Department of Public Health
University of Helsinki,
Finland

Mitigating bias and dealing with multiple time scales in cohort studies

Studying medications and complications of diabetes

Anna But

ACADEMIC DISSERTATION

To be presented, with the permission of the Faculty of Medicine of the University of Helsinki, for public examination in Auditorium XII, University main building, Unioninkatu 3, on 30 November 2018, at 12 noon.

Helsinki 2018

Dissertationes Scholae Doctoralis Ad Sanitatem Investigandam Universitatis Helsinkiensis

Doctoral Programme in Population Health (DocPop)

Supervisors

Docent Jari Haukka
Clinicum, Department of Public Health, University of Helsinki, Finland

Docent Tommi Härkänen
Department of Public Health Solutions, National Institute for Health and Welfare, Finland

Reviewers

Professor Kari Auranen
Department of Mathematics and Statistics, University of Turku, Finland

Docent Maarit Korhonen
Institute of Biomedicine, University of Turku, Finland

Opponent
Professor Paul Dickman
Karolinska Institutet, Sweden

ISBN 978-951-51-4718-9 (paperback)
ISBN 978-951-51-4719-6 (PDF)

ISSN 2342-3161 (paperback)
ISSN 2342-317X (online)

Unigrafia
Helsinki 2018

'*If you don't know where you are going any road can take you there*.'

Lewis Carroll, Alice in Wonderland

Contents

Abstract

Cohort studies are an important and powerful tool of epidemiologic research. When based on a representative cohort, observational cohort studies provide results of a high external validity given that the internal validity is not impaired by bias. Bias can be introduced at any stage of research and there are numerous sources of bias. Pharmacoepidemiological observational studies are often threaten by selection bias, time-related biases and bias by confounding. Bias, however, can be avoided or mitigated by using appropriate research methods.

Diabetes and cancer represent two prevalent, complex, diverse and potentially fatal chronic diseases. Among individuals with diabetes, cancer occurs more often than could be expected by chance only. Cancer and diabetes share common risk factors, such as obesity and smoking. Diabetes is characterized by hyperinsulinemia, hyperglycemia and inflammation, which may favour the development and / or progression of cancer. In addition, antidiabetic medications may contribute to the association between diabetes and cancer. The empirical part of this thesis comprised two pharmacoepidemiological observational cohort studies (Studies I and II) which were conducted retrospectively to address the relationship between the use of antidiabetic medications and cancer risk.

Due to their longitudinal nature, cohort studies involve at least one time scale and, therefore, allows for studying time-dependent dynamics of a phenomenon. There is often more than one relevant time scale, for instance, the risk of long-term complications of diabetes may vary with age, duration of diabetes and calendar time. However, the traditional statistical methods of survival analysis, such as Cox proportional hazards model, rely on a single time scale. In the methodological part of this thesis (Studies III and IV), I addressed the issue of multiple time scales in cohort studies.

In Study I, I studied the risk of cancer in 23 394 individuals from the National FINRISK cohorts that were linked to the register data on prescriptions (Prescription Register), cancer (Finnish Cancer registry) and death (Statistics Finland). Prevalent users of antidiabetic medication and those with history of cancer at baseline were excluded. I assessed the variation of the cancer risk along time since initiation of anti-diabetic medication, when controlling for several potential confounders, including smoking and body mass index. After a median follow-up period of 9 years, 1081 individuals were diagnosed with cancer, of which 53 in 1301 users of antidiabetic medication. After adjustment for potential confounders, there was no association between the cancer risk and use of antidiabetic medication. However, the small number of cancer cases among users precluded firm conclusions.

In Study II on the CARING (CAncer Risk and INsulin Analogues) five-country (Denmark, Finland, Norway, Sweden, UK) cohort of 327 112 new insulin users identified from the national prescription registers, the risk of ten site-specific cancers and any cancer was scrutinized by contrasting the cumulative exposures to human insulin and insulin analogues glargine and detemir. A particular emphasis of this work was on mitigating biases involved in previous observational studies. During a median follow-up of 3.7 years, 21 390 individuals were diagnosed with cancer. We found no evidence of consistent differences in the studied risks as assessed for insulin glargine or insulin detemir use relative to that of

human insulin. The results of this study are of particular clinical relevance because they imply that none of the studied insulin treatments should be preferred over others as being safer with respect to cancer risk.

In Study III, I addressed the issue of multiple time scales by introducing and evaluating a nonparametric Bayesian model for estimation of intensity on two time scales jointly. Evaluation of the method using simulated data demonstrated its superiority over two other methods. A better performance of the model arises from the flexibility, which is attributable to both Bayesian and nonparametric approaches. In addition, even with the limited data, the model yields accurate results due to the built-in smoothing and borrowing of strength in two dimensions.

In Study IV, I used the Bayesian model to explore the time-dependent dynamics of end-stage-renal-disease and death without end-stage-renal disease in 11 810 individuals with type 1 diabetes from the nationwide FinDM study, which is aimed at monitoring the incidence and prevalence of diabetes and its complications in Finland. I modelled the time-dependent dynamics of these outcomes on two and three time scales jointly, including age, diabetes duration and calendar time. I demonstrated that the two-dimensional Bayesian model can be easily extended to the model allowing for the multiplicativity of time-scale-specific hazards and to the model incorporating more than two time scales. These models can be used to address both empirical and methodological questions. To facilitate the interpretation of results, I used informative graphical outputs, such as surface plots and heatmaps, which illustrate the overall time-dependent dynamics at one glance but also allow for scanning patterns.

Abstrakti

Kohorttitutkimukset ovat epidemiologisen tutkimuksen tärkeä ja tehokas väline. Edustavaan otokseen perustuvassa havainnoivassa kohorttitutkimuksessa saatujen tulosten ulkoinen luotettavuus on korkea edellyttäen, ettei sisäinen luotettavuus ole harhan heikentämä. Harha voi syntyä missä tahansa tutkimuksen vaiheessa ja on olemassa lukuisia harhan lähteitä. Lääke-epidemiologiassa useat harhat, kuten valintaharha, aikaan liittyvät harhat ja sekoittuneisuus, uhkaavat havaintotutkimuksen sisäistä luotettavuutta. Harha voidaan kuitenkin estää tai vähentää oikeilla tutkimusmenetelmillä.

Diabetes ja syöpä ovat kaksi yleistä, kompleksista, monimuotoista ja mahdollisesti hengenvaarallista kroonista sairautta. Diabetesta sairastavilla henkilöillä syöpä esiintyy odotettua useammin. Syövällä ja diabeteksella on yhteisiä riskitekijöitä, kuten liikalihavuus ja tupakointi. Diabetekselle ovat ominaisia hyperinsulinemia, hyperglykemia ja tulehdus, jotka voivat edistää syövän kehittymistä ja / tai etenemistä. Lisäksi diabeteslääkkeiden käyttö saattaa selittää diabeteksen ja syövän välistä yhteyttä. Tämän väitöskirjan empiirinen osa koostui kahdesta havainnoivasta lääke-epidemiologisesta kohorttitutkimuksesta (Työt I ja II), jotka tehtiin retrospektiivisesti tarkastellakseen diabeteslääkkeiden käytön ja syöpäriskin välistä suhdetta.

Kohorttitutkimukset perustuvat seurantaan ja näin niihin liittyy vähintään yksi aikaskaala, jolla voidaan tutkia ilmiön ajasta riippuvaa dynamiikkaa. Usein ilmiöön liittyy useampi kuin yksi relevantti aikaskaala, esimerkiksi diabeteksen pitkäaikaiskomplikaatioiden riski voi vaihdella iän, sairauden keston ja kalenteriajan mukaan. Perinteiset elinaika-analyysin menetelmät, kuten Coxin suhteellisten hasardien malli, perustuvat yhteen skaalaan. Väitöskirjan menetelmällisessä osassa (Työt III ja IV) käsittelin kohorttitutkimukselle ominaista aikaan liittyvää moniulotteisuutta.

Työssä I tarkastelin syöpäriskiä 23 394 yksilön kohortissa, joka perustui kansallisiin FINRISK-kohortteihin ja oli yhdistetty syöpä- (Suomen syöpärekisteri) ja kuolematietoihin (Tilastokeskuksen kuolemansyyrekisteri) ja diabeteslääkitystä koskeviin tietoihin (KELA:n lääkekorvausrekisteri). Poissulkukriteereinä olivat aiempi diabeteslääkkeiden käyttö ja aiemmin sairastettu syöpä. Tarkastelin syöpäriskiä suhteessa aikaan diabeteksen lääkehoidon aloittamisesta samalla ottaen huomioon sekoittavat tekijät, kuten painoindeksi ja tupakointi. Seuranta-ajan mediaani oli 9 vuotta ja tutkimuspopulaatiossa todettiin 1081 syöpätapausta, joista 53 diagnosoitiin niiden 1301 joukossa, jotka aloittivat diabeteslääkkeiden käytön. Kun otettiin huomioon syövän ja diabeteksen yhteiset riskitekijät, mitään yhteyttä ei löytynyt diabeteslääkkeiden ja syöpäriskin välillä. Tulosten perusteella ei kuitenkaan voida tehdä varmoja johtopäätöksiä johtuen syöpätapausten vähäisestä määrästä diabeteslääkkeiden käyttäjien joukossa.

Työssä II viiden maan (Tanska, Suomi, Norja, Ruotsi, Iso-Britannia) CARING (CAncer Risk and INsulin analoGues) rekisteripohjaisessa tutkimuksessa tarkastelin insuliinihoitoa aloittaneen 327 112 yksilön kohortissa kokonaissyöpäriskiä ja kymmenen eri syöpätyypin riskiä suhteessa ajassa kertyvään insuliinialtistukseen ja vertailemalla insuliinianalogi glargiinia ja detemiria ihmisinsuliiniin. Tämän osatyön pääpaino oli aiempien havainnoivien tutkimusten harhojen välttämisessä ja pienentämisessä. Seuranta-ajan mediaani oli 3,7 vuotta, jonka aikana syöpä diagnosoitiin 21 390 yksilössä. Syöpäriskissä

ei havaittu johdonmukaista eroa eri insuliinityyppien välillä. Tutkimuksen tuloksilla on tärkeää käytännön merkitystä, koska kaikki tutkitut insuliinityypit ovat yhtä turvallisia syöpäriskiin suhteen.

Työssä III esitin parametrittoman Bayes-päättelyyn perustuvan mallin intensiteettifunktion estimoimiselle kahdella aikaskaalalla. Arvioin menetelmän toimivuutta soveltamalla malli simuloituun aineistoon. Vertailujen perusteella Bayes-malli osoittautui kahta muuta menetelmää tarkemmaksi. Bayes-mallin parempi toimivuus perustuu sen joustavuuteen, joka on sekä bayesiläisen että parametrittoman lähestymistavan ominaisuuksia. Lisäksi, koska malli perustuu silotukseen ja voiman lainaamiseen kahdessa aikaulottuvuudessa, malli antaa tarkat tulokset myös aineiston ollessa pieni välttäen samalla vääriä, satunnaisuudesta johtuvia positiivisia tuloksia.

Työssä IV tarkastelin ajasta riippuvaa dynamiikkaa loppuvaiheen munuaistaudin ilmaantuvuudessa ja kuolleisuudessa soveltamalla Bayes-malli 11 810 tyypin 1 diabetesta sairastavan henkilön aineistoon. Aineisto pohjautui maanlaajuiseen FinDM tutkimukseen, jonka tavoitteena on diabeteksen ja sen lisäsairauksien esiintyvyyden ja ilmaantuvuuden rekisteripohjainen tutkiminen. Mallinsin kummankin vastemuuttujan ajasta riippuva dynamiikka kahden ja kolmen aikaskaalaan suhteen (ikä, diabeteksen keston, kalenteriaika). Näytin, että kaksiulotteista Bayes-mallia voi laajentaa multiplikatiiviseksi malliksi, jolla on mahdollista mallintaa aikaskaalakohtaiset hasardit, että malliksi, jolla voidaan mallintaa hasardia useammalla kuin kahdella aikaskaalalla. Näin ollen, soveltamalla eri malleja on mahdollista vastata joustavasti sekä empiirisiin että metodologisiin kysymyksiin. Havainnollistin tulokset graafisesti riskipintoina ja lämpökarttoina, jotka sekä antavat kokonaisvaltaisen kuvan ajasta riippuvasta hasardin dynamiikasta että mahdollistavat myös riskiprofiilin tarkastelun.

List of original publications

This thesis is based on the following publications:

I But A, Wang H, Männistö S, Pukkala E, Haukka J. Assessing the effect of treatment duration on the association between anti-diabetic medication and cancer risk. PLoS One 2014; 9(11):e113162.

II But A, De Bruin ML, Bazelier MT, Hjellvik V, Andersen A, Auvinen A, Starup-Linde J, Schmidt MK, Furu K, de Vries F, Karlstad O, Ekström N, Haukka J. Cancer risk among insulin users: comparing analogues with human insulin in the CARING five-country cohort study. Diabetologia 2017; 60(9):1691–1703.

III Härkänen T, But A and Haukka J. Non-parametric Bayesian Intensity Model: Exploring Time-to-Event Data on Two Time Scales. Scandinavian Journal of Statistics 2017; 44(3):798–814.

IV But A, Sund R, Arffman M, Helve J, Finne P, Haukka J, Härkänen T. Bayesian Modelling of Time-To-Event Data on Multiple Time Scales: Exploring the Hazard of End-Stage Renal Disease. Submitted manuscript.

The publications are referred to in the text by their Roman numerals.

1 Introduction

In the developed countries, chronic diseases, such as cardiovascular diseases, cancers, diabetes and chronic respiratory diseases, are considered important public health issues, as these are the leading cause of illness, disability and premature death (OECD/EU 2016). Chronic diseases share common risk factors, such as smoking, obesity, physical inactivity, and, therefore, are likely to coexist (Kivimäki *et al*., 2017; Tu *et al*., 2017; Klil-Drori *et al*., 2017).

Chronic diseases are often long-lasting, have persistent health effects, require continuous treatment and monitoring and induce short- and long-term complications. Use of the appropriate, effective and safe medications plays a central role in avoiding and postponing complications associated with chronic diseases. However, exposure to medications may also be associated with adverse effects, including cancer.

Among individuals with diabetes, cancer occurs more often than could be expected by chance only (Carstensen *et al*., 2012; Carstensen *et al*., 2016). Diabetes is also known for its long-term complications, which include myocardial infarction, stroke and chronic kidney disease (Fowler, 2008; Arffman *et al*., 2014). In some persons with diabetes, chronic kidney disease progress to the end-stage renal disease (ESRD), a life-threatening condition with poor prognosis, requiring treatment by dialysis and kidney transplantation. Given a constantly increasing incidence and prevalence of diabetes (World Health Organization, 2016) and the fact that diabetes burdens both an individual and populations, treatment of diabetes using effective and safe medications and monitoring and preventing its complications are of public health importance.

Cohort studies – studies tracking a group of individuals over time – have been traditionally used for monitoring and studying chronic diseases and their complications, assessing impact of known risk factors and interventions as well as identifying novel risk factors (Brennan *et al*., 2017). Observational pharmacoepidemiological studies have been used to derive evidence about the drug safety issues after drug marketing (Garbe and Suissa, 2014).

When conducted rigorously, observational cohort studies provide a powerful epidemiological tool and add valuable real-world evidence (Concato, 2000). One of the advantages of observational cohort studies is their external validity, particularly when conducted using population-based or nation-wide cohorts (Szklo, 1998). In contrast, internal validity of observational cohort studies has largely been viewed as a common concern (Grimes and Schulz, 2002). Indeed, observational cohort studies are prone to various types of bias, which can occur at any stage of research (Grimes and Schulz, 2002; Delgado-Rodriguez and Llorca, 2004). Observational pharmacoepidemiolgical studies are subject to the specific biases, including prevalent user bias, indication bias and time-related biases (Suissa and Azoulay, 2012).

I will address the limitations and biases involved in previous observational pharmacoepidemiological studies on the association between the use of anti-diabetic medications, particularly long-acting insulins, and cancer risk (Renehan, 2012; Walker *et al*., 2013; Wu *et al*., 2016). I will highlight the importance of using the appropriate methodological and analytical approaches, including the active comparator new-user design

(Ray, 2003, Yoshida *et al*., 2015) and time-varying exposure definition (Zhou *et al*., 2005; Stricker and Stijnen, 2010).

A long follow-up time is a prerequisite for studying the effect of exposure on the outcome with a long latency, such as cancer, or lifetime complications of chronic disease, such as diabetes. Many phenomena exhibit complex time-dependent dynamics, evaluation of which may provide additional insights to the underlying mechanisms.

Data arising in cohort studies are called time-to-event data and involve at least one time scale that is time-on-study. The statistical methods of survival analysis are used to describe and analyse time-to-event data (Kalbfleisch and Prentice, 2002). Time-to-event data from long-term cohort studies often include several relevant time scales, such as age, calendar time, time since diagnosis or initiation of treatment. However, the traditional survival analysis methods for analysis of time-to-event data, such as Cox proportional hazards regression model (Cox 1972), are not suitable for modelling time-to-event data on several time scales jointly.

The Bayesian approach to statistical inference offers a coherent and versatile framework, which has been increasingly used in epidemiological and medical research (Dunson, 2001; Ashby, 2006). I will present the general aspects of using time scales in the analysis of time-to-event data and I will introduce a nonparametric Bayesian model, which allows modelling time-to-event data on two and more time scales jointly. I will also demonstrate the applicability of the model by exploring the time-dependent dynamics of end-stage renal disease and death without it in individuals with type diabetes. By extending the model, I will address both empirical and methodological questions, which may arise in cohort studies with multiple relevant time scales.

Abbreviations

ADM	anti-diabetic medication
AG	Arjas and Gasbarra (prior)
APC	age-period-cohort
ATC	anatomical therapeutic chemical
BMI	body mass index
CARING	CAncer Risk and INsulin analoGues
CI	confidence interval
CKD	chronic kidney disease
CPRD	Clinical Practice Research Datalink
DDD	defined daily dose
DIC	deviance information criterion
DM	diabetes mellitus
ENCePP	European Network of Centres for Pharmacoepidemiology and Pharmacovigilance
ESRD	end-stage renal disease
HR	hazard ratio
ICD	International Classification of Diseases
IR	incidence rate
LOO	leave-one out (cross-validation)
MLE	maximum likelihood estimation
NIADM	non-insulin antidiabetic medication
NHPP	non-homogeneous Poisson process
MCMC	Markov chain Monte Carlo
PP	point process
RCT	randomized control trial
RR	rate ratio
SII	Social Insurance Institution
T1D	type 1 diabetes
T2D	type 2 diabetes
WAIC	Watanabe-Akaike (or widely applicable) information criterion

2 Review of the literature

2.1 Cohort studies

Cohort studies are used to track a group of individuals over time to monitor for changes in their physical, physiological or other characteristic(s) of interest or change in their (health) state as specified by occurrence of the event of interest. In this work, I will focus on the latter case, the defining characteristic which is that the outcome of interest is not present in the individuals at the start of follow-up (Grimes and Schulz, 2002).

In cohort studies, individuals are often selected of those with similar backgrounds or those who experienced a particular event within a certain timeframe. The common determinant can be such as having been born during the same decade (birth cohort), practicing the same profession (cohort of nurses), having been exposed to the same risk factor (cohort of atomic-bomb survivors or nickel refinery workers) or diagnosed with the same disease (cohort of diabetes patients).

Perhaps the most appreciated feature of cohort studies is preserving the chronological order of observations as this allows evaluation of the relationship between exposure to a putative causal factor and the outcome of interest. Therefore, cohort studies provide one of the major investigative approaches of etiological epidemiology (Goldacre, 2001). For instance, cohort studies have played an important role in cancer epidemiology, as these allowed to establish link between the risk of cancer and many occupational, lifestyle and medicinal factors (Breslow and Day, 1987).

2.1.1 Measures of frequency

Assessment of frequencies of outcome of interest, often disease or death, is a major aim of epidemiological research. In cohort studies, the frequency at which the outcome of interest occurs can be quantified by two fundamental measures, incidence rate (IR, force of morbidity, incidence density) and risk (cumulative incidence, average risk) of a given outcome (Benichou and Palta, 2014). The cumulative incidence is calculated as the proportion of individuals of the initially disease-free population, who developed disease or other condition of interest within a stated period of time. The cumulative incidence is non-decreasing and varies between zero and one, and being an overall measure, provides no detailed information on changes that potentially occurred during the studied timeframe. In contrast, the incidence rate always contains a dimension of time as it measures the number of new cases per unit of person-time, which is a combination of the number of people and the amount of time they were followed. The incidence based on person-time expresses the instantaneous rate of change or the pace at which individuals develop the diseases or other condition in the population.

2.1.2 Cohort study designs

2.1.2.1 Prospective and retrospective design

Cohort studies can be divided according to the chronology in collection of follow-up data into prospective and retrospective studies (Doll, 2001a, 2001b). In a prospective cohort study, baseline information is collected from all individuals at the time the study starts, and individuals are then followed up from that point over a period of time to identify new events of interest. The Framingham Heart Study, which began in 1948 and is still ongoing, exemplifies the application of sound, prospective epidemiological design (Mahmood et al. 2014). Another well-known prospective cohort study is the British Doctors Study that has examined the effect of smoking on mortality over a period of decades (Doll et al. 2004).

Retrospective cohort studies, which are also known as historical cohort studies, are conceived after the baseline information was measured in the past, and some individuals have already developed the outcomes of interest. Retrospective cohort studies can be completed relatively fast as compared with prospective cohort studies. Among disadvantages of retrospective cohort studies, is the use of data, collection and quality of which is not under the control of the researcher (Sørensen *et al*., 1996).

Examples of the early retrospective cohort studies in medical research include studies on tuberculosis spread and mortality (Frost, 1933; Morabia and Guthold, 2007), and studies on cancer risk with respect to occupational exposures, such as those involved in chemical and nickel refining industry in the first half of the twentieth century (Breslow and Day, 1987; Doll, 2001b).

2.1.2.2 Experimental and observational design

Cohort studies are used in both experimental and observational research. Randomized clinical trials, also referred as randomized controlled trials (RCT), are experimental by their nature. RCTs are often based on cohort of individuals randomly assigned to the experimental and control groups, which are then followed up prospectively to see if there are any differences between these groups in outcome. Randomization, the corner stone of RCTs, is aimed at the random allocation of exposure (treatment, intervention) and balancing the groups with respect to the important prognostic factors (Concato, 2000). The most fundamental difference between experimental and observational research concerns exposure. Whereas in RCTs exposure is randomly assigned, in observational studies it should be ascertained. In observational research, there is no control on the allocation of exposure and all information is simply recorded (prospective design) or derived from already available records (retrospective design). RCTs are primarily used for demonstrating efficacy of a treatment or intervention, whereas observational cohort studies are usually aimed at assessing association between exposure and outcome over time (Booth and Tannock, 2014).

Both RCTs and observational studies have strengths and limitations (Booth and Tannock, 2014). Irrespectively of design, each study needs to be evaluated in terms of its internal and external validity. The former refers to the ability of study to measure what it set out to, the former refers to the generalizability of results to the target population (Grimes and Schulz, 2002).

On the one hand, RCTs are advantageous over observational studies in terms of internal validity (Booth and Tannock, 2014) because RCTs are most likely to be free of bias as compared to observational studies, which are prone to bias due to their non-experimental nature. On the other hand, the external validity of RCTs is considered to be low because RCTs are usually conducted using highly selected populations (Booth and Tannock, 2014). In contrast, properly conducted observational studies, especially when conducted using population-based cohorts, are considered to be of high external validity (Szklo, 1998, Booth and Tannock, 2014).

There are also other differences between RCTs and observational studies. Observational studies avoid problems of feasibility and ethical aspects, which are involved in RCTs. Evaluation of rare outcome, such as cancer of a specific type, requires a large study population and a long follow-up, which are not affordable with RCTs. In addition, observational design allows to assess multiple exposures and outcomes using the same cohort.

2.1.3 Bias

Bias refers to the presence of systematic error or deviation from the truth, which can yield to the distorted results, undermining therefore internal validity of study (Grimes and Schulz, 2002). Observational cohort studies are subject to various forms of bias, which can occur at any stage of research (Grimes and Schulz, 2002; Sedgwick, 2014a, 2014b; Delgado-Rodriguez and Llorca, 2004). The biases can be classified into selection bias, information bias and bias by confounding (Grimes and Schulz, 2002; Delgado-Rodriguez and Llorca, 2004). In addition, immortal time bias, time-lag and time-window biases, although being of different type, are also referred collectively as time-related biases (Suissa and Azoulay, 2012). Time-related biases are particularly problematic in the observational studies using the data from secondary sources, such as registries and databases.

Selection bias refers to distortions of the relation between exposure and outcome of interest due to the procedures or sources used to select study population or / and due to factors, which influence participation (Rothman and Greenland, 2014). As a result, the study population is not representative of the target population. Selection bias can be produced by an inappropriate definition of the eligible population (ascertainment bias), lack of accuracy of sampling frame or uneven diagnostic procedures in the target population (Delgado-Rodriguez and Llorca, 2004). Among others, selection bias includes healthcare access bias and prevalent user bias.

Information bias can occur due to an inaccurate measurement or, as applied to discrete variables, misclassification of exposure, outcome, or confounding variables, when the availability, measurement, interpretation or definition of the needed information is distorted

(Grimes and Schulz, 2002; Gerhard, 2008; Rothman and Greenland, 2014). Two different types of misclassification are distinguished. Differential misclassification arises when the proportion of individuals misclassified on outcome depends on exposure or vice versa, whereas non-differential misclassification represents an even noise. The effect of information bias depends on its type. Differential misclassification can distort the results in either direction, towards or away null, whereas for non-differential misclassification the direction is usually towards null, although the latter does not apply universally but exceptions occur (Rothman and Greenland, 2014).

Bias by confounding occurs when the relation between an exposure and an outcome is distorted by a third factor, which is associated with both the exposure and the outcome without being an intermediate link in the causal pathway between them (Grimes and Schulz, 2002). Confounding is a problem of non-comparability of groups being studied and leads to mixing or blurring of effects (Pearce and Greenland, 2014).

I will focus on the types of bias, which were addressed in the empirical part of this work. In the following paragraphs, I will give definitions of these biases and outline the methodological and analytical approaches to disentangle and mitigate them. It should be, however, noted that the classification and definitions of these biases are not always consistent across the research areas and study designs. I will provide the definitions relevant to the field of pharmacoepidemiology and observational cohort design. The majority of the biases I will describe are specific to the observational pharmacoepidemiological research but some of them, such as healthcare access bias, detection bias and residual confounding, are common in the observational research in general. The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP) provides a methodological guidance in pharmacoepidemiology, which includes also descriptions of many but not all biases encountered in pharmacoepidemiological research (ENCePP, 2018).

2.1.3.1 Healthcare access bias

Healthcare access bias is a selection bias that occurs when a study population is based on healthcare data (hospital discharge registry, primary care records etc.), in which compared groups are represented at differing proportions as compared to the target population (Delgado-Rodriguez and Llorca, 2004). For instance, as compared to non-users, users of a drug can be overrepresented when the study population is selected based on primary care records. According Delgado-Rodriguez and Llorca (2004) healthcare access bias arises when compared groups are drawn from the healthcare organizations of different level (primary, secondary, tertiary care). In addition, some subgroups can be under- or overrepresented because of socioeconomic, cultural, geographical and other factors when these factors are related to the healthcare access. Healthcare access bias can be avoided by selecting the study population based on the nationwide registries.

