Pharmacoepidemiologic studies on prescription drugs and cancer risk

Kristensen, Kasper Bruun

DOI:
10.21996/rw46-dn96

Publication date:
2022

Document version:
Final published version

PhD Thesis

Pharmacoepidemiologic studies on prescription drugs and cancer risk

Kasper Bruun Kristensen

Clinical Pharmacology, Pharmacy and Environmental Medicine
Department of Public Health
University of Southern Denmark

SDU

PhD Thesis

Pharmacoepidemiologic studies on prescription drugs and cancer risk

Kasper Bruun Kristensen

Clinical Pharmacology, Pharmacy and Environmental Medicine
Department of Public Health
University of Southern Denmark

Main supervisor

Anton Pottegård

Co-supervisors

Søren Friis

Joshua J. Gagne

Laurel A. Habel

Funding

Independent Research Fund Denmark

Region of Southern Denmark

Preface

This thesis was carried out during my appointment at Clinical Pharmacology and Pharmacy, University of Southern Denmark. The work in this thesis was funded by Independent Research Fund Denmark and the Region of Southern Denmark. Many people have contributed to this thesis for which I am deeply thankful. Even though the lack of space prevents me from thanking you all by name, I wish to express a special thanks to Anton who have coped with me through the entire thesis. You have patiently as well as energetically introduced me to pharmacoepidemiology and, not least, to many other skills that are useful in and out-side academia. I also wish to thank my co-supervisors who have provided invaluable support. Søren, thank you for the numerous inputs with regards to writing style and your insights into cancer diseases and cancer registries. Laurie and Josh, thank you for supporting me throughout many of these studies and your solid methodological input. Annmarie, thank you for introducing me to research as a medical student and for being a role model as a clinician and researcher. A special thanks to the entire unit of Clinical Pharmacology and Pharmacy for providing a friendly, open atmosphere. I am going to miss working with all of you. Lastly, to my wife and two children, thank you for reminding me of the important things in life.

Odense, December 2021

Kasper Bruun Kristensen

Table of contents

Abbreviations

ATC	Anatomic Therapeutic Chemical group
CCI	Charlson Comorbidity Index
CI	Confidence Interval
DCR	Danish Cancer Registry
DDD	Defined Daily Dose
ECI	Elixhauser Comorbidity Index
IARC	International Agency for Research on Cancer
ICD-10	International Classification of Disease 10th version
ICD-O-3	International Classification of Disease for Oncology, 3rd version
NCI	Nordic Comorbidity Index
OR	Odds Ratio
RCT	Randomized Clinical Trial
TNM	Tumor Node Metastasis
VKA	Vitamin K antagonists
WHO	World Health Organization

List of studies

I. **Use of vitamin K antagonists and risk of prostate cancer: Meta–analysis and nationwide case–control study** Kristensen KB, Jensen PH, Skriver C, Friis S, Pottegård A. Int J Cancer 2019, 144: 1522-1529.

II. **Use of antiepileptic drugs and risk of skin cancer: A nationwide case-control study** Kristensen KB, Pedersen SA, Schmidt SAJ, Pottegård A. J Am Acad Dermatol 2020, 82(2):326-335.

III. **Risk of Renal Cell Carcinoma Associated with Calcium Channel Blockers: A Nationwide Observational Study Focusing on Confounding by Indication** Kristensen KB, Habel LA, Gagne JJ, Friis S, Andersen KK, Hallas J, Pottegård A. Epidemiology 2020, 31(6):860-871.

IV. **Identification of drug-cancer associations: A nationwide screening study** Kristensen KB, Friis S, Lund LC, Hallas J, Cardwell CR, Andreassen BK, Habel LA, Pottegård A. Submitted for publication.

V. **Development and validation of a Nordic Comorbidity Index based on hospital diagnoses and filled prescriptions** Kristensen KB, Lund LC, Jensen PB, Broe A, Rotbain E, Damkier P, Pottegård A, Andersen JH, Højlund M, Olesen M, Rasmussen L, Hansen MR, Ernst MT, Wesselhoeft R, Henriksen DP, Reilev M, Bliddal M, Hallas J. Submitted for publication.

English summary

Exogenous exposures such as prescription drugs may affect the risk of cancer. Pharmacoepidemiology, the study of drugs in groups of people, can be used to study these effects, however, drug-cancer studies pose a particular challenge since it may take years or even decades from the initial exposure until a cancer is diagnosed. This thesis includes five studies, of which three examined specific drug-cancer associations, one was a hypothesis-free screening study for carcinogenic effects of drugs, and the last study focused on developing a comorbidity summary score for use in pharmacoepidemiologic studies.

In the first study, we summarized existing evidence on vitamin K antagonists and their association with prostate cancer risk and conducted a case-control study using Danish registries to add to the existing evidence. Taken together, the identified studies reported heterogenous results, however, we concluded that a clinically relevant preventive effect of vitamin K antagonists against prostate cancer was unlikely.

In the second study, we examined whether antiepileptic drugs were associated with non-melanoma skin cancer and malignant melanoma in a series of case-control studies. Reassuringly, associations were close to unity for most antiepileptic drugs and outcomes except for a positive association between squamous cell carcinoma and lamotrigine and carbamazepine. Due to the hypothesis-generating nature of the study, we concluded that the results had no direct clinical implications and that further research was needed to qualify these signals further.

In the third study, we examined whether use of calcium channel blockers was associated with increased risk of kidney cancer. We illustrated methods to identify and account for confounding by indication including the assessment of cumulative

dose-response relationships, adjusting for severity of disease, using negative control exposures, and using active comparators. We concluded that the observed association between calcium channel blockers and kidney cancer was at least partly explained by confounding by indication.

The fourth study, a hypothesis-free screening study, aimed to identify carcinogenic effects of drugs by examining all drugs and cancers in a series of case-control studies. We identified individuals with incident cancer of 33 sites and 85 histological subtypes. Approximately 14,000 drug-cancer pairs were evaluated, and the estimates were made available online for research purposes. We identified known drug-cancer associations e.g., azathioprine and non-Hodgkin lymphoma, and highlighted a number of drug-cancer associations that deserved further scrutiny. We concluded that hypothesis-free screening of drug-cancer associations were feasible, however, signal identification remained an issue and the hypothesis-generating nature of the results should be stressed when communicating the findings of the study.

