cancer as a first
and subsequent tumour is assessed by semi-parametric models for relative and cause-specific survival. Survival from cancer as a first and subsequent tumour is also assessed simultaneously with proportions cured from cancer using a parametric mixture model for relative survival. Survival estimates from cancer are compared between the models, and between a first and subsequent tumour of the same site. A detailed list of the aims is presented in section 4.

The proportion of patients with more than two primary cancers among patients with multiple primary cancers is rather small even if it is increasing with calendar time: In 1970-79, 92 % of subsequent primary cancers were secondary cancers whereas in 1990-99, 84 % were secondary. This study concentrates on assessing survival from one subsequent primary cancer of interest. Patients with two primary cancers are from now on called patients with multiple (primary) cancers or multiple-cancer patients, and patients with one (primary) cancer are referred to as single-cancer patients.

3 LITERATURE REVIEW

3.1 Survival of multiple-cancer patients

Studies of the survival of multiple-cancer patients have been shown conflicting results: Patients diagnosed with a subsequent cancer may survive better than [1, 2, 3, 4], worse than [5] or similarly to [6, 7, 8] patients a first cancer in the same site. Results may be dependent on the site of cancer but even the studies of survival of patients with multiple breast cancers have been non-conclusive [2, 5, 7, 8]. The number of patients with subsequent cancer may also have been too small to draw definitive conclusions [9].

One possible reason for these non-conclusive results may be due methodological drawbacks, as discussed earlier by Sankila and Hakulinen [4, 10]. These methodological problems include the lack of adjustment for prognostic factors such as age and stage of cancer [6, 11], ignoring the time interval between the diagnoses of a first and subsequent cancer [1, 11], estimating overall survival instead of survival from cancer [8, 11, 12], and most of all, not appropriately accounting for the mortality associated with the first primary cancer.
To avoid the methodological difficulties in the previously published studies, Sankila and Hakulinen [4, 10] proposed a new method for estimating relative survival of multiple-cancer patients. The method is an extension to the analysis of relative survival of single-cancer patients based on life-table estimates and grouped data. The method is presented briefly in [4] and with details in [10]. Sankila and Hakulinen studied survival from a subsequent colorectal cancer, diagnosed after a first breast cancer, and compared it with that from a first colorectal cancer. Survival from subsequent colorectal cancer was adjusted for the prognostic factors and for the fixed excess hazard of the first breast cancer. The method by Sankila and Hakulinen and the new models are compared in section 8.1.

### 3.2 Survival of single-cancer patients

The new models presented in this thesis are extensions and modifications to existing models for analysing survival from cancer of single-cancer patients. Survival from cancer has usually been assessed with the analysis of relative or cause-specific survival [13, 14] with so-called semi-parametric models. In these semi-parametric models, prognostic factors are incorporated with some parametric form but nothing, or almost nothing, has been assumed on the survival distribution. Survival from cancer has also been estimated with proportion cured from cancer using parametric mixture [15] and non-mixture [16] models.

In the analysis of relative survival of patients with a single cancer c1, the total mortality hazard \( \lambda_T \) is assumed to consist of an expected mortality hazard \( \lambda_e \), usually estimated from a representative general population group, and an excess hazard \( \lambda_1 \) attributable to cancer c1. In the analysis of cause-specific survival, the deaths from cancer c1 contribute to the cause-specific hazard \( \lambda_1 \).

The relative and cause-specific survival at follow-up time \( t_1 \), \( S_1(t_1) \), is a function of the cumulative hazard \( \Lambda_1(t_1) \) until \( t_1 \), namely \( S_1(t_1) = \exp(-\Lambda_1(t_1)) \).

The analysis of relative survival is thus possible if patients are known to have died or been censored but in the analysis of cause-specific survival, a cause of death must be defined to be either cancer c1 or non-c1.

The analysis of relative and cause-specific survival are traditionally based on grouped data. Using the method by Estève et al. [14], survival from cancer...
can also be estimated using individual patient data. The original method was applicable only when the effect of prognostic factors was proportional to a baseline (proportional hazards) but it has later been modified to include non-proportional hazards [17].

In parametric mixture and non-mixture models, survival is estimated simultaneously with a proportion cured from cancer. In parametric mixture models, a patient is assumed, at the diagnosis of his or her cancer, to belong to an unknown proportion of patients cured from cancer $P$ or to that of patients who are bound to die from cancer $1 - P$. Those belonging to the proportion cured are bound to die only due to the other causes of death similarly [15], or proportionally similarly [18], to those in the corresponding general population group, and those belonging to the proportion non-cured are assumed to survive from cancer according to a pre-defined parametric survival distribution $S_c$. In a parametric mixture model, survival from cancer of all patients $S$ is thus estimated with $S = P + (1 - P)S_c$. The proportion cured is considered to be a function of prognostic factors [19], and survival $S_c$ is often assumed to follow a Weibull [15, 16, 19] or a log-normal [20, 21] distribution. The parametric non-mixture models are based on an assumption that the proportion cured $P$ is an asymptotic limit of survival from cancer $S$ [16, 22, 23].
4 AIMS OF THE STUDY

The specific aims of the study are as follows:

1. To develop concepts for assessing survival of patients with multiple cancers [I-IV].
2. To develop flexible and easily applicable models for estimating survival from cancer as a first and subsequent tumour [I-IV] based on the analysis of the relative and the cause-specific survival.
3. To develop models for estimating cause-specific survival of multiple-cancer patients based on available information on the cause of death, either as an official cause of death or as a cancer-specific (tumour-specific) cause of death [II].
4. To develop a model for assessing proportions cured from cancer as a first and subsequent tumour [III].
5. To compare a new relative survival model with the respective method by Sankila and Hakulinen, and to apply the new model with large empirical data on patients with multiple cancers [IV].
6. To include the effects of prognostic factors into survival in all the models [I-IV].
7. To compare survival from cancer between a first and subsequent primary tumour [I-IV].
8. To study possible associations between a first and subsequent tumour of multiple-cancer patients using time interval between the diagnoses as a possible prognostic factor for survival from a subsequent cancer [I-IV].

In all of these aims, the precision of all the parameters should be estimable and the individual variability should be taken into account as well as possible.
5 MATERIALS

Population-based data on cancer patients have been available at the Danish Cancer Registry since 1943 and at the Finnish and Norwegian cancer registries approximately a decade later. In all of these countries, the notification of new cancer cases is compulsory, in Denmark since 1987 [24]. Reporting is based on multiple sources including physicians, hospitals, institutions with hospital beds, pathological and cytological laboratories and death certificates. Coverage of almost 100% is achieved in all of these countries [25].

Since each resident of the Nordic countries has a unique personal identification number, linking of individual records is simple and reliable. The reliable identification and the follow-up of patients with multiple primary cancers was essential for this study. The Nordic cancer registries link their records of individual cancer patients also with the official cause of death [25]. At the Finnish Cancer Registry, the relationship between the cancer and the official cause of death is additionally recorded with a special code, a cancer- or tumour-specific cause of death, for each primary cancer.

Publications I-III are based on Finnish data supplied by the Finnish Cancer Registry and publication IV on Danish, Finnish and Norwegian data supplied by the respective cancer registries. For publications
- I-II, the data consisted of Finnish female patients diagnosed with one or two primary breast cancers;
- III, the data consisted of Finnish male patients diagnosed with first primary localized colorectal cancer possibly followed by a second primary lung cancer and, for survival comparisons, of Finnish male patients diagnosed with a first primary lung cancer; and
- IV, the data consisted of Danish, Finnish and Norwegian male patients diagnosed with subsequent primary prostate cancer after the first primary colorectal cancer, and for comparisons, of respective male patients with a first primary prostate cancer.