2.1.3.2 Prevalent user bias

Inclusion of prevalent users can lead to prevalent user bias, often classified as a selection bias (Danaei *et al*., 2012; ENCePP, 2018), leads to a number of forms of bias, including depletion of those susceptible to the outcome of interest, immortal time bias, bias due to uneven presentation of early and late drug effects (Gerhard, 2008; Yoshida *et al*., 2015; ENCePP, 2018). Moreover, the baseline characteristics of prevalent users can be affected by the treatment, distorting the association between the outcome of interest and confounders. Thus, mixing prevalent users and new-users distorts the association between exposure and outcome and may obscure excess harm because of weighting toward continuation of use. Use of new-user design allows to avoid most of the biases involved due to inclusion of prevalent users (Ray, 2003; Yoshida *et al*., 2015). Use of new-user design reduces but does not prevent immortal time bias, which can be eluded only by classifying the follow-up correctly.

2.1.3.3 Detection bias

Detection bias is a form of information bias, also known as surveillance bias, which arises when individuals in one group have a different probability of having the outcome of interest detected (Haut and Pronovost, 2011; ENCePP, 2018). For instance, comparison of users and non-users may be hampered by detection bias. This bias can be mitigated by using active-comparator design (Yoshida *et al*., 2015).

2.1.3.3 Protopathic bias

Protopathic bias, also known as reverse causality, refers to a reversal of cause and effect and occurs when the symptoms treated by a drug are a manifestation of the yet undiagnosed disease (Gerhard, 2008). This type of bias is likely to arise in studies on associations between the drug use and cancer risk (ENCePP, 2018). By studying the variation in the risk of outcome by duration of drug use, the risk patterns attributable to the protopathic bias can be detected (Korhonen *et al*., 2009; Carstensen *et al*., 2012). In the presence of risk patterns, which cannot be attributed to the drug itself, a specific initial period of use should be either excluded by using lag-time (Tamim *et al*., 2007) or separated analytically from the rest follow-up either through stratification or by using time-dependent exposure definition.

2.1.3.5 Immortal time bias

Immortal time bias, sometimes referred to as survival bias or survival treatment bias, occurs due to exclusion or misclassification of the follow-up time between cohort entry and date of first exposure to a drug when the former precedes the latter (Delgado-Rodriguez and Llorca, 2004; Suissa, 2007; Suissa, 2008; ENCePP, 2018). The period between entering the

study and starting medication is called immortal time. During this period, to be classified as exposed, the individual has to remain alive (and event free if the event of interest is other than death) until start of exposure (Suissa, 2007; Suissa, 2008). For example, immortal time bias arises when information on the future exposure is used to classify individuals to users and non-users already at cohort entry. Observational cohort studies with time-based, event-based or exposure-based design comparing users and non-users of a drug are particularly prone to immortal time bias, which hampers the results in favour of the treatment (Suissa, 2008).

Although the immortal time bias was first described in the early 1970s and has been repeatedly highlighted in the scientific publications since then it continues to be overlooked (Glesby and Hoover, 1996; Suissa, 2008; Lévesque *et al*., 2010; Lange and Kielding, 2014). Potential for immortal time bias can be reduced by using new-user design (Yoshida *et al*., 2015). Irrespectively of design, all immortal time should be accounted for (Suissa, 2008). Zhou et al. (2005) studied three different approaches to deal with immortal time bias and found that matching on time-to-treatment and use of a time-dependent exposure definition to be appropriate methods to control for this type of bias.

2.1.3.6 Time-lag bias

Time-lag bias arises when compared treatments are commonly used at the different stages of the disease, for example when the first-line therapy is compared to the second- or third-line therapies (Suissa and Azoulay, 2012). Individuals treated with the second- or third-line therapy are unlikely to be at the same stage of disease as compared to those treated with the first-line therapy. When the risk of outcome under study varies with duration of disease, such a comparison leads to time-lag bias. In the presence of time-lag bias, the results are biased in favour of the first-line therapy as compared to a subsequent one when the risk of outcome increases with increasing duration of disease (Suissa and Azoulay, 2012). For a decreasing risk with increasing duration of disease, the results favour the second- or third-line therapies over the first-line therapy. This bias can be avoided by matching on diseases duration (Suissa and Azoulay, 2012) or by adjusting for its effect. Naturally, studies comparing two first-line (or second-line etc.) therapies avoid immortal time bias.

2.1.3.8 Confounding by indication

Confounding by indication appears when the reason of prescription is associated with the outcome of interest (ENCePP, 2018). In such a situation, compared groups differ with respect to the individual's condition or characteristics related to condition, which determine the choice and initiation of a specific drug (Gerhard, 2008). Confounding by indication can be avoided comparing groups of individuals sharing similar indications, including condition (disease) itself, its severity and presence of comorbidities (Gerhard, 2008). An active-comparator design, which refers to the comparison of two active drugs with the same or similar indications, increases the overlap in important characteristics between the compared groups (Yoshida *et al*., 2015).

2.1.3.9 Residual and unmeasured confounding

Both residual and unmeasured confounding can mix the effects between the exposure being studied and the outcome of interest (ENCePP, 2018). The former refers to confounding that remains after controlling for confounders due to their misclassification, the latter arises when important confounders cannot be controlled because they are not measured (Fewell et al., 2007). In the pharmacoepidemiological register-based studies, important confounders, such as clinical parameters and lifestyle factors are often not measured. Unmeasured confounding can be reduced by applying active-comparator design (Yoshida *et al*., 2015).

2.1.4 Register-based studies

In a retrospective study, the information necessary to determine exposure and disease status is often obtained from the secondary data sources, such as national health and administrative registers. Such data have been proved as having a great value and utility beyond the purpose for which they have been originally established (Gissler and Haukka, 2004).

There are several reasons behind the increasing popularity of conducting studies based on secondary data sources. First, data are readily available as well as relatively fast and inexpensive to acquire. Second, there exists a wide range of essential and reliable information often collected on large populations and over long periods. Gathering information from secondary data sources allows for the use of broader inclusion criteria and fewer exclusion criteria. This allows constructing comprehensive real-life cohorts and, therefore, leads to studies with greater generalizability. For instance, in pharmacoepidemiology, the majority of studies today are performed as observational research using the secondary data sources to obtain information on both drug exposure and health outcome (Andersen, 2014).

Use of secondary data sources implies, however, translation of administrative and clinical questions into exposures and outcomes that can be reliably measured using the available information (Sund, 2003). The definitions of study subjects, exposure, and outcome measures are, therefore, guided not only by specific questions of interest, but also by characteristics of the available data. In such settings, it is important to evaluate accuracy and completeness of the available data and to take into account other important aspects, such as information retrieving processes, the size of the data sources, registration periods (Sørensen *et al*., 1996).

2.1.4.1 Nordic registries

Nordic countries, including Denmark, Finland, Iceland, Norway and Sweden, have a long tradition of registry-based epidemiological research (Gissler and Haukka, 2004; Furu *et al*., 2010; Schmidt *et al*., 2014; Ludvigsson *et al*., 2016). All five Nordic countries have National Health and Administrative registries, most of which are of high completeness and contain data of good to high quality. Moreover, in the Nordic countries, each resident is issued a

personal identity number. Using personal identity numbers, information from different registries can be linked.

The Nordic National registries cover very similar periods of data collection as well as have similar design and contents (Maret-Ouda *et al*., 2017; Pukkala *et al*., 2018). At present, these registries cover 26 million people making it possible to form large and statistically powerful cohorts. Such settings create opportunities for conducting nationwide cohort studies of high external validity and for studying rare exposures as well as outcomes with long latency.

In cancer research, there is a long history of using the register data from several Nordic countries to form the cohort as well as to evaluate the exposure and outcome (Andersen *et al*., 1999; Pukkala *et al*., 2009; Engholm *et al*., 2010; Kvåle *et al*., 2017; Andersson, 2017). Although multi-country register-based cohort studies have been proved useful, data sharing initiatives are still rare in many research areas, including pharmacoepidemiology. For instance, a systematic literature review found that among pharmacoepidemiological register-based studies from the Nordic countries only four of 515 published during 2005–2010 used data from more than one country (Wettermark *et al*., 2013).

There are, however, some challenges that should be taken into account when planning a Nordic register-based cohort study, including differences in coding systems, requirements and procedures regarding ethical vetting, acquisition, management and sharing of the data (Ludvigsson *et al*., 2015; Maret-Ouda *et al*., 2017; Pukkala *et al*., 2018). For instance, In Denmark no data retrieved from the registries are allowed to leave the country (Maret-Ouda *et al*., 2017). In the Nordic countries, different versions of the International Classification of Diseases (ICD versions 7–10; ICD-O for oncology, versions 1–3) have been used across the countries and over time. Therefore, recoding of the data variables into the same coding system is usually an unavoidable step, which can be facilitated by compilation of coding dictionaries.

2.1.4.2 Clinical Practice Research Datalink in the UK

The Clinical Practice Research Datalink (CPRD) of the UK is another well-known source of secondary data. The clinical practice research database was established in 1987 for routine recording of the patient-level information from the participating general practices. Currently, 4.4 million individuals, 6.9% of the UK population, meet the quality criteria and are broadly representative of the entire population with regard to demographic characteristics (Herrett *et al*., 2015). The CPRD database contains anonymized patient-level data from primary care, including demographics, prescriptions and cancer diagnoses. The data on cancer diagnoses are considered to be in general of good quality (Boggon *et al*., 2013). Extensive use of the CPRD in the observational research has yielded over 1000 studies across a broad range of health outcomes (Herrett *et al*., 2015).

2.2 Survival analysis

Survival analysis refers to the application of statistical methods to the time-to-event data that arise from cohort studies when the occurrence of a specific event is of interest. Survival statistical methods include methods for summarizing data, hypothesis testing as well as modelling the survival times and incidence statistics, hazard rate and risk. Survival analysis methods account for the features, such as censoring, which are often encountered in time-to-event data.

Survival analysis is a statistical discipline with the history dating back to demography and actuarial science (Dickman, 2014). The development of demographic and actuarial techniques started already in the seventeenth century, and in the mid-twentieth century there existed a well-established methodology. However, the methods of actuarial statistics were based on life tables, in which birth and mortality data are aggregated by 1- or 5-year age and calendar time intervals and precise event times are not necessarily known nor are of interest.

In the 1950's, clinical trials, an emerging research area, called for techniques for the analysis of data on much smaller numbers of individuals followed on day by day basis yielding detailed observations. These data included exact event times but were also subject to censoring due to which some event times remained unobserved. In clinical trials, the major interest was in the differences between studied groups in terms of survival time, and therefore the exact event times provided valuable information. In 1958, this demand for new analytical techniques was addressed by Kaplan and Meier who introduced a non-parametric tool for estimation of survival function from incomplete observations (Kaplan and Meier, 1958). This method, today known as the Kaplan-Meier estimator, opened a new research area that advanced rapidly during the following decades.

Major advances in analytical techniques, among which was a model proposed by Cox for estimation of the hazard function (Cox, 1972), created a need for a unifying theoretical basis. The development of the underlying theory started in 1975 with the PhD thesis by Aalen, who studied the basic nonparametric statistical problems for censored data in terms of the conditional intensity of a counting process. This was followed by the formal introduction of martingales, i.e. differences between the counting process and the integrated intensity process, into survival theory (Aalen, 1978). The martingale concept and viewing time-to-event data as a result of an underlying stochastic process turned out to be a useful framework for the general theory. Further developments in the area resulted in an elaborate theory presented along with its mathematical details in the textbook by Andersen *et al.* (1993). A non-homogeneous Poisson process (NHPP) is a generalization of homogeneous Poisson process, in which the average intensity of arrivals is allowed to vary with time.

2.2.1 Time-to-event data

To construct time-to-event data, one must have a clear definition of the event of interest as well as clearly defined start- and endpoints at which the individuals enter and exit the study. Time-to-event or survival data include at least one time origin that is the start of follow-up, which creates time scale known as time-on-study. In addition, time-to-event data

incorporate event times, which are often recorded as time since the start of follow-up. In case of recurrent event, such as epileptic seizure, more than one event can be observed for each individual. In this work, I consider events that can occur only once. The classic example of such an event is death. For some individuals, the event of interest remains unobserved. This situation is referred to as censoring.

2.2.2 Censoring

A key characteristic that distinguishes time-to-event data from data arising in other study designs is the occurrence of censoring. Censoring means that event times are incompletely observed and, to avoid bias, analysis should be then performed by taking into account censored event times.

There are three general scenarios leading to censored times: right-censoring, left-censoring, and interval-censoring (Kalbfleisch and Prentice, 2002, p. 12-14). The most common type of censoring is right-censoring, which occurs when the individual leaves the study before the occurrence of the event of interest or study ends before the event has occurred. The follow-up time is left-censored when the event occurred before some lower time bound and the actual event time is unknown. Interval-censoring refers to a situation where the event time is known to lie within an interval instead of being observed exactly. In addition, there can be delayed entry or left truncation, in which the exposure or other defining event, after which the individual is considered at risk, precedes the entry to the study.

Beside the censoring types described above, several underlying censoring mechanisms are distinguished. In survival analysis, standard analytical techniques consider right-censored data assuming an independent and non-informative censoring. Independent censoring means that at any time the event process is not altered by censoring experience. In other words, the event process is independent of the censoring process. The assumption of non-informative censoring means that the censoring mechanism contains no information about the distribution of the event times.

A mathematical definition of censoring mechanisms along with some intuitive examples of different censoring types and mechanisms is provided by Andersen *et al.* (1993, pp. 135–152 on right-censoring).

The basic methods of survival analysis are designed for independent right-censored observations, but methods for interval and left-censored data are also available. In the following sections, I consider the basic concepts and some survival analysis methods for right-censored time-to-event data, when assuming that the incompleteness of observations is caused by independent and non-informative censoring. In such a scenario, there is no need to model censoring because the parameters of process causing incompleteness in observations can be viewed as nuisance parameters and the event process can be entirely described in terms of hazard and survival functions.

2.2.3 Hazard and survival functions

Let T denote the random variable representing time to event of interest, with the probability density function $f(t)$ and cumulative distribution function $F(t) = Pr(T \leqslant t)$, such that $f(t) = F'(t)$. The distribution of the time to event is mostly described by the survival function $S(t) = P(T > t) = 1 - F(t)$, whereas statistical models for time-to-event data are often based on the hazard function $h(t)$ for T defined as

$$h(t) = \frac{\lim_{\Delta t \to 0} Pr(t \leqslant T < t + \Delta t | T \geqslant t)}{\Delta t}$$

The hazard function describes the conditional probability that the event of interest will occur in the interval $[t, t + \Delta t)$, given that it has not occurred before time t. The hazard function is both a theoretical and descriptive tool. The hazard function can be seen as a statistical definition of an instantaneous incidence rate (density) used by epidemiologists and is also called theoretical rate, hazard rate or instantaneous conditional incidence.

Many statistical models for time-to-event data are based on the hazard function, whereas the distribution of the event times is mostly described by the survival distribution function. If one of the three functions, the probability density, survival and hazard function, is known, the other two can be derived using known relations between them including

$$h(t) = \frac{f(t)}{S(t)}$$

and

$$S(t) = exp\big(-H(t)\big),$$

where $H(t) = \int_0^t \quad h(t)dt$ is the cumulative hazard function.

2.2.4 The counting process approach

Since the work of Aalen (1978) the statistical theory of survival analysis has been based on the probabilistic theory of counting processes. Describing the of the occurrence of random events in terms of counting processes and martingales unified the previously scattered results and provided a basis for both parametric and nonparametric estimation and hypothesis testing in the setting of survival analysis. In a simple survival analysis, individuals can experience event of one type only. Counting process on some fixed continuous-time interval $[0, t]$

$$N(t) = \sum_{i=1}^{K} \quad 1\{T_i \leq t\}$$

is then counting the number of discrete events as they occur among K individuals at the time $T_i,\ i = 1, \ldots, K$. For any $0 < t_1 < \ldots < t_j,\ N(0) = 0 < N(t_1) < \ldots < N(t_j)$.

Many counting processes can be split into a random or martingale process $M(t)$, and a systematic or predictable process $\Lambda(t)$

$$N(t) = M(t) + \Lambda(t).$$

2.2.4.1 Intensity function

The systematic part of the counting process $\Lambda(t)$, also called as the cumulative intensity process (Andersen *et al*., 1993), can be represented by an intensity function of time

$$\Lambda(t) = \int_0^t \quad \lambda(s)ds.$$

The intensity $\lambda(t)$ represents the rate at which the events are expected to occur at the time t or soon after it, conditional on the history before this time point. The relation between the intensity function and the hazard function is given by

$$\lambda(t) = Y(t)h(t),$$

where $Y(t)$, is the number at risk (the size of risk set) just before time t for failing in the time interval $[t, t + \Delta t)$. Obviously, the intensity functions equals zero when the risk set includes no individuals.

2.2.4.2 Poisson process

One of the most important point processes is Poisson process. A homogeneous Poisson process describes a sequence of events over time and is specified by a non-negative intensity function of time. In the homogenous Poisson process, the interarrival times, the intervals between the consecutive event times, are independent and obey the exponential distribution with the same parameter λ.

A non-homogeneous Poisson process is a generalization of homogeneous Poisson process. In the NHPP, the average intensity of arrivals is allowed to vary over time and the process is specified by a non-negative intensity function $\lambda(t) \geqslant 0$ of time (Figure 1). The process generates no events when the intensity parameter equals zero and the number of event times generated by the process per time unit increases with the increasing value of $\lambda(t)$.

A counting process N is a NHPP process with intensity function $\lambda(t)$ for all $t \geqslant 0$, if

1. The counting process N has independent increments $N(t_j) - N(t_{j-1})$ and
2. $N(t_j) - N(t_{j-1}) \sim Poisson\left(\int_{t_{j-1}}^{t_j} \quad \lambda(t)dt\right)$ for all $0 < t_{j-1} < t_j$,

Where the second condition means that in any interval, the expected number of events is calculated as the area under the intensity curve bounded by the time axis and the end points of the interval (Figure 1).

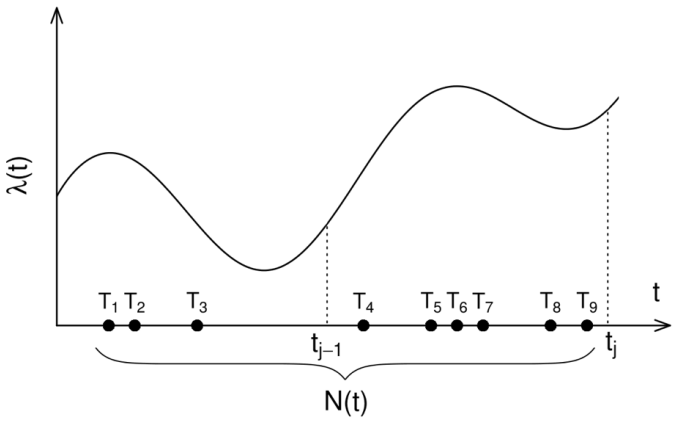

Figure 1 *The increments $N(t_j) - N(t_{j-1})$ of the non-homogeneous Poisson process N which is specified by the intensity function $\lambda(t)$. T_1 , T_2, …, T_9 are event times.*

2.2.5 Regression models

Most often, data comprise information on a set of covariates Z and it is generally of interest not only to estimate the hazard, but also to describe the relationship between a factor of interest and the time to event, when controlling for a set of other potentially confounding factors. Regression modelling of time-to-event data is commonly used to study the relationships of interest and is based on either the density or hazard function. A number of survival models are available to analyse right-censored survival data.

2.2.5.1 Likelihood function

The likelihood function is a key element of statistical inference. The likelihood function describes a statistical model given observed data D, which can be a scalar, vector or matrix. When assuming some model $D \sim f(D|\theta)$, where f is a density function with parameter θ, the likelihood function $L(D|\theta)$ is any function of θ proportional to $f(D|\theta)$. The likelihood function, therefore, does not obey the laws of probability but it is proportional to the probability of the observed data. In case of time-to-event data, there are seldom settings in which censoring is not encountered. To account for the effect of censoring on inference, censoring must be considered as a random variable contributing to the likelihood function. This means that the relationship between two random variables, time to event and time to

censoring, may affect inference about event time mechanism. This is why the assumptions of independent and non-informative censoring are essential.

Maximum likelihood estimation (MLE) is one way to use the likelihood function to extract the information on the model parameters (Tanner, 1994, p. 9–18). The method of maximum likelihood selects the set of values of the model parameters that maximizes the likelihood function. MLE provides estimators that have many desirable statistical properties allowing calculation of standard errors and statistical tests. The natural logarithm of $L(D|\theta)$, which is called log-likelihood function, is typically used to derive the maximum likelihood estimator of the parameter, because working with the log-likelihood $l(D|\theta)$ is more convenient.

2.2.5.2 Cox proportional hazards model

The proportional hazards model proposed by Cox (1972) for the analysis of data from clinical trials is the most commonly used method to analyse time-to-event data in medical research. Based on the Cox model, the hazard at time t is defined as product of baseline hazard λ_0 and exponential transformation of a linear combination of covariates Z and corresponding coefficients β

$$\lambda(t|Z) = \lambda_0(t)exp(\beta'Z),$$

where the baseline hazard λ_0 is a function of time t and $exp(\beta'Z)$ is independent of time. The baseline hazard function is not requested to obey any preset statistical distribution and is then the nonparametric component of the semi-parametric Cox model. The parametric part of the Cox model includes coefficients β, which are the model parameters to be estimated. At any point in time, the covariates Z act multiplicatively on the baseline hazard λ_0. The hazard ratio (HR) associated with a covariate is given by the exponent of its coefficient.

The estimation of the model parameters β is performed by maximizing the likelihood function. In case of the Cox model, MLE is based on a partial likelihood, also called a profile likelihood for the model parameters β (Clayton and Hills, 1993), which was introduced by Cox (1975). In the Cox model, the baseline hazard λ_0 is allowed to vary continuously over time by dividing the follow-up time into clicks, the intervals with no more than one event, and by assigning each click the hazard parameter for the corresponding hazard level. In the partial likelihood, these parameters are treated as nuisance parameters and are substituted by their most likely values.

The Cox proportional hazards model allows for modelling the baseline hazard function on a single time scale only. This underlying time scale, also called a primary time scale, determines the sequence of event times as well as the size of risk population for each click in the partial likelihood, and, therefore, affects the shape of the baseline hazard.

2.2.5.3 Poisson regression model for empirical rates

Poisson regression model for event rates is an important alternative to the Cox proportional hazards model (Frome, 1983; Breslow and Day, 1987; Clayton and Hills, 1993, pp. 227–229). Hereafter, I will refer to such a model simply as the Poisson regression. To recall, the event rate (mortality, incidence rate) is an empirical quantity used in epidemiology to describe the density of the event occurrences in a prespecified population followed over some period of time, during which each individual contributed some amount of person-time. Therefore, the incidence rate is the density measure in an accumulated amount of person-time (Benichou and Palta, 2014).

Time-to-event data can be organized according to the categorical covariates into a format similar to that of a life-table with cells including the total numbers of events and the total amount of person-time. The Poisson regression model, when applied to the tabulated time-to-event data, builds on the assumption of a constant hazard rate λ for each cell. The incidence rate represents a valid estimate of the hazard rate when the assumption of the constant hazard rate can be done (Benichou and Palta, 2014). Such an assumption is often realistic when considering the hazard in a short time interval.

The individual follow-up time can be divided into J small intervals, which contribute d_j events and person-time Y_j to the corresponding cell of life-table. At its simplest, these intervals can be of the same length y and the individual contributions to the cells can be treated as independent observations from the Bernoulli distribution with probability of event (i.e. success) λy. The log likelihood of observing independent empirical rates is given by

$$l(\lambda|D,Y) = Dln(\lambda) - Y\lambda,$$

where $D = \sum_{j=1}^{j=J} d_j$ and $Y = \mathrm{J}y$, when assuming that the empirical rates have the same hazard rate λ. Carstensen (2005) provides a detailed derivation of the above log likelihood. Except a constant Y, the resulting log likelihood is equivalent to the Poisson log-likelihood that would arise if the event counts in the cells were independent Poisson observations $D_j \sim Poisson(\lambda Y)$. Actually, contributions provided by an individuals to the cells and, hence, to the log-likelihood, are not independent but can be treated as conditionally independent. Importantly, the Poisson likelihood for a set of empirical rates equals the likelihood from the Cox regression model (Clayton and Hills, 1993, pp. 298–299).

The Poisson regression model can be specified as an additive or multiplicative model. The additive and multiplicative models are used to quantify an excess risk in terms of absolute and relative risks, respectively. The multiplicative Poisson regression model including the covariates Z is fitted as log-linear regression

$$E(D_j) = \lambda Y = exp(Z'\beta)Y = exp(Z'\beta + \log(Y)),$$

where coefficients β are the model parameters and the natural logarithm of person-time Y as an offset term, for which the coefficient is set to one. The model parameters β are estimated using MLE for generalized linear models (Frome, 1983) and the exponents of β give rate ratios (RR).

2.2.6 Multiple time scales in survival analysis

Epidemiologic cohorts usually constitute individuals who are the subject of multiple and varying biological and environmental circumstances, such as aging, diseases, exposure to some medications, toxins or interventions. For some of the involved factors, the time origin can be determined and represents the point at which an individual experiences a defining event, such as birth, disease or smoking onset or initiation of treatment. Each of the time origins creates a time scale, which represents the time elapsed since its defining event. In considering time-to-event data, both time origins and time scales play an important role. The time origin determines the time scale and should be defined in a clear and unambiguous way (Kalbfleisch and Prentice, 2002). Event times are recorded along one time scale and the sequence of event times depends on the scale that is used to measure time.

In cohort studies, there are usually several time scales and these may be relevant when considering the variation in the hazard of the event of interest. Although measuring time is a common feature for all time scales, their importance pertains not only in their chronology-preserving character. Many time scales are appealing because they can serve as a proxy measure of some exposure or experience. For instance, progression on the age scale corresponds to aging, proceeding with calendar time is often associated with changes in treatment methods and time elapsed from the onset of diabetes reflects a cumulative glycaemic burden.

The survival analysis methods allowing for graphical representation of time-to-event-data on more than one time scale include descriptive tools such as a Lexis diagram and Lexis surface plot. Analytical approaches of dealing with multiple time scales include the use of time scales as covariates in the Cox or Poisson regression model and age-period-cohort (APC) models. In the following chapters I overview these alternatives.

2.2.6.1 Lexis diagram

An inevitable involvement of the age and calendar time in demographic research created a need for a simple chart to represent the underlying population dynamics. Around the 1870's, this need was addressed by various graphical techniques that were developed by several German scientists in the field of population statistics, primarily by Knapp, Zeuner and Lexis (Vandeschrick, 2001; Keiding, 2011). A German statistician, economist, and social scientist Wilhelm Lexis introduced a diagram as a solution to the problem of locating deaths on one plane according to three demographic co-ordinates: the moment of death; the age of a deceased and the moment of birth of the deceased (Lexis 1875). Although the modern age-period-cohort chart is nowadays known as the Lexis diagram, it is not exactly the same plot as introduced by Lexis, suggesting that there probably were several scientists who contributed to the development of the tool (Vandeschrick, 2001; Keiding, 2011).

The Lexis diagram is a two-dimensional graphical representation of individual follow-up times on a plane formed by two time scales, originally by age on the vertical axis and calendar time on the horizontal axis (Figure 2). Each individual's trajectory is represented by a diagonal line, a life line, which allows for keeping track of the individual progressing through time. Moreover, the life line preserves the correspondence between the two time

scales: as the life line proceeds, the same amount of time passes on both time scales. The Lexis diagram is used to visualize the experience of an entire cohort or its subgroup.