In the fifth study, we developed a numerical score to adjust for comorbidity in pharmacoepidemiologic studies, the Nordic Comorbidity Index. We developed the index in a population-based cohort of randomly sampled Danish residents and the index included a total of 50 diagnoses or drugs that predicted 5-year mortality. We concluded that the index was superior to the Charlson and Elixhauser comorbidity scores in predicting mortality and that it could be preferred as a summary score in studies utilizing Nordic registry data, however, it remains to be validated in specific patient populations and other Nordic countries.

Dansk resumé

Lægemidler kan, som andre udefrakommende eksponeringer, påvirke risikoen for kræft. Farmakoepidemiologi, studiet af lægemidler i populationer, er et nyttigt redskab til at undersøge sådanne effekter, men det er en udfordring, at det kan tage adskillige år fra lægemiddeleksponeringen indtil en eventuel kræftsygdom opstår. Denne afhandling inkluderer fem studier, hvoraf tre undersøger specifikke lægemidler og kræfttyper, et studie er et hypotesegenerende screeningsstudie for hidtil ukendte sammenhænge mellem lægemidler og kræft og det sidste studie præsenterer et komorbiditetsindeks til brug i farmakoepidemiologiske studier.

I det første studie sammenfattede vi de eksisterende studier omhandlende vitamin K antagonister og prostatakræft og udførte selv et case-kontrolstudie med danske registerdata. Konklusionen var, at de eksisterende studier var heterogene, men at en klinisk relevant effekt af vitamin K antagonister i forhold til forebyggelse af prostatakræft var usandsynlig.

I det andet studie undersøgte vi, om brug af antiepileptika var forbundet med en øget risiko for hudkræft og modermærkekræft i en række case-kontrol studier. Resultaterne var overordnet betryggende, idet vi ikke observerede en sammenhæng mellem de undersøgte kræftformer og de fleste antiepileptiske lægemidler. Vi fandt dog en sammenhæng mellem spinocellulært karcinom, en sjælden form for hudkræft, og brug af carbamazepin og lamotrigin. På grund af studiets hypotese-generende natur, har disse fund ingen direkte kliniske konsekvenser og flere studier er nødvendige for at karakterisere disse signaler yderligere.

I det tredje studie undersøgte vi, om brug af calcium kanal blokkere var forbundet med øget risiko for nyrekræft. Vi demonstrerede samtidig metoder til at identificere og imødegå `confounding by indication´ inklusive evaluering af kumulative dosis-

respons sammenhænge, justering for sværhedsgrad af sygdom, brug af andre eksponeringer som negative kontroller, og brug af aktive komparatorer. Vi konkluderede, at sammenhængen mellem calcium kanal blokkere og nyrekræft til dels var påvirket af `confounding by indication´.

Det fjerde studie, et hypotese-generende screeningsstudie, havde til formål at identificere lægemidler med mulige karcinogene egenskaber. Vi undersøgte sammenhænge mellem lægemidler og kræft i 33 forskellige organer delt på 85 forskellige histologiske typer i en række case-kontrol studier. Vi undersøgte cirka 14,000 associationer, og alle resultater blev gjort tilgængelige til forskningsmæssige formål. Vi identificerede kendte sammenhænge mellem lægemidler og kræft, eksempelvis brug af azathioprin og non-Hodgkin lymfom og beskrev en række sammenhænge, der ikke tidligere er beskrevet og fortjener at blive undersøgt nærmere. Studiet demonstrerede, at det er muligt at udføre sådanne hypotese-generende screenings studier, men også at den efterfølgende sortering af signaler skal udvikles og at det er vigtigt at understrege den hypotese-generende natur af studiet når resultater herfra præsenteres.

I det femte studie udviklede vi et numerisk komorbiditetsindeks, med det formål at justere for og beskrive den samlede sygdomsbyrde i farmakoepidemiologiske studier. Indekset blev udviklet i en kohorte udtrukket tilfældigt fra den samlede population af indbyggere i Danmark. Indekset inkluderede 50 tilstande defineret ved diagnose eller lægemiddelbrug og blev udviklet til at forudsige 5-års dødelighed. Vi konkluderede at vores indeks var bedre til at forudsige dødelighed end de hyppigt brugte Charlson og Elixhauser komorbiditesindekser og at indekset kan bruges som en indikator for sygelighed i studier baseret på Nordiske registre. Vi konkluderede dog også at indekset skulle valideres i specifikke patientpopulationer og øvrige Nordiske lande.

Introduction

The motivation of this thesis is that, while a few drugs are currently known to increase the risk of cancer in humans [1], the possibility to identify and describe a potential increased risk of cancer due to use of drugs continues to grow. How do we know whether a drug cause cancer? A part of this answer is epidemiology; the study of the occurrence and reasons for disease in groups of people [2].

To understand how epidemiology played a pivotal role in identifying causes of cancer, we must go back 250 years to the streets of London long before the biological mechanisms of cancer was known. The main heating source at that time was indoor fireplaces. The houses had narrow chimneys and only boys, often working naked in order not to get stuck, were able to get into the chimneys to clean them [3]. The surgeon and early epidemiologist Percival Pott worked in London and had noted a, often fatal, cancer of the scrotum of young men that worked as chimney sweepers. Pott described this in 1775 in the first text that linked an occupational substance to cancer as cited by Melicow: "*[…] there is a disease as peculiar to a certain set of people, which has not, at least to my knowledge, been publickly noticed; I mean the chimney-sweepers' cancer. It is a disease which always makes its first attack on, and its first appearance in, the inferior part of the scrotum […]. The fate of these people seems singularly hard; in their early infancy, they are most frequently treated with great brutality and almost starved with cold and hunger; they are thrust up narrow and sometimes hot chimneys where they are bruised, burned and almost suffocated and when they get to puberty, become peculiarly liable to a most noisome, painful and fatal disease"* [3, 4]. Pott's thoughts were of epidemiologic nature in that he showed that a disease was occurring with high prevalence in a specific group of people. We know now that the chimney-sweepers' cancer is an aggressive type of skin cancer (squamous cell carcinoma) caused by direct contact to soot namely the carcinogenic compounds benzo(a)pyrenes that was isolated in coal tar in 1933 [3].