Since the Nordic cancer registries collaborate on standardization of registration and classification [25], the Danish, Finnish and Norwegian data were easy to standardize with the major exception being the stage of cancer: The coding practice for the stage of cancer varies from one Nordic country to another.
For the estimation of relative survival (I,III-IV), the expected general population mortality hazards $\lambda_e$ were needed. The $\lambda_e(s, a, p, C)$ were tabulated by gender $s$, one-year age group $a$ ($a=0,1,...,99$), 5-year calendar time period $p$ and country $C$ (Denmark, Finland, Norway). For patients aged more than 99 years at the end of the follow-up, the $\lambda_e(s,99,p,C)$ were used. In Denmark, the $\lambda_e(s,90+,p,Dk)$ were received as pooled, 90+ referring to those at least 90 years old. The one-year age-specific $\lambda_e(s,m,p,Dk)$, $m=90,...,99$, were calculated by assuming that the ratio between the Danish $\lambda_e(s,85−89,p,Dk)$, pooled over ages 85-89, and the Norwegian $\lambda_e(s,85−89,p,N)$ was the same also for age $m$ given $s$ and $p$.

### 6 METHODS

#### 6.1 General background

In the analysis of relative survival of patients with multiple cancers $c_1$ and $c_2$, an excess hazard $\lambda_{1,2}$ was assumed to be due to both cancers $c_1$ and $c_2$. In the corresponding analysis of cause-specific survival, deaths from either of the cancers $c_1$ or $c_2$ contribute to the cause-specific hazard $\lambda_{1,2}$.

According to the theory of competing risks one may decompose for a patient group of single- and multiple-cancer patients that

$$\lambda_{1,2}(t_1, t_2) = \lambda_1(t_1) + c_2\lambda_2(t_2),$$

where $c_2$ is an indicator whether a patient is a single-cancer ($c_2 = 0$) or a multiple-cancer ($c_2 = 1$) patient, and $\lambda_2(t_2)$ is a hazard related to a subsequent cancer $c_2$ at time $t_2$ since the diagnosis of $c_2$.

Although $t_1$ and $t_2$ represent times since diagnoses of $c_1$ and $c_2$, respectively, $\lambda_{1,2}(t_1, t_2)$ has only one time dimension: If $x_1$ and $x_2$ are the ages at diagnoses of cancers $c_1$ and $c_2$, respectively, then $t_2 = t_1 − (x_2−x_1)$ with $0 ≤ t_2 ≤ t_1 < \infty$.

The additivity of the hazards shown in equation (1) does not imply that the hazards are independent [26]. For the estimation of proportions cured from cancers $c_1$ and $c_2$ with a parametric mixture model (III), an assumption of
independence is required: given the values of prognostic factors, survival from cancer \( c_2 \) should be independent of that from cancer \( c_1 \). This assumption must be and has been motivated further with the presentation of the data used in III.

The \( \lambda_1 \) in equation (1) is assumed to be the same for single- and multiple-cancer patients with their first cancer \( c_1 \). Since it is not known in advance who is to become a multiple-cancer patient and who is not, the inclusion of all single-cancer patients is required for the estimation of \( \lambda_1 \). Moreover, when the estimation of the \( \lambda_1 \) is based on all possible data, its estimate will be precise. It could of course be studied whether \( \lambda_1 \) for single-cancer patients equals that for multiple-cancer patients. However, the number of multiple-cancer patients is small for the simultaneous estimation of parameters related to two hazards, and thus the statistical power to study this question is rather limited.

It is of primary interest to study whether the hazard \( \lambda_2 \) of a subsequent cancer \( c_2 \) of site \( c \) is the same, or proportional to the corresponding hazard \( \lambda_1 \) of a first cancer \( c_1 \) of site \( c \). If multiple-cancer patients have their cancers at the same primary site, only the hazards \( \lambda_1 \) and \( \lambda_2 \) are estimated. Usually, however, cancers of multiple-cancer patients are not at the same site. In this case the underlying hazard \( \lambda_0 \) of cancer of site non-\( c \) is also estimated. Let \( \lambda_v \) denote for all the hazards, \( v=1,2 \) for patients with multiple cancers of the same site, and \( v=0,1,2 \) otherwise, and let \( \lambda_f \) be the hazard of the first cancer, \( f=0,1 \). The hypothesis of the proportionality of hazards between a first and subsequent tumour at site \( c \) can also be expressed formally with

\[
(2) \quad H_0 : \lambda_2 = p\lambda_1 \quad vs \quad H_1 : \lambda_2 \neq p\lambda_1,
\]

where \( p \) is a constant to be estimated.

To use all the possible information, especially on patients with a subsequent cancer \( c_2 \), the estimation of \( \lambda_v \) is based on individual patient data. The semi-parametric approach presented by Estève et al. [14] for estimating survival from cancer of single-cancer patients was chosen as a basis for the further development (I,II,IV). The survival and proportions cured from cancer as a first and subsequent tumour (III) were estimated using an extension to a parametric mixture model which was introduced in personal discussion with Dr. Arduino Verdecchia, Instituto Superiore di Sanita’, Rome, Italy, and which is
6.2 Adjustment for prognostic factors

The $\lambda_v$ are known to vary with follow-up time $t_v$ and with prognostic factors $z_v$, and should therefore be more accurately denoted as $\lambda_v(t_v, z_v)$.

In applications I, II and IV, $\lambda_v$ is assumed to be piecewise constant with follow-up time $t_v$ for patients with $z_v=0$. The effect of prognostic factors $z_v$ on $\lambda_v$ is assumed to be proportional with respect to follow-up time $t_v$.

In the semi-parametric method following Estève et al. [14], a hazard can be written as

$$
\lambda_v(t, z_v) = \exp(\beta_v^T z_v) \sum_{k=1}^{m_v} \tau_{vk} I_{vk}(t),
$$

where $\beta_v$ is a vector of parameter estimates for prognostic factors, $\tau_{vk} > 0$ is a baseline hazard during the discrete time interval $k$ for patients with $z_v = 0$, $m_v$ is the number of discrete time intervals after the diagnosis of cancer $c_v$, and $I_{vk}(t)$ is an indicator function of the $k$th interval.

In application III based on a parametric mixture model, survival of those non-cured $S_{1-P_v}$, $v=0,1,2$, is assumed to follow a Weibull distribution, and prognostic factors $z_v$ are assumed to affect both the proportion cured $P_v$ from cancer $c_v$ and survival $S_{1-P_v}$. The adjustment of prognostic factors can be understood through survival from cancer $S_v$ which can be estimated with

$$
S_v(t, z_v) = [P_v + (1 - P_v) \exp(-(\alpha_v t)^{\gamma_v})]^{\exp(\beta_v^T z_v)}, \alpha_v, \gamma_v > 0,
$$

where $\alpha_v$ and $\gamma_v$ are shape and scale parameters, respectively, and $\beta_v$ is a vector of prognostic factors.

The hazard $\lambda_v$ can then be calculated as

$$
\lambda_v = \frac{-d \log(S_v(t, z_v))}{dt} = \exp(\beta_v^T z_v) \frac{(1 - P_v) \exp(-(\alpha_v t)^{\gamma_v})((\alpha_v t_v)^{\gamma_v-1}) \alpha_v \gamma_v}{P_v + (1 - P_v) \exp(-(\alpha_v t)^{\gamma_v})}.
$$
The time interval between the diagnoses of the subsequent cancer \( c_2 \) and the underlying first cancer \( c_f \), and tumour rank indicator (first or subsequent) were considered as potential prognostic factors for the \( \lambda_2 \). The time interval between the diagnoses was classified into categories based on the behaviour of the underlying hazard \( \lambda_f, f = 0,1 \) and was used as a categorical variable in modelling. The effect of the tumour rank indicator was tested only when the effect of follow-up time \( t_v \) on the \( \lambda_v \) was proportional between cancers \( c_2 \) and \( c_1 \).