Nevertheless, plotting life lines is not meaningful for large populations, of which aggregated data, such as data on counts and person-years, are used to represent the raw and smoothed death rates and ratios and other demographic parameters by means of the Lexis surface plots, contour maps and heatmaps (Arthur and Vaupel, 1984; Vaupel *et al*., 1987; Schöley and Willekens, 2017; Rau *et al*., 2018). In demography, these graphical tools are used for detection of patterns and trends at the population level. As such, these approaches are not applicable in epidemiology, where the focus is in evaluating the individual-level observations. Moreover, time-to-event data from cohort studies are often limited in the number of events and the size of risk population, and, therefore, the evaluation of uncertainty is essential.

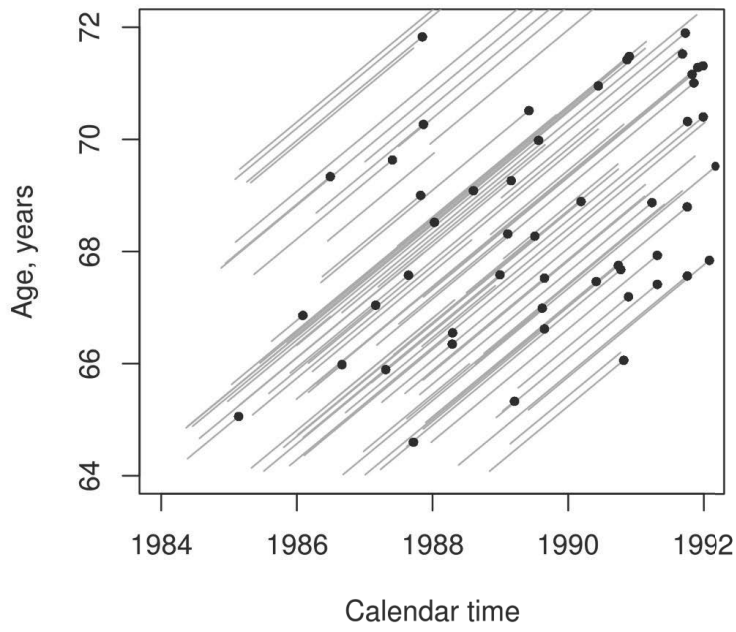

Figure 2 *The Lexis diagram depicts by life lines the follow-up of a sample of women from the gbcs dataset (Hosmer et al., 2008) who were diagnosed with breast cancer at the age of 64–69 years and were followed up until recurrence (dot) or censoring due to death.*

2.2.6.2 Multiple time scales in the Cox model

The Cox regression model relies on a single time scale, but multiple time scales can be dealt with by introducing time scales other than a primary one to the Cox model as fixed baseline or time-dependent covariates and / or by stratifying on one or more time scales. Though these solutions seem to be straightforward, their implementation is not a simple task in practice. First, in the presence of multiple time scales, the modelling with the Cox method involves the choice of primary time scale, an aspect that is not routinely considered in epidemiological studies (Cologne *et al*., 2012). Second, questions arise regarding the modelling of the effect of the time scales other than the primary one. In the Cox model, the baseline hazard is modelled nonparametrically. This implies a flexible modelling of the relationship between the primary time scale and the baseline hazard. In contrast, modelling of the effect of time scales that enter the parametric part of the Cox model requires a careful consideration of the relationship between these time scales and the baseline hazard.

The Cox model was originally proposed for the analysis of time-to-event data from clinical trials. In such settings, time-on-study as the primary time scale is an appropriate choice both in the biological (clinical) and analytical sense. The Cox model has been routinely used in the analysis of time-to-event data from observational cohort studies, where time scales other than time-on-study may be of biological relevance. However, there is no well-established procedure for choosing the primary time scale of several relevant time scales.

Farewell and Cox (1979) defined the primary time scale as one that accounts for as much of the variation in hazard as possible and suggested that in some contexts the most informative time scale is equivalent to the one on which event times form a Poisson process. Clayton and Hills (1993, pp. 311) suggested the choice to be done based on the way the baseline hazard vary along each time scale. If a rapid and irregular variation is expected along one of the time scales this should be used as the primary one. If variation is smooth along all the time scales it is better to use the time scale with the strongest relationship to the baseline hazard. According to Pencina *et al*. (2007), the choice of primary time scale can be addressed by considering both mathematical equivalence and correctness of models resulting from different choices, by comparing the estimated regression coefficients and/or predicting accuracy as well as by searching for the time scale better capturing the nature of the data.

In addition to time-on-study, also age (attained age, age at risk) has been used as the primary time scale. Commenges et al. (1998) pointed out that using age as the primary time scale may be of particular epidemiologic interest when studying the disease incidence, because with this choice the baseline hazard is equivalent to the age-specific incidence. In observational studies, individuals enter the study at different ages rather than at their birth that is the time origin for age.

Entering the study after the time origin is called late entry or left truncation if individuals, while being at risk for the event of interest, remain unobserved and have to survive until some point of time to be sampled. When those entered the study and those who precluded from entering differ in their risk, left truncation may introduce bias (Matsuura and Eguchi, 2005; Cain *et al*., 2011). In many situations, the assumption of independent and non-informative left truncation can be done. It is, however, necessary to account for late entry

itself by using the software that provides procedures capable to incorporate both entry and exit times in the Cox model formula (Clayton and Hills, 1993, pp. 302; Cain *et al*., 2011).

Several studies compared the Cox regression models that were based either on time-on-study or age and were applied to simulated and/or empirical data (Korn *et al*., 1997; Thiébaut and Bénichou, 2004; Pencina *et al*., 2007; Cologne *et al*., 2012; Chalise *et al*., 2016). Some of the compared models were based on attained age, whereas other models were based on time-on-study and used age at entry as a fixed covariate or stratified on it (Table 1). Altogether, ten different models were investigated but the variety of possible alternatives can be further extended by the models with age as the primary time scale and time-on-study as a time-dependent covariate and vice versa, and models that allow for capturing the functional form of the time scale that is used as a covariate. Obviously, in the presence of more than two relevant time scales, there are many more models to consider and choose of. Accounting for multiple time scales with the Cox model is further complicated by contradictory recommendations regarding the choice of the primary time scale.

Table 1 *The Cox proportional hazards models (Weibull proportional hazards models in Pencina* et al*) fitted to the empirical and/or simulated data using two alternative time scales, time-on-study and attained age (age at risk), as the primary time scale, with and without adjustment for age at entry as a fixed covariate or with stratification on birth cohort or age at entry.*

	Primary time scale					
	Time-on-study			Attained age		
	Unadjusted	*Adjusted for entry age*	*Stratified on birth cohort or entry age*	*Unadjusted*	*Adjusted for entry age*	*Stratified on birth cohort*
Korn *et al*. (1997): *empirical data*	Fitted		Fitted[a)]		Fitted	
Thiébaut and Bénichou (2004): *simulated and empirical data*	Fitted	Fitted[b)]	Fitted[c)]	Fitted		
Pencina *et al*. (2007): *simulated and empirical data*		Fitted[d)]	Fitted[a)]	Fitted	Fitted[e)]	Fitted
Cologne *et al*., (2012): *empirical data*		Fitted			Fitted[e)]	
Chalise et al. (2016): *empirical data*		Fitted		Fitted	Fitted[e)]	

a) Stratified on birth cohort; b) Two models, one adjusted for continuous and one adjusted for categorical age at entry; c) Stratified on age at entry; d) Two models, one adjusted for linear and one for quadratic age at entry; e) Accounted for left truncation by conditioning on age at entry.

All studies presented in Table 1 recommended to use age rather time-on-study as the primary time scale when analysing time-to-event data from epidemiological cohort studies. The studies disagreed on the conditions upon which the age-at-entry-adjusted Cox model with time-on-study as the primary time scale provides approximately unbiased estimates in situations when the Cox model based on age is the correct one. Korn *et al*. (1997) found the estimated regression coefficients to be different in the model that was based on time-on-study and adjusted for age at entry, when compared to results from two other models that were based on age with and without stratification on birth cohort (these two latter models provided similar results). Korn *et al*. suggested that at least one of two conditions should be satisfied to ensure that the Cox model based on time-on-study provides approximately unbiased results: an exponential age-specific baseline hazard and statistical independence between covariates and age at entry.

Thiébaut and Bénichou (2004) and Pencina *et al*. (2007) performed simulation studies to investigate the effect of the choice of time scale on the estimated regression coefficients when the conditions postulated by Korn *et al*. (1997) are satisfied. Thiébaut and Bénichou found that the unadjusted Cox model with age as the primary time scale performed without large bias regardless of the distribution of the age-specific baseline hazard and that any of the four Cox models with time-on-study as the primary time scale performed similarly when the age-specific baseline hazard was exponential. In contrast, Pencina *et al*. observed an inferior performance of the unadjusted Cox model with age as the primary time scale as compared to any of the five other models, which were based either on age or time-on-study and were either adjusted for age at entry or stratified on birth cohort. Chalise et al. found the Cox model based on time-on-study and adjusted for age at entry and the Cox model based on age to provide significantly different results in 40 cohorts of 54 studied when accounting for left truncation was ignored in the later model. These differences appeared independently of the magnitude of correlation between the covariate and age at entry. When left-truncation was accounted for, two models agreed in 51 cohorts of 54.

Table 1 lacks models with the time scale as a time-dependent covariate. In the Cox model, time proceeds on the primary time scale but not on the time scale included into the model as a fixed (baseline) covariate. The relation between the baseline hazard and the time scale included in the Cox model as a fixed (baseline) covariate is taken to be log-linear and the relative change in hazard is assumed to be the same and for a single time unit increase and the effect of such a covariate is assumed to be constant over time. Moreover, the baseline hazard is assumed to vary over time (as measured on the primary time scale) similarly for all the baseline values of the time scale included as a fixed covariate.

Violation of the parametric assumptions may result in an inappropriately modelled effect of the covariates included in the model, and, therefore may introduce bias (Thiébaut and Bénichou, 2004; Cologne *et al*., 2012). It is possible to relax the aforementioned assumptions by modelling the effect of time scales included into the parametric part of the Cox model as time-dependent covariates (Fisher and Lin, 1999; Lehr and Schemper, 2007) and by using polynomial functions, which are the flexible mathematical functions defined by piecewise polynomials. (Durrleman and Simon, 1989; Gray, 1992; Berger *et al*., 2003). Modelling of the covariate effect in a time-dependent manner and / or modelling its functional form provides greater flexibility and allows for capturing the changes in the effect

over time but also implies an additional level of methodological complexity and increases the number of parameters to be estimated.

Also stratification by age at entry or other relevant baseline time determinant can be used to account for the potentially different baseline hazard functions. However, the relevant controlling for the differences may require a fine stratification that may result in strata with the sparse data, and therefore, may yield inadequate estimates of both the baseline hazard and covariate estimates (Cologne *et al*., 2012) .

2.2.6.3 Multiple time scales in the Poisson model

In the Poisson regression model, there is no need to choose the primary time scale but all the relevant time scales can be included in the model as covariates. The time scales included in the model may represent time at entry (baseline) or time at risk measured during the follow-up. When modelling the rates with the Poisson regression, time-to-event data can be prepared for the analysis by splitting the follow-up time into smaller bands along one or more time scales (Carstensen, 2012). Thus, the Poisson regression model, when fitted to the split data, provides an efficient and intuitive method for dealing with time-dependent covariates, such as time scales.

The issues that need to considered include the ways of adequate modelling of the effect of time scales. The effect of a particular time scale can be flexibly modelled by using polynomial functions. For instance, restricted cubic splines are an easy way of including covariates in a smooth non-linear way (Carstensen, 2012).

2.2.6.4 Age-period-cohort models

The Lexis diagram has traditionally been used for analysing incidence and mortality rates with the APC models (Smith and Wakefield, 2016). Since the mathematical theory for the relationship of rates as described in continuous time in the presence of age-, period- and cohort-effects was introduced by Keiding (1990), the Lexis diagram has been extensively used to analyse data for the presence of these effects using various APC models (Carstensen, 2007; Held and Riebler, 2012; Brinks *et al*., 2014; Christiansen *et al*., 2015). Models using Bayesian inference have been also developed for the Lexis diagram. Two APC models introduced for the Lexis diagram with fixed grid used a random walk prior of first and second order for all APC effects, assuming a constant or linear trend over time, respectively (Berzuini *et al*., 1993; Berzuini and Clayton, 1994). Several multivariate APC models using the Bayesian approach with smoothing priors have been introduced for the estimation of relative risk (Riebler and Held, 2010; Riebler *et al*., 2012).

APC models are subject to identifiability problem because of over-parametrization due to having three time variables, one of which is a sum of two others (Clayton and Schifflers, 1987a, 1987b; Holford, 1991). Although various solutions have been proposed to address the identifiability problem, these solutions are not necessarily intuitive nor applicable

without further consideration of the involved assumptions and constrains with respect to the specific research question (Bell and Jones, 2015).

2.3 Bayesian analysis

2.3.1 Bayes' theorem

Bayes' theorem shows the relation between two conditional probabilities that are the reverse of each other. This theorem dates back to the original 1763 paper by Thomas Bayes (1701-1761) and is also referred to as Bayes' law or Bayes' rule. Bayes' theorem provides an expression for the conditional probability of A given B, which is equal to

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}.$$

Bayesian inference refers to statistical inference, which are based on the use of Bayes' theorem. The methods that are known today as Bayesian were, however, brought under this name relatively recently in terms of their 260-year history (Fienberg, 2006a).

2.3.2 Model-based Bayesian inference

In the Bayesian approach, the most basic model consists of two parts, a likelihood specification $D|\theta \sim L(D|\theta)$ summarizing the evidence about parameter θ provided by the data D and a prior $\theta \sim \pi(\theta)$ reflecting uncertainty about the parameter before the data are observed. Bayes' formula provides an expression for updating the prior density into the posterior $p(D|\theta)$ by taking into account the observed data D:

$$p(\theta|D) = \frac{L(D|\theta)\pi(\theta)}{m(D)} \sim L(D|\theta)\pi(\theta),$$

where the proportionality coefficient

$$m(D) = \int L(D|\theta)\pi(\theta)d\theta$$

is the marginal density of the data. The proportionality coefficient can be obtained analytically only for some special choices of L and π that are known as conjugate distributions (Gutiérrez-Peña *et al*., 1997). Results from the Bayesian models are usually reported in the form of (marginal) posterior expectations and probabilities of the model parameters.

2.3.3 Prior distributions

The prior can play many roles in Bayesian inference (Gelman *et al*., 2017). First, the prior provides a formal way to incorporate already available information, to quantify hypotheses or to represent researchers' degree of belief in a statement regarding the problem being analysed. Second, the prior can be used to regularize and stabilize inferences. Third, the prior can be treated as a necessary but noninformative part of the analysis, its only role being to fulfil the formal requirements of Bayesian analysis while having a minimum impact on the posterior distribution.

In applied Bayesian research, these three prior types are called informative, weakly informative and noninformative priors (Syversveen, 1998; Gelman *et al*., 2008; Kerman, 2011; Hamran *et al*., 2013). Within Bayesian inference, the different ways to treat the prior have yielded two opposed strands of Bayesian inference, objective and subjective (Berger, 2006; Goldstein, 2006; Fienberg, 2006b; Ghosh, 2011). In the former, inference is based on the use of an objective prior with a minimal impact on the posterior distribution. In the latter, the prior is defined to reflect the researcher's subjective beliefs. However, the convenience of this distinction between objective and subjective priors is considered to be dispensable in practice (Gelman and Henning, 2017).

Developing prior distributions is an inevitable and undoubtedly the most controversial aspect of any Bayesian analysis. Inappropriate choices of priors can lead to incorrect inferences, and the methods for choosing priors are an issue of considerable debate. During the last decade, the problem of specifying sensible priors has become even more pressing due to emergence of increasingly complex statistical models.

A recently published article by Gelman *et al*. (2017) called into question the standard Bayesian workflow, in which the prior distribution should come before the data model and with no reference to the data. Gelman *et al*. (2017) provided some recommendations regarding the choice of prior in real-world modelling and pointed out that in practice constructing of the prior often depends on the data and, therefore on the assumed likelihood for the data. Another recent work provided a broad framework for building informative priors for a large class of hierarchical models (Simpson *et al*., 2017). Stan Development Team (2017) has collected prior choice recommendations for some specific problems, and Zondervan-Zwijnenburg *et al*. (2017) provided a guideline for constructing informative priors in small sample research.

2.3.4 Bayesian – frequentist debate

Although some of the Bayesian statistical methods and ideas were introduced and used already in the 19th century, decades before the frequentist techniques were developed, the latter became a dominating approach for over half of the 20th century owing to the influential work by statisticians R. A. Fisher, J. Neyman and E. Pearson (Greenland, 2006). As the development and spread of frequentist methods were stimulated by use of experimental design, the frequentist inference is driven by the idea of a repeatable experiments and null hypothesis testing. In the frequentist paradigm, the presumption is that the data constitute a random sample from a distribution (model) controlled by an unknown

but fixed distribution (model) parameter(s). However, data generating mechanisms are often poorly understood, especially in the non-experimental design. Bayesian inference accounts for the modelling parameters as for random variables and, therefore, allows modelling the uncertainty in the parameters related with the sampling scheme and the data generating process.

Both the Bayesian and frequentist points of view have been upheld by philosophical, practical and pedagogical issues with aim to demonstrate superiority of the one approach over another (Mayo and Cox, 2006; Gelman, 2008; Gelman and Shalizi, 2013). Although the potential for a fruitful synthesis of these two fundamental statistical ideas is almost completely ignored in this debate, some discussion has emerged (Bayarri and Berger, 2004; Little, 2006). According to Efron (2005), Bayesian-frequentist dispute "reflects two different attitudes to the process of doing science, both quite legitimate". Little (2006) suggested that a compromise involving Bayesian inference for models and frequentist ideas on model assessment may be beneficial for the teaching and practice of statistics.

2.3.5 Hierarchical Bayesian estimation

In Bayesian modelling, model parameter θ is often given a probabilistic specification $\pi(\theta|\eta)$ by introducing further parameters η known as hyperparameters. The basic Bayesian model can be extended by adding additional levels or sub-models, which then form a hierarchical or multi-stage model. For instance, to quantify the uncertainty regarding hyperparameters η in $\theta|\eta$ an additional prior $h(\eta|\lambda)$ can be set. Each new prior, which is called a hyperprior, forms a new level in the model hierarchy. Estimating priors through hyperpriors is known as hierarchical Bayes estimation and is a method to elicit the optimal prior distributions.

Such hierarchical thinking allows to construct flexible models for describing complex settings and phenomena from classical multilevel modelling for nested data to problems involving multiple dimensions or complicated dependence structures. Gelman *et al*. (2014) and Craigmile *et al*. (2009) provide some examples on use of hierarchical models, including model building and fitting.

2.3.6 Nonparametric Bayesian inference

Statistical models can be divided into parametric, semi-parametric and nonparametric models. In nonparametric models, the parameters of the model are infinite in dimensions. In the Bayesian framework, this is achieved by setting priors on probability models with infinitely many parameters (Müller and Quintana, 2004; Müller and Mitra, 2013). These priors are known as nonparametric Bayesian priors, and inference on such models is referred as nonparametric Bayesian inference. Bayesian nonparametric models provide a simple framework for modelling complex data. In general, nonparametric models are more robust against modelling errors and are therefore more realistic and flexible than parametric models, which are likely to understate uncertainties and miss some important structure with

more complex data. Examples of nonparametric Bayesian models include approximation of intensity functions of a point processes, density estimation, spline regression models and multivariate regressions (Arjas and Gasbarra, 1994; Müller *et al*., 2017).

2.3.7 Markov chain Monte Carlo methods

The use of unappealing approximations was inevitably involved in applied Bayesian inference on the non-conjugate distributions before development of Monte Carlo statistical methods (Robert and Casella, 2004). MCMC algorithms are used in Bayesian inference to sample from the posterior probability distributions when the marginal density is analytically intractable. In MCMC methods, a sequence of random quantities is generated using Markov chain that is a random process, which has the Markov property of memoryless. Markov chains are characterized by the state space, index state and transition probabilities between the states. The state space is either a finite (countable) or infinite set of values (states), which the chain can take. The index state usually refers to time and can be either discrete or continuous. In case of memoryless, the probability of the future state of the chain can be affected by the more recent state(s) only and as time goes by, the process loses the memory of the past.

Markov chain can be used to draw random sample from some target probability distribution. In Bayesian inference, the target distribution is usually the posterior distribution of the model parameters. Each Markov chain begins with an initial value or state and the algorithm iterates or transits from the current state to the next one thereafter. By iterating Markov chain attempts to converge to the target probability distribution. To achieve the convergence, it is necessary to construct a transition operator, which after long run makes the (stationary) distribution of the chain to match the target one.

There are two widely used versions of MCMC, the Gibbs sampler and Metropolis-Hastings algorithms (Robert and Casella, 2004, pp. 270–272, 337–343). Standard MCMC sampling algorithms cannot be used for the models of varying dimension, such as nonparametric models. Reversible jump algorithms enable sampling from the posterior distribution in such models by providing moves between submodels of varying dimensions (Robert and Casella, 2004, p. 429–433).

2.3.8 Bayesian methods in practice

Computational advances, such as Markov chain Monte Carlo methods, emerged during the computer era for and from Bayesian analysis (Andrieu *et al*., 2004) and have led to widespread use of Bayesian inference. In the two past decades, the emergence of Bayesian analysis was seen in many methodological and applied fields, including medicine, public health and epidemiology (Etzioni and Kadane, 1995; Gurrin *et al*., 2000; Dunson, 2001; Spiegelhalter *et al*., 2004; Ashby, 2006). However, for instance in epidemiologic research, the number of studies in which Bayesian techniques have been used for primary data analysis has remained constant over the years (Reitbergen *et al*., 2017).

2.4 Diabetes Mellitus

Diabetes mellitus is a group of chronic, progressive diseases characterized by elevated levels of blood glucose as caused by deficient insulin production in type 1 diabetes (T1D) (Atkinson *et al*., 2014) or by body's ineffective use of insulin in type 2 diabetes (T2D) (Kahn *et al*., 2014). A constantly increasing prevalence of type 1 and 2 diabetes has been reported worldwide (Lammi *et al*., 2008; Tuomilehto, 2013; Harjutsalo *et al*., 2013; NCD-RisC, 2016; Ogurtsova *et al*., 2017). According to the report of World Health Organization (2016), there were 422 million people with diabetes in the world in 2014. The number of people with diabetes aged 20–79 years was predicted to rise to 642 million in 2040 (Ogurtsova *et al*., 2017).

2.4.1 Treatment of diabetes

Controlling blood glucose levels is the ultimate goal of diabetes management and requires a lifelong treatment. The goals of diabetes treatment regardless of its type include elimination of short-term risk of high or low glycaemic levels as well as prevention of short- and long-term complications of diabetes (American Diabetes Association, 2016). There are different classes of anti-diabetic medications (ADM), which can be divided into injected drugs, such as different insulins, and oral drugs, such as sulfonylureas, biguanides (metformin), thiazolidinediones (pioglitazone, rosiglitazone) and glucosidase inhibitors. All ADMs are targeted at lowering blood glucose levels but different drug classes achieve the target through different physiological actions (American Diabetes Association, 2016)

Since the discovery of insulin in 1921–22 and its rapid translation to practice, insulin administration has become a fundamental treatment of T1D (Polonsky, 2012). Preventive strategies and treatments for T2D have evolved especially during the last decades (Kahn *et al*., 2014). At its early stages, T2D is usually treated by lifestyle modifications along with oral ADMs. Due to the progressive character of T2D, initiation of insulin therapy is often required at later stages.

2.4.2 Diabetes in Finland

In Finland, individuals with specified diseases and conditions, including diabetes, are entitled to special reimbursement for drug costs in outpatient treatment. Entitlement for special reimbursement is granted to the person by the Finnish Social Insurance Institution (SII) after evaluation for eligibility according to the application completed by a physician certificate. All individuals entitled to special reimbursement are then recorded in the Reimbursement Register maintained by SII and can be identified from there by their unique identification numbers. The register was established in 1964 and holds information on dates when the reimbursement entitlement started (and possibly ended), and the code of the disease or condition due to which the reimbursement entitlement was granted. This nationwide register provides a reliable source for assessment of disease rates and is often used to define study populations in health care research.

In the Reimbursement Register, the reimbursement entitlement due to diabetes is recorded regardless of the type of diabetes. In 2017, diabetes medications other than insulin were transferred from the higher to the lower special reimbursement category (65% reimbursement rate, special reimbursement code 215). Before 2017, any diabetes medication belonged to the higher special reimbursement category (100% reimbursement rate, special reimbursement code 103), but since 2017 the highest reimbursement rate is applied to insulin only.

In Finland, the number of people with diabetes trebled from 122 675 people in 1990 to 368 314 in 2011, and this increase was due to the increasing prevalence of both type 1 and T2D (Arffman *et al*., 2014). During the period 2009–2011, the age-adjusted prevalence rate increased from 74 to 98 per 10 000 persons for T1D, and from 310 to 761 per 10 000 persons for T2D (Arffman *et al*., 2014). Based on the statistics of the Reimbursement Register, there were altogether 336 401 people entitled for insulin, and 336 406 people entitled for other diabetes medicines at the end of 2017 (Kelasto-reports, 2018). The lower category includes people with T2D, whereas the higher category includes people with type 1 or 2 diabetes, and the majority of the individuals included in the lower category are also entitled to the higher special reimbursement for insulin.

2.4.2.1 Finnish nationwide programs and studies on diabetes

In Finland, large-scale nationwide programs and studies, such as FINRISK (Borodulin *et al*., 2017), FINDRISC (Lindström and Tuomilehto, 2003), DPS (Lindström *et al*., 2003), FIN-D2D (Saaristo *et al*., 2010), FinDM (Sund and Koski, 2009), have been implemented during the last decades to prevent diabetes, to study causes and prognosis of the disease, as well as to monitor changes in its prevalence and incidence over time. In addition, register-based observational studies outside these programs have contributed to the extensive and high-quality research on diabetes in Finland. Indeed, data from the administrative health and welfare registers, such as Reimbursement and Prescription Registers (SII), Finnish Hospital Discharge Register and Cancer Register (National Institute for Health and Welfare), have been intensively used for research purposes (Gissler and Haukka, 2004). The Finnish registers have a long recording history and have been shown to be of good quality in general (Gissler and Haukka, 2004; Sund, 2012).

2.4.3 Morbidity and mortality

Despite the advances in diabetes treatment that emerged during the last decades, diabetes is still associated with an increased overall risk of dying prematurely, being ranked the eighth leading cause of death in 2012 (WHO, 2016).

Acute complications of diabetes, such as ketoacidosis and hypoglycemia, remain a significant contributor to the increased mortality (Groop *et al*., 2018). Over time, all types of diabetes can lead to chronic complications, many of which are caused by damage to large (macrovascular) and small (microvascular) blood vessels, thereby affecting the function of

many organs, including heart, kidneys, eyes, and nervous system (Fowler, 2008). The Finnish study reported that in 2011 about 17% of persons with diabetes had at least one serious complication, of which the most common were myocardial infarction (46%) and stroke (45%) and the least common were amputation (6%) and chronic kidney disease (3%) (Arffman *et al*., 2014). A recent Finnish study, which covered the period of 1994–2011, reported a decreasing risk of diabetes-related complications among persons with T2D, but an increasing risk of multiple serious complications in those with T1D (Forssas *et al*., 2016). Moreover, this study found that having diabetes-related complications was associated with an increased risk of death irrespective of diabetes type.