It was not until the 20th century, however, that epidemiology became an academic discipline and its underlying methods such as the case-control and cohort studies were refined and formalized [5]. Two studies from the 1950's played a major role in attracting public attention and stimulating further epidemiologic research into the causes of cancer. The studies, originating from Doll and Hill in the UK and Wynder and Graham in the United States [6, 7], were published at a time where the health effects of smoking was heatedly debated. The studies showed an increased and dose-dependent risk of lung cancer associated with smoking and were subject to considerable academic debate, however, several forthcoming studies did confirm the findings [8]. Despite initial skepticism, smoking became established as the most important preventable cause of lung cancer as well as a range of other cancers [9]. Hill described his considerations on when and how to draw causal inferences from epidemiological studies based on, among others, his experience with tobacco and lung cancer, in his famous presidential address at a meeting of the Royal Society of Medicine in 1965 [10]. Hill's considerations, later known as the Bradford Hill criteria, remain relevant today and are also used in this thesis. Indeed, although new data sources and computational methods are becoming increasingly available, current epidemiological research still relies on the basic principles developed in classic epidemiology during the mid-20th century [5].

Drugs are an extrinsic exposure that could influence cancer risk, however, unlike many environmental and lifestyle related exposures, such as soot and smoking, data on drug exposure is readily available in administrative registries such as the Danish National Prescription Registry [11]. Together with the other health registries, the prescription registry, allows us to utilize epidemiologic principles to study potential carcinogenic effects of drugs. Observational studies are needed since randomized controlled trials may fail to detect late and rare adverse effects with their limited

sample size and follow-up length. However, observational studies must account for the fact that drug treatment is not a random event and that the groups that are being compared are not necessarily comparable. This thesis illustrates methodological considerations related hereto, specifically with regards to confounding by indication.

In general, studies on drug-cancer associations may have three outcomes and three potential perspectives. First, they may return a positive association. If the observed association is due to a carcinogenic effect of the drug, such a finding will allow us to reduce the burden of cancer attributable to the drug by withdrawing the drug, using safer alternatives, or screening treated individuals for cancer. Second, and perhaps most common, a neutral association or null finding is reported. We examined numerous drug-cancer associations in this thesis and most were neutral or were interpreted as less likely to be due to a carcinogenic drug effect but rather due to a specific bias. This corresponds well with the fact that relatively few drugs have been established as carcinogenic with sufficient evidence in humans by the International Agency for Research on Cancer [1]. These studies may also be conducted in response to findings from smaller or flawed studies and holds value in reassuring the public and health care professionals about the safety of drugs. Third, an inverse association may be found. Depending on the specific study, this may be interpreted as due to a protective effect of a drug towards cancer. A such finding does not necessarily mean that the drug is useful for cancer prevention, however, these findings could advance our understanding of tumor biology and help to identify novel cancer treatments.

In this thesis, we examine three specific drug-cancer hypotheses, present a hypothesis-free screening study with the aim of generating new hypotheses, and present a tool to adjust for confounding in drug-cancer studies as well as pharmacoepidemiologic studies in general.

Aims

In study I, we addressed the hypothesis that use of vitamin K antagonists reduce the risk of prostate cancer. We identified existing literature in a systematic review, conducted a case-control study and pooled the available evidence in a meta-analysis.

In study II, we conducted a focused screening study on antiepileptic drugs and their association with non-melanoma skin cancer and malignant melanoma.

In study III, we illustrated methods to identify confounding by indication using the case of calcium channel blockers' association with kidney cancer.

In study IV, we conducted a hypothesis-free screening study to be used for hypothesis generation and as a pharmacovigilance tool. Drugs were screened for their association with cancer outcomes and selected results were manually reviewed.

In study V, we developed a numerical comorbidity index to summarize the burden of disease and adjust for confounding in epidemiologic studies based on Nordic health registries.

Methods

The Danish registries

Denmark has a long-standing tradition for maintaining administrative health registries that are well suited for research purposes. The registries are linkable on individual level with a unique personal identification number assigned to all Danish residents at birth or immigration [12]. For example, the unique personal identification number is used when a patient sees a general practitioner (as gatekeepers for secondary care, general practitioners are often the patient's first contact with the health care system), when a prescription drug is filled at the pharmacy, when a patient is seen at the hospital for inpatient and outpatient care, and when a biopsy from a resected tumor is examined by the pathologist. The unique personal identification number enables linkage between individual registries and allows for follow-up of all Danish residents. In other words, the Danish registries provide an open cohort with accurate dates of entry at birth or immigration and exit at death or emigration [13]. Coupled with the many registries with, e.g., pathology reports and hospital diagnoses, this enables us to use the entire Danish population as a cohort for health research. The key registries used in this thesis are the Danish National Patient Registry, the Danish Cancer Registry, and the Danish Prescription Registry.

The Danish National Patient Registry

The Danish National Patient Registry was founded in 1977 and has provided nationwide coverage on somatic inpatient contacts since 1978. Since 1995, all somatic and psychiatric in- and out-patient encounters and emergency department contacts are recorded in the patient registry [14]. The registry holds information on contact date, contact type (admission or ambulatory care), primary and secondary

discharge diagnoses, and surgical procedures and hospital-based treatments. The patient registry is also the primary data source for recording of incident cancer in the Danish Cancer Registry [15].

The Danish Cancer Registry

The Danish Cancer Registry (DCR) was founded in 1942 [15]. The establishment of the DCR was a remarkable achievement considering the fact that classic epidemiology as a research field was in its infancy at that time [5]. The DCR has been an important tool to study the prevalence, incidence, and prognosis of cancers over time and has enabled comprehensive epidemiologic research into the causes of cancer [15, 16]. The DCR has recorded nationwide data on incident cancers in Denmark since 1943 with almost complete coverage and high histological verification [15, 17]. The data has been based on mandatory electronic reporting of incident cancers by public hospital departments, pathology institutes, private hospitals and specialists, and general practitioners; from 1943 by notification forms and since 2004 electronically primarily via the Danish National Patient Registry [15, 18]. The reports are coupled with pathologic-anatomical diagnoses from the Danish National Pathology Registry and cancers are validated using a combination of automatic algorithms and manual checks [19, 20]. Incident cancers from 1978 and onwards are classified according to the International Classification of Diseases version 10 (ICD-10) and International Classification of Diseases for Oncology version 3 (ICD-O-3) [15]. Since 2004, the DCR has recorded cancer stage according to the Tumor, Node, Metastasis (TNM) classification [21] although the completeness of information regarding cancer stage vary considerably with the specific cancer in question [22].