### 0.1 Models for estimating survival from cancer as a first and subsequent tumour

Usually the hazard \( \lambda_{1,2} > \lambda_2 \) unless the time interval between the diagnoses of cancers \( c_1 \) and \( c_2 \) is so long that \( \lambda_1 \approx 0 \) and thus \( \lambda_{1,2} \approx \lambda_2 \). In practice, survival from cancer \( c_v \) is of primary interest when the \( \lambda_v > 0 \), that is, when there is additional mortality due to cancer \( c_v \). The mathematical restrictions, \( \lambda_v \geq 0 \) needed later in model (6) and \( \lambda_v > 0 \) needed in (7-9), are therefore not limiting in practice.

Let us consider survival of patients with multiple cancers at different sites. Let the number of patients diagnosed with a first cancer \( c_1 \) at site \( c \) be \( N_1 \) and the number with a first cancer \( c_0 \) at site non-\( c \) be \( N_0 \). Each patient \( i \) has prognostic factors \( z_{vi} \), follow-up times \( t_{vi} \) up to death or censoring, an individual hazard \( \lambda_{vi}(t_{vi}, z_{vi}) \), an individual cumulative hazard \( \Lambda_{vi}(t_{vi}, z_{vi}) \) and a fixed expected mortality hazard \( \lambda_{ei}(t_{fi} + x_{fi}, z^*_{fi}) \); \( x_{fi} \) is the age at diagnosis of cancer \( c_f \), \( f=0,1 \), and \( z^*_{fi} \) may be regarded as a sub vector of \( z_{fi} \). The arguments for follow-up time \( t_{vi} \) and covariates \( z_{vi} \) as well as the possible constants have been omitted from the log-likelihoods.

#### 0.1.1 Semi-parametric model for relative survival

For comparisons of survival from cancer as a first and subsequent tumour, \( c_1 \) and \( c_2 \), individual contributions to the log-likelihood \( L_r \) can be written as
where $\delta_i$ and $\delta_j$ are indicators whether patients $i$ and $j$, respectively, died ($\delta_i = \delta_j = 1$) or were censored ($\delta_i = \delta_j = 0$) in the study period; and $c_2i$ is an indicator whether patient $i$ was diagnosed with a subsequent cancer $c2$ ($c_{2i} = 1$) or not ($c_{2i} = 0$) in the study period.

### 6.3.2 Semi-parametric models for cancer-specific survival

In the analysis of cause-specific survival of patients with multiple cancers, analogously to that of single-cancer patients, a death from cancer should be attributed to a first tumour $cf$ or a subsequent tumour $c2$. Given that the official cause of death exists, such a distinction is usually possible, i.e, cancer-specific (tumour-specific) causes of death can be derived. For patients with multiple cancers of the same site, however, this may not be possible. The analysis of cancer-specific survival of patients with multiple cancers should thus be possible in two alternative ways depending on whether the cancer-specific causes of death are known or not.

The concept of cancer-specific survival is liable to misunderstanding especially when data consist of patients with multiple cancers of the same primary site. The concepts of cancer-specific (tumour-specific) and $c1$- and $c2$-specific survival were therefore introduced to refer to survival from cancers $c1$ and $c2$, respectively.

#### Cancer-specific causes of death known (Cancer-specific survival model 1)

If a death from cancer can be distinguished between a first and subsequent tumour, individual contributions to the log-likelihood $L_{cs1}$ can be written as

$$
L_{cs1} = \sum_{i=1}^{N_0} [-\Lambda_{0i} + \delta_{0i}\log(\lambda_{0i}) + c_{2i}(-\Lambda_{2i} + \delta_{2i}\log(\lambda_{2i}))] + \sum_{j=1}^{N_1} [-\Lambda_{1j} + \delta_{1j}\log(\lambda_{1j})]
$$

where $\Lambda_{0i}$ and $\Lambda_{1j}$ are the cumulative hazard functions for the time to death from the first and subsequent tumours, respectively. $\delta_{0i}$ and $\delta_{2i}$ are indicators of death and diagnosis with a subsequent cancer, respectively. $\lambda_{0i}$ and $\lambda_{2i}$ are the hazard rates for death from the first and subsequent tumours, respectively.
where $\delta_{vp}$ indicates whether patient $p, p = i, j$ died from his or her cancer $cv$ in the study period ($\delta_{vp}=1$) or not ($\delta_{vp}=0$).

Cancer-specific causes of death not known (Cancer-specific survival model 2)

If a death from cancer cannot be distinguished between a first and subsequent tumour, or one is not willing to do so, the individual contributions to the log-likelihood $L_{cs2}$ of the model can be written as

\[(8) \quad L_{cs2} = \sum_{i=1}^{N_0} [-\Lambda_0i - c_{2i}\Lambda_2i + \delta_{ci} \log(\lambda_0i + c_{2i}\lambda_2i)] + \sum_{j=1}^{N_1} [-\Lambda_{1j} + \delta_{1j} \log(\lambda_{1j})],\]

where $\delta_{ci}$ indicates whether patient $i$ died ($\delta_{ci}=1$) either from his or her cancers, $c_0$ or $c_2$, in the study period or was censored ($\delta_{ci}=0$); and $\delta_{1j}$ indicates whether patient $j$ died ($\delta_{1j}=1$) from his or her cancer $c_1$ in the study period or was censored ($\delta_{1j}=0$).

6.3.3 Parametric mixture model for relative survival

Survival from cancer $S_v(t_{vi}, z_i), v=0,1,2$, can be defined for patient $i$ using equation (4). The individual contributions to the log-likelihood $L_p$ of the model can then be written as

\[(9) \quad L_p = \sum_{i=1}^{N_0} [\log(S_{0i}) + c_{2i}\log(S_{2i}) + \delta_{i} \log(\lambda_{ei} + \lambda_0i + c_{2i}\lambda_2i)] + \sum_{j=1}^{N_1} [\log(S_{1j}) + \delta_{j} \log(\lambda_{ej} + \lambda_{1j})].\]

For the corresponding survival analysis of patients with multiple cancers of the same site, an underlying cancer $c0$ should be renamed as cancer $c1$ and individuals $j$ should be excluded.

6.4 Implementation of the models

The $-\log$-likelihoods $L_{cs1}$ and $L_{cs2}$ of the models (7) and (8) were minimized using the ms() -function in Splus5 [27] under Linux, and $L_r$ and $L_p$ of the models (6) and (9) with the CML (Constrained Maximum Likelihood) function in Gauss 4.0 [28]. For models (7) and (8) the approximative standard errors of
the parameters were calculated from the inverse of the observed information matrices. The first and second derivatives of $L_{cs1}$ and $L_{cs2}$ are presented in the Appendix.

In some situations, especially when data consist of patients with multiple cancers at different sites, the simultaneous estimation of all parameters related to cancers $cv, v=0,1,2$ can be somewhat unnecessary and overly time-consuming. Often the parameter estimates related to the underlying first cancer $c0$ are not of primary interest and do not change whether they are estimated simultaneously with the other parameters or not. One may therefore first estimate the parameters related to the first underlying cancer $c0$, calculate model-based hazards $\hat{\lambda}_0i$ and $\hat{\Lambda}_0i$ for each patient $i$ and regard these hazards fixed when parameters related to cancers $c1$ and $c2$ are estimated simultaneously. For the estimation of the parameters related to cancer $c0$ alone, the log-likelihoods (6)-(9) should be modified by excluding the contributions due to cancer $c1$, by censoring the follow-up times $t_{0i}$ at ages $x_{2i} - x_{0i}$ for those with $c2i = 1$ and by setting $c2i = 0$ for all patients $i$. This approach was used in application IV with the log-likelihood (6). The possible drawbacks of this approach are discussed in section 8.1.1.