2.4.3.1 End-stage renal disease in type 1 diabetes

One of the most common comorbidities in patients with T1D is chronic kidney disease (CKD), which can progress to ESRD, a life-threatening condition with poor prognosis, requiring treatment by dialysis and kidney transplantation. The cumulative incidence of ESRD varies between 0.7% and 9.3% after 20-30 years of diabetes duration (Finne *et al*., 2005; Lecaire *et al*., 2014; Helve *et al*., 2017; Gagnum *et al*., 2017). Observational studies have reported an increasing risk of ESRD with increasing duration of T1D (Raile *et al*., 2007; Lecaire *et al*., 2014; Helve *et al*., 2017). Several studies observed a decreasing cumulative risk of ESRD with increasing calendar year of T1D onset (Finne *et al*., 2005; Lecaire *et al*., 2014; Helve *et al*., 2017).

Age at onset of T1D has been suggested as another potential factor influencing the risk for ESRD. However, the findings regarding the association between age at onset of T1D and the risk of ESRD have been inconsistent. Several studies observed the lowest risk of ESRD among those aged 0–4 or <6–10 years at onset (Finne *et al*., 2005; Svensson *et al*., 2006; Gagnum *et al*., 2017; Costacau and Orchard, 2018), whereas two studies found the highest cumulative hazard for the onset ages of 5–9 years (Helve *et al*., 2017) and 10–14 years (Gagnum *et al*., 2017).

Presence of CKD in persons with T1D increases the risk not only for ESRD but also for cardiovascular death, being a strong determinant of excess mortality associated with T1D (Groop *et al*., 2009; Forsblom *et al*., 2011). Some observational follow-up studies on ESRD risk accounted for death without ESRD as a competing event (Finne *et al*., 2005; Forsblom *et al*., 2011; Lecaire *et al*., 2014; Helve *et al*., 2017).

2.4.4 Diabetes and cancer risk

Large-scale epidemiological studies conducted in various populations have demonstrated an association between diabetes and increased risk for any cancer as well as for certain site-specific cancers (Carstensen *et al*., 2012; Harding *et al*., 2015; Carstensen *et al*., 2016; Ballotari *et al*., 2017).

Observational studies have reported statistically significant excess risks of 8-25% for any cancer among men and women with T2D and both types combined when compared to

the general population (Harding *et al*., 2015; Ballotari *et al*., 2017). Two studies found statistically significant differences in the incidence of any cancer in women with T1D but not in men with T1D when compared to those without T1D (Harding *et al*., 2015; Carstensen *et al*., 2016).

For both type 1 and 2 diabetes, the excess cancer risk has been reported for pancreas, liver, endometrium (corpus uteri), kidney and stomach (Harding *et al*., 2015; Carstensen *et al*., 2016; Ballotari *et al*., 2017). Among individuals with T2D, the excess risks have been observed also for cancer of gallbladder, colon and rectum, lung, breast, bladder, and thyroid, for ovarian and cervical cancer, for Hodgkin's and Non-Hodgkin's lymphoma, and multiple myeloma (Harding *et al*., 2015; Ballotari *et al*., 2017). An observational study in the Australian population found no difference in the incidence of breast cancer among women with T1D relative to the general female population (Harding *et al*., 2015), whereas a five-country study on T1D reported a reduced risk (Carstensen *et al*., 2016). For prostate cancer, a reduced risk for type 1 and 2 diabetes has been reported by many (Carstensen *et al*., 2012; Harding *et al*., 2015; Carstensen *et al*., 2016), but not all epidemiological studies (Ballotari *et al*., 2017).

A consensus report by Giovannucci *et al*. (2010) listed three major mechanisms, which may underlie the association between diabetes and cancer: 1) risk factors common to both diseases, including obesity, diet, smoking, physical inactivity; 2) specific metabolic disturbances typical of diabetes, such as hyperglycemia, insulin resistance, hyperinsulinemia; 3) ADMs. It has been also suggested that the association between diabetes and excess cancer risk can be only partially explained by a detection bias and/or reverse causation.

2.4.4.1 Shared risk factors

A recent Taiwanese study on a cohort of 405,878 subjects participating to a standard medical screening program explored the relationship between several chronic diseases and their markers and the risk for cancer (Tu *et al*., 2018). This study found that diabetes was associated with an increased risk for any cancer (HR 1.10, 96% CI 1.03–1.18) even after adjustment for known risk factors (age, education, occupation, smoking status and pack years of smoking, alcohol consumption, body mass index (BMI), physical activity, fruit and vegetable intake). It is, however, unlikely that the extent to which the risk factors contribute to the association between diabetes and cancer is uniform across diabetes types and different cancer sites.

In recent years, the role of obesity in the association between diabetes and cancer has been reviewed (Garcia-Jiménez *et al*., 2016; Klil-Drori *et al*., 2017) and scrutinized by observational studies. A cohort study in 300 039 CPRD patients with T2D found an increased risk of colorectal cancer for a cumulative obesity duration of 4–8 years (HR 1.19, 95% CI 1.06–1.34) and 8 years or more (HR 1.28, 95% CI 1.11–1.49]), as compared with non-obese persons with T2D (Peeters *et al*., 2015). The results from a cohort study, which comprised 88 107 postmenopausal women from the Women's Health Initiative, suggested

that the association between diabetes and incidence of endometrial cancer may be largely confounded by body weight (Luo *et al*., 2014).

2.4.4.2 Potential biological mechanisms

The complex process of the formation of cancer is termed carcinogenesis and can be divided into several, though rather simplified steps: initiation, promotion and progression. A number of plausible biological mechanisms, including the effects of hyperglycemia, hyperinsulinemia, and chronic inflammation on the promotion and progression of cancer, have been suggested as potential pathways linking diabetes and cancer (Giovannucci *et al*., 2010; Johnson *et al*., 2012; Gallagher and LeRoith, 2015). These metabolic abnormalities may contribute directly or indirectly through several mechanisms. However, as cancers are a group of heterogeneous diseases, it appears unlikely that these diabetes-related factors act uniformly across cancer sites. Despite multiple studies elucidating various mechanisms, through which diabetes-related factors are likely to influence the neoplastic process, our current understanding of the association between diabetes and cancer relies on hypothesized rather than established biological links.

Although both experimental and epidemiological evidence is more consistent with the hyperinsulinemia hypothesis (Giovannucci *et al*., 2010), the hyperglycemia hypothesis cannot be ruled out (Stattin *et al*., 2007; Ryu *et al*., 2014). Hyperglycemia and hyperinsulinemia refer to high circulating glucose and insulin levels, respectively. Hyperglycemia is the hallmark of both type 1 and 2 diabetes, whereas endogenous hyperinsulinemia, the one caused by disease itself, is associated with T2D. However, regardless of the diabetes type, diabetic patients with insulin treatment may have exogenous hyperinsulinemia due to insulin treatment. It is possible that hyperinsulinemia, both endogenous and exogenous, could promote carcinogenesis directly through the insulin and insulin-like growth factor-1 receptors, which are expressed by most cancer cells and stimulate cell metabolism and cell growth and mitosis; or/and indirectly by increasing circulating levels of bioactive insulin-like growth factor-1, which has higher tumour favouring activity than insulin, or/and by causing elevated levels of sex steroids, which are associated with a higher risk of certain cancers (Giovannucci *et al*., 2010).

2.4.4.3 Antidiabetic medications

Epidemiological evidence, though inconsistent, has suggested that different ADMs may modulate the risk of cancer (Smith and Gale, 2009, Giovannucci *et al*., 2010; Tokajuk *et al*., 2015). A meta-analysis of 265 of observational studies and RCTs, reported a lower risk of cancer for the use of metformin and thiazolidinediones and an increased risk of cancer for the use of insulin, sulfonylureas and alpha glucosidase inhibitors (Wu *et al*., 2015). However, the results from the observational studies regarding specific ADMs have been inconsistent.

Lewis *et al*. (2011) found an increased cancer risk for the use of pioglitazone, whereas more recent studies found no association between the use of pioglitazone and the risk of cancer (Lewis *et al*., 2015, Kowall *et al*., 2015; Levin *et al*., 2015; Korhonen *et al*., 2016) and rosiglitazone (Tuccori et al. 2016). A reduced risk of cancer has been reported for the use of metformin by numerous observational studies but only three of 27 studies reviewed by Suissa and Azoulay (2012) avoided time-related biases.

Results from four observational studies (Colhoun, 2009; Currie *et al*., 2009; Hemkens *et al*., 2009; Jonasson *et al*., 2009) raised concerns on the potential association between insulin analogue glargine and increased cancer risk (Pocock and Smeeth, 2009; Giovannucci *et al*., 2010; Johnson and Yasui, 2010). These concerns appeared to be well-founded in the light of preclinical safety evaluations suggesting an increased mitogenic potency of some insulin analogues due to their greater binding affinity for the insulin and insulin-like growth factor-1 receptors as compared to human insulin (Baricevic *et al*., 2015).

Findings from the numerous observational studies, which emerged after the publication of four initial studies, provided no conclusive evidence on the relationship between the use of insulin analogue glargine and cancer risk (Karlstad *et al*., 2013; Badrick and Renehan, 2014). The majority of these observational studies have been criticized for limitations, methodological drawbacks and biases, including short follow-up, prevalent-user design, confounding by indication and time-related biases (Pocock and Smeeth, 2009; Johnson *et al*., 2012; Wu *et al*., 2016). In addition, the findings of several observational studies have suggested that reverse causality and detection bias may explain an increased cancer risk at the time of diabetes onset or initiation of ADMs (Carstensen *et al*., 2012; De Brujin, 2014).

There are only few observational studies with new-user design and/or assessment of the cancer risk by treatment duration or cumulative dose (Suissa *et al*., 2011; Ruiter *et al*., 2012; Fagot *et al*., 2013; Stürmer *et al*., 2013; Peeters *et al*., 2016; Wu *et al*., 2017). These studies compared use of insulin glargine to human insulin with respect to the incidence of breast cancer (Suissa *et al*., 2011; Peeters *et al*., 2016; Wu *et al*., 2017), any cancer and breast cancer (Ruiter *et al*., 2012; Fagot *et al*., 2013), any cancer, breast cancer, prostate cancer, and colon cancer (Stürmer *et al*., 2013). For any cancer, no association with use of glargine was seen in two study (Fagot *et al*., 2013; Stürmer *et al*., 2013), and a decreased risk was found in the study by Ruiter *et al*. (2012). For breast cancer, two studies observed an increased risk for the use of insulin glargine compared with that of human insulin (Ruiter *et al*., 2012; Wu *et al*., 2017), and three found no association (Suissa *et al*., 2011; Stürmer *et al*., 2013; Peeters *et al*., 2016). Stürmer *et al*. (2013) observed no difference in the risk of prostate and colon cancer.

The need for robustly designed and conducted observational studies on well-powered cohorts with long follow-up has been acknowledged and calls have been made for second-generation observational studies (Renehan, 2012; Johnson *et al*., 2012). Among principles and recommendations regarding the appropriate methodological and analytical approaches, the use of time-varying (cumulative) exposure definition, the new-user cohort design and evaluation of site- and sex-specific cancer endpoints have been emphasized (Johnson *et al*., 2012; Renehan, 2012; Walker *et al*., 2013; Badrick and Renehan, 2014; Wu *et al*., 2016).

3 Aims

The present study was undertaken to investigate the relationship between the use of anti-diabetic medications, including different insulin types, and the risk of cancer when addressing methodological shortcomings and mitigating bias potentially involved in previous observational pharmacoepidemiological studies. In addition, the present study aimed at addressing the issue of multiple time scales, which are often present in observational cohort studies, by introducing a novel method and by applying it to a real-world problem.

Study-specific objectives were as follows.

Study I

To evaluate the relationship between anti-diabetic medication and risk for any cancer in the population-based FINRISK cohorts, when accounting for the effect of duration of anti-diabetic treatment and controlling for important confounders, such as smoking, use of alcohol and BMI.

Study II

To investigate the relationship between use of certain insulins and risk for any cancer and ten specific cancers in a five-country cohort study on new insulin users, when addressing the limitations and biases involved in previous studies.

Study III

To introduce a nonparametric Bayesian model for the estimation of intensity function on two time scales jointly.

Study IV

To demonstrate that exploring time-to-event data on several time scales jointly may provide additional insights to a phenomenon of interest and to provide a motivating example by modelling the time-dependent dynamics of the hazard of ESRD and death without ESRD with the models based on the method from Study III.

4 Materials and methods

The empirical part of this thesis includes two studies, in which the relationship between the cancer risk and the use of ADM (Study I) and different insulins (Study II) was studied using the FINRISK cohort and the CARING cohort, respectively. In addition, in Study IV, a cohort of T1D patients (IV) was studied to demonstrate several applications of the Bayesian intensity model introduced in Study III. The study cohorts are presented graphically in Figure 3.

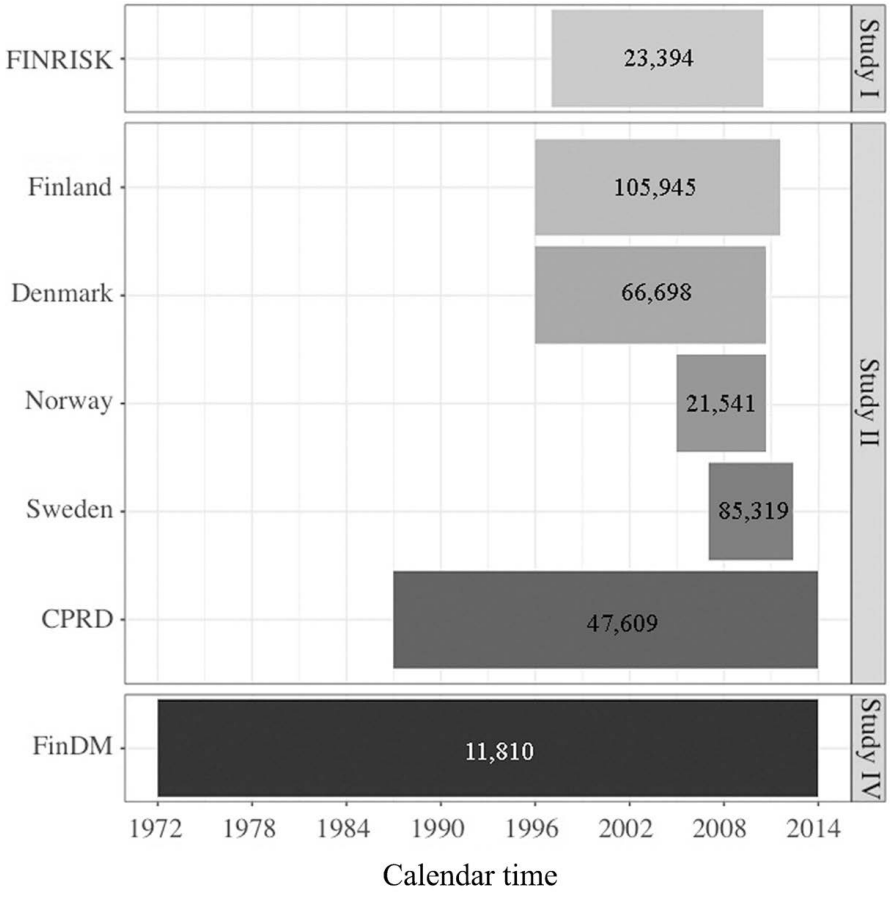

Figure 3 *Bars show the study periods and the size of study populations investigated in this work: the FINRISK cohorts (Study I), cohorts of new insulin users from Nordic countries and the CPRD from the UK (Study II) and the FinDM (Diabetes in Finland) cohort of individuals diagnosed with T1D (Study IV).*

4.1 Empirical studies (I,II)

4.1.1 FINRISK cohort (I)

The study population comprised three FINRISK cohorts on respondents of representative, cross-sectional population surveys carried out in 1997, 2002 and 2007 in several areas in Finland (Borodulin *et al*., 2017). Data on potential confounders, such as smoking, alcohol consumption and BMI were measured as part of the FINRISK survey and were augmented by data on incident cancers from the Finnish Cancer Registry, death records from Statistics Finland, and ADMs from the Prescription Register (SII). Individuals entered the cohort at the date they visited the FINRISK study site and were followed until December 31, 2010, the date of diagnosis of any cancer or death, whichever occurred first. After exclusion of individuals with any history of cancer at baseline (N=870) and prevalent users of ADM (N=548) identified using a half-year wash-out period, the study population comprised 23,394 individuals.

4.1.1.1 Outcomes, exposures and confounders

Prescription data included the date of purchase, anatomical therapeutic chemical (ATC) code (A10 for any ADM, A10A for insulin etc.), the number of packages purchased and the code of reimbursement entitlement. In this study, the exposure to any and specific ADM was measured in terms of time since date of the first purchase. Individuals who purchased any ADM during the study period were considered as new ADM users after the first purchase and, according to intention-to-treat approach, were regarded as users thereafter. In this manner, users of specific ADM, including metformin, sulfonylurea and insulin, were identified.

The outcome of interest was any incident cancer defined as the first primary cancer of any site except for skin cancer other than melanoma. To avoid uncertainty in the sequence of the initiation of ADM and cancer diagnoses, we regarded those diagnosed with cancer within the first month of using ADM as being diagnosed as non-users.

According to the alcohol consumption reported at baseline, the individuals were divided into non-users, moderate users (<14 and <7 portions per week for men and women, respectively) and heavy users for the higher consumption. Based on the baseline smoking status, individuals were divided into never, former and current smokers. For BMI (weight in kilograms divided by square of height in meters), four categories were used: underweight (<18.5kg/m^2), normal (18.5–24.9 kg/m^2), overweight (25–29.9 kg/m^2) and obese (30 kg/m^2 and over). Because there were missing values in alcohol consumption (4%), smoking status (1.3%) and BMI (11%), the categorical variables included also a "missing" category.

4.1.1.1 Methods

To avoid immortal time bias, the individual follow-up time was cut into intervals according to the date of the first purchase of any ADM. As a result, the follow-up time of user was split in two intervals, the time before and after the initiation of ADM. For each specific group of ADM, the individual follow-up was treated in the same manner. Further, the time since initiation of any and specific ADM was split into smaller intervals at 3, 6, 12, 24, and 48 months. Therefore, each user contributed multiple intervals of their follow-up time (i.e. person-time) to the split data.

Based on the split data, I constructed several variables describing this follow-up time in terms of ADM use. The time before and after the initiation of ADM was described by a binary variable with two categories, 'no ADM' and 'ADM'. To specify the time elapsed since the initiation of ADM (i.e. duration of use), we used a time-varying variable with the category 'no ADM' and three categories '≤ 1 year', '1–4 years' and '>4 years' for ADM use. For each specific group of ADM, otherwise similar exposure variables were constructed except they included an additional category 'other ADM' to denote the time on other ADM(s) that possibly preceded the initiation of the specific ADM of interest.

For instance, if an individual used ADM for five years, and started with metformin that was then augmented by insulin three years later, his/her exposure was described by the following variables:

1. Any ADM: the binary variable and time-varying variables as described above
2. Metformin: the categorical variable with categories 'no metformin', 'other ADM' and 'metformin' so that s(he) contributed no person-time to the category 'other ADM'; in the time-varying variable, s(he) contributed person-time to the categories 'no metformin' and '≤ 1 year', '1–4 years', '>4 years' on metformin
3. Insulin: the categorical variable with categories 'no insulin', 'other ADM' so that s(he) contributed person-time to all categories; in the time-varying variable, s(he) contributed person-time to the categories 'no insulin', 'other ADM' and '≤ 1 year', '1–4 years' on insulin.

Using the univariate logistic regression, we assessed the effect of each confounder on the probability of starting ADM. The incidence rate (IR) of cancer was modelled by using Poisson regression model for rates with the offset term for the logarithm of person-time. The crude and adjusted rate ratios (RR) with 95% confidence intervals (CIs) for the use of (specific) ADM relative to non-use were assessed by using univariate and multivariate (sex, age modelled by the cubic spline, calendar time, BMI, smoking) Poisson regression models, respectively. The covariates other than those for ADM exposure were selected into the model based on the Akaike information criterion and the deviance test. As accounting for the baseline use of alcohol did not improve the multivariate model, this covariate was not included into the final model. In addition, I examined the variation in RR graphically by using restricted cubic splines for the duration of (specific) ADM with knots set at quartiles.

The data were processed using the Lexis machinery (Carstensen, 2012) available in the Epi package (Plummer and Carstensen, 2011) for the R statistical software (R Core Team, 2017). The p-values corresponding to the z ratio were calculated and $p<0.05$ were considered statistically significant.

4.1.2 CARING cohort (II)

An observational cohort study on new users of insulin was conducted as a part of the five-country CARING project to study the relationship between the use of different insulins and cancer risk. Population-based cohorts were formed by using the prescription data from the Norwegian, Swedish, Danish and Finnish National Prescription Registries and from the CPRD in the UK. For each country, the start of the study period was defined as the year when the collection of prescription data started. The study period was 1996–2010 for Denmark, 1996–2011 for Finland, 2005–2011 for Norway, 2007–2012 for Sweden and 1987–2013 for the UK.

The Nordic prescription data were linked to the register data on cancer, death and emigration. For the UK cohort, the data on cancer and death were retrieved from the CPRD. The data sources and formation of the study cohorts are described in more detail in the study protocol, which was registered in the ENCePP electronic register of studies (CARING Consortium, 2015*)*.

The individuals entered the cohort on the date of the first prescription of any insulin (index date). The exclusion criteria comprised any history of cancer at baseline and prevalent use of insulin as defined by a 1-year wash-out period. After applying the exclusion criteria, the study population comprised 327 040 new insulin users.

4.1.2.2 Outcomes, exposures and confounders

The primary outcome of interest included incident cancer at ten cancer sites: trachea and lung, breast, endometrium, prostate, colon and rectum (colorectal), liver, pancreas, bladder, melanoma of skin and non-Hodgkin lymphoma. The secondary outcome of interest was any incident cancer except non-melanoma skin cancer. Diagnoses are recorded in the Nordic countries using the ICD codes (revisions 7, 9, 10 and O-3) and in the CPRD according to the Read code system. Identification of the incident cancers from the data was performed relying on coding dictionaries compiled according to the different coding systems. To achieve concordance between the diagnosis codes, the code lists were carefully revised for equivalence and completeness.

In the Nordic Prescription Registers, purchased medicines are recorded using ATC codes, in the CPRD with the British National Formulary codes. The primary interest was on the long-acting insulins (human insulin, insulin glargine and insulin detemir) and all other insulins were considered as one group. The Nordic prescription data included the date of purchase and the amount purchased in defined daily doses (DDD) but no information on individual dosage. The CPRD prescription data included the date of prescription and the dosage information (substance strength and amount), from which the DDD was derived. By assuming a dose of 1 DDD per day, the exposure to insulins of interest was assessed in a time-dependent manner as the cumulative treatment time. For each insulin type of interest, the exposure started at the date of first purchase, and an individual was considered exposed from that point onward. To capture the changes in exposure status during the follow-up period, the individual follow-up time was split into 120-day intervals. The exposure was updated at the start of each interval. Time on a particular insulin cumulated until exposure

stopped due to switching to another insulin type or discontinuation (or the end of follow-up) and cumulative exposure remained at the same level unless the treatment was resumed. In episodes of repeated prescriptions, possible gaps between the periods covered by each prescription did not accumulate treatment time. Cumulative treatment time was divided into half-year categories for the first year, followed by 1-year categories for longer exposure, with the last being >6 years for the broadly, and >12 years for the finely categorized long-term exposure. In addition, each exposure variable incorporated a non-exposed status.

We only considered confounders available in all five datasets: sex, age (time-varying, 10-year groups), calendar time. The baseline co-medication (non-insulin ADM, statins, nonsteroidal anti-inflammatory drugs, hormone replacement therapy) was defined as at least one prescription within one year before the index date. Individuals aged 30 years or younger at index date with no oral ADMs were considered to have T1D, whereas those aged 40 years or older were considered to have T2D. The rest of cohort was assigned unspecified diabetes type. Duration of insulin-treated diabetes was defined as a time since index date in 1-year intervals. Menopausal status (no/yes) was evaluated time-dependently. Women were assumed to reach menopause at the age of 50 years. Furthermore, country of the data source was used as a covariate. In order to tabulate the data, only categorical variables were used.

4.1.2.3 Methods

Table 2 summarizes the approaches, including the study design, methodological and analytical methods, used in this study to mitigate different types of bias. For each cohort, the data were tabulated and the number of cancer cases and person-years were aggregated by categorical variables. Tabulated data on five cohorts were combined and IRs were estimated by fitting multivariate Poisson regression models on tables of event numbers using the log of person-years as an offset. Each model incorporated all three insulin exposures, and was adjusted for age, calendar time, duration of insulin-treated diabetes, country, baseline use of non-insulin ADM and other co-medication. The RRs for cancer incidence with 95% CIs were evaluated by contrasting rates in the same exposure categories of glargine, detemir, and human insulin (glargine vs. human insulin, detemir vs. human insulin, glargine vs. detemir). In the primary analyses, sex- and site-specific endpoints were examined using insulin exposures with a broader category for longer duration. For the secondary analyses, similar evaluations were performed without stratifying on sex and using the insulin exposures with the finer categories of cumulative duration.

To evaluate the robustness of the results, several sensitivity analyses were performed. Potential confounding effect of diabetes type was evaluated by restricting the study population to those who fulfilled the definition of T2D. To account for possible changes in the profile of new insulin users after the launch of glargine in 2000, we excluded all individuals with the index date before 2000. Given that recording of cancer diagnosis in the CPRD is based on a different approach and coding system than in the Nordic Cancer Registries, we repeated the primary analysis using the Nordic cohorts only. For breast and endometrial cancer, further adjustment for menopausal status was performed.

Table 2 *Study II: methodological shortcomings and biases mitigated, addressed by checking the robustness of the results or disentangled by the design and analytic features used in the study.*

	Selection bias		Information bias			Time-related bias		Confounding bias	
Design / analytic feature	Health care access bias	Prevalent user bias	Misclassification of exposure	Misclassification of outcome, detection bias	Protopathic bias (reverse causation)	Immortal time bias	Time-lag bias	Confounding by indication	Residual or unmeasured confounding
Adjustment for duration of insulin treated diabetes	-	-	mitigated	-	-	-	mitigated	-	mitigated
Active comparator approach	-	-	-	mitigated[b]	-	mitigated	mitigated	mitigated	mitigated
Cumulative exposure definition	-	-	mitigated	-	disentangled	-	-	mitigated	-
Nationwide Nordic drug registers	mitigated	-	-	-	-	-	-	-	-
Nationwide Nordic cancer registers	-	-	-	mitigated[c]	-	-	-	-	-
New-user design	-	mitigated	mitigated	-	-	mitigated	mitigated	mitigated	-
Sensitivity analysis	checked[a]	-	-	checked[c,a]	-	-		-	checked[e]
Time-varying exposure definition	-	-	mitigated	-	disentangled	mitigated	mitigated[d]	mitigated	

a – restriction to the Nordic cohorts, b – detection bias, c – outcome misclassification, d – restriction to the calendar period from 2000 onward, e – restriction to the individuals with T2D.

4.3 Bayesian intensity model (III, IV)

4.3.1 Two time scales (III)

The model proposed in this work is based on the point process on the Lexis diagram and allows for modelling the right-censored time-to-event data on two time scales jointly. By viewing the intensity process as a non-homogeneous Poisson process, the model yields the Poisson likelihood for statistical inference. A less complex parametrization of the model is achieved by discretizing the process with respect to one of the two time scales. First, the Lexis diagram is transformed into its isomorphic representation, in which the life lines become horizontal (Figure 4). Then, the observational space defined on this isomorphic representation is divided into K strips or strata A_k, within which the life lines remain over the entire observation period. These strata are assigned the intensity functions λ_k, each of which can vary with respect to the one time scale only that is along life lines. These strata-specific intensity functions are modelled by piecewise constant functions $g_k, k = 1, \ldots, K$, which are specified by jump points $\varepsilon_{k,j}$ and hazard levels $h_{k,j}$ (Figure 5).