The Danish National Prescription Registry

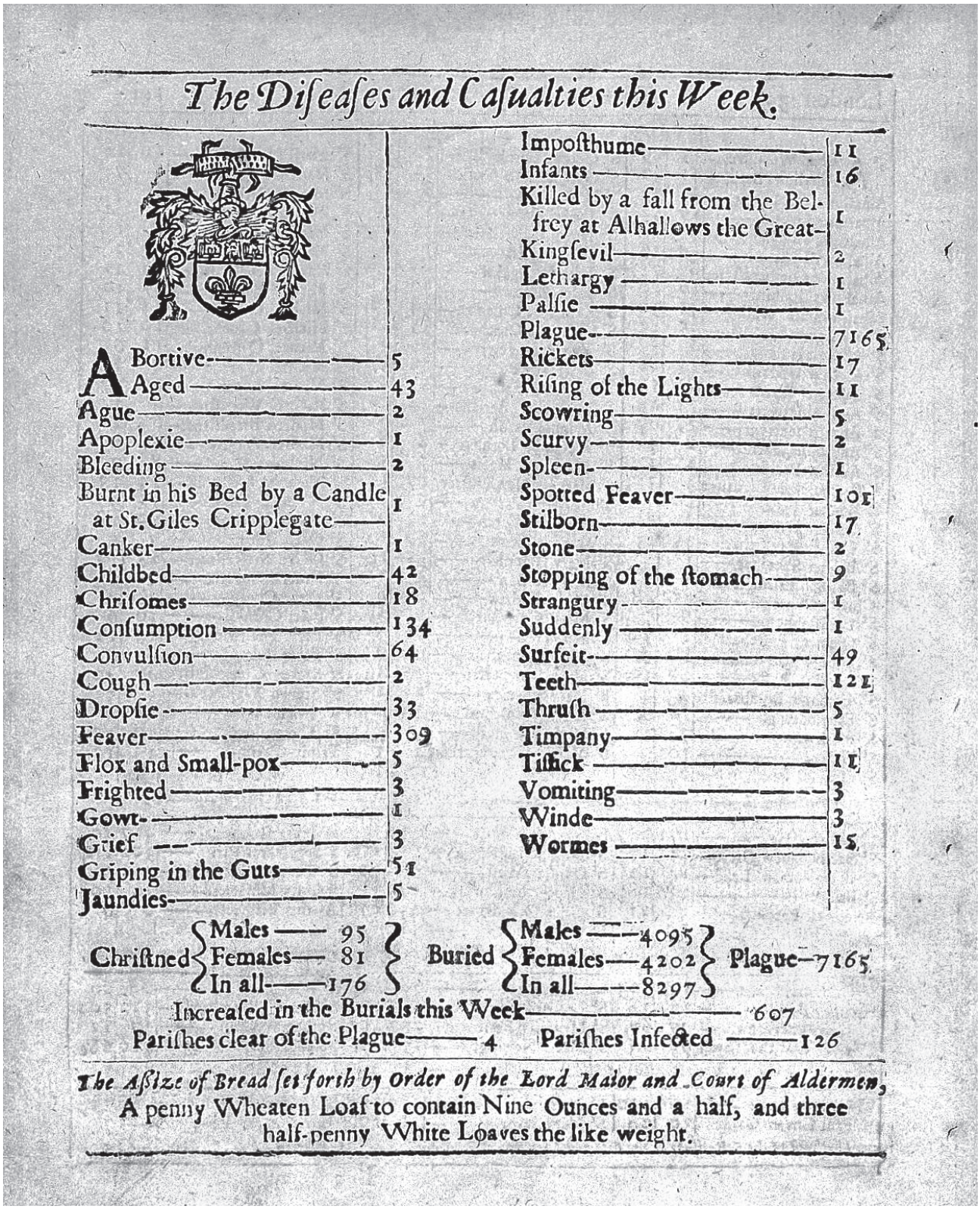
The Danish National Prescription Registry holds data on all filled prescriptions at community pharmacies in Denmark since 1995 [11]. The data that enter the prescription registry are born when a patient visits a community pharmacy. Here, the pharmacist scans the patient's health insurance card with the personal identification number. Together with the barcode of the sold drug package, this enables automatic reporting to the prescription registry of the patient id, date of dispensing, number of packages sold, and the product code of the sold package [11]. Using the product code, detailed information on the contents of the drug package can be obtained including the name of the drug, the ATC code of that drug, the package size (e.g., number of tablets in each package), the strength (e.g., the amount of active substance in each tablet) and the number of defined daily doses in each package. Because of the automatic data entry, data errors are minimized. Information on the intended duration and dose of the drug is not readily available so researchers must make educated assumptions about the actual dose and/or duration for each prescription fill. A widely used method is to assume that the patient uses the whole package with a daily dose corresponding to the defined daily dose of the drug (DDD). Another approach is to estimate the duration of a prescription fill empirically using the reverse waiting time distribution [23]. Around 90% of all prescriptions in Denmark are issued by general practitioners [24]. Thus, insight into the clinical reasoning and circumstances in general practice is essential to conduct pharmacoepidemiologic studies based on Danish data. The Danish National Prescription registry cover prescriptions that are bought by the patient after being prescribed. This is in contrast to physician-based registries e.g., the Clinical Practice Research Datalink in UK where all prescribed drugs are recorded including those not filled at the pharmacy [25]. One in ten prescriptions that are prescribed in

Denmark are never filled (primary non-adherence) and if these prescriptions were included in the exposure assessment it would introduce misclassification [26]. We do have to assume that the patient uses the drug after filling the prescription and the extent of misclassification introduced by secondary non-adherence is difficult to quantify.