7 RESULTS

The effect of stage on the hazard $\lambda_f$ was found to be non-proportional with respect to follow-up time $t_f, f=0,1$ in all studies. To allow comparisons between the $\lambda_1$ and $\lambda_2$ of interest, all hazards $\lambda_v$ were assumed to be non-proportional between the stages of cancer $cv$. In I-III, the non-proportionality was handled by stratifying the $\lambda_v$ by stage. In IV, the hazards $\lambda_f, f=0,1$ were non-proportional also between the countries and were modelled by country with time-dependent covariates for the first years of the follow-up after which they were assumed to be proportional.

Survival from cancer $v, v=0,1,2$ is illustrated with model-based relative or cancer-specific survival $\hat{S}_v$ by stage. Survival from cancers $cf, f=0,1$ and $c2$ together is illustrated with model-based overall relative or cancer-specific survival $\hat{S}_{f2}, f=0,1$, by stage.
The data on Finnish female patients with one or two primary breast cancers were analysed with the cancer-specific survival models 1 (7) and 2 (8) and the relative survival model (6). The comparisons between the cancer-specific survival models 1 and 2 showed that death from breast cancer, either from the first or subsequent tumour, provided sufficient information to enable estimation of the $c_1$- and $c_2$-specific (cancer-specific) survival (II). This conclusion was based on the comparison between the values of the log-likelihoods and the Akaike Information Criterion.

Cancer-specific survival models 1 and 2, and the relative survival model gave similar estimates for survival from breast cancer as a first and subsequent tumour with few exceptions. Model-based survival from the subsequent breast cancer based on cancer-specific survival model 1 tended to be lower than those based on the other two models, and for the oldest patients (70+) with the first breast cancer, model-based relative survival was higher than the corresponding cancer-specific survival.

*Figure 1* Model-based survival from breast cancer as a first and subsequent tumour by stage in age groups 50-59 and 70+ in calendar year 1992. Survival from the first breast cancer is estimated with relative and cancer-specific survival model 1 and that from the subsequent breast cancer with relative and cancer-specific survival models 1 and 2.
Figure 1 illustrates model-based estimates of survival from breast cancer as the first and subsequent tumour by stage in age group 50-59 and 70+ for calendar year 1992. For the first breast cancer, model-based cancer-specific survival is practically identical between models 1 and 2, and thus only estimates based on cancer-specific survival model 1 are shown.

All three models (6-8) lead to the conclusion that survival from subsequent breast cancer was different from that from the first breast cancer (I,II). The quantification of the difference was, however, difficult since the effect of follow-up time on the stage-specific hazard $\lambda_v$ was non-proportional between the first and subsequent breast cancer, $v=1,2$, especially for patients with non-localized cancer.

In the parametric mixture model, model-based overall relative survival (and proportions cured) of multiple-cancer patients are products of model-based relative survival of (and proportions cured from) the first and subsequent cancer. The left-hand side of Figure 2 illustrates survival from localized cancer in the form of first lung cancer $c_1$, as a first colorectal cancer $c_0$, and as a first colorectal cancer $c_0$ and a second lung cancer $c_2$ together, subsequent being diagnosed 2 or 5 years apart. The left-hand side also shows how the hazard of dying from the underlying cancer $c_0$ diminishes and disappears with follow-up time whereafter the long-term model-based relative survival of single- and multiple-cancer patients become practically identical. The right-hand side of Figure 2 illustrates the need for the adjustment: survival from subsequent cancer $c_2$ should not be addressed unless it has not been adjusted for the hazard of the underlying first cancer $c_0$. Model-based relative survival of multiple-cancer patients ($c_0\&c_2$, 2 and 5 years apart) can be considered unadjusted or overall model-based relative survival for subsequent cancer $c_2$ whereas survival from cancer ($c_2$) refers to adjusted survival. Survival from lung cancer as a first and subsequent tumour was not found to be different but there was suggestive evidence that survival from subsequent cancer could be higher than that from the corresponding first cancer (III). The right-hand column of Figure 2 shows that the unadjusted survival from subsequent cancer is lower than the adjusted one.
Figure 2  Model-based survival from the first localized colorectal cancer (c0), the first localized lung cancer (c1), the subsequent localized lung cancer (c2), and the first localized colorectal cancer c0 and the subsequent localized lung cancer c2 together, subsequent being diagnosed 2 (c0&c2, 2 years apart) or 5 years (c0&c2, 5 years apart) after the first colorectal cancer. In the plots on the left-hand side, follow-up time is given since the diagnosis of the first cancer (c0 or c1) and in the plots on the right-hand side, since the diagnosis of lung cancer (c1 or c2). Model-based survival has been calculated with the parametric mixture model for Finnish male patients diagnosed with their cancer (c0, c1 and c2) between ages 60-69 during the calendar period 1983-96.
The need for adjustment for the underlying first cancer $c_2$ is equally evident when semi-parametric models are used. Figure 3 shows model-based relative survival for prostate cancer in Norway as a first cancer $c_1$, as a subsequent cancer $c_2$ after colorectal cancer $c_0$ ($c_2$, adjusted), and as first colorectal $c_0$ and subsequent prostate cancer $c_2$ together ($c_0&c_2$, unadjusted) with diagnoses 3 years apart.

![Figure 3](image)

**Figure 3** Model-based survival from the first localized prostate cancer ($c_1$), from the subsequent localized prostate cancer ($c_2$), and from the first localized colorectal cancer $c_0$ and the subsequent localized prostate cancer $c_2$ ($c_0&c_2$) together, subsequent being diagnosed 3 years after the first colorectal cancer. Model-based survival is calculated with the relative survival model using Norwegian data on patients diagnosed with their prostate cancer between ages 70-72 in calendar period 1985-97.

The effect of time interval between the diagnoses of cancers $c_f, f=0,1$ and $c_2$ on the $\lambda_2$ was tested in all studies. It was found to be clinically small and statistically non-significant at the 5% level in all studies and was therefore usually excluded from the final analyses (I-III). In other words, the additional risk for death from subsequent cancer was not affected by the time interval since the underlying first cancer. When the effect of follow-up time $t_v$ on $\lambda_v$ was proportional between the cancers $c_1$ and $c_2$ (III-IV), the effect of tumour rank on the $\lambda_2$ was tested. It was also clinically small and statistically non-significant at the 5% level in both studies.
8 DISCUSSION

Stage-specific survival from breast cancer was different between the first and subsequent cancer (I,II). In the studies of patients with multiple cancers at different sites (III,IV), stage-specific survival from cancer as a first and subsequent tumour was not found to be different. One reason for the lack of a difference in survival in these latter studies could be the lack of statistical power. The data on female patients with multiple breast cancers were the largest Finnish data with respect to the number of subsequent primary cancers. For breast cancer (I-II), survival comparisons were also based on within-patient variation whereas the other comparisons were based on between-patient variation (III,IV). One of the aims of the Nordic study (IV) was the pooling of Danish, Finnish and Norwegian data to enable derivation of clinically meaningful results with adequate statistical precision. However, the hazards of the first prostate and colorectal cancer were non-proportional between the stages of cancer and between the countries, so direct pooling could not be done.