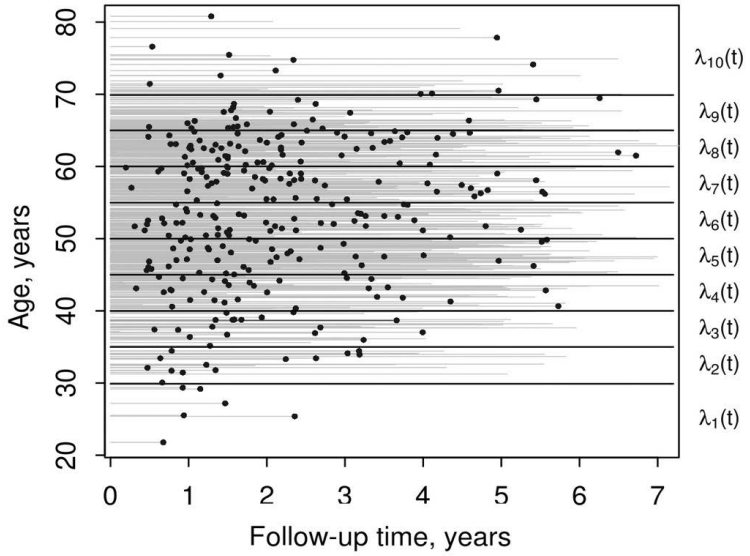

Figure 4 *The Lexis diagram as divided into ten strata and transformed to its isomorphic representation in which the life lines become horizontal. The follow-up of each women from the gbcs cohort is represented by a life line with or without dot at the end to denote recurrence and censoring, respectively. The life lines proceed along time since mastectomy. The strata are formed according to age at mastectomy and are assigned the intensity functions* λ_k, k=1, …, 10. *The stratum-specific intensity function* λ_k *is a function of time t that is measured as time since mastectomy.*

4.3.1.1 Prior for piecewise hazard functions

Jump points $\varepsilon_{k,j}$ and hazard levels $h_{k,j}$ are the model parameters to estimate. The Bayesian modelling proceeds by assigning a prior distribution to the model parameters. Figure 5 visualizes the structure of the prior. Arjas and Gasbarra (1994) proposed a nonparametric Bayesian approach to the estimation of the intensity function of a non-homogenous Poisson process on the real line. The prior proposed by Arjas and Gasbarra for the hazard levels, hereafter referred as the AG prior, assumes no trend a priori. The prior for the hazards levels of the two-dimensional model, hereafter referred as the Lexis prior, was constructed by extending the idea of Arjas and Gasbarra to apply the two-dimensional case.

The prior distribution of the jump points $\varepsilon_{k,j}$ is assumed to be a time-homogenous Poisson-process with intensity parameter μ. Figure 5 visualizes the structure of the Lexis prior. The Lexis prior for the hazard levels assumes no trend a priori but additionally incorporates smoothing and borrowing of strength in two dimensions, within and over strata as shown by curved arrows in Figure 5.

Let $\gamma(\cdot,\cdot)$ denote a Gamma prior with the shape and scale parameter. Within the first stratum A_1, the prior for the hazard levels is assumed to be $\gamma(\alpha_0,\beta_0)$ and $\gamma(\alpha,\alpha/h_{1,j-1})$, $j>0$, for $h_{1,0}$ and $h_{1,j}$, respectively. Within the strata A_k, $k>1$, the prior for the hazard levels $h_{k,j}$ is assumed to be $\gamma(\alpha,\beta_{k,j})$, where $\beta_{k,j}$ is calculated by dividing α by the weighted mean of the previous hazard level $h_{k,j-1}$, if exists, and the average of the neighbouring hazard levels of A_{k-1}. The weighted mean is calculated by using a weight parameter $\phi>0$.

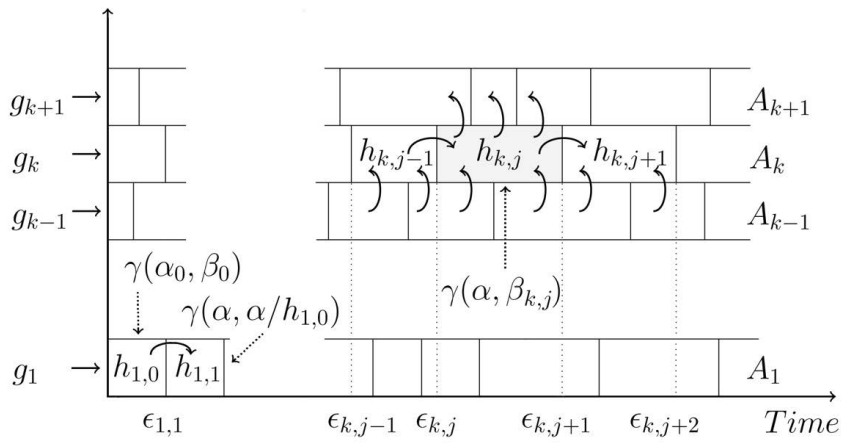

Figure 5 *The data are divided into the strata A_k, $k=1,\ldots,K$, and the stratum-specific piecewise constant functions g_k are specified by the jump points $\varepsilon_{k,j}$ and the corresponding hazard levels $h_{k,j}$,* j= 0,1, … . *Within A_1, the prior for the hazard levels is assumed to be $h_{1,0} \sim \gamma(\alpha_0,\beta_0)$ and $h_{1,j} \sim \gamma(\alpha,\alpha/h_{1,j-1})$ when $j>0$. Within A_k, $k>1$, the hazard levels $h_{k,j}$ are assigned the prior $\gamma(\alpha,\beta_{k,j})$ that tightens $h_{k,j}$ with the previous hazard level $h_{k,j-1}$ within A_k and the neighbouring hazard levels of A_{k-1} as shown by curved arrows.*

Altogether, the Lexis prior comprises five hyperparameters, μ, α_0, β_0, α and ϕ. The small values for μ imply few jump points in the hazard rate and vice versa. The initial hazard levels of the first and subsequent strata are controlled by α_0 and α, respectively, that regulate the looseness of the hazards levels with the larger values resulting in a stronger smoothing. The direction and strength of smoothing is regulated by ϕ, with $\phi = 1$ standing for equally strong impact of $h_{k,j-1}$ and the average of neighbouring hazard levels of A_{k-1}, and $\phi = 0$ standing for a priori independent hazard functions λ_k and λ_{k-1}. For the latter case, the Lexis prior simplifies to the AG prior. The hyperparameters can be given some values or be set priors.

4.3.1.2 Inference

Inference under the proposed nonparametric Bayesian model is accomplished using the MCMC method, namely the reversible jump Metropolis-Hastings algorithm described by Härkänen (2003). The reversible jump Metropolis-Hastings algorithm allows for sampling from the posterior distribution with the parameter space of variable dimension.

The analyses were performed using the BITE software (Härkänen, 2003), which incorporates a set of nonparametric Bayesian intensity models, and provides the MCMC procedure for drawing the samples from posterior distribution (Haukka *et al.*, 2016). The CODA software package (Plummer *et al.*, 2006) for R (R Core Team, 2015) was used to assess the convergence of parameter estimates.

4.3.1.3 Simulated data and gbcs dataset

To evaluate the adequacy of the proposed method, survival data were generated by specifying a hazard function incorporating abrupt changes in hazard levels. Five hundred simulated data sets were generated to contain each 5000 individuals with the event or censoring times defined on two time scales.

To illustrate the feasibility of the proposed model, it was fitted to the *gbcs* dataset available on https://www.umass.edu/statdata/statdata/data/ (University of Massachusetts Amherst). The data originate from a trial started in 1984 and conducted by the German Breast Cancer Study Group (Schumacher *et al.*, 1994). The dataset comprises 686 women diagnosed with primary node-positive breast cancer and followed from mastectomy until recurrence (N=299), death or end of follow-up. There are several prognostic factors available in the *gbc*s data, including age at operation and number of lymph nodes involved at the time of operation. To model the hazard rate with respect to age and time since mastectomy, we divided data into ten strata (5-year strata except the youngest <30 and the oldest ≥70), in which the number of patients varied from 6 to 137, and the number of events from 6 to 58. Similarly, to explore the changes in the effect of positive lymph nodes over time, the data were divided into 22 strata (by one node for the first 20 nodes, 21-24 and ≥25 nodes), in which the number of patients varied from 3 to 187, and the number of events from 2 to 59.

4.3.1.4 Evaluation of the method

Several nonparametric Bayesian models with various values for the hyperparameters α and ϕ were applied to 500 simulated datasets, each of which was divided into 15 strata. For each stratum, the posterior expectations of the strata-specific hazards with 95% credible intervals were calculated for 99 equally spaced points. The adequacy of the method was examined by assessing the coverage probability. For each of these 1485 points, the coverage probability was calculated as the proportion of pointwise credible intervals containing the true value. For each model, we reported the proportion of credible intervals for which the coverage probability was within the [0.94, 0.96]. In addition, we assessed the median of mean squared errors (MSEs) obtained for each of the 1485 points. The MSEs were calculated by taking the average of the squared difference between the expectations and the true values of the hazard function.

In addition, we compared the performance of our model with two other methods. The nonparametric Bayesian model with the AG prior was fitted by setting $\phi = 1$ that implies no smoothing over a priori independent strata-specific hazard functions λ_k. In addition, each hazard function λ_k was modelled using the Poisson regression model with the cubic spline for the effect of time.

The proposed method was used to model the hazard rate of recurrence of breast cancer in the *gbcs* dataset (Hosmer et al., 2008). In the first model, the hazard of recurrence was estimated on two time scales, age and time since diagnosis of breast cancer. In the second model, the hazard of recurrence was estimated with respect to the number of lymph nodes involved at the diagnosis and time since the diagnosis of breast cancer.

4.3.2 Application of the method (IV)

4.3.2.1 FinDM cohort

To demonstrate the advantages of exploring the hazard on two or more time scales jointly, the models based on the method from Study III were applied to the data from the nationwide register-based FinDM cohort study. The FinDM study is aimed at monitoring the incidence and prevalence of diabetes and its complications in Finland (Sund and Koski, 2009). In this study, 11,810 individuals diagnosed with T1D before the age of 30 years between 1972 and 1991 were followed up from diabetes onset until the end of follow-up at 31 December 2014 for the incident onset of ESRD or death without ESRD. The relevant time scales to study the hazard of ESRD and death without ESRD included age, diabetes duration and calendar time. Three models were applied to the data to explore the time-dependent dynamics of the hazard.

4.3.2.2 Two-dimensional hazard

The two-dimensional hazard of each outcome was modeled as a function of age and diabetes duration. The former was used to form six diabetes-onset-age strata (0–3, 4–8, 9–13, 14–18, 19–23, 24–29 years). Within the strata, the sequence of event times and size of the risk set at each time point are determined by the time scale that is used in the analysis as a continuous one. The intensity model with the Lexis prior ($\mu =, \alpha_0 = 1, \beta_0 = 1, \alpha = 0.1$ and $\phi = 0.5$) was applied to estimate the stratum-specific hazard functions either as functions of diabetes duration or age.

4.3.2.2 Multiplicative model for two time scales

By assuming the time-scale-specific hazard components to act multiplicatively, the assumption involved in the modelling of time-to-event data with the Cox regression, I studied the individual contribution of diabetes duration and age to the time-dependent dynamics of the hazard of both outcomes. Alternatively, an additive model can be applied. The time-scale-specific components of the two-dimensional hazard of death without ESRD were modelled as a product of two-dimensional function of diabetes duration using broader strata (0–8, 9–18, 19–29 years) and one-dimensional function of age. The individual contribution of diabetes duration and age to the hazard of ESRD was studied by modelling both components as one-dimensional functions.

4.3.2.3 Three-dimensional hazard

The three-dimensional hazard was modelled to explore whether and how the time-dependent dynamics of the two-dimensional hazard vary with the diabetes cohort (1972--1975, 1976--1979, 1980--1983, 1984--1987, 1988--1991). The intensity model with the Lexis prior with two weight parameters ($\mu =, \alpha_0 = 1, \beta_0 = 1, \alpha = 0.1$ and $\phi = (0.5,0.5)$) was applied. In this model, borrowing of power and smoothing within both the diabetes-onset-age strata and diabetes cohorts as well as between the diabetes-onset-age strata and diabetes cohorts.

4.3.2.3 Inference and graphical output

The models were fitted to the data using the BITE software (Härkänen, 2003), the convergence of the MCMC simulation results was assessed using the CODA software package (Plummer *et al*., 2006) for R (R Core Team, 2015). Based on the trace plots, we found the convergence to be reasonable. The *lattice* package (Sarkar, 2008) for R was used to plot the estimated hazard surfaces by means of wireframe plots and heatmaps. The comparison of the stratum-specific hazard rates were based on the visual examination of the 95% credible intervals for overlapping.

5 Results

5.1 Empirical studies (I, II)

5.1.1 Cancer risk and duration of ADM (I)

In the FINRISK cohort, during a median follow-up of 9 years 1301 individuals of 23 394 started ADM. The mean baseline age of users of ADM was 61 years and of non-users 48 years. Table 3 shows other baseline characteristics of the FINRISK participants. Among users, there were more men, more obese (BMI≥30 kg/m^2) individuals and less those who never smoked. The probability of starting ADM was higher for men, overweight (25-29 kg/m^2) and obese individuals and among current and former smokers (Table 3). Similar associations were found between the baseline characteristics and the risk of cancer, when measured using the univariate Poisson regression models.

Table 3 *Distribution of the baseline characteristics as number of individuals (%) of the FINRISK participants divided into users of ADM and non-users.* The crude odds ratio and 95% CI for starting ADM according to the baseline characteristics.

	Users of ADM N = 1 301	Non-users N = 22 093	Crude odds ratio (95% CI)
Sex	764 (58.7)	10 428 (47.2)	reference
Men	537 (41.3)	11 665 (52.8)	0.61 (0.58-0.64)
Women			
FINRISK year			
1997	633 (48.5)	6 928 (31.4)	reference
2002	472 (36.2)	8 256 (38.6)	0.67 (0.64-0.71)
2007	199 (15.3)	6 906 (31.3)	0.47 (0.43-0.52)
BMI (kg/m^2)			
<18.5	-	158 (0.7)	-
18.5-24	96(7.4)	8 026 (36.4)	reference
25-29	458 (35.2)	7 958 (36.0)	5.24 (4.71-5.84)
≥30	725 (55.8)	3 429 (15.5)	20.27 (18.28-22.54)
Missing	22 (1.7)	2 522 (11.4)	1.44 (1.16-1.78)
Smoking			
Never	594 (45.7)	11 525 (52.2)	reference
Former	346 (26.6)	4 861 (22.0)	1.39 (1.31-1.48)
Current	323 (24.8)	5 436 (24.6)	1.12 (1.05-1.19)
Missing	38 (2.9)	271 (1.2)	2.94 (2.52-3.41)
Alcohol consumption			
Non-user	525 (40.3)	7 820 (35.4)	reference
Moderate	499 (38.4)	10 375 (47.0)	0.72 (0.63-0.82)
Heavy	3 136 (14.2)	3 136 (14.2)	0.79 (0.73-1.87)
Missing	762 (3.4)	762 (3.4)	1.36 (1.21-1.52)

In the FINRISK cohort, 1071 individuals were diagnosed with cancer during the follow-up period, 53 cancers occurred in users and 1028 in non-users. For the use and non-use of ADM, there accumulated 5.3 and 192.2 thousand person-years, respectively. The crude IR was 9.93 (95% CI 7.44–13.00) per 1000 person-years of using ADM and 5.35 (95% CI 5.02-5.69) per 1000 person-years not using ADM, yielding a crude RR of 1.86 (95% CI 1.39–2.42). The association attenuated after adjustment for age, sex and calendar time (RR 1.08, 95% CI 0.81–1.42) and further for BMI and smoking (RR 1.01, 95% CI 0.75–1.33).

As compared to non-use, a higher incidence rate was observed for 1–4 years' duration of the use of any ADM (i.e. duration of treated diabetes) yielding the crude RR of 2.44 (95% CI 1.67–3.34). However, the association attenuated after adjustment for calendar time, age and sex (RR=1.47, 95% CI 1.00–2.06) and further after adjustment for baseline BMI and smoking status, (RR 1.37, 95% CI 0.94–1.94). No association was found between cancer risk and durations of the use of ADM longer or shorter than 1–4 years. The crude RR varied with time since initiation of ADM but this variation flattened after adjustment (Figure 6). No association between the cancer risk and various durations of drug use was seen in the similar analyses performed for metformin, sulfonylurea, insulin and any oral ADM.

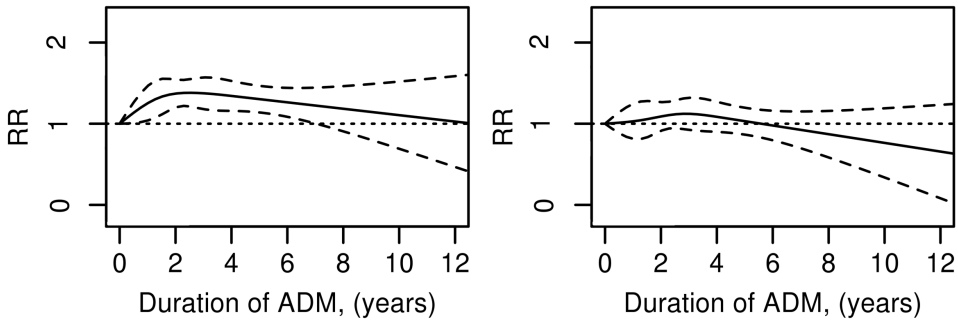

Figure 6 *Rate ratio of cancer for the use of DM as compared to non-use as modelled along time since initiation of the first ADM.*

5.1.2 Cancer risk and use of insulin (II)

The CARING cohort comprised 327,112 new users of any insulin, 66,698 from Norway, 105 945 from Finland, 21,541 from Norway, 85 319 from Sweden, and 47,609 from the UK (CPRD). After the mean follow-up of 4.6 years, there were 212,848, 82,851 and 46,721 ever-users of human insulin, insulin glargine and insulin detemir, respectively.

A total of 1.47 million person-years (54.7% in men) accumulated and 21 390 new cancer cases occurred during the follow-up. There were 2812 prostate, 2423 colorectal, 2311 pancreatic, 2233 lung, 1793 breast (women only), 809 liver, 634 endometrial cancer cases, 612 cases with Non-Hodgkin lymphoma, and 584 with melanoma of skin. Prostate cancer (IR=3.50, 95% CI 3.37–3.63) and breast cancer (2.69, 2.57–2.82) showed the highest IR per 1000 person-years in men and women, respectively, and were followed by colorectal (1.65, 1.58–1.72), pancreatic (1.57, 1.51–1.64) and lung (1.52, 1.46–1.58) cancer. About 32% of

all cancer cases and the majority of pancreatic cancer cases (63%) were diagnosed during the first year of insulin treatment.

In the main analysis performed by sex and cancer site, a few increased and decreased risks but no systematic differences in the risk for studied cancers was found, when comparing the cumulative treatment time (≤0.5, 0.5–1, 1–2, 2–3, 3–4, 5–6, >6 years) on insulin analogue glargine to that on human insulin (Figures 7 and 8). Comparisons of insulin detemir to human insulin and insulin glargine to insulin detemir also showed no consistent differences in incidence rates of sex- and site-specific cancers (results not shown).

In women, we observed an increased risk for colorectal and endometrial cancer for the first half-year, and for melanoma of skin for 2–3 and 4–5 years of the cumulative treatment time on insulin glargine relative to that of human insulin. In men, we observed a decreased risk for pancreatic cancer for 2–3 years, for liver cancer for 3–4 years and >6 years of the cumulative treatment time on insulin glargine relative to that on human insulin.

For any cancer, we found an increased risk in women for 0.5 year and a decreased risk in men for 0.5–1 year, 1–2 years and >6 years of the cumulative treatment time on insulin glargine relative to that on human insulin. The results of the main analysis were robust across a range of sensitivity analyses.

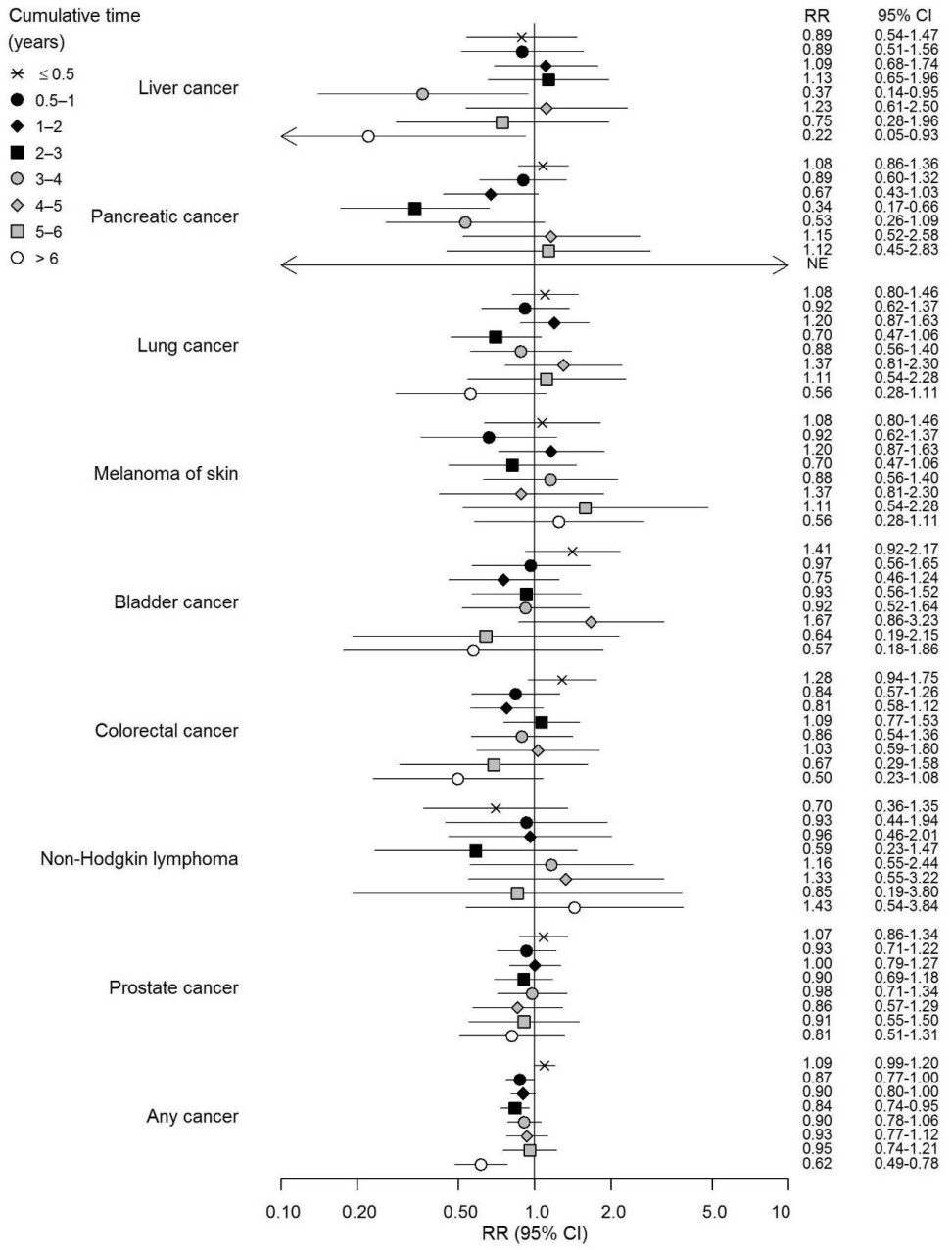

Figure 7 *Adjusted rate ratios (95% CI) of cancer occurrence (eight site-specific cancers and any cancer) in male insulin users when calculated by cumulative treatment time on insulin glargine as compared to human insulin.*

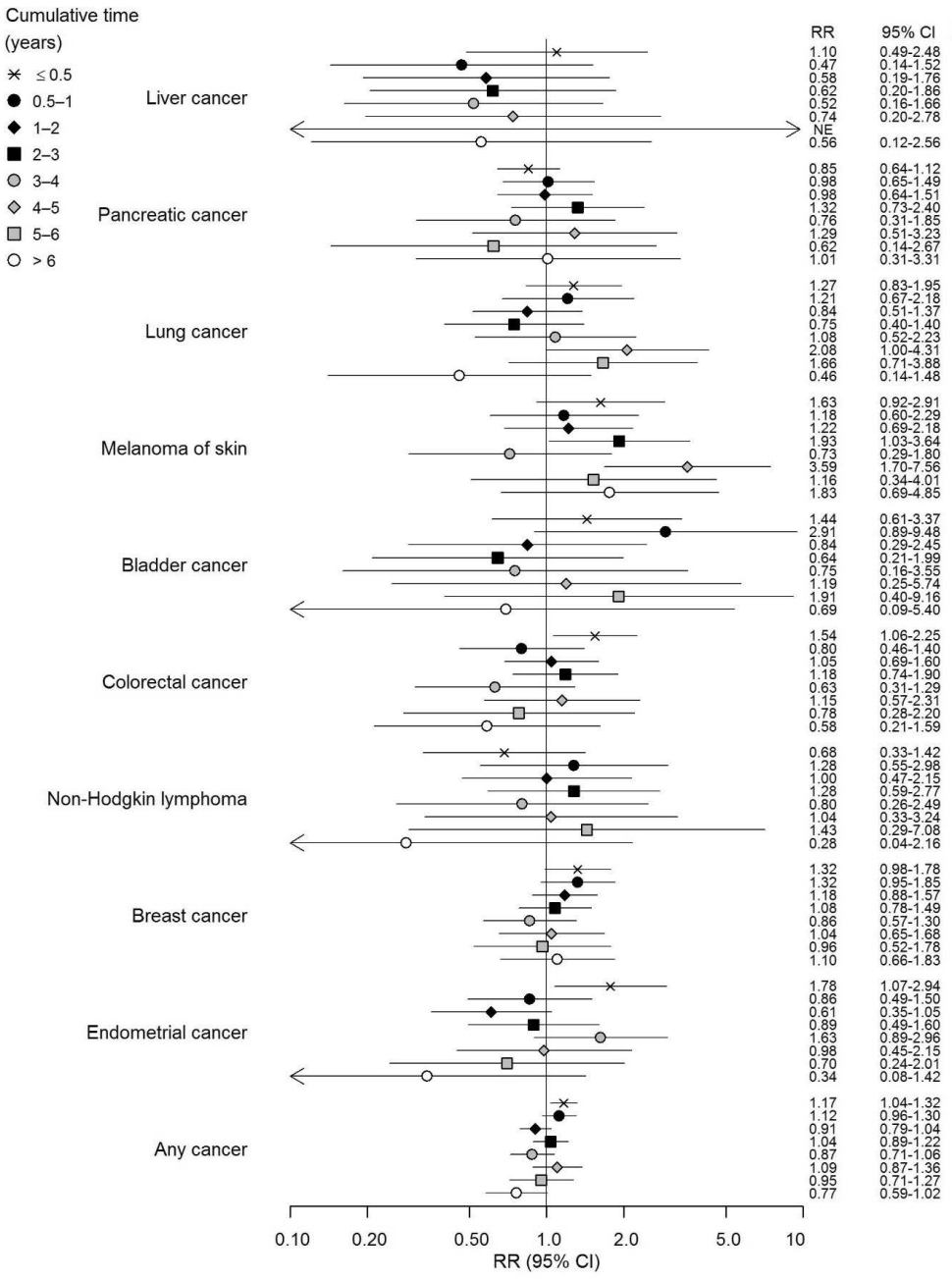

Figure 8 *Adjusted rate ratios (95% CI) of cancer occurrence (nine site-specific cancers and any cancer) in female insulin users when calculated by cumulative treatment time on insulin glargine as compared to human insulin.*

5.2 Bayesian intensity model (III, IV)

5.2.1 Performance and the use of the model (III)

Based on the 95% coverage probability, the Bayesian intensity model with hyperparameters $\alpha = 0.1$ and $\beta = 0.5$ demonstrated the best fit with the coverage probability falling into the interval $[0.94, 0.96]$ in 39% of the examined points. The corresponding figure for the strata-specific Poisson models and the Bayesian model with the AG prior (i.e. without smoothing over strata) was 26% and 6%, respectively.