From an epidemiological perspective, the Danish data with hospital diagnoses since 1977 and prescription drugs since 1995 are unique [13]. Fortunately, other population-based registries are beginning to accumulate enough follow-up time to study late drug effects such as cancer, for example in Norway and Sweden where the nationwide prescription registries has coverage from 2004 and 2005, respectively [27]. This allows for replication studies and more widespread examination of drug-cancer associations which is important to advance the field.

Classification of diseases and drugs

Classification systems enable us to communicate unambiguously and precisely about drugs, cancers, and diseases. In Denmark, the International Classification of Diseases version 10 is used to classify diseases [28, 29], the International Classification of Diseases for Oncology version 3 is used to classify cancers [30], and the Anatomical Therapeutic Chemical codes are used to classify drugs [31]. Disease classification systems are essential tools in research and their initial development was driven by the need of clinicians to describe diseases in a uniform way at a time where there was a myriad of different disease names but no accepted disease definitions [32]. An early motivation to classify disease was to keep track of deaths and their causes. Early death counts, as shown in the 17th century London mortality bill on the next page, did not follow a standard nomenclature and the same cause of death could have many different names.



The Diſeaſes and Caſualties this Week.

Abortive	5	Impoſthume	11
Aged	43	Infants	16
Ague	2	Killed by a fall from the Belfrey at Alhallows the Great	1
Apoplexie	1	Kingſevil	2
Bleeding	2	Lethargy	1
Burnt in his Bed by a Candle at St. Giles Cripplegate	1	Palſie	1
Canker	1	Plague	7165
Childbed	42	Rickets	17
Chriſomes	18	Riſing of the Lights	11
Conſumption	134	Scowring	5
Convulſion	64	Scurvy	2
Cough	2	Spleen	1
Dropſie	33	Spotted Feaver	101
Feaver	309	Stilborn	17
Flox and Small-pox	5	Stone	2
Frighted	3	Stopping of the ſtomach	9
Gowt	1	Strangury	1
Grief	3	Suddenly	1
Griping in the Guts	51	Surfeit	49
Jaundies	5	Teeth	121
		Thruſh	5
		Timpany	1
		Tiſſick	11
		Vomiting	3
		Winde	3
		Wormes	15

Chriſtned	Males — 95		Buried	Males — 4095	
	Females — 81			Females — 4202	Plague — 7165
	In all — 176			In all — 8297	

Increaſed in the Burials this Week — 607

Pariſhes clear of the Plague — 4 Pariſhes Infected — 126

The Aſſize of Bread ſet forth by Order of the Lord Maior and Court of Aldermen,
A penny Wheaten Loaf to contain Nine Ounces and a half, and three half-penny White Loaves the like weight.

London's dreadful visitation: or, a collection of all the Bills of Mortality for this present year: beginning the 27th of December 1664 and ending the 19th of December following: as also the general or whole years bill. According to the report made to the King's most excellent Majesty / by the Company of Parish-Clerks of London. Provided by the Wellcome Collection: https://wellcomecollection.org/works/bqxkq9yy licensed under Public Domain Mark.

The need of a uniform way to describe causes of death as well as diseases was acknowledged during the 18th and 19th century where taxonomical schemes were widely used to organize knowledge [32]. Taking the London mortality bill as example, it's utility would be greatly enhanced if a hieratical classification was used instead of ordering the mortality causes alphabetically. For example, *"Burnt in his Bed by a Candle at St. Giles Cripplegate*" would be classified under the ICD-10 code X05: *Exposure to ignition or melting of nightwear* and *"Killed by a fall from the Belfrey at Alhallows the Great"* would be classified as W13: *Fall from, out of or through building or structure*. Both ICD-10 codes belong to the same ICD-10 section V01-X59: *Accidents*, thus, if we were interested in the number of deaths due to accidents in London in the 17th century, a hierarchal classification of death causes would make this task much easier. Founded on groundwork by many others, Jacques Bertillon published the International List of Causes of Death in Paris in 1893 [32]. This list was revised to become the first version of the International Classification of Diseases (ICD-1) in 1901 and by 1909, the ICD-1 was in use throughout the world [32]. The ICD has since been continuously revised. The 9th version included major revisions, among others a more precise adaptation for cancers with the ICD for Oncology (ICD-O) [32]. The ICD-10 has been used in Denmark since 1994 as part of the Danish Health Care Classification System [28]. The ICD-10 is organized as a four-character coding system with a letter followed by up to three digits, e.g.:

Chapter II (C00-D48)	Neoplasms
C69-C72	Malignant neoplasms of eye, brain and other parts of central nervous system
C70	Malignant neoplasm of meninges
C70.1	Spinal meninges

Modified after World Health Organization: International Statistical Classification of Diseases and Related Health Problems 10th Revision: icd.who.int/browse10/2016/en

The widely used classification of drugs, the Anatomic Therapeutic Chemical (ATC) system was introduced in 1981 as a response to the need of uniform definitions of drugs in research [33]. The ATC is a hierarchal classification with five levels with 14 main groups (the 1st level) down to the individual chemical substance (5th level) as illustrated for metformin below.

A	Alimentary tract and metabolism
A10	Drugs used in diabetes
A10B	Blood glucose lowering drugs, excl. insulins
A10BA	Biguanides
A10BA02	metformin

Modified after: WHO Collaborating Centre for Drug Statistics and Methodology, Norwegian Institute of Public Health. ATC, Structure and principles: www.whocc.no/atc/structure_and_principles

Along with the ATC system, a standardized unit of measurement, the defined daily dose, was developed. The defined daily dose is defined as "*the assumed average maintenance dose per day for a drug used for its main indication in adults*" [34]. The DDD is not necessarily equal to the recommended dose since the therapeutic doses vary depending on the indication of the drug (including off-label indications), interactions with other drugs, and patient characteristics including age, weight, kidney function, and liver function. Nevertheless, the DDD serves as a standardized measure enabling comparison of drug consumption between countries over time in drug utilization studies [34].