8.1 Comparison between the survival models

8.1.1 Sankila and Hakulinen’s method vs. new survival models

All the log-likelihoods of the new models (6)-(9) were based on individual patient data. With such data, proper testing of the equality and proportionality of the hazards was possible and all subsequent primary cancers could be included in the analysis. In the method by Sankila and Hakulinen [4, 10] based on grouped data, survival from an underlying first cancer is considered fixed and thus proper testing of alternative hypotheses is not possible for reasons discussed later in this chapter. In practice the effect of fixing depends on whether multiple-cancer patients have their cancers at the same site or not. If multiple-cancer patients’ cancers are at different sites, survival from an underlying cancer $c_0$ is usually not of interest. Thus, given that survival from cancer $c_0$ is modelled adequately, the conclusions on survival from cancer of site $c$ are unlikely to be affected by the fixed effect of the underlying first cancer $c_0$. However, if multiple-cancer patients’ cancers are at the same site, a case of special interest here (I-II), the ability to appropriately test hypotheses of interest is essential.
The method proposed by Sankila and Hakulinen [4, 10] could be considered somewhat impractical and inflexible. The estimation of the life-table estimates can be time-consuming and needs to be repeated with each change in the distribution of prognostic factors. Furthermore, if data on subsequent primary cancer are sparse with respect to the distribution of prognostic factors, as they are likely to be, life-table estimates can be unstable or non-consistent leading to exclusion of some valuable subsequent primary cancers or to widening of the categories of prognostic factors.

The use of individual patient data, on the other hand, may also have drawbacks. When estimating survival from cancer as a first and subsequent tumour, especially when they are not of the same site, the data may comprise tens of thousands of individuals. If the hazards are also non-proportional and the non-proportionality is handled by stratification, the number of parameters easily multiplies to tens. Maximum likelihood estimation based on large data with tens of parameters may be slow and even impossible due to limitations of the software Splus5 under Linux and/or the computer used. Even in this case, the data can be analysed, alike Sankila and Hakulinen, by fixing the hazard of the underlying first cancer as discussed at the end of chapter 6.3.3 (III). Note, however, that especially in this case, the estimate for $\lambda_2$ can be highly dependent on whether the model for the underlying hazard $\lambda_f, f=0,1$ fits, and the standard errors of the parameters related to $\lambda_2$ can be underestimated. However, since the precision of the parameters related to $\lambda_f$ is high, the size of the underestimation is unlikely to have an effect in practice. For the semi-parametric models (6)-(8), a flexible parametric form for the baseline hazard $\tau$ could probably be used in reducing the number of parameters.

Since the method by Sankila and Hakulinen is semi-parametric, it can also be included in the semi-parametric models when they are compared with the parametric mixture model in the following section.

8.1.2 Semi-parametric models vs parametric mixture model

For the estimation of survival of multiple-cancer patients, the semi-parametric models (6)-(8) (I,II,IV) and the method by Sankila and Hakulinen can generally be considered more useful than the parametric model (model (9) in III). Some of the assumptions behind the parametric mixture model can be quite restrictive and its use is thus limited. The hazard of a Weibull distribution
(III) can be either decreasing ($\gamma < 1$), increasing ($\gamma > 1$) or constant ($\gamma = 1$) with respect to follow-up time. These alternatives cover the majority of the cancers [29] but they do not include, for example, breast cancer for which the lognormal distribution is more convenient [21, 30]. In addition, the assumption of the independence of survival rates, given the values of prognostic factors, can be questionable. On the other hand, when applicable, the parametric mixture model can have advantages over the semi-parametric ones. For example, if the number of subsequent primary cancers is small during some short follow-up time interval, the estimates for the piecewise constant baseline hazards can be unstable (models (6)-(8)) whereas the estimates for $\alpha_v$ and $\gamma_v$ are likely to be more stable given that the assumptions underlying model (9) are appropriate.

8.1.3 Relative survival model vs cause-specific survival model

The estimation of survival from cancer always requires more information than when estimating overall survival: Either the cause of death or an expected mortality hazard should be available for all patients. The analysis of relative survival is often preferred over the cause-specific survival. An expected general population mortality hazard may not, however, be an appropriate choice for $\lambda_c$. As shown by Phillips et al. [18], cancer patients’ hazard of dying from other causes can be different from that of the corresponding general population group. For stomach cancer, for example, the ratio between the hazards (cancer patients/general population) was reported to be 1.4 (95% CI 1.1-1.8). Thus if one does not adjust cancer patients’ hazard of dying from other causes to be higher (or lower) than that of the corresponding general population group, the excess hazard may be overestimated (or underestimated). On the other hand, due to continuous selection of the most robust individuals, the excess hazard may underestimate the cause-specific hazard [31] and thus the cause-specific survival may be lower than the respective relative survival.

In the Finnish data on patients with single and multiple breast cancers, estimates of model-based survival related to the first breast cancer were not generally affected by the method of estimation; only among the oldest patients (at least 70 years at diagnosis) with a first localized breast cancer was model-based relative survival higher than the respective survival based on the cancer-specific survival models (7) and (8). This inequality between the ex-
cess and cause-specific hazards has also been reported earlier for Norwegian breast cancer patients aged at least 70 years [31]. A possible reason for this difference could be that a death is too often coded as being due to cancer, especially among the elderly.

8.1.4 Cancer-specific survival model 1 vs cancer-specific survival model 2

The analysis of cause-specific survival of multiple-cancer patients raised a need for new concepts: Survival from cancer should be distinguished between a first and subsequent tumour. The need was especially evident when survival of patients with one of two primary breast cancers was studied. Cancer- or tumour-specific survival as well as $c_1$ and $c_2$-specific survival were introduced to refer directly to survival from specific cancers or tumours.

If data on cause of death are available, the analysis of cause-specific survival of patients with multiple primary cancers at different sites should be possible analogously to that of patients with single cancer. However, for a corresponding analysis of patients with multiple primary cancers of the same site, the cause of death should be defined to be either a first or subsequent cancer, or non-cancer, i.e., the cancer-specific causes of death must be available (II). These data were fortunately available at the Finnish Cancer Registry but are often not available in other cancer registries. However, the analysis of Finnish data on breast cancer showed that survival from the first and subsequent tumour can be reliably estimated even if data on the cancer-specific causes of death are not available. It can even be that subsequent breast cancer was too often coded as a cause of death (model 1 in Figure 1).

8.2 Survival as a function of prognostic factors

Age at and calendar period of diagnosis and stage of cancer were considered common prognostic factors and used in assessing survival from first and subsequent cancer in all the studies. In addition, time interval between the diagnoses of the first and subsequent cancer and tumour rank were considered as potential prognostic factors for explaining survival from the subsequent cancer. The former was thought to describe a possible association between the cancers and the latter a ratio of the survival rates between a first and subsequent cancer of
the same primary site. The effects of these prognostic factors on subsequent
cancer were clinically small (and statistically non-significant at the 5 % level)
in all the studies (I-IV). Nevertheless, it would be plausible to think that pa-
tients’ multiple cancers are associated with survival, especially if they are of
the same primary site. This association could not be assessed in these studies
or it does not have any relationship with time interval between the diagnoses.
On the other hand, the lack of the association could also be considered as a
consequence of the sufficient modeling of survival from first and subsequent
cancer.