The results from the Bayesian model depicted how the hazard varies both with the different values of the prognostic factors (i.e. the variation over strata) and along time-on-study (i.e. the variation within strata). The highest hazard of recurrence was associated with the age of 20–30 years at mastectomy (Figure 9). The hazard decreased with older ages at mastectomy, achieving the lowest levels at the age of 45–50 years after which it increased again. A fairly constant hazard of recurrence was seen throughout the follow-up period for all ages at mastectomy except the age of 55–65 years that was associated with the lower hazards levels during the first year after mastectomy. The hazard of recurrence varied with both the number of involved nodes and time since mastectomy, being the lowest during the first year after mastectomy in women with 1–4 involved nodes (Figure 9). Starting from 5 nodes, the hazard of recurrence increased gradually with the increasing number of nodes, achieving the highest value and levelling off around 15 nodes.

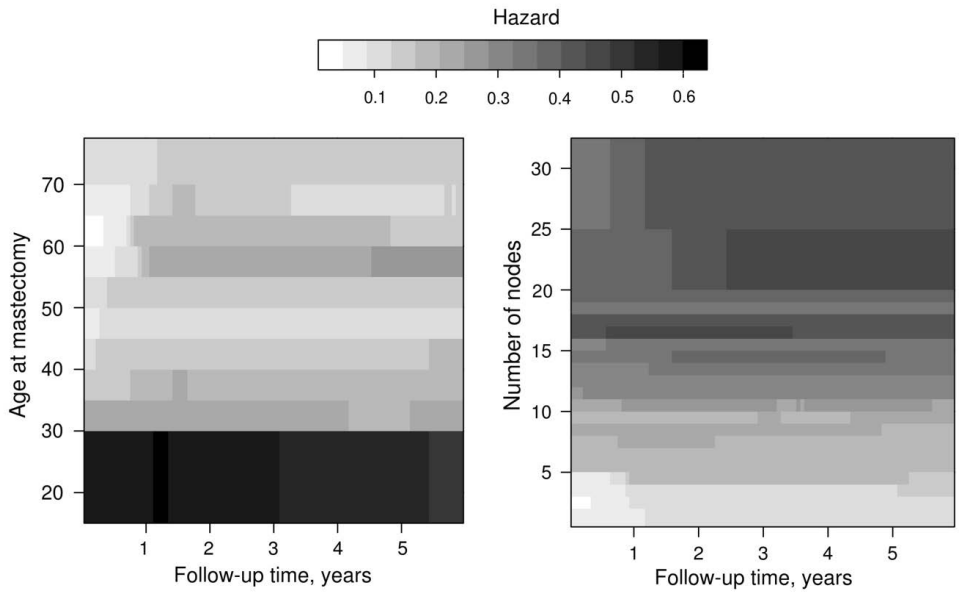

Figure 9 *The posterior expectation of the hazard of recurrence by time since mastectomy and age at mastectomy (left) and by number of positive lymph nodes (right).*

5.2.2 Exploring the multidimensional hazard (IV)

After a mean follow-up of 26.7 years, a total of 338,493 person-years accumulated and 844 individuals developed ESRD, yielding a crude incidence rate of 2.53 (95% confidence interval 2.36-2.71) per 1000 person-years. There were 1905 deaths among persons without ESRD, resulting in a crude mortality rate of 5.71 (95% CI 5.46-5.97) per 1000 person-years.

The hazard of ESRD varied mainly with attained age, while the hazard of death without ESRD was strongly influenced by diabetes-onset age, being higher for older diabetes-onset ages (Figure 10). The most apparent differences in the hazard of both outcomes appeared between individuals with diabetes-onset ages of 9–13 and 24–29 years.

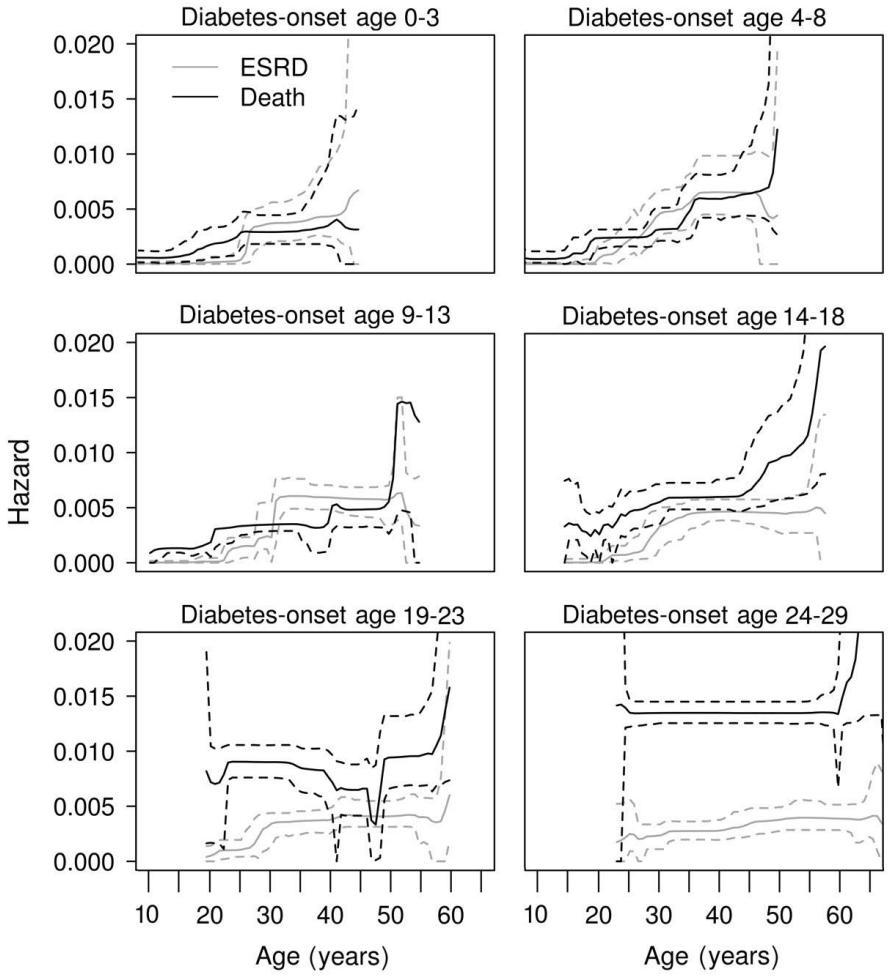

Figure 10 *The posterior expectation of hazard for ESRD (grey line) and death without ESRD (black line) by diabetes-onset-age strata. 95% credible intervals are shown by dashed lines.*

By fitting the multiplicative model, it was possible to separate the hazard components related to diabetes duration and age. For the hazard of ESRD, the former component diverged from zero already at diabetes onset and showed a slow and smooth increase after 15 years of diabetes duration, while the latter component diverged from zero around the age of 20 years and after a steady rise remained constant for 10-15 years (Figure 11). The dynamics of the hazard of death was different across the diabetes-onset-age strata during the initial period of 10-20 years but stabilized at the same level thereafter. For the hazard of death without ESRD, the age-specific component demonstrated a J-curve.

Based on the results from the three-dimensional model (not shown), the hazard of both outcomes also varied with the diabetes cohort.

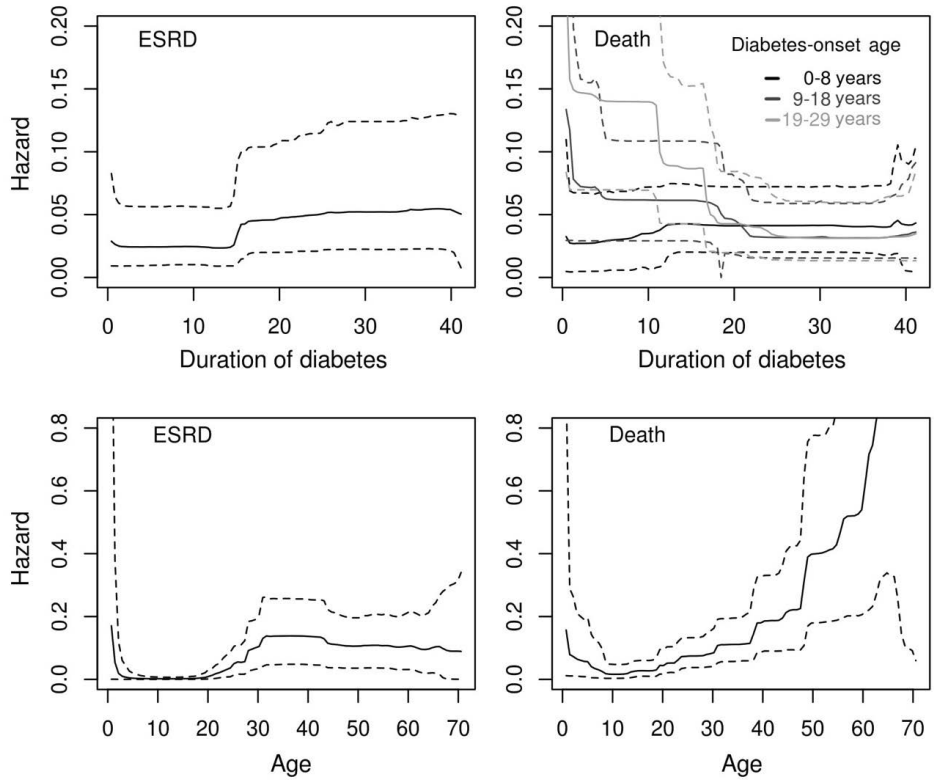

Figure 11 *The posterior expectation of the age-scale-specific components of the hazard of ESRD and death without ESRD with 95% credible intervals (dashed lines) that were estimated by using the multiplicative model. For example, for an individual diagnosed with diabetes at the age of 10 years, the hazard of ESRD at the age of 30 years (i.e. after 20 years since diabetes onset) is calculated by multiplying the corresponding values of the age-specific and diabetes-duration-specific hazard functions. Since the contribution of diabetes duration to the hazard of death without ESRD varied across three diabetes-onset-age strata, the hazard of death for the same individual should be calculated according to his/her diabetes-onset-age stratum (9–18 years).*

6 Discussion

6.1 The use of ADM and cancer risk (I and II)

In the study on the FINRISK cohort, the primary focus was to assess the variation of the cancer risk with respect to time since starting ADMs when accounting for important confounders. Several studies have highlighted the importance of considering time-varying representation of exposure when exploring the link between ADM and cancer risk (Carstensen *et al*., 2012; Renehan, 2012; Walker *et al*., 2013). However, this approach has not been routinely used by studies on the relationship between the use of ADM and cancer risk. Suissa and Azoulay (2012) reviewed 27 observational studies on the relationship between use of metformin and cancer risk and found that 24 studies were subject to time-related biases. Time-related biases can be avoided by applying with new-user design and by using time-varying representation of exposure (Suissa, 2008; Yoshida *et al*., 2015).

There was a considerable discrepancy in the baseline characteristics between the FINRISK participants who started the use of ADM as compared to non-users. We observed that several known risk factors, including age, BMI, smoking and alcohol use, influence in a highly similar way the probability of starting ADM and the risk of cancer. Therefore, these risk factors are likely to confound the relation between the use of ADM and cancer risk.

Only a limited number of observational studies on the link between the use of ADM and cancer risk have accounted for the effect of common risk factors (Habel *et al*., 2013; Onitilo *et al*., 2013; Peeters *et al*., 2015; Peeters *et al*., 2016; Wu *et al*., 2017). Therefore, there is still a need for the further evaluation of the relationship between cancer risk and ADM by accounting for important risk determinants. Recently, a Dutch research group published a study protocol concerning the within-class examination of differences in cancer risk for sulfonylurea treatment in a prospective cohort of patients with T2D (Schrijnders *et al*., 2017). According to the protocol, covariates collected at cohort entry and annually thereafter include among others Hemoglobin A1c, BMI, smoking, diabetes duration, serum creatinine, and use of other ADMs than sulfonylurea.

In the CARING cohort study of 327 112 new insulin users, no consistent differences were found in the risk of any cancer and ten site-specific cancers for the use of insulin glargine or insulin detemir relative to that of human insulin, when examining by the cumulative treatment duration. Of the 136 associations tested in the main analysis, only a few increased and decreased risks were found.

The findings of previous observational studies on the link between cancer risk and use of insulin glargine have been conflicting, at least partly due to methodological shortcomings and biases (Walker *et al*., 2013; Karlstad *et al*., 2013; Wu *et al*., 2016). In addition, most of the previous studies lacked important features considered by Renehan (2012) as important components of an appropriately conducted pharmacoepidemiological study on the link between insulin analogues and cancer risk. Among others, Renehan (2012) listed new-user design, sufficiently long follow-up, evaluation of sex- and site-specific cancer outcomes, time-varying representation of exposure, cumulative duration or dose of insulin treatment.

The results of the CARING study suggested a shift towards increased risk for breast cancer for the initial year of glargine use compared to that of human insulin (RR=1.32, 95% CI 0.98–1.78 ≤0.5 years, RR=1.32, 95% CI 0.95–1.85 for 0.5-1 years of use), possibly suggesting presence of detection or protopathic bias. We found no association between the risk of breast cancer and longer treatment durations with insulin glargine as compared to human insulin.

Two studies on the CPRD cohorts and with a similar study design reported contradictory findings regarding the link between the use of insulin glargine and risk of breast cancer among persons with T2D (Peeters *et al*., 2016; Wu *et al*., 2017). When studying new insulin users, Peeters *et al*. (2016) found no association between the incidence of breast cancer and the overall use (HR=0.99, CI 95% 0.71–1.37) and various cumulative treatment durations of insulin glargine as compared to human insulin. However, among users of insulin glargine with past use of >3 years of other insulins, Peeters *et al*. (2016) observed an increased risk of breast cancer (HR=3.17, 95% CI 1.28–7.84). Wu *et al*. (2017) observed an increased risk of breast cancer for the overall use (HR=1.44, 95% CI 1.11–1.85) and for the treatment duration >5 years (HR= 2.29, 95% 1.26–4.16) of insulin glargine as compared to that of human insulin.

The heterogeneity of the above results may be at least partly due to the complex nature of breast cancer, which is not a single disease but involves different subtypes with the potentially different response to the insulin exposure (Bronsveld et all., 2015). Therefore, further research into the relation between the use of insulin and risk of breast cancer is important. Preferably prospective, large-scale studies with long follow-up should be conducted to examine the relationship separately for premenopausal and postmenopausal breast cancer and also by diabetes type.

6.1.1 Strengths and limitations

In the study on the FINRISK cohorts, we used high-quality register data from the Finnish Cancer Registry and Prescription Register. By using the new-user design and by employing time-varying definition of exposure, we eluded several biases, including prevalent user bias and immortal time bias. The potential for detection bias was addressed by evaluating the variation of the cancer risk by time since initiation of treatment. Moreover, we controlled for confounding through adjustment for known risk factors available from the FINRISK survey.

However, there were also limitations in the study on the FINRISK cohorts. The major limitation was a relatively small number (N=53) of cancer cases in those who started ADM during the follow-up period, precluding examination of the risk of site-specific cancers. In addition, we lacked information on the important confounders, including the type of diabetes and duration of diabetes. In addition, in cohort studies with long follow-up, defining the confounders based on their baseline values may result in misclassification. Therefore, the presence of both residual and unmeasured confounding cannot be ruled out.

The CARING cohort was fivefold in size compared to the largest new-user cohort of those studied up to date (Fagot *et al*., 2013) and had, therefore, enough statistical power for

the assessment of both sex- and ten site-specific cancer outcomes by cumulative treatment durations. By using the new-user active comparator study design and time-varying definition of exposure (Suissa, 2008; Yoshida *et al*., 2015), we avoided and mitigated several biases, including prevalent user bias, immortal time bias, confounding by indication and unmeasured confounding. Moreover, use of similar data sources minimized pitfalls arising in multi-country cohort studies, and application of the same design and analytical approaches across populations provided apparent benefits (Bazelier et al. 2015).

The CARING study, however, involved some limitations. The most substantial limitation was lack of information on confounders, such as smoking, BMI, type and duration of diabetes, comorbidities. To reduce the unmeasured confounding, we adjusted for the duration of insulin-treated diabetes and used the active-comparator design (Yoshida *et al*., 2015). Although we adjusted for country, we could not rule out any potential confounding effects resulting from the differences in insulin user profiles between the countries. In addition, examining a large number of potential associations is likely to produce some false positive results.

In the context of register-based research, which always has some limitations, the internal validity of the CARING study can be considered good. As the CARING study was based on the nationwide (Nordic countries) and representative population-based (CPRD) cohorts, the findings of the study are generalizable to the studied populations.

6.2 The Bayesian intensity model in practice (III and IV)

Application of the Bayesian model to the empirical data yielded reasonable results and provided additional insights to the phenomena of interest.

6.1.1 Application to empirical data

In Study III, I used the *gbcs* dataset to examine the hazard of breast cancer recurrence with the Bayesian intensity model and found a considerable variation in the hazard according to the number of involved positive lymph nodes and age at mastectomy. Similar relationships were found in a study that investigated the functional form of relationships between the prognostic factors and the 5-year recurrence-free survival using the same data from the German Breast Cancer Study (Sauerbrei *et al*., 1999). Sauerbrei *et al*. modelled the average effect of the prognostic factors on the hazard over the 5-year period, whereas estimation of the two-dimensional hazard with the Bayesian model allowed for studying the variation in the hazard as a function of the age or the number of nodes at mastectomy as well as exploring potential changes in this variation over time. Such an analysis corresponds to the evaluation of time-varying effects.

Gray (1992) studied the functional form of the effect of several prognostic factors, including age and the number of positive lymph nodes at diagnosis, on the hazard of breast cancer recurrence in 2,404 women. In the preliminary analysis, Gray addressed potential time-varying effects of the prognostic factors by performing a time-varying coefficient

analysis and observed time-varying effects for all the factors. However, Gray found the variation to be moderate for all the factors except oestrogen receptor status. To account for the nonproportionality of the effect of oestrogen receptor status, Gray used proportional hazards models with strata by oestrogen receptor status and assessed the average effect of the other factors over the period of 12 years. Natarajan *et al*. (2009) studied the effect of nine predictor factors on the recurrence-free survival in a cohort of 3,088 breast cancer patients by applying the Gray's 10-knot spline models. Natarajan *et al*. found time-varying effects for oestrogen receptor status, with the assumption of proportional hazards being violated for the later.

With the Bayesian intensity model, the effect of variables on the hazard can be explored or accounted for in two different ways. In fact, the baseline hazard can be modelled as a function of time (measured on one of the involved time scales) and one or more ordinal variables. When studying the effect of the determinant of interest, the usual practice is to adjust for potential confounders, such as sex and baseline age. Instead of adjusting, these variables can be used to model the multidimensional baseline hazard. It is also possible to explore whether and how the baseline hazard is modified by the effect of other variables by including them in the Bayesian model as covariates. Then, the (multidimensional) baseline hazard and covariate effect can be combined in a multiplicative manner similar to the that of the Cox model or in an additive manner. Importantly, estimation of these effects with the Bayesian intensity model involves no assumption of proportional hazards.

Moreover, because of the multidimensional smoothing and borrowing of strength, which are built-in features of the Bayesian model, it yields accurate results even when the data are limited. For instance, the cohorts studied by Gray (1992) and Natarajan *et al*. (2009) were three to four times as large as compared to the *gbcs* dataset that included 686 women. Estimation of the two-dimensional hazard with the Bayesian model yielded, however, reasonable results. Indeed, based on the comparisons in Study III, the Bayesian model outperforms the methods that approximate the two-dimensional hazard by a collection of one-dimensional functions without smoothing and borrowing of strength.

In Study IV, by applying the Bayesian models to the data on individuals with T1D from the FinDM study, I studied the time-dependent dynamics of the hazard of ESRD and death without ESRD on two and three time scales jointly. The results are consistent with previous studies but also provide an additional insight to the nature of previous findings.

I used the two-dimensional model to depict the variation of the hazard according to diabetes duration and diabetes-onset age. For ESRD, I observed the similarly shaped stratum-specific hazards, which, however, involved the lag period of different length and differed slightly in their magnitudes. For death without of ESRD, the stratum-specific hazards differed by both their shape and magnitude.

When trying to interpret the results from such a model, it should be noted that the time scales are not truly different variables but are measuring time from different origins and time proceeds on all time scales at the same pace. In the two-dimensional model, one of the time scales is used to form the strata of life lines and another to proceed along life lines. Obviously, as the follow-up of an individual proceeds along the time scale that is used as a continuous one, the same amount of time proceeds on all time scales. For instance, the more time elapses from diabetes onset the older individual becomes. Although the hazard is being

assessed as a function of diabetes duration, age is likely to contribute to the observed variation. Therefore, the two-dimensional model provides a descriptive rather than analytical tool. However, it is important to explore the time-to-event data in a such way because the hazard pattern itself may guide further analysis.

I explored the individual role of each time scale with the multiplicative model. Based on these results, diabetes duration and attained age but not diabetes-onset age contributed to the variation in the two-dimensional hazard of ESRD. In contrast, the two-dimensional hazard of death without ESRD was influenced by all the three time determinants, suggesting an interaction between diabetes-onset age and diabetes duration.

The individual contribution of diabetes duration is likely to reflect an impact of the increasing cumulative glycaemic burden (Writing Team for the Diabetes Control and Complications Trial, 2003; de Boer, 2011). In addition, the hazard of ESRD varied with attained age but not with diabetes-onset age. Findings of previous studies have, however, suggested that diabetes-onset age is an important determinant for the hazard of ESRD (Finne *et al*., 2005, Svensson *et al*., 2006; Costacau and Orchard, 2017; Gagnum *et al*., 2017; Helve *et al*., 2017). I observed the longer lag periods for younger diabetes-onset ages and the similarly shaped stratum-specific hazards. These results suggest that the age-specific hazard is intensified by the effect of cumulative glycaemic burden and the differences attributed to diabetes-onset age arise from the differences in diabetes duration.

There were differences in the hazard of death without ESRD between the diabetes-onset-age strata. During the first 15-25 years after diabetes onset, the hazard was increasing among those with younger diabetes-onset age (0–8 years) and was decreasing among individuals with older diabetes-onset age. The study on the cohort of Finnish T1D patients without albuminuria observed that during the first 10 years after diabetes diagnosis the excess mortality was driven by acute complications (Groop *et al*., 2018). The study on mortality before age of 30 years among patients with childhood-onset T1D reported that acute complications were the main cause of excess mortality (Wasag *et al*. 2018). In addition, Wasag *et al*. (2018) reported an increasing risk of death with increasing diabetes-onset-age and older age during the follow-up. A J-shaped age-specific hazard, which was observed for death without ESRD when modelled with the multiplicative model, is a typical pattern seen for all-cause age-specific mortality in the populations of developed countries (Siegel 2012, p. 80–82).

Based on the results from three-dimensional model, which was based on diabetes duration, diabetes-onset age and calendar time, the hazard of ESRD and death without ESRD also varied with diabetes cohort. I observed a delayed onset of ESRD and a decreasing hazard of death without ESRD for the latter cohorts.

Along with the empirical questions, examination of the hazard using the Bayesian model allows for addressing methodological issues, such as the choice of the primary time scale and competing risks. So far, the focus has been on the evaluation of the impact of the primary time scale on the hazard ratio instead of exploring the actual contribution of each time scale to the hazard. As a result, the recommendations have been contradictory (Korn *et al*., 1997; Thiébaut and Bénichou, 2004; Pencina *et al*., 2007; Cologne *et al*., 2012; Chalise *et al*., 2016). Obviously, there is no single time scale that fits all situations, because different phenomena involve different time scales and the contribution of each time scale

and their interplay is not uniform across phenomena. Modelling the hazard on two or more time scales jointly avoids confusion regarding the choice of time scale, reduces the potential for the misspecification and inferior performance of the model as well as allows for the informed choice of the most informative time scale, if is of interest. Previous epidemiological studies on the risk of ESRD have accounted for death as a competing event by using the competing risks model by Fine and Gray (1999) which is based on the subdistribution hazards and estimates the cumulative hazard of the event of interest by conditioning on the risk of the competing event. As pointed out by Andersen *et al*. (2012), it is important to report both the cumulative hazard from the subdistribution hazards analysis and the actual hazard rates of both the event of interest and a competing event.

6.1.2 Other methods

To our knowledge, there is no methods for the estimation of multidimensional hazard function equivalent to that proposed in this study. Some generic methods can be developed using survival analysis methods tailored for the estimation of one-dimensional hazard. In study III, I approximated the two-dimensional hazard function by a collection of one-dimensional hazard functions modelled using Poisson regression models with splines for the effect of time or the Bayesian intensity models with the AG priors. The two-dimensional model outperformed these methods, suggesting that smoothing and borrowing the strength, the built-in features of the proposed model, are crucial to the multidimensional modelling. In addition, since the nonparametric Bayesian model is based on piecewise constant hazard functions, it adapts to data flexibly and is advantageous for the modelling of the hazard with a complex time-dependent dynamics and for detection of change points.

6.1.3 Future development

Application of the proposed method demonstrated that it can be used as a powerful tool which that allows for exploring the hazard with a complex time-dependent dynamics. At present, the proposed method lacks desirable statistical features that allows for making inference from the model. There are several directions for the further development of the proposed method.

Data modelling can be considered as an iteration of four steps: (1) model building, (2) model assessment, (3) model inference, (4) prediction. Selection of the model and assessment of the model accuracy are therefore essential steps towards the model inference. Model selection involves comparison of the fitted models. The general approach for assessment and comparison of the fitted models is to evaluate their predictive accuracy. Out-of-sample checks using within-sample fits are employed for this purpose. Out-of-sample accuracy can be estimated through the cross-validation or by means of an information criterion.

Among various methods proposed for the model comparison and assessment in the Bayesian framework, the deviance information criterion (DIC) is the most popular option

(Spiegelhalter *et al*., 2002). Popularity of this method stems from its simplicity and link with cross-validation, which make the method advantageous in practical use. However, this option is not applicable to nonparametric models due to the poor performance of DIC in more complex models (Plummer, 2008). According to Gelman et al. (2014), DIC provides unreasonable results in the models yielding the posterior distribution, which is not well summarized by its mean.

There are several other options. Watanabe–Akaike information criterion or widely applicable information criterion (WAIC). WAIC can be viewed as an improvement of the DIC, because it closely approximates Bayesian cross-validation, and, unlike DIC, is invariant to parametrization and also works for singular models (Watanabe, 2010; Watanabe, 2013). Gelman *et al*. (2014) pointed that data partition, on which WAIC relies, is likely to cause difficulties with structured models such as spatial models. Another option is Bayesian cross-validation methods, including k-fold cross-validation and leave-one-out (LOO) cross-validation (Vehtari and Ojanen, 2012; Gelman *et al*., 2014). LOO cross-validation is based on the log-likelihood evaluated at the posterior simulations of the model parameters. For the proposed model, LOO cross-validation appears to be the most appealing alternative. Vehtari *et al*. provided (2017) a fast and stable computation for LOO cross-validation in the case of Monte Carlo posterior inference. Based on this method, Vehtari *et al*. (2018) have developed a *loo* R package, which provides the efficient approximate LOO cross-validation.