Carcinogenesis following drug exposure

It often takes several years from the initial transformation of normal cells to cancer cells until the cancer is diagnosed. Cancer development is a multistage process where normal cells progress to clonally expanding cancer cells through different stages [35]. The induction period is the time from the initial cellular changes until a clone of dividing cancer cells has developed and the latent period is the time from onset of the cancer clone until the cancer is diagnosed [36]. Drugs may exert a carcinogenic effect as tumor initiators by initiating the development of cancer in the induction period and as tumor promoters by increasing the rate of cancer expansion in the latent period.

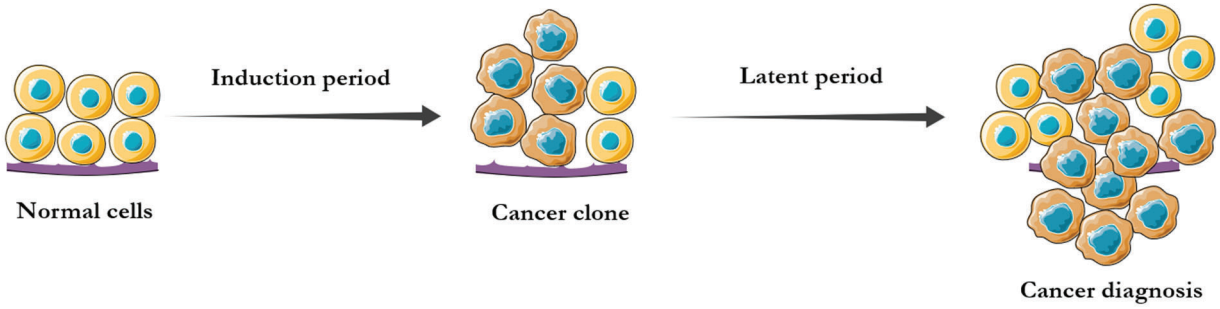


Illustration of the development of a clone of cancer cells from normal cells (induction period) and the time from cancer development until clinical manifestation or diagnosis of the cancer (latent period). The figure is adapted from illustrations provided by Servier Medical Art: smart.servier.com licensed under CC BY 3.0

It is impossible to distinguish between these periods empirically and the periods may be combined to one term, the latency period, covering the time from the first exposure to the initial cellular changes and clonal expansion until the cancer is diagnosed. The exact duration of the latency period is generally unknown, however, it likely spans several years for most cancers. Nadler and Zurbenko estimated that the time from initiation of a cancer until it was diagnosed often exceeded 10 years but with variability between cancer types [37]. While the majority of cancers seem to develop over years, some cancers, mainly hematological cancers, have shorter latency periods illustrated by the notorious sudden increase in leukemia incidence within 2 years after the bombing of Nagasaki and Hiroshima in 1941 [38].

The case-control study

As described in the introduction, it was two large, methodologically sound case-control studies that established the first solid evidence of smoking as a cause of lung cancer [6, 7]. The case-control study continues to be a fundamental method in epidemiology and relies on the basic principle of identifying individuals with a disease of interest (cases) and comparing their exposure to that of individuals without disease (controls).

Identification of cases and sampling of controls

In this thesis, we identified cases using the Danish Cancer Registry [15]. We used a combination of ICD-10 codes to describe the cancer site and ICD-O-3 morphological codes to describe the histological type of cancer. We restricted the cancer outcomes to those that were histologically verified to ensure validity of the cancer outcomes and because histological subtypes of cancers from the same organ may represent different diseases with regards to etiology and prognosis. We did not include individuals with cancer before the age of 18 since pediatric cancers represent a specific disease entity that may be due to genetic exposure and early environmental exposures and because use of prescription drugs is rare in children. We excluded patients with previous cancer, partly because oncologic treatment regimens may increase risk of subsequent cancers and partly to reduce the risk that the cancer of interest was an incorrectly diagnosed primary tumor or a metastasis from an earlier tumor [39]. We also excluded individuals with recent migrations to ensure capture of drug use and covariates. We matched each case to a number of controls on age and sex using risk-set sampling. The strategy consists of identifying a risk-set for each case consisting of individuals who reside in Denmark at the index-date and are at risk of the given cancer. A number of controls is then randomly selected from

each risk-set to form the final study population. We applied the same exclusion criteria to the controls as described above and cases were eligible as controls before their cancer diagnosis. With this sampling scheme, the odds ratios reflect incidence rate ratios of a cohort study of the entire danish population [40].

Classification of drug exposure

After identifying cases and sampling controls, we retrieved data on filled prescription drugs for cases and controls. We included data on prescription fills prior to the index date, however, we disregarded drug fills in a period just before the index date. This lag-time was most often two years and we ignored drug use in the two years leading up to the index date since recent exposure is not likely to be causally related to the development of the cancer in question [41]. Further, the lag-time may mitigate reverse causation and surveillance bias as discussed later in this thesis. We assessed cumulative dose-response patterns in all studies. A cumulative dose-response association is plausible for most cancers since cancer development is a multistep process driven by accumulation of mutations and increasing malignant potential [42]. The thoughts set forth by Hill in 1965 specifically mentions dose-response as an argument for causation that was labeled biological gradient; "*For instance, the fact that the death rate from cancer of the lung rises linearly with the number of cigarettes smoked daily, adds a very great deal to the simpler evidence that cigarette smokers have a higher death rate than non-smokers*" [10].

Analysis

The analyses followed the conventional analyses for a matched case-control study using conditional logistic regression, hence the matching variables were adjusted for by design (age, sex, and calendar time). Other potential confounders were included

as covariates in a multivariate conditional logistic regression model. We conducted a range of supplementary analyses that included e.g., examining effect measure modification by patient characteristics such as age and sex and stratified analyses by cancer stage (e.g., localized disease vs. non-localized disease). If an association was present for localized disease but not for non-localized disease, this could indicate greater health care surveillance in the exposed group, i.e., surveillance bias. Other supplementary analyses evaluated the robustness of our results towards our analytical choices. For example, we varied the lag-time to examine whether this affected the results and associations for similar drugs were assessed to evaluate whether an association could be due to confounding by indication.