The possible effect of screening with mammography on survival from the first
breast cancer was controlled for by excluding all first breast cancers which
may have been found due to the nationwide screening programme [32]. Since
individual data on the screened women were not available, the exclusion was
based on generalizations and assumptions: All breast cancers diagnosed among
the women invited for screening were assumed to be associated with screening
and the first breast cancers were diagnosed due to the screening programme if
they were diagnosed among the women scheduled to be screened in that year.
Furthermore, it was assumed that the screening programme was performed
similarly in each municipality. The assessment of screening might thus well be
incorrect at an individual level, but at the group level it lead to meaningful
results as interactions between age groups and calendar time were no longer
evident after making this exclusion. One could have also incorporated an addi-
tional covariate into the analysis for describing survival from screen-detected
first breast cancers. Both of these approaches were attempted and lead to
identical model-based survival from the first non-screen detected breast can-
cer.

In the parametric mixture model, prognostic factors were assumed to affect
both survival and proportion cured. In several other parametric mixture mod-
eels, prognostic factors are assumed to affect survival through the scale param-
eter $\alpha_v = \exp(\beta_1 v z_v)$ and the proportion non-cured through the logistic link
function, i.e., $1 - P_v = \exp(\beta_2 v z_v)/(1 + \exp(\beta_2 v z_v))$ [15]. The parametric mix-
ture model applied was suitable for lung cancer with poor survival but could
be too inflexible for other sites of cancer.
8.3 Non-proportionality of the hazards

The hazard $\lambda$ due to cancer is often found to be non-proportional with respect to stage of cancer and/or age of the patient. If the hazard $\lambda_1$ of the first cancer of site $c$ is non-proportional with respect to stage of cancer and meaningful comparisons are to be done, the hazard $\lambda_2$ needs to be similarly non-proportional. If the non-proportionality is accounted for by stratifying with respect to a covariate $z_{v1}$ with non-proportional effects, the number of parameters may increase extensively, especially in applications (I-II) and (IV). The increase in the number of parameters with a simultaneous increase in the number of small strata may lead to problems in the estimation of $\lambda_2$. The number of small strata may affect not only the choice between the semi-parametric and parametric models, discussed in 8.1.2, but also the formulation of the model for $\lambda_2$. If the parameters of $\lambda_2$ are estimated freely, estimation may fail by producing obscure estimates. For this reason, the parameters of $\lambda_2$ were often assumed to be dependent on those of a respective first cancer, that is, that the effects of the common prognostic factors have to be considered to be at least proportionally similar.

The non-proportionality of the hazards may not, however, last for the whole follow-up time (III, [17]). In this case, time-dependent covariates are often needed only for the first few years of the follow-up thus minimizing the problem with large numbers of parameters and small strata. A drawback of this approach is that the number of time-dependent covariates included into a model is likely to be data-driven.

According to Zahl and Tretli [31], the non-proportionality of the $\lambda$ with respect to stage of cancer and age group should be considered in the estimation of long-term survival. For first cancers, the stratification is not a problem but for subsequent cancers it can be, especially when cancer- and country-specific data are used. Since stratification with respect to two or more factors with non-proportional effects will probably lead to problems in the estimation of the parameters related to subsequent cancer, the non-proportionality of the $\lambda_v$ with respect to two prognostic factors was not considered. For the same reason, the classification of stage of cancer was relatively raw, non-localized stage consisting of regional and distant metastases (II-IV). On the other hand, the stratification by stage and age group might not always be necessary. The
analysis of data on lung cancer with a parametric mixture model (III) showed that if the excess hazard due to the first lung cancer was stratified by stage, further stratification by age group did not have much effect on the results and none on the conclusions.

8.4 Proportion cured from cancer as a first and subsequent tumour

In the parametric mixture model (III), a patient is assumed, at the time of diagnosis, to belong to an unknown proportion cured from cancer or to a group who will die from the cancer. In reality, the effectiveness of the cancer treatment, for example, is likely to affect the group to which patient eventually belongs.

The proportions cured from cancer were assessed in the parametric mixture model but not in the other models. The proportions cured from cancer can, however, be considered as limits of the long-term survival from cancer although this interpretation is debated. The improved survival from cancer would lead to an elevated proportion cured even if improved survival was due to an increase in recurrence free survival time among the non-cured [33]. Different alternatives for the relationship between the proportion cured and mean survival time have also been illustrated, for example, by Verdecchia et al. [34]. In addition, the proportion cured from cancer is a limit of the long-term survival from cancer only in a small class of proper distributions, and the class of improper distributions is much wider [35].

8.5 Coding of multiple primary cancers

The conclusions from the results of these studies are dependent on the coding of multiple primary cancers. Even if special attention has been paid to the rules according to which a subsequent cancer is defined a new and independent primary cancer, rather than a metastasis of a previous one, practices are likely to vary. If the coding practice varies between countries, with calendar time, time interval between the diagnoses of consecutive cancers and/or with age of the patient, the comparability of the results may lack confidence. The reliability of the coding of multiple primary cancers could be studied on a ‘case...
by case’-basis but that was not done since the coding as such was not of primary interest in this study. The coding practices of multiple primary cancers were studied, however, by analysing the standardized incidence rates (SIRs) of being diagnosed with a subsequent primary cancer with Poisson regression. For comparisons of SIRs between the countries and with calendar periods of diagnosis, the SIRs of a subsequent primary prostate cancer among the patients with a first colorectal cancer were analysed using the Danish, Finnish and Norwegian data. For comparisons of SIRs between the calendar periods of and age groups at diagnosis, and time intervals between the diagnoses, SIRs of a subsequent primary breast cancer among the patients with a first primary breast cancer were analysed using the Finnish data.

The SIRs of being diagnosed with a subsequent primary prostate cancer after a first primary colorectal cancer are presented for each calendar time period for Denmark, Finland and Norway, respectively, in Table 1. In Finland and Norway, the SIRs were constant with the calendar time periods but they were significantly different between the countries, about 1.19 in Finland and 1.55 in Norway. In Denmark, the SIRs varied significantly with calendar time periods but were, however, at the same level as in Finland in calendar periods 1965-74 and 1985-96. Since the frequency of diagnosing prostate cancer is not known to be affected by any extraneous factor, such as screening, during the study period, there seemed to be a difference in the coding of multiple primary cancers between Norway and the other countries. Consequently, if the coding of multiple primary cancers is not comparable between the countries, one should indeed be careful in comparing survival from subsequent cancer between the countries.