Various statistical tests are used to draw inference from the model. In the case of our model, statistical inference would concern testing for differences in the hazard between the strata and testing for trend within the strata. A preliminary evaluation of differences in the hazard between the strata can be done by using the graphical output and based on credible intervals. Overlapping credible intervals suggest that data do not provide sufficient evidence for differences, whereas the absence of overlap advocates an opposite conclusion. Preliminary checks for an increasing (decreasing) trend within a stratum can be done by calculating the posterior probability of increasing (decreasing) trend. Although, these preliminary checks are informative, statistical testing is needed for further examination.

Last but not the least, statistical programs or packages are needed to perform modeling of the time-to-event data using our method. The BITE software (Härkänen 2003) is an already available tool, incorporating a set of nonparametric Bayesian models for the estimation of both one-dimensional and multidimensional intensities and providing the Markov chain Monte Carlo (MCMC) procedure for drawing samples from the posterior distribution. The software is freely available and can be downloaded from the BITE website (Haukka et al. 2016). The reference manual for BITE provides instructions on the installation and running BITE, as well as some background theory and model examples (Härkänen and But 2016). In the R statistical environment (R Core Team 2017), a user-friendly package can be developed based on the BITE software. This would allow linking BITE with other useful R packages such as the CODA package (Plummer et al 2006) for the assessment of the convergence of MCMC iterates, the *lattice* package (Sarkar 2008) for visualization of results with three-dimensional plots and heatmaps, the *loo* R package, the *loo* package for the LOO cross-validation (Vehtari *et al*., 2018).

7 Conclusions

Given an increasing prevalence and incidence of diabetes, I have addressed an important clinical and public health question whether the use of ADM influences the risk of cancer. As there is a variety of different glucose lowering medications and even more different cancer types, answering this question is not straightforward and requires a rigorous and systematic research. I have contributed to the research on this topic by addressing a specific question whether some of commonly used insulin treatments should be preferred over others as safer with respect to the cancer risk.

In the rigorously performed five-country CARING study, no persistent differences in the risk for the ten cancers and any cancer was found, when comparing the cumulative use of insulin analogues glargine or detemir to that of human insulin. These results add to the conclusive evidence on the absence of the relationship between the cancer incidence and use of insulin analogues at follow-up exceeding five years.

I have learned that there are several simple but effective methods by applying which it is possible to avoid or mitigate bias in observational pharmacoepidemiological studies. These methods include new-user active-comparator study design and time-varying definition of exposure. It is also important to account for confounding whenever it is possible.

I have demonstrated that analysis of the time-to-event data on multiple time scales jointly may provide an additional insight to the real-life phenomenon. Cohort studies on chronic diseases with long follow-up and multiple time scales would benefit from the use of nonparametric Bayesian intensity model that was introduced in this work. It is also possible to study the time-varying effects of prognostic factors with the method. Overall, the proposed approach provides an appealing and flexible framework for modelling time-to-event data on multiple time scales.

Acknowledgments

In 2012 I became a doctoral student in the doctoral programme in population health, Faculty of Medicine, University of Helsinki. One year earlier, in 2011, I graduated with a Master degree in biometry, statistics, from the University of Helsinki, where I started studies in 2001 at the Faculty of Mathematics. In 1996 I met my husband, 1998 I had my first child, 2007 my second child and 2009 my third child. In 1993 I came to Finland with my mother and sister, my father followed us two years later. I was born in 1977. Why am I going back in my history? It is a kind of prospective study and I shall analyze these observations.

Each of the listed years is the date when an important event took place in my life. Each of these events gave a rise to something new, started a new era, created a new time scale. Time, timelines and time scales are an exciting matter to think about and to study. This can be done in many ways, including empirical, philosophical, mathematical and statistical.

I am sure a game of chance has played a huge role in what has happened to me so far and I am grateful for the strokes of fortune. Since these events have taken place within one individual, these cannot be assumed to be independent of each other. Perhaps, the pattern reflects my personality but I am not, however, going to analyze myself because such an evaluation will be biased (towards or away from null) any way. Instead, I would like to convey my gratitude to people who have contributed to this story.

I would like to express my deepest gratitude to my supervisors Jari Haukka and Tommi Härkänen for their invaluable input. Under Jari's supervision I have had an opportunity to grow as a researcher at my own pace but with the confidence that help, good advice and experience of life is always available upon asking. Tommi has presented me challenges, shared his knowledge and practical experience and guided me gently but firmly on my path.

I owe my gratitude to all the co-authors of the original publications. This thesis would not have been possible without their input. It was the pleasure to work with all the collaborators form the CARING project, the team of warm people and skilled researchers coordinated by Professor Marie Luise De Bruin. I also sincerely thank those who provided me the data: Satu Männistö for providing the FINRISK data, Reijo Sund and Martti Arffman for providing the FinDM data and the collaborators from the CARING project for acquiring and preparing the data from the Nordic National registries and the CPRD.

I also thank my thesis committee members, Professor Jukka Corander and Professor Nea Malila for willing to be involved. I sincerely thank Professor Peter Muller for helpful comments on the draft of the third publication. I greatly appreciate the precise and educational comments and thoughts the preliminary examiners, Professor Kari Auranen and Docent Maarit Korhonen, provided me with.

There would not have been this thesis without the financial support. The CARING project was funded by the European Community's Seventh Framework Programme (FP-7). My work on finalizing the thesis was supported by Finnish Cultural Foundation.

During these years, I has been lucky to work at the Department of Public Health alongside the great people and experienced researchers. I am grateful to Professor Ossi Rahkonen for encouraging, asking questions and giving advices. It has been the pleasure to work with my colleague Paula Bergman who has an enthusiastic and professional grasp on work. I am thankful to Paula for having a good laugh almost every day and for interesting

conversations and I am sorry for my absences and for being sometimes absent even when present during the last year.

There would not be me like I am without my family. I am grateful to my husband for his daily care to me and our children and for his support during this long way. My children have always motivated me to go ahead, to learn new things and to grow. They have kept my feet steadily on the ground but also energized me and have made sure that my inner child is alive and thriving with them. My sister has been a peer support of mine, we share not only many genes but also many thoughts. My parents have been overwhelmingly supportive, understanding and loving throughout my life. They have taught me invaluable lessons about life, including love, patience, parenting and work.

Helsinki, 11 November 2018

Anna But

References

Aalen O (1978). Nonparametric inference for a family of counting processes. *The Annals of Statistics* **6**:701–726.

American Diabetes Association (2016). Standards of medical care in diabetes—2016. *Diabetes Care* **39**(supp 1).:S1–S109.

Andersen PK, Borgan Ø, Gill RD and Keiding N (1993). *Statistical models based on counting processes*. Springer Series in Statistics, Springer-Verlag, New York.

Andersen A, Barlow L, Engeland A, Kjaerheim K, Lynge E and Pukkala E (1999). Work-related cancer in the Nordic countries. *Scandinavian Journal of Work, Environment & Health* **25**(suppl 2):1–116.

Andersen PK, Geskus RB, and de Witte Hein Putter (2012). Competing risks in epidemiology: possibilities and pitfalls. *International Journal of Epidemiology* **41**(3):861–870.

Andersen M (2014). Research on drug safety and effectiveness using pharmacoepidemiological databases. *Journal of Internal Medicine* **275**(6):548–50.

Andersson TM, Weiderpass E, Engholm G, Lund AQ, Ólafsdóttir E, Pukkala E, Stenbeck M and Storm H (2017). Avoidable cancer cases in the Nordic countries - the impact of overweight and obesity. *European Journal of Cancer* **79**:106–118.

Andrieu C, Doucet A and Robert CP (2004). Computational advances for and from Bayesian Analysis. *Statistical Science* **19**(1):118–127.

Arffman M, Forssas E, Ruuth I, Keskimäki I, Manderbacka K and Sund R (2014). Changes in multimorbidity among people with diabetes in Finland 1990-2011. *European Journal of Public Health*. **24** doi:cku165–084, https://doi.org/10.1093/eurpub/cku165.084

Arjas E and Gasbarra D (1994). Nonparametric Bayesian inference from right censored survival data, using the Gibbs sampler. *Statistica Sinica* **4**(2):505–524.

Arthur WB and Vaupel JW (1984). Some general relationships in population dynamics. *Population Index* **50**(2): 214–226.

Ashby D (2006). Bayesian statistics in medicine: a 25 year review. *Statistics in Medicine* **25**:3589–3631.

Atkinson MA, Eisenbarth GS and Michels AW (2014). Type 1 diabetes. *Lancet* **383**(9911):69–82.

Badrick E and Renehan AG (2014). Diabetes and cancer: 5 years into the recent controversy. *European Journal of Cancer* **50**:2119–2125.

Ballotari P, Vicentini M, Manicardi V, Gallo M, Ranieri SC, Greci M and Rossi PG (2017).Diabetes and risk of cancer incidence: results from a population-based cohort study in northern Italy. *BMC Cancer* **17**:703.

Baricevic I, Jones DR, Roberts DL, Lutzen A, Lundby A, Worm J, Hansen BF and Renehan AG (2015). A framework for the in vitro evaluation of cancer-relevant molecular characteristics and mitogenic potency of insulin analogues. *Carcinogenesis* 36:1040–1050.

Bayarri MJ and Berger JO (2004). The interplay of Bayesian and frequentist analysis. *Statistical Science* **19**:58–80

Bazelier MT, Eriksson I, de Vries F, Schmidt MK, Raitanen J, Haukka J, Starup-Linde J, De Bruin ML and Andersen M (2015). Data management and data analysis techniques in pharmacoepidemiological studies using a pre-planned multi-database approach: a systematic literature review. *Pharmacoepidemiology and Drug Safety* **24**(9):897–905.

Bell A and Jones K (2015). Age, period and cohort processes in longitudinal and life course analysis: a multilevel perspective. In: Burton-Jeangros C, Cullati S, Sacker A, Blane D

(eds) *A Life Course Perspective on Health Trajectories and Transitions*. Life Course Research and Social Policies, vol 4. Springer, Cham.

Benichou J and Palta M (2014). Rates, risk, measures of association and impact. In Ahrens W and Pigeot I (Ed.), *Handbook of epidemiology*. 2nd Edition. New-York, Springer.

Berger J (2006). The Case for Objective Bayesian Analysis. *Bayesian Analysis* **1**:385–402.

Berger U, Schäfer J and Ulm K (2003). Dynamic Cox modelling based on fractional polynomials: time-variations in gastric cancer prognosis. *Statistics in Medicine* **22**(7).1163–1180.

Berzuini C and Clayton D (1994). Bayesian analysis of survival on multiple time scales. *Statistics in Medicine* **13**(8):823–838.

Berzuini C, Clayton D and Bernardinelli L (1993). Bayesian inference on the Lexis diagram. *Bulletin of the International Statistical Institute*, **55**, 149–164.

Boggon R, van Staa TP, Chapman M, Gallagher AM, Hammad TA and Richards MA (2013). Cancer recording and mortality in the general practice research database and linked cancer registries. *Pharmacoepidemiology and Drug Safety* **22**:168–175.

Booth CM and Tannock IF (2014). Randomized controlled trials and population-based observational research: partners in the evolution of medical evidence. *British Journal of Cancer* **110**:551–555.

Borodulin K, Tolonen H, Jousilahti P, Jula A, Juolevi A, Koskinen S, Kuulasmaa K, Laatikainen T, Männistö S, Peltonen M, Perola M, Puska P, Salomaa V, Sundvall J, Virtanen M and Vartiainen E (2017) Cohort Profile: The National FINRISK Study. *International Journal of Epidemiology* doi:10.1093/ije/dyx239.

Brennan P, Perola M, van Ommen GJ and Riboli E; European Cohort Consortium (2017). Chronic disease research in Europe and the need for integrated population cohorts. *European Journal of Epidemiology* **32**(9): 741–749.

Breslow NE and Day *NE* (1987). *Statistical methods in cancer research*, Vol. II: *The design and analysis of cohort studies*. IARC Scientific Publications 82, International Agency for Research on Cancer, Lyon.

Brinks R, Landwehr S, Fischer-Betz R, Schneider M and Giani G (2014). Lexis diagram and illness-death model: simulating populations in chronic diseases epidemiology. *PloS One* **9**(9): e106043

Bronsveld HK, ter Braak B, Karlstad Ø, Vestergaard P, Starup-Linde J, Bazelier MT, De Bruin ML, de Boer A, Siezen CL, van de Water B, van der Laan JW and Schmidt MK (2015). Treatment with insulin (analogues) and breast cancer risk in diabetics; a systematic review and meta-analysis of in vitro, animal and human evidence. Breast Cancer Research **5**;17:100. doi: 10.1186/s13058-015-0611-2.

Cain KC, Harlow SD, Little RJ, Nan B, Yoseff M, Taffe JR and Elliott M (2011). Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *American Journal of Epidemiology* **173**(9):1078–1084.

CARING Consortium (2015). Common study protocol version 3.3. Available from www.encepp.eu/encepp/openAttachment/fullProtocolLatest/963

Carstensen B (2005). Demography and epidemiology: practical use of the Lexis diagram in the computer age or who needs the Cox-model anyway? Research report (Department of Biostatistics, University of Copenhagen) 06/2. Available from http://biostat.ku.dk/reports/2006/rr-06-2.pdf

Carstensen B (2007). Age-Period-Cohort models for the Lexis diagram. *Statistics in Medicine* **26**(15):3018–3045.

Carstensen B. (2012). DCRC: Tabulation outline. Steno Diabetes Center, Gentofte, Denmark http://www.bendixcarstensen.com/DCRC/Tabulate/tabulate.pdf: Diabetes and Cancer Reasearch Consortium.

Available: http://www.bendixcarstensen.com/DCRC/Tabulate/tabulate.pdf.

Carstensen B, Witte DR and Friis S (2012). Cancer occurrence in Danish diabetic patients: duration and insulin effects. *Diabetologia* **55**:948–958.

Carstensen B, Read SH, Friis S, Sund R, Keskimäki I, Svensson AM, Wild SH, Kerssens JJ, Harding JL, Magliano DJ and Gudbjörnsdottir S (2016). Diabetes and Cancer Research Consortium. Cancer incidence in persons with type 1 diabetes: a five-country study of 9,000 cancers in type 1 diabetic individuals. *Diabetologia* **59**(5):980–988.

Chalise P, Chicken E and McGee D (2016). Time scales in epidemiological analysis: an empirical comparison. *International Journal of Statistics and Probability* **5**(3):91–101.

Chien LC, Wu YJ, Hsiung CA, Wang LH, Chang IS (2015). Smoothed Lexis diagrams with applications to lung and breast cancer trends in Taiwan. Journal of the American Statistical Association **110**:1000–1012.

Christiansen MC, Niemeyer A and Teigiszerova L (2015). Modeling and forecasting duration-dependent mortality rates. *Computational Statistics & Data Analysis* **23**:65–81.

Clayton D and Hills M (1993). *Statistical models in epidemiology*. Oxford University Press, Oxford.

Clayton D and Schifflers E (1987a). Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *Statistics in Medicine* **6**(4):449–467.

Clayton D and Schifflers E (1987b). Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine* **6**(4):469–481.

Colhoun HM, SDRN Epidemiology Group (2009). Use of insulin glargine and cancer incidence in Scotland: a study from the Scottish diabetes research network epidemiology group. *Diabetologia* **52**:1755–1765.

Cologne J, Hsu WL, Abbott RD, Ohishi W, Grant EJ, Fujiwara S and Cullings HM (2012). Proportional hazards regression in epidemiologic follow-up studies: an intuitive consideration of primary time scale. *Epidemiology* **23**:565–573. doi: 10.1097/EDE.0b013e318253e418

Commenges D, Letenneur L, Joly P, Alioum A and Dartigues JF (1998). Modelling age-specific risk: application to dementia. Statistics in Medicine **17**(17):1973–1988.

Concato J, Shah N and Horwitz RI (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *New England Journal of Medicine* **342**:1887–1892.

Costacau T and Orchard TJ (2018). Cumulative kidney complication risk by 50 years of type 1 diabetes: the effects of sex, age, and calendar year at onset. *Diabetes Care* **41**(3):426–433.

Cox DR (1972). Regression models and life tables (with discussion). *Journal of Royal Statistical Society* Series B **34**:187–220.

Cox DR (1975). Partial likelihood. *Biometrika* **62**(2): 269-276.

Craigmile PF, Calder CA, Li H, Paul R and Cressie N (2009). Hierarchical model building, fitting, and checking: a behind-the-scenes look at a Bayesian analysis of arsenic exposure pathways. *Bayesian Analysis* **4**(1):1–36.

Currie CJ, Poole CD, Gale EAM (2009). The influence of glucose-lowering therapies on cancer risk in type 2 diabetes. *Diabetologia* **52**:1766–1777.

Danaei G, Tavakoli M and Heman MA (2012). Bias in observational studies of prevalent users: lessons for comparative effectiveness research from a meta-analysis of statins. *American Journal of Epidemiology* **175**(4):250–262.

de Boer IH, Sun W and Cleary PA; The DCCT/EDIC Research Group (2011). Intensive diabetes therapy and glomerular filtration rate in type 1 diabetes. *New England Journal of Medicine* **365**(25):2366– 2376. doi:10.1056/NEJMoa1111732.

De Brujin KM, Ruiter R, de Keyser CE, Hofman A, Stricker BH and van Eijck CH (2014). Detection bias may be the main cause of increased cancer incidence among diabetics: results from the Rotterdam Study. *European Journal of Cancer* **50**:2449–2455.

Delgado-Rodriguez M and Llorca J (2004). Bias. *Journal of Epidemiology and Community Health* **58**:635–641.

Dickman P (2014). Survival analysis, overview. Available from http://www.pauldickman.com/survival/handouts/21%20%20EoB%20Survival%20analysis%20overview.pdf.

Doll R (2001a). Cohort studies: history of the method. I. Prospective cohort studies. *Soz Praventivmed* **46**(2):75–86.

Doll R (2001b). Cohort studies: history of the method. II. Retrospective cohort studies. *Soz Praventivmed* **46**(3):152–160.

Dunson DB (2001). Commentary: practical advantages of Bayesian analysis of epidemiologic data. *American Journal of Epidemiology* **153**(12):1222–1226

Durrleman S and Simon R (1989). Flexible regression models with cubic splines. *Statistics in Medicine* **8**:551–561.

Efron B (2005). Bayesians, frequentists, and scientists. Division of Biostatistics, Stanford University. Available from https://courses.physics.ucsd.edu/2015/Fall/physics210b/REFERENCES/Efron_Bayesians_Frequentists.pdf

ENCePP (2018). ENCePP Guide on methodological standards in Pharmacoepidemiology. 7th Revision. Available from http://www.encepp.eu/standards_and_guidances/methodologicalGuide.shtml.

Engholm G, Ferlay J, Christensen N, Bray F, Gjerstorff ML, Klin A, Køtlum JE, Ólafsdóttir E, Pukkala E and Strom HH (2010). NORDCAN – a Nordic tool for cancer information, palnning, quality control and research. *Acto Oncologica* **49**(5):725-36. doi: 10.3109/02841861003782017

Etzioni RD and Kadane JB (1995). Bayesian statistical methods in public health and medicine. *Annual Review of Public Health* **16**:23–41.

Fagot JP, Blotière PO, Ricordeau P, Weill A, Alla F and Allemand H (2013). Does insulin glargine increase the risk of cancer compared with other basal insulins? *Diabetes Care* **36**:294–301.

Farewell V and Cox D (1979). A note on multiple time scales in life testing. *Journal of the Royal Statistical Society Series C (Applied Statistics)* **28**:73–75.

Fewell Z, Davey SG and Sterne JA (2007). The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. American Journal of Epidemiology **166**(6):645–655.

Fienberg SE (2006a). When did Bayesian inference become "Bayesian". *Bayesian Analysis* **1**:1–40

Fienberg SE (2006b). Does it Make Sense to be an "Objective Bayesian"? (Comment on Articles by Berger and by Goldstein). *Bayesian Analysis* **1**:429–432

Fine JP and Gray RJ (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of American Statistical Association* **94**:496–509.

Finne P, Reunanen A, Stenman S, Groop P-H and Grönhagen-Riska C (2005). Incidence of end-stage renal disease in patients with type 1 diabetes. *JAMA* **294**(14):1782–1787.

Fisher L and Lin L (1999). Time-dependent covariates in the Cox proportional hazards regression model. *Annual Reviews in Public Health* **10**:145–157.

Forsblom C, Harjutsalo V, Thorn LM, Wadén J, Tolonen N, Saraheimo M, Gordin D, Moran JL, Thomas MC and Groop PH; FinnDiane Study Group (2011) Competing-risk analysis of ESRD and death among patients with type 1 diabetes and macroalbuminuria. *Journal of the American Society of Nephrology* **22**(3):537–44.

Forssas E, Arffman M, Manderbacka K, Keskimäki I, Ruuth I and Sund R (2016). Multiple complications among people with diabetes from Finland: an 18-year follow-up 1994–2011. *BMJ Open Diabetes Res Care*. **4**(1):e000254

Fowler MJ (2008). Microvascular and macrovascular complications of diabetes. *Clinical Diabetes* **26**(2):77–82.

Frome EL (1983). The analysis of rates using Poisson regression models. *Biometrics* **39**:665–674.

Frost WH (1933). Risk of persons in familial contact with pulmonary tuberculosis. *American Journal of Public Health* **23**:426–32.

Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB and Sørensen HT. (2010). The Nordic countries as a cohort for pharmacoepidemiological research. *Basic Clin Pharmacol Toxicol* **106**(2):86–94.

Gagnum V, Saeed M, Stene LC, Leivestad T, Joner G and Skrivarhaug T (2017). Low incidence of end-stage renal disease in childhood-onset type 1 diabetes followed for up to 42 years. *Diabetes Care* doi: 10.2337/dc17-0906

Gallagher EJ and LeRoith D (2015). Obesity and diabetes: the increased risk of cancer and cancer-related mortality. *Physiological Reviews* **95**:727–748.

Garbe E and Suissa S (2014). Pharmacoepidemiology. In Ahrens W and Pigeot I (Ed.), *Handbook of epidemiology*. 2nd Edition. New-York, Springer.

Garcia-Jiménez C, Gutiérrez-Salmerón M, Choccarro-Calvo A, Garcia-Martinez JM, Castano A and De la Vieja A (2016). From obesity to diabetes and cancer: epidemiological links and role of therapies. *British Journal of Cancer* **114**(7):716–722.

Gelman A (2008). Objections to Bayesian statistics. *Bayesian Analysis* **3**(3):445–450

Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A and Rubin DB (2014). *Bayesian Data Analysis*. Third Edition. Chapman and Hall/CRC. Boca Raton. London. New York.

Gelman A, Jakulin A, Pittau MG and Su YS (2008). A weakly informative default prior distribution for logistic and other regression models. *Annals of Applied Statistics* **2**(4):1360–1383

Gelman A and Henning C (2017). Beyond subjective and objective in statistics. *Journal of The Royal Statistical Society*. Series A **180**(4): 967–1033; doi:10.1111/rssa.12276

Gelman A, Hwang J and Vehtari A (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing* **24**:997–1016.

Gelman A and Shalizi CR (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology* **66**:8–38

Gelman A, Simpson D and Betancourt M (2017). The prior can often only be understood in the context of the likelihood. *Entropy* **19**(10), 555; doi:10.3390/e19100555

Gerhard T (2008). Bias: considerations for research practice. *American Journal of Health-System Pharmacy* **65**:2159–2168.

Gerstein HC, Bosch J, Degenais GR, Diaz R, Jung H, Maggioni AP, Pogue J, Probstfield J, Ramachandran A, Riddle MC, Ryden LE and Yusuf S for the ORIGIN Trial Investigators. (2012). Basal insulin and cardiovascular and other outcomes in dysglycemia. *New England Journal of Medicine* **367**(4):319–328.

Giovannucci E, Harlan DM, Archer MC, Bergenstal RM, Gapstur SM, Habel LA, Pollak M, Regensteiner JG, Yee D (2010). Diabetes and cancer: a consensus report. *Diabetes Care* **33**(7):1674–1685.

Gissler M and Haukka J (2004). Finnish health and welfare registers in epidemiological research. *Norsk Epidemiologi* **14** (1):113–120.

Ghosh M (2011). Objective priors: an introduction for frequentists. *Statistical Science* **26**(2): 187–202.

Glesby MJ and Hoover DR (1996). Survivor treatment selection bias in observational studies: examples from the AIDS literature.

Goldacre M.(2001). The role of cohort studies in medical research. *Pharmacoepidemiology and Drug Safety* **10**:5–11

Goldstein M (2006). Subjective Bayesian analysis: principles and practice. *Bayesian Analysis* **1**: 403–420.

Gray R (1992). Flexible methods for analyzing survival data using splines, with applications to breast cancer prognosis. *Journal of the American Statistical Association* **87**(420): 942–951.

Greenland S (2006). Bayesian perspectives for epidemiological research: I Foundations and basic methods. *International Journal of Epidemiology* **35**(3):765–775.

Grimes DA and Schulz KF (2002). Cohort studies: marching towards outcomes. *Lancet* **359**(9303):341–345.

Groop PH, Thomas MC, Moran JL, Wadèn J, Thorn LM, Mäkinen VP, Rosengård-Bärlund M, Saraheimo M, Hietala O and Forsblom C; FinnDiane Study Group (2009). The presence and severity of chronic kidney disease predicts all-cause mortality in type 1 diabetes. *Diabetes* **58**(7):1651–1658.

Groop PH, Thomas M, Feodoroff M, Forsblom C and Harjutsalo V (2018). Excess mortality in patients with type 1 diabetes without albuminuria – separating the contribution of early and late risks. *Diabetes Care* doi: 10.2337/dc17–1618.

Gurrin LC, Kurinczuk JJ and Burton PR (2000). Bayesian statistics in medical research: an intuitive alternative to conventional data analysis. *Journal of Evaluation in Clinical Practice* **6**(2):193–204.

Gutiérrez-Peña E, Smith AFM, Bernardo JM, Consonni G, Veronese P, George EI, Girón FJ, Martínez ML, Letac G and Morris CN (1997). Exponential and bayesian conjugate families: Review and extensions. *Test* **6**(1):1–90.

Habel LA, Danforth KN, Quesenberry CP, Capra A, Van Den Eeden SK, Weiss NA and Ferrara A (2013) Cohort study of insulin glargine and risk of breast, prostate, and colorectal cancer among patients with diabetes. *Diabetes Care* **36**:3953–3960.

Hamran G, Richardson D, MacLebose R and Wing S (2013). Integrating informative priors from experimental research with Bayesian methods. An example from radiation epidemiology. *Epidemiology* **24**(1):90–95.

Hanley JA and Foster BJ (2014). Avoiding blunders involving 'immortal time'. *International Journal of Epidemiology* **43**(3):949–961.

Harding JL, Shaw JE, Peeters A, Cartensen B and Magliano DJ (2015). Cancer risk among people with type 1 and type 2 diabetes: disentangling true associations, detection bias, and reverse causation. *Diabetes Care* **38**(2):264–270.

Harjutsalo V, Sund R, Knip M and Groop P-H (2013). Incidence of type 1 diabetes in Finland. *JAMA* **310**:427–428.

Haukka J, But A and Härkänen T (2016). BITE: Bayesian intensity estimator. http://blogs.helsinki.fi/ bayesian- intensity/.