Screening studies

The hypothesis-free screening study in this thesis was based on the idea of conducting several case-control studies, in principle, one for each of the included cancer outcomes. For each outcome, we estimated associations for long-term use of the drug compared to never-use when there were 25 or more long-term users among cases. Long-term use was practically defined as 8 or more prescriptions. The cutoff of 25 was chosen based on the recently proposed *bottleneck analysis* [43]. In practice, the number of exposed cases will be the limiting factor for precision, and the bottleneck analysis is used to estimate the width of the confidence intervals (CI) for a given bottleneck parameter. For example, with 25 exposed cases, the expected width of the CI for a null result (odds ratio = 1) would be a CI ranging from 0.7 to 1.5. We chose to apply this, agreeably arbitrary, cutoff arguing that more statistical imprecision would make interpretation of the thousands of associations even harder. The main effect estimate of interest was the odds ratio (OR) for long-term use and to further qualify the assessment of each drug-cancer association we provided ORs

for each doubling of cumulative dose in users. In these analyses, users were compared to users within different levels of cumulative dose instead of comparing users to never-users. Hence, in addition to describing cumulative dose-response patterns, these estimates may be less prone to confounding. We also estimated the OR for low use of the drug (1-2 prescriptions) compared to never use. Since this cumulative dose is too low to plausibly affect cancer development, a positive association with low use could indicate bias by e.g., confounding. Lastly, we showed the population attributable fraction for long-term use of the drug. The population attributable fraction is estimated from the OR for long-term use and the prevalence of long-term among cases and can be interpreted as the hypothetical fraction of cancers that would be prevented if long-term use of the drug was eliminated [44]. This interpretation assumes that the OR are estimates of the causal effect of the drug on cancer development and should not be interpreted literally in our study. However, it may serve as an indicator of the potential contribution of a drug to the prevalence of a given cancer on population level.

We analyzed approximately 14,000 drug-cancer associations and with a traditional significance threshold of 0.05, 700 of these would be false positives. The number of false positives could be reduced by imposing a penalty for multiple testing for example by requiring a lower significance threshold [45]. However, this would inevitably reduce our ability to identify associations that were due to a carcinogenic effect of the drug. Instead, we chose a more liberal way of highlighting associations that were then subject to further evaluation in a manual review. We highlighted associations for manual review based on the strength of association with long-term use and presence of a cumulative dose-response association. The manual review of these signals was based on subject matter knowledge regarding risk factors for the cancer and indications for the drug in question and served to classify the associations

as to their credibility and potential for bias. The overall assessment of the likelihood that a given association may be due to a causal effect of the drug may include the before mentioned Bradford Hill criteria [10]. For example, one Bradford Hill criterion is related to the strength of the association. For drug-cancer associations with ORs far away from unity it seems less likely that some other factor explains the association. Another criterion is related to the biological gradient, that is if a cumulative dose-response pattern is observed this may be indicative of a cumulative, causal effect of the drug on cancer risk. Another criterion is related to the specificity of the association. We assessed specificity by evaluating whether drugs with similar indications but different mechanisms of actions were associated with the cancer as well. We also assessed specificity by evaluating whether the drug of interest was associated with several different cancer sites and/or histologic subtypes.

Comorbidity indices and predictive models

In study V, we aimed to develop a numerical score that reflected the comorbidity, frailty, and general health state of an individual for use specifically in research based on the Nordic registries. Such indices are used widely because they are easy to use, they can summarize the burden of disease with a single variable, and they can be used in a standardized way across studies. The most used scores are the Charlson and Elixhauser comorbidity indexes [46, 47], however, these were developed to predict in-hospital mortality or other outcomes related to hospital admissions. We developed our index score to predict 5-year mortality in a Danish population of community dwelling adults. Contrary to the other studies in this thesis, the objective was to predict as opposed to describe or explain. Our aim was to build a model that, based on a set of conditions, predicted 5-year mortality. This model could be constructed by using artificial intelligence or machine learning

algorithms or it could be based on traditional logistic regression, predicting the logit transformed probability of death within 5 years based on the predictors x_1-x_k and their parameters:

$$logit(p_{death}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_k x_k .$$

This model is easy to describe and interpret and methods based on artificial intelligence or machine learning have not been shown to outperform logistic regression models [48, 49]. The included predictors were selected based on a comprehensive review of diagnoses and prescriptions with a frequency above 1:1000 in the Danish adult population. We only included outcomes that we deemed meaningful and had face validity. As such, this step of the variable selection was based on subject matter knowledge and resulted in 150 potential predictors. Because we aimed to develop a simple model that was easy to apply, and where it was possible to assess the individual components, we reduced the model further using automatic variable selection. We used a traditional backwards elimination approach where we fitted all 150 predictors along with age and sex in a logistic regression model. The predictor with the highest p-value was dropped and the model was fitted anew with the 149 remaining predictors. This process was repeated until 50 predictors remained. To assess the performance of the index, we evaluated how the index discriminated between patients who died and survived using the concordance (c) statistic [50]. We assessed calibration by comparing the predicted mortality rates to the observed mortality rates and conducted a comprehensive validation in cohorts different from the cohort used to develop the model. We examined the performance in a temporal validation cohort and in six cohorts of new users of selected drugs.

Summary of findings

In this section, the background, main findings, and interpretation of each study are described. For more details, please refer to the appendix where all papers are included in their full length.

Study I

Prostate cancer is the most prevalent cancer among men in the Nordic countries and only non-modifiable risk factors (age, family history, ethnicity, and genetics) have been established [51, 52]. Thus, a drug that can be used to prevent or treat prostate cancer would have a big impact on public health. In 2000, a secondary analysis of a randomized clinical trial (RCT) published in the New England Journal of Medicine pointed towards a possible anticarcinogenic effect of the anticoagulant drugs, Vitamin K Antagonists (VKA) [53]. The study reported a reduced incidence of `all cancers´ mainly driven by urogenital cancers in patients randomized to 6 months of VKA treatment compared to patients randomized to 6 weeks of VKA treatment followed for a mean of 8.1 years [53]. Since then, observational studies have produced conflicting results as evidenced by the systematic review that was part of study I. We included 8 studies in the meta-analysis with a pooled relative risk estimate of 0.86 (95% CI 0.70 to 1.05). Of note, we observed considerable between study heterogeneity and the pooled estimate should be interpreted with caution. The large degree of heterogeneity may be explained by differences in study populations, exposure definitions, prostate cancer definitions, and statistical analyses. A particular study was prone to immortal time bias [54–56], and when this study was excluded from the meta-analysis, the pooled estimate was 0.94 (95% CI, 0.83–1.07). In our case-control study including 38,832 men with prostate cancer, we found no association with an OR for 3 or more years of VKA use compared to never-use of