<table>
<thead>
<tr>
<th>Country</th>
<th>OBS (1965-74)</th>
<th>SIR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>42</td>
<td>1.19</td>
<td>(0.86,1.61)</td>
</tr>
<tr>
<td></td>
<td>179</td>
<td>1.67</td>
<td>(1.44,1.93)</td>
</tr>
<tr>
<td></td>
<td>260</td>
<td>1.24</td>
<td>(1.09,1.40)</td>
</tr>
<tr>
<td>Finland</td>
<td>8</td>
<td>1.17</td>
<td>(0.51,2.31)</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>1.21</td>
<td>(0.89,1.62)</td>
</tr>
<tr>
<td></td>
<td>182</td>
<td>1.19</td>
<td>(1.02,1.37)</td>
</tr>
<tr>
<td>Norway</td>
<td>47</td>
<td>2.09</td>
<td>(1.54,2.78)</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>1.52</td>
<td>(1.29,1.79)</td>
</tr>
<tr>
<td></td>
<td>476</td>
<td>1.52</td>
<td>(1.38,1.66)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Country</th>
<th>OBS (1965-74)</th>
<th>SIR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Denmark</td>
<td>42</td>
<td>1.19</td>
<td>(0.86,1.61)</td>
</tr>
<tr>
<td></td>
<td>179</td>
<td>1.67</td>
<td>(1.44,1.93)</td>
</tr>
<tr>
<td></td>
<td>260</td>
<td>1.24</td>
<td>(1.09,1.40)</td>
</tr>
<tr>
<td>Finland</td>
<td>8</td>
<td>1.17</td>
<td>(0.51,2.31)</td>
</tr>
<tr>
<td></td>
<td>46</td>
<td>1.21</td>
<td>(0.89,1.62)</td>
</tr>
<tr>
<td></td>
<td>182</td>
<td>1.19</td>
<td>(1.02,1.37)</td>
</tr>
<tr>
<td>Norway</td>
<td>47</td>
<td>2.09</td>
<td>(1.54,2.78)</td>
</tr>
<tr>
<td></td>
<td>145</td>
<td>1.52</td>
<td>(1.29,1.79)</td>
</tr>
<tr>
<td></td>
<td>476</td>
<td>1.52</td>
<td>(1.38,1.66)</td>
</tr>
</tbody>
</table>
The SIRs of being diagnosed with a subsequent breast cancer after the first primary breast cancer are presented in Table 2. If the time interval between the diagnoses of the first and subsequent breast cancer is short, less than two years, the SIRs were significantly lower than in the other time intervals. This result coincides with the practice of coding multiple primary breast cancers at the Finnish Cancer Registry (II and personal discussion with Professor Lyly Teppo, MD, Chief medical officer at the Finnish Cancer Registry 1972-2001). For patients aged 50-59, the SIRs decreased significantly with calendar time. With these exceptions, the SIRs support the conclusion that the coding of multiple primary cancers has been consistent with calendar period and time interval between the diagnoses of the first and subsequent breast cancer, and increasing with age at diagnosis. The SIRs are known to be highest among the youngest and decreasing with age [36]. Survival from a first and subsequent breast cancer should thus be comparable (I,II) with respect to coding practice of multiple primary cancers.

Table 2: Observed number (OBS) and standardized incidence ratio (SIR), with 95% confidence interval (95%CI), of being diagnosed with subsequent breast cancer among patients with a first breast cancer by age at diagnosis of the first cancer (<50, 50-59, 60-69, 70+ years), time interval between the diagnoses (<24, 24-59, 60-119,120+ months) and calendar period of diagnosis of the first cancer (1968-86, 1987-96).

<table>
<thead>
<tr>
<th>Time interval</th>
<th>Calendar period</th>
<th>1968-86</th>
<th>1987-96</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;24</td>
<td>&lt;50</td>
<td>OBS</td>
<td>SIR</td>
</tr>
<tr>
<td></td>
<td></td>
<td>24</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>50-59</td>
<td>23</td>
<td>2.0</td>
</tr>
<tr>
<td></td>
<td>60-69</td>
<td>19</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td>70+</td>
<td>23</td>
<td>1.1</td>
</tr>
<tr>
<td>24-59</td>
<td>&lt;50</td>
<td>38</td>
<td>8.3</td>
</tr>
<tr>
<td></td>
<td>50-59</td>
<td>47</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>60-69</td>
<td>53</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>70+</td>
<td>45</td>
<td>2.1</td>
</tr>
<tr>
<td>60-119</td>
<td>&lt;50</td>
<td>21</td>
<td>8.8</td>
</tr>
<tr>
<td></td>
<td>50-59</td>
<td>38</td>
<td>3.6</td>
</tr>
<tr>
<td></td>
<td>60-69</td>
<td>35</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>70+</td>
<td>50</td>
<td>2.6</td>
</tr>
<tr>
<td>120+</td>
<td>&lt;50</td>
<td>1</td>
<td>2.7</td>
</tr>
<tr>
<td></td>
<td>50-59</td>
<td>21</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>60-69</td>
<td>27</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>70+</td>
<td>30</td>
<td>3.4</td>
</tr>
</tbody>
</table>
8.6 Conclusions

Survival from subsequent cancer should be examined only if survival from the underlying first cancer is controlled for. The overall survival of multiple-cancer patients with respect to both cancers together will generally be worse than that of single-cancer patients with either of these cancers as a first tumour. The conclusions whether survival from cancer as a first and subsequent tumour are really different (I-II) or equal (III,IV) should also be confirmed later with a further increase in numbers of subsequent primary cancers. Such confirmation can be possible, for example, if the international project EUROCARE [37] is enlarged to consider survival of multiple-cancer patients.

Further development of the models for estimating survival from cancer as a first and subsequent tumour would be worth considering. For the semi-parametric models (I-II,IV), an alternative parametric formulation of the $\lambda_v$ could be useful. The measures for assessing the possible association with survival between the first and subsequent cancer should also be improved.

With the current increase in the numbers of subsequent cancers, models for estimating survival of multiple cancer patients are of growing need and importance. With the concepts introduced and models developed in this thesis, survival of patients with multiple cancers can be assessed and estimated with respect to a first and subsequent tumour. The use of individual data on single- and multiple-cancer patients enables analyses where survival from cancer as a first or subsequent tumour can be estimated simultaneously. There are only few assumptions underlying the first three models (6)-(8) and thus these models are generally applicable for any site of cancer. If needed, the presented models can be modified for estimating observed survival and are also expandable for estimating survival of patients with more than two primary cancers.
References


9 APPENDIX

THE FIRST AND SECOND DERIVATIVES OF THE LOG-LIKELIHOODS
L_{cs1} AND L_{cs2} OF MODELS (7) AND (8)

In the notations $R_{vk} = \exp(\beta_v z_{vm})$, $R_{vk} = \exp(\beta_v z_{vm})$, $v=0,1,2$, $m=i,j$; indices $a,b = 1, \ldots, r$ stand for the elements of the parameter vectors $\beta_v$, and $k,h =1,\ldots,m_v$ for the follow-up intervals.

**Cancer-specific survival model 1 (model (7))**

Cancer-specific cause of death known

The first derivatives of $L_{cs1}(\beta_0, \beta_1, \beta_2, \tau_0, \tau_1, \tau_2)$ are

\[ \frac{\partial L_{cs1}}{\partial \beta_{0a}} = \sum_{i=1}^{N_0} z_{0ia}(-\Lambda_{0i} + \delta_{0i}), \quad \frac{\partial L_{cs1}}{\partial \tau_{0k}} = \sum_{i=1}^{N_0} R_{0i}(-t_{0ki} + \frac{\delta_{0i} I_{0k}(t_{0i})}{\lambda_{0i}}), \]

\[ \frac{\partial L_{cs1}}{\partial \beta_{2a}} = \sum_{i=1}^{N_0} c_{2i} z_{2ia}(-\Lambda_{2i} + \delta_{2i}), \quad \frac{\partial L_{cs1}}{\partial \tau_{2k}} = \sum_{i=1}^{N_0} c_{2i} R_{2i}(-t_{2ki} + \frac{\delta_{2i} I_{2k}(t_{2i})}{\lambda_{2i}}), \]

\[ \frac{\partial L_{cs1}}{\partial \beta_{1a}} = \sum_{j=1}^{N_1} z_{1ja}(-\Lambda_{1j} + \delta_{1j}), \quad \frac{\partial L_{cs1}}{\partial \tau_{1k}} = \sum_{j=1}^{N_1} R_{1j}(-t_{1kj} + \frac{\delta_{1j} I_{1k}(t_{1j})}{\lambda_{1j}}). \]