Haut ER and Pronovost PJ (2011). Surveillance bias in outcomes reporting. *JAMA* **305**(23):2462–2463.

Held L and Riebler A (2012). A conditional approach for inference in multivariate age-period-cohort models. *Statistical Methods in Medical Research* **21**(4): 311–329.

Helve J, Sund R, Arffman M, Harjutsalo V, Groop PH, Grönhagen-Riskaand C and Finne P (2017). Incidence of end-stage renal disease in patients with type 1 diabetes. *Diabetes Care* Doi:dc172364

Hemkens LG, Grouven U, Bender R, Gunster C, Gutschmidt S, Selke GW and Sawicki PT (2009). Risk of malignancies in patients with diabetes treated with human insulin or insulin analogues: a cohort study. *Diabetologia* **52**:1732–1744.

Herrett E, Gallagher AM, Bhaskaran K, Forbes H, Mathur R, van Staa T and Smeeth L (2015) Data resource profile: clinical practice research datalink (CPRD). *International Journal of Epidemiology* **44**:827–836.

Holford TR (1991). Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annual Rev Public Health* **12**:425–457.

Hosmer DW, Lemeshow S and May S (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data:* Second Edition, John Wiley and Sons Inc., New York.

Härkänen T (2003). BITE: Bayesian intensity estimator. *Computational Statistics* **18**:565–583.

Härkänen T and But A (2016). BITE: Bayesian intensity estimator. Reference manual. Available from https://blogs.helsinki.fi/bayesian-intensity/files/2016/12/bite.pdf

Johnson JA and Yasui Y (2010). Glucose-lowering therapies and cancer risk: the trials and tribulations of trials and observations. *Diabetologia* **53**:1823–1826.

Johnson JA, Carstensen B, Witte D, Bowker SL, Lipscombe L and Renehan AG; Diabetes and Cancer Research Consortium (2012). Diabetes and cancer (1): evaluating the temporal relationship between type 2 diabetes and cancer incidence. *Diabetologia* **55**(6):1607–1618.

Jonasson JM, Ljung R, Talbäck M, Haglund B, Gudbjömsdòttir S and Steineck G (2009). Insulin glargine use and short-term incidence of malignancies – a population-based follow-up study in Sweden. *Diabetologia* **52**:1745–1754.

Kahn SE, Cooper ME and Del Prato S (2014). Pathophysiology and treatment of type 2 diabetes: perspectives on the past, present, and future. *Lancet* **383**(9922):1068-1083.

Kaplan EL and Meier P (1958). Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association* **53**:457–481.

Kalbfleisch JD and Prentice RL (2002). *The Statistical Analysis of Failure Time Data*. Hoboken, New Jersey, John Wiley & Sons.

Karlstad O, Starup-Linde J, Vestergaard P, Hjellvik V, Bazelier MT, Schmidt MK, Andersen M, Auvinen A, Haukka J, Furu K, de Vries F and De Bruin ML (2013) Use of insulin and insulin analogs and risk of cancer – systematic review and meta-analysis of observational studies. *Current Drug Safety* **8**:333–348

Keiding N (1990) Statistical inference in the Lexis diagram. *Philosopfical Transaction of the Royal Society of London,* Series A. **332**:487–509.

Keiding N (2011). Age–period–cohort analysis in the 1870s: Diagrams, stereograms, and the basic differential equation. *The Canadian Journal of Statistics* **39**(3):405–420.

Kelasto-reports | Kela (2018): Existing, new and withdrawn entitlements to reimbursement of drug expenses. Available from (last accessed 11 November 2018): http://raportit.kela.fi/ibi_apps/WFServlet?IBIF_ex=NIT084AL&YKIELI=E

Kerman J (2011). Neutral noninformative and informative conjugate beta and gamma prior distributions. *Electronic Journal of Statistics* **5**:1450–1470.

Kivimäki M, Kuosma E, Ferrie JE, Luukkonen R, Neberg ST, Alfredsson L, Batty GD, Brunner EJ, Fransson E, Golgberg M, Knutsson A, Koskenvuo M, Nordin M, Oksanen T, Pentti J, Rugulies R, Shipley MJ, Singh-Manoux A, Steptoe A, Suominen SB, Theorell T, Vahtera J, Virtanen M, Westerholm P, Westerlund H, Zins M, Hamer M, Bell JA, Tabak AG and Jokela M (2017). Overweight, obesity, and risk of cardiometabolic multimorbidity: pooled analysis of individual-level data for 120 813 adults from 16 cohort studies from the USA and Europe. *Lancet Public Health* **2**(6):e277-e285. doi: 10.1016/S2468-2667(17)30074-9

Klil-Drori AJ, Azoulay L and Pollak MN (2017). Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? *Nature Reviews Clinical Oncology* **14**:85–89.

Kooman JP, Kotanko P, Schols AM, Shiels PG and Stenvinkel P (2014). Chronic kidney disease and premature ageing. *Nature Reviews Nephrology* **10**(12):732–742.

Korhonen MJ, Huupponen R, Ruokoniemi P and Helin-Salmivaara A (2009). Protopathic bias in observational studies on statin effectiveness. *European Journal of Clinical Pharmacology* **65**:1167–1168.

Korhonen P, Heintjes EM, Williams R, Hoti F, Christopher S, Majak M, Kool-Houweling L, Strongman H, Linder M, Dolin P and Bahmanyar S (2016). Pioglitazone use and risk of bladder cancer in patients with type 2 diabetes: retrospective cohort study using datasets from four European countries. *BMJ* **354**:i3903.

Korn EL, Graubard BI and Midhune D (1997). Time-to-event analysis of longitudinal follow-up of a survey: choice of time-scale. *American Journal of Epidemiology* **145**:72–80 doi: 10.1093/oxfordjournals.aje.a009034

Kowall B, Rathmann W and Kostev K (2015). Are sulfonylurea and insulin therapies associated with a larger risk of cancer than metformin therapy? A retrospective database analysis. *Diabetologia* **38**(1):59–65. doi: 10.2337/dc14-0977.

Kvåle R, Myklebust TÅ, Engholm G, Heinävaara S, Wist E and Møller B (2017). Prostate and breast cancer in four Nordic countries: A comparison of incidence and mortality trends across countries and age groups 1975-2013. *International Journal of Cancer* **141**(11):2228–2242

Lammi N, Blomstedt PA, Moltchanova E, Eriksson JG, Tuomilehto J and Karvonen M (2008). Marked temporal increase in the incidence of type 1 and type 2 diabetes among young adults in Finland. *Diabetologia* **51**:897–899.

Lange T and Kielding N (2014). Letter: Skin cancer as a marker of sun exposure: a case of serious immortality bias. *International Journal of Epidemiology* **43**:971.

Lecaire TJ, Klein BE, Howard KP, Lee KE, Klein R (2014). Risk for end-stage renal disease over 25 years in the population-based WESDR cohort. *Diabetes Care* **37**(2):381–388.

Lehr S and Schemper S. (2007). Parsimonious analysis of time-dependent effect in the Cox model. *Statistics in Medicine* **26**(13):2686–2698.

Lévesque LE, Hanley JA, Kezouh A and Suissa S (2010). Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *BMJ* 2010;340:b5087

Levin D, Bell S, Sund R, Hartikainen SA, Tuomilehto J, Pukkala E, Keskimäki I, Badrick E, Renehan AG, Buchan IE, Bowker SL, Minhas-Sandhu JK, Zafari Z, Marra C, Johnson JA, Stricker BH, Uitterlinden AG, Hofman A, Ruiter R, de Keyser CE, MacDonald TM, Wild SH, McKeigue PM and Colhoun HM (2015). Pioglitazone and bladder cancer risk: a multipopulation pooled, cumulative exposure analysis. *Diabetologia* **58**(3):493–504 . doi: 10.1007/s00125-014-3456-9.

Lewis JD, Ferrara A and Peng T (2011). Risk of bladder cancer among diabetic patients treated with pioglitazone: interim report of a longitudinal cohort study. Diabetes Care **34**:916–22.

Lewis JD, Habel LA, Quesenberry CP, Strom BL, Peng T, Hedderson MM, Ehrlich SF, Mamatani R, Bilker W, Vaughn D, Nessel L, Van Den Eeden SK and Ferrara A (2015). Pioglitazone use and risk of bladder cancer and other common cancers in persons with diabetes. JAMA **314**:265–77

Lexis, W (1875). Einleitung in die theorie der bevölkerungstatistik, Trubner, Strasbourg. In *Mathematical Demography: Selected Papers* (eds Smith, D. & Keyfitz, N.), Biomathematics 6, (1977) Springer-Verlag, Berlin; 39–41.

Lindström J and Tuomilehto J (2003). The diabetes risk score: a practical tool to predict type 2 diabetes risk. *Diabetes Care* **26**(3):725–31

Lindström J, Louheranta A, Mannelin M, Rastas M, Salminen V, Eriksson J, Uusitupa M and Tuomilehto J; Finnish Diabetes Prevention Study Group (2003). The Finnish Diabetes Prevention Study (DPS): Lifestyle intervention and 3-year results on diet and physical activity. *Diabetes Care* **26**(12):3230–6

Little RJ (2006). Calibrated Bayes: A Bayes/frequentist roadmap. *American statistician* **60**(3):213–223

Ludvigsson JF, Håberg SE, Knudsen GP, Lafolie P, Zoega H, Sarkkola C, von Kraemer S, Weiderpass E and Nørgaard M (2015). Ethical aspects of registry-based research in the Nordic countries. *Clinical Epidemiology* **7**: 491–508

Ludvigsson JF, Almqvist C, Bonamy AK, Ljung R, Michaëlsson K, Neovius M, Stephansson O and Ye W (2016). Registers of the Swedish total population and their use in medical research. *European Journal of Epidemiology* **31**(2):125–36

Luo J, Beresford S, Chen C, Chlebowski R, Garcia L, Kuller L, Regier M, Wactawski-Wende J and Margolis KL (2014). Association between diabetes, diabetes treatment and risk of developing endometrial cancer. *British Journal of Cancer* **111**:1432–1439.

Mahmood SS, Levy DL, Vasan RS and Wang TJ (2014). The Framingham Heart Study and the Epidemiology of cardiovascular diseases: a historical perspective. *Lancet* **383**(9921):999–1008.

Maret-Ouda J, Wenjing T, Wahlin K and Lagergren J (2017). Nordic registry-based cohort studies: possibilities and pitfalls when combining Nordic registry data. *Scandinavian Journal of Public Health* **45**(17_suppl):14–19 .

Matsuura M and Eguchi S (2005). Modeling late entry bias in survival analysis. *Biometrics* **61**:559–566.

Mayo D and Cox D (2006). Frequentist Statistics as a theory of inductive inference. *Institute of Mathematical Statistics Lecture Notes* - Monograph Series **49**:77–67.

Miller AB, Goff Jr. DC, Bammann K and Wild P (2014). Cohort studies. In Ahrens W and Pigeot I (Ed.), *Handbook of epidemiology*. 2nd Edition. New-York, Springer.

Morabia A and Guthold R (2007). Wilhelm Weinberg's 1913 Large Retrospective Cohort Study: A Rediscovery. *American Journal of Epidemiology* **165**(7):727–733

Müller P and Mitra R (2013). Bayesian nonparametric inference - why and how. *Bayesian Analysis* **8**(2). doi: 10.1214/13-BA811

Müller P and Quintana FA (2004). Nonparametric Bayesian data analysis. *Statistical Science* **19**(1):95–110.

Müller P, Quintana FA and Page G (2017). Nonparametric Bayesian inference in applications. *Statistical Methods and Applications*. https://doi.org/10.1007/s10260-017-0405-z

Natarajan L, Pu M, Parker BA, Thomson CA, Caan BJ, Flatt W, Madlensky L, Hajek RA, Al-Delaimy WK, Saquib N, Gold B and Pierce JP (2009). Time-Varying Effects of Prognostic Factors Associated With Disease-Free Survival in Breast Cancer. *American Journal of Epidemiology* **169**(12): 1463–1470.

NCD Risk Factor Collaboration (NCD-RisC) (2016). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4.4 million participants. *Lancet* **387**(10027):1513–30.

OECD/EU (2016). Health at a glance: Europe 2016 – State of Health in the EU cycle, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264265592-en

Ogurtsova K, da Rocha Fernandes JD, Huang Y, Linnenkamp U, Guariguata L, Cho NH, Cavan D, Shaw JE and Makaroff LE (2017). IDF Diabetes Atlas: Global estimates for the prevalence for 2015 and 2040. *Diabetes Research and Clinical Practice* **128**:40–50.

Onitilo AA, Stankowski RV, Berg RL, Engel JM, Glurich I, Williams DN and Doi SA (2013) Type 2 diabetes mellitus, glycemic control, and cancer risk. *European Journal of Cancer Prevention* **23**(2):134–140.

Pearce N and Greenland S (2014). Confounding and interaction. In Ahrens W and Pigeot I (Ed.), *Handbook of epidemiology*. 2nd Edition. New-York, Springer.

Peeters PJHL, Bazelier MT, Leufkens HGM, de Vries F and De Bruin ML (2015). The risk of colorectal cancer in patients with type 2 diabetes: associations with treatment stage and obesity. *Diabetes Care* **38**(3): 495–502.

Peeters PJHL, Bazelier MT, Leufkens HGM, Auvinen A, van Staa TP, de Vries F, and De Bruin ML (2016). Insulin glargine use and breast cancer risk: Associations with cumulative exposure. *Acta Oncologica* **55**(7): 851–858.

Pencina MJ, Larson MG and D'Agostino RB (2007). Choice of time scale and its effect on significance of predictors in longitudinal studies. *Statistics in Medicine* **26**:1343–1359. doi: 10.1002/sim.2699

Plummer M (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics* **9**:523–539.

Plummer M, Best N, Cowles K and Vines K (2006). CODA: convergence diagnosis and output analysis for MCMC. R News. 6:7–11. http://CRAN.R-project.org/doc/Rnews/Rnews_2006-1.pdf.

Plummer M and Carstensen B (2011). Lexis: an R class for epidemiological studies with long-term follow-up. *Journal of Statistical Software* **38**:1175–1179.

Polonsky KS (2012). The past 200 years in diabetes. *New England Journal of Medicine* **367**:1332–1340.

Pocock SJ and Smeeth L (2009). Insulin glargine and malignancy: an unwarranted alarm. *Lancet* **374**:511–513

Pukkala E, Martinsen JI, Lynge E, Gunnarsdottir HK, Sparen P, Tryggvadóttir L, Weiderpass E and Kjaerheim K (2009). Occupation and cancer – follow-up of 15 million people in five Nordic countries. *Acta Oncologica* **48**(5): 646–790.

Pukkala E, Engholm G, Højsgaard Schmidt LK, Storm H, Khan S, Lambe M, Pettersson D, Ólafsdóttir E, Tryggvadóttir L, Hakanen T, Malila N, Virtanen A, Johannesen TB, Laronningen S and Ursin G (2018). Nordic Cancer Registries – an overview of their procedures and data comparability. *Acta Oncologica* **57**(4):440–455.

Raile K, Galler A, Hofer S, Herbst A, Dunstheimer D, Busch P and Holl RW (2007). Diabetic nephropathy in 27,805 children, adolescents, and adults with type 1 diabetes: effect of diabetes duration, A1C, hypertension, dyslipidemia, diabetes onset, and sex. *Diabetes Care* **30**(10):2523–2528

Rau R, Bohk-Ewald C, Muszyńska MM and Vaupel JW (2018) Visualizing Mortality Dynamics in the Lexis Diagram. The Springer Series on Demographic Methods and population Analysis, Springer, Cham.

Ray WA (2003). Evaluating medication effects outside of clinical trials: new-user design. *American Journal of Epidemiology* **158**:915–920.

R Core Team (2017). R: a language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Reitbergen C, Debray TPA, Klugkist I, Janssen KJM, and Moons KGM (2017). Reporting of Bayesian analysis in epidemiologic research should become more transparent. *Journal of Clinical Epidemiology* **86**:51–58.e2. doi: 10.1016/j.jclinepi.2017.04.008

Renehan AG (2012). Insulin analogues and cancer risk: the emergence of second-generation studies. *Diabetologia* **5**5:7–9

Riebler A and Held L (2010). The analysis of heterogeneous time trends in multivariate age-period-cohort models. *Biostatistics* **11**(1):57–69.

Riebler A, Held L and Rue H (2012). Estimation and extrapolation of time trends in registry data – Borrowing strength from related populations. *Annals of Applied Statistics* **6**(1):304–333.

Robert CP and Casella G (2004). *Monte Carlo Statistical Methods*. New York: Springer-Verlag

Rothman KJ and Greenland S (2014). Basic concepts. In Ahrens W and Pigeot I (Ed.), *Handbook of epidemiology*. 2nd Edition. New-York, Springer.

Ruiter R, Visser LE, van Herk-Sukel MPP, Coebergh JW, Haak HR, Geelhoed-Duijvestijn PH, Straus SM, Herings RM and Stricker BH (2012). Risk of cancer in patients on insulin glargine and other insulin analogues in comparison with those on human insulin: results from a large population-based follow-up study. *Diabetologia* **55**:51–62.

Ryu TY, Park J and Scherer PE (2014). Hyperglycemia as a Risk Factor for Cancer Progression. *Diabetes and Metabolism Journal* **38**(5):330–336.

Saaristo T, Moilanen L, Korpi-Hyövälti E, Vanhala M, Saltevo J, Niskanen L, Jokelainen J, Peltonen M, Oksa H, Tuomilehto J, Uusitupa M and Keinänen-Kiukaanniemi S (2010). Lifestyle Intervention for Prevention of Type 2 Diabetes in Primary Health Care. One-year follow-up of the Finnish National Diabetes Prevention Program (FIN-D2D). *Diabetes Care* **33**(10):2146–2151.

Sarkar D (2008). *Lattice: Multivariate Data Visualization with R*, Springer, New York. ISBN 978-0-387-75968-5. Available at http://lmdvr.r-forge.r-project.org.

Sauerbrei W, Royston P, Bojar H, Schmoor C, Schumacher M and the German Breast Cancer Study Group (GBSG) (1999). Modelling the effects of standard prognostic factors in node-positive cancer. British Journal of Cancer 79: 1752–1760.

Schmidt M, Pedersen L and Sorensen HT (2014). The Danish Civil Registration System as a tool in epidemiology. *European Journal of Epidemiology* **29**(8):541–549.

Schrijnders D, de Bock GH, Houweling ST, van Hateren KJJ, Groenier KH, Johnson JA, Bilo HJG, Kleefstra N and Landman GWD (2017). Within-class differences in cancer risk for sulfonylurea treatments in patients with type 2 diabetes (ZODIAC-55) – a study protocol. *BMC Cancer* **17**: 444 doi: 10.1186/s12885-017-3433-z

Schumacher M, Bastert G, Bojar H, Hübner K, Olschewski M, Sauerbrei W, Schmoor C, Beyerle C, Neumann RLA and Rauschecker HF (1994). Randomized 2x2 trial evaluating hormonal treatment and the duration of chemotherapy in node-positive breast cancer patients. *Journal of Clinical Oncology* **12:**2086–2093.

Schöley J and Willekens F (2017). Visualizing compositional data on the Lexis surface. *Demographic Research* **36**(21):627–658.

Sedgwick P (2014a). Retrospective cohort studies: advantages and disadvantages. *BMJ* **348**:g1072

Sedgwick P (2014b). Bias in observational study designs: prospective cohort studies. *BMJ* **349**:g7731

Simpson D, Rue Rue H, Riebler A, Martins TG and Sørbye SH (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science* **32**(1):1–28.

Smith U and Gale EAM (2009). Does diabetes therapy influence the risk of cancer? *Diabetologia* **52**(9):1699–1708.

Smith T and Wakefield J (2016). A review and comparison of age-period-cohort models for cancer incidence. *Statistical Science* **31**:591–610.

Sørensen HT, Sabroe S and Olsen J (1996). A Framework for evaluation of secondary data sources for epidemiological research. *International Journal of Epidemiology* **25**(2):435–442.

Spiegelhalter DJ, Best NG, Carlin BR and van der Linde A (2002). Bayesian measures of model complexity and fit (with discussion). *Journal of the Royal Statistical Society*. Series B. Statistical Methodology **64**:583–639.

Stan Development Team (2017). Prior choice recommendations. Available from https://github.com/stan-dev/stan/wiki/Prior-Choice-Recommendations

Stattin P, Björ O, Ferrari P, Lukanova A, Lenner P, Lindahl B, Hallmans G and Kaaks R. (2007). Prospective study of hyperglycemia and cancer risk. *Diabetes Care* **30**(3):561–567.

Spiegelhalter DJ, Abrams KR and Myles JP (2004). *Bayesian Approaches to Clinical Trials and Health-Care Evaluation*. Wiley, Chichester.

Stricker BHC and Stijnen T (2010). Analysis of individual drug use as a time-varying determinant of exposure in prospective population-based cohort studies. *European Journal of Epidemiology* **25**(4): 245–251

Suissa S (2007). Immortal time bias in observational studies of drug effects. *Pharmacoepidemiology and Drug Safety* **16**:241–249.

Suissa S (2008). Immortal time bias in pharmacoepidemiology. *American Journal of Epidemiology* **167**(4):492–499

Suissa S, Azoulay L, Dell'Aniello S, Evans M, Vora J and Pollak M (2011). Long-term effects of insulin glargine on the risk of breast cancer. *Diabetologia* **54**:2254–2262.

Suissa S and Azoulay L (2012). Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes Care* **35**(12):2665–2673.

Suissa S and Garbe E (2007). Primer: administrative health databases in observational studies of drug effects--advantages and disadvantages. Nature Clinical Practice. *Rheumatology* **3**(12):725–732.

Sund R (2003). Utilisation of administrative registers using scientific knowledge discovery. *Intelligent Data Analysis* **7**:501–519.

Sund Rand Koski S. FinDM II (2009.) On the register-based measurement of the prevalence and incidence of diabetes and its long-term complications. A technical report.Tampere: Finnish Diabetes Association; Available from: https://www.diabetes.fi/files/1167/DehkoFinDM_Raportti_ENG.pdf

Sund R (2012). Quality of the Finnish Hospital Discharge Register: A systematic review. *Scandinavian Journal of Public Health* **40**:505–515.

Svensson M, Nyström L, Schön S and Dahlquist G (2006). Age at onset of childhood-onset type 1 diabetes and the development of end- stage renal disease. *Diabetes Care* **29**:538–542.

Syversveen AR (1998). Noninformative Bayesian Priors. Interpretation And Problems With Construction And Applications. Available from: https://pdfs.semanticscholar.org/a518/5cbbe2f271455f2cd87e9ed594b0b1820d91.pdf

Szklo M (1998). Population-based cohort studies. *Epidemiologic Reviews* **20**(1):81–90.

Tanner MA (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. Springer Series in Statistics. New York: Springer-Verlag.

Tamim H, Monfared AA and LeLorier J (2007). Application of lag-time into exposure definitions to control for protopathic bias. *Pharmacoepidemiology and Drug Safety* **16**:250–258.

Thiébaut ACM and Bénichou J (2004). Choice of time-scale in Cox's model analysis of epidemiologic cohort data: a simulation study. *Statistics in Medicine* **23**:3803–3820. doi: 10.1002/sim.2098

Tokajuk A, Kryźanowska-Grycel E, Tokajuk A, Grycel S, Sadowska A and Car H (2015). Antidiabetic drugs and risk of cancer. *Pharmacological Reports* **67**(6):1240–1250.

Tu H, Wen CP, Tsai SP, Chow W-H, Wen C, Ye Y, Zhao H, Tsai MK, Huang M, Dinney CP, Tsao CK and Wu X (2018). Cancer risk associated with chronic diseases and disease markers: prospective cohort study. *BMJ* **60**:k134

Tuccori M, Filion KB, Yin H, Yu OH, Platt RW and Azoulay L (2016). *BMJ* **352**:i1541.

Tuomilehto J (2013). The emerging global epidemic of type 1 diabetes. *Current Diabetes Reports* **13** (6):795–804.

Vandeschrick C (2001). The Lexis diagram, a misnomer. *Demographic research* **4**(3):97–124.

Vaupel JW, Gambill BA and Yashin AI. Thousands of Data at a Glance: Shaded Contour Maps of Demographic Surfaces. IIASA Research Report. 1987; IIASA, Laxenburg, Austria: RR-87-016

Vehtari A, Gelman A and Gabry J (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing* **27**(5):1413–1432.

Vehtari A, Gabry J, Yao Y and Gelman A (2018). "loo: Efficient leave-one-out cross-validation and WAIC for Bayesian models." R package version 2.0.0, https://CRAN.R-project.org/package=loo.

Vehtari A and Ojanen J (2012). A survey of Bayesian predictive methods for model assessment, selection and comparison. *Statistics Surveys* **6**:142–228.

Walker JJ, Johnson JA and Wild SH (2013). Diabetes treatments and cancer risk: the importance of considering aspects of drug exposure. *Lancet Diabetes and Endocrinology* **1**(2):132–139.

Wasag DR, Gregory JW, Dayan C and Harvey JN (2018). Excess all-cause mortality before age 30 in childhood onset type 1 diabetes: data from the Brecon Group Cohort in Wales. *Archives if Disease in Childhood* **101**(1):44–48.

Watanabe S (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research* **11**:3571–3594.

Watanabe S (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research* **14**:867–897.

Wettermark B, Zoëga H, Furu K, Korhonen M, Hallas J, Nørgraad M, Almarsdottir A, Andersen M, Andersson Sundell K, Bergman U, Helin-Salmivaara A, Hoffman M, Kieler H, Martikainen J, Mortensen M, Petzold M, Wallach-Kildemoes H, Wallin C and Sørensen H (2013) The Nordic prescription databases as a resource for pharmacoepidemiological research – a literature review. *Pharmacoepidemiol Drug Saf*. **22**(7):691–699.

Writing team for the diabetes control and complications trial/Eepidemiology of diabetes interventions and complications research group (2003) Sustained effect of intensive treatment of type 1 diabetes mellitus on development and progression of diabetic nephropathy: the Epidemiology of Diabetes Interventions and Complications (EDIC) study. *JAMA* **290**:2159–2167.

World Health Organization (2016). Global report on diabetes. Available at http://apps.who.int/iris/bitstream/handle/10665/204871/9789241565257_eng.pdf;jsessionid=38CD26325B62654B77ABCA7AA26DC6E8?sequence=1

Wu JW, Azoulay L, Maidan A, Boivin JF, Pollak W and Suissa S (2017). Long-term use of long-acting insulin analogs and breast cancer incidence in women with type 2 diabetes. *Journal of Clinical Oncology* **35**(32):3647–3653.

Wu JW, Filion KB, Azoulay L, Doll MK and Suissa S (2016). Effect of long-acting insulin analogs on the risk of cancer: A systematic review of observational studies. *Diabetes Care* **39**(3):486–494.

Wu JW, Zhu J, Prokop LJ and Murad MH (2015). Pharmacologic Therapy of Diabetes and Overall Cancer Risk and Mortality: A Meta-Analysis of 265 Studies. Scientific Reports. 5:doi:10.1038/srep10147.

Yoshida K, Solomon DH and Kim SC (2015) Active-comparator design and new-user design in observational studies. *Nature Reviews Rheumatology* **11**:437–441.

Zhou Z, Rahme E, Abrahamovicz M and Pilote L (2005). Survival bias associated with time-to-treatment initiation in drug effectiveness evaluation: a comparison of methods. *American Journal of Epidemiology* **162**(10):1016–1023.

Zondervan-Zwijnenburg M, Peeters M, Depaoli S and van de Schoot R (2017). Where do priors come from? Applying guidelines to construct informative priors in small sample research. *Research in Human Development* **14**(4):305–320.