The second derivatives of $L_{cs1}$ are

\[ \frac{\partial^2 L_{cs1}}{\partial \beta_{0a} \partial \beta_{0b}} = \sum_{i=1}^{N_0} -z_{0ia} z_{0ib} \Lambda_{0i}, \quad \frac{\partial^2 L_{cs1}}{\partial \beta_{0a} \partial \tau_{0k}} = \sum_{i=1}^{N_0} -z_{0ia} R_{0i} t_{0ki}, \]

\[ \frac{\partial^2 L_{cs1}}{\partial \tau_{0k} \partial \tau_{0h}} = \sum_{i=1}^{N_0} -\delta_{0i} I_{0k}(t_{0i}) I_{0h}(t_{0i}) R_{0i}^2, \]

\[ \frac{\partial^2 L_{cs1}}{\partial \beta_{0a} \partial \beta_{1b}} = \sum_{i=1}^{N_0} -c_{2i} z_{2ia} z_{2ib} \Lambda_{2i}, \quad \frac{\partial^2 L_{cs1}}{\partial \beta_{2a} \partial \tau_{2k}} = \sum_{i=1}^{N_0} -c_{2i} z_{2ia} R_{2i} t_{2ki}, \]

\[ \frac{\partial^2 L_{cs1}}{\partial \tau_{2k} \partial \tau_{2h}} = \sum_{i=1}^{N_0} -c_{2i} \delta_{2i} I_{2k}(t_{2i}) I_{2h}(t_{2i}) R_{2i}^2, \]

\[ \frac{\partial^2 L_{cs1}}{\partial \tau_{2k} \partial \tau_{1k}} = \sum_{i=1}^{N_0} \lambda_{2i}^2, \quad \frac{\partial^2 L_{cs1}}{\partial \tau_{2k} \partial \tau_{1h}} = \sum_{i=1}^{N_0} \lambda_{2i}^2, \]

\[ \frac{\partial^2 L_{cs1}}{\partial \beta_{2a} \partial \beta_{1b}} = \sum_{i=1}^{N_0} -z_{1ja} z_{1jb} \Lambda_{1j}, \quad \frac{\partial^2 L_{cs1}}{\partial \beta_{1a} \partial \tau_{1k}} = \sum_{j=1}^{N_1} -z_{1ja} R_{1j} t_{1kj}, \]

\[ \frac{\partial^2 L_{cs1}}{\partial \tau_{1k} \partial \tau_{1h}} = \sum_{j=1}^{N_1} \lambda_{1j}^2. \]
The second derivatives of $L_{cs2}$ (model (8))

Cancer-specific survival model 2 (model (8))

Cancer-specific cause of death not known

The first derivatives of $L_{cs2}(\beta_0, \beta_1, \beta_2, \tau_0, \tau_1, \tau_2)$ shown in (8) are

\[
\frac{\partial L_{cs2}}{\partial \beta_0} = \sum_{i=1}^{N_0} z_{0ia}(-\Lambda_{0i} + \frac{\delta_{ci} \lambda_{0i}}{\lambda_{0i} + c_{2i} \lambda_{2i}}), \quad \frac{\partial L_{cs2}}{\partial \beta_1} = \sum_{i=1}^{N_0} R_{0i}(-t_{0ki} + \frac{\delta_{ci} I_{0k}(t_{0i})}{\lambda_{1i} + c_{2i} \lambda_{2i}}),
\]

\[
\frac{\partial L_{cs2}}{\partial \beta_2} = \sum_{i=1}^{N_0} c_{2i} z_{2ia}(-\Lambda_{2i} + \frac{\delta_{ci} \lambda_{2i}}{\lambda_{0i} + \lambda_{2i}}), \quad \frac{\partial L_{cs2}}{\partial \tau_0} = \sum_{i=1}^{N_0} c_{2i} R_{2i}(-t_{2ki} + \frac{\delta_{ci} I_{2k}(t_{2i})}{\lambda_{0i} + \lambda_{2i}}),
\]

\[
\frac{\partial L_{cs2}}{\partial \tau_1} = \sum_{j=1}^{N_1} z_{1ja}(-\Lambda_{1j} + \delta_{1j}), \quad \text{and} \quad \frac{\partial L_{cs2}}{\partial \tau_1} = \sum_{j=1}^{N_1} R_{1j}(-t_{1kj} + \frac{\delta_{1j} I_{1k}(t_{1j})}{\lambda_{1j}}).
\]

The second derivatives of $L_{cs2}$ are

\[
\frac{\partial^2 L_{cs2}}{\partial \beta_0 \partial \beta_0} = \sum_{i=1}^{N_0} z_{0ia} z_{0ib}(-\Lambda_{0i} + \frac{c_{2i} \delta_{ci} \lambda_{0i} \lambda_{2i}}{\lambda_{0i} + (c_{2i} \lambda_{2i})^2}),
\]

\[
\frac{\partial^2 L_{cs2}}{\partial \beta_0 \partial \tau_0} = \sum_{i=1}^{N_0} z_{0ia} R_{0i}(-t_{0ki} + \frac{c_{2i} \delta_{ci} I_{0k}(t_{0i}) \lambda_{2i}}{\lambda_{0i} + \lambda_{2i}^2}),
\]

\[
\frac{\partial^2 L_{cs2}}{\partial \tau_0 \partial \tau_0} = \sum_{i=1}^{N_0} \frac{-\delta_{ci} I_{0k}(t_{0i}) \delta_{ci} I_{0k}(t_{0i}) R_{0i}^2}{\lambda_{0i}^2 + (c_{2i} \lambda_{2i})^2}, \quad \frac{\partial^2 L_{cs2}}{\partial \beta_0 \partial \beta_2} = \sum_{i=1}^{N_0} \frac{-c_{2i} \delta_{ci} z_{2ia} z_{2ib} \lambda_{0i} \lambda_{2i}}{\lambda_{0i} + \lambda_{2i}^2},
\]

\[
\frac{\partial^2 L_{cs2}}{\partial \tau_0 \partial \tau_2} = \sum_{i=1}^{N_0} \frac{-c_{2i} \delta_{ci} I_{0k}(t_{0i}) I_{2k}(t_{2i}) R_{0i} R_{2i}}{\lambda_{0i} + \lambda_{2i}^2}, \quad \frac{\partial^2 L_{cs2}}{\partial \tau_2 \partial \tau_2} = \sum_{i=1}^{N_0} \frac{\delta_{ci} I_{2k}(t_{2i}) I_{2k}(t_{2i}) R_{2i}^2}{\lambda_{0i}^2 + \lambda_{2i}^2},
\]

\[
\frac{\partial^2 L_{cs2}}{\partial \tau_1 \partial \tau_1} = \frac{\partial^2 L_{cs2}}{\partial \tau_1 \partial \tau_1} = \frac{\partial^2 L_{cs2}}{\partial \tau_1 \partial \tau_1} = 0,
\]

\[
\frac{\partial^2 L_{cs2}}{\partial \beta_0 \partial \beta_1} = \frac{\partial^2 L_{cs2}}{\partial \tau_0 \partial \tau_1} = \frac{\partial^2 L_{cs2}}{\partial \tau_2 \partial \tau_1} = \frac{\partial^2 L_{cs2}}{\partial \tau_2 \partial \tau_1} = 0,
\]

\[
\frac{\partial^2 L_{cs2}}{\partial \beta_1 \partial \beta_1} = \sum_{j=1}^{N_1} z_{1ja} z_{1jb} \Lambda_{1j}, \quad \frac{\partial^2 L_{cs2}}{\partial \beta_1 \partial \beta_1} = \sum_{j=1}^{N_1} z_{1ja} R_{1j} t_{1kj}, \quad \text{and}
\]

\[
\frac{\partial^2 L_{cs2}}{\partial \tau_1 \partial \tau_1} = \sum_{j=1}^{N_1} \frac{\delta_{1j} I_{1k}(t_{1j}) I_{1k}(t_{1j}) R_{1j}^2}{\lambda_{1j}^2}.
\]