1.03 (95% CI 0.97 to 1.10). When stratifying by calendar time, we noted a tendency towards an inverse association early in the study period 2005 to 2008 (OR 0.93, 95% CI 0.82 to 1.06) that was reversed in the latter period 2012 to 2015 (OR 1.11, 95% CI 1.01 to 1.22). After publication of our study, a study demonstrated how detection bias could introduce an inverse association between VKAs and prostate cancer, the magnitude of which was reduced over calendar time [57]. The study was based on US Medicare beneficiaries from 1999 to 2015 and compared the incidence of prostate cancer in men receiving anticoagulants to men from the general population. The incidence of prostate cancer was reduced in the latter group, however, the risk difference decreased over time corresponding to a decreased use of prostate biopsies in the general population. The authors suggested that detection bias due to a lower rate of prostate cancer biopsies in anticoagulated patients could be an explanation for the inverse association between VKA and prostate cancer observed in earlier studies [57]. These observations illustrate that the evidence base is dynamic and ongoing critical appraisal of the literature is important. It can be argued that detection bias may also act in the opposite direction. Because patients receiving VKA have more frequent healthcare encounters they may be more likely to undergo screening for prostate cancer. If this was the case, we would expect that the association was stronger for localized disease compare to advanced disease, which was not the case. However, information on clinical stage of prostate cancer was not complete in the DCR and the missing information may be influenced by age and comorbidities [58]. Thus, this observation must be taken with some precaution.

We concluded that the existing evidence was conflicting but that a clinically relevant protective effects of VKAs against prostate cancer was unlikely.

Study II

This study originated from a conversation with a physician about a patient who had developed numerous squamous cell carcinomas of the skin while being treated with valproic acid. Valproic acid may increase the skin's sensitivity to sunlight and hence theoretically increase skin cancer risk where squamous cell carcinomas of the skin are particularly dependent on cumulative UV exposure [59]. Although several antiepileptic drugs are established as photosensitizing, only limited evidence with regards to skin cancer risk was available [60, 61], and we decided to conduct a study of antiepileptic drugs and their association with skin cancer. We included as outcomes the two most common types of non-melanoma skin cancer, basal cell carcinoma and squamous cell carcinoma as well as malignant melanoma. We examined the 11 most used antiepileptic drugs in Denmark and defined the main exposure of interest as a cumulative dose of 500 or more defined daily doses. We took interest in cumulative doses since antiepileptic drugs were expected to influence skin cancer risk through sensitization to UV exposure [59]. The results were overall reassuring and none of the 11 antiepileptic drugs were associated with an increased risk of basal cell carcinoma or malignant melanoma. Two antiepileptic drugs were associated with squamous cell carcinoma, namely carbamazepine (OR 1.88, 95% CI 1.42 to 2.49) and lamotrigine (OR 1.57, 95% CI 1.12 to 2.22). There is a biological rationale for the observed associations since carbamazepine and lamotrigine are known to cause photosensitivity reactions and have chemical properties compatible with photosensitizing potential [62–66]. Due to the nature of the study as an applied screening-study, we stressed that these associations should be considered hypothesis-generating and examined further. The main limitation of the study was the lack of data on UV exposure and skin type. Based on the indications for some antiepileptic drugs (e.g., phenobarbital for alcohol withdrawal

symptoms and valproic acid for bipolar disorder) it may be hypothesized that these drugs are negatively associated with outdoor activities and UV exposure. Further, users of antiepileptic drugs may be less aware of skin changes and less likely to undergo diagnostic work-up. This potential detection bias would move the estimates towards a seemingly protective effect of antiepileptic drugs, a tendency we noted for malignant melanoma, a disease where the incidence has increased rapidly during the last decades due to increased diagnostic intensity and biopsies [67].

To summarize, we concluded that most antiepileptic drugs were not associated with skin cancer and that the observed associations between lamotrigine and carbamazepine and squamous cell carcinoma had no clinical implications but should be evaluated in other studies.

Study III

In the third study, we aimed to describe the association between calcium channel blockers and renal cell carcinoma (the most common type of kidney cancer) while illustrating methods to account for confounding by indication. Hypertension is in itself a risk factor for kidney cancer and it has been difficult to disentangle the effect of hypertension from the effect of its treatment [68]. We aimed to evaluate the extent of this source of confounding by indication in several ways. First, we identified all incident cases of renal cell carcinoma in Denmark and matched each case with up to 20 controls. Compared to never-use, long-term use (1000+ defined daily doses) of calcium channel blockers was associated with renal cell carcinoma with an OR of 1.76 (95% CI 1.63 to 1.90). The association was weakened but did not disappear when adjusting for indicators for hypertension severity e.g., use of other antihypertensive drugs and a hospital diagnosis of hypertension (OR 1.37, 95% CI 1.25 to 1.49). We repeated the analyses for other first-line antihypertensive drugs, and they were all associated with renal cell carcinoma (after adjusting for calcium channel blockers). We interpreted the lack of specificity with regards to treatment as indicative of confounding by indication since the drugs have different mechanisms of action and a carcinogenic effect of all these drugs seem unlikely. We then compared calcium channel blockers to other first-line antihypertensive drugs directly, that is we changed the comparator group from never-users of calcium channel blockers to users of other first-line antihypertensive drugs. We chose three active comparators: angiotensin converting enzyme inhibitors, angiotensin receptor blockers, and thiazides. We did this by conducting three case-control studies nested in new users of calcium channel blockers and the active comparator drug. As expected, the point estimates were closer to unity and the confidence intervals were wider interpreted as a gain in study validity at the cost of statistical precision. Since

we did not have continuous data on blood pressure levels, we had to rely on indirect methods to account for confounding by high blood pressure. Another limitation was that we lacked data on the two other known modifiable risk factors for renal cell carcinoma: smoking and elevated body mass index [68].

We concluded that there was evidence of confounding by indication for the association between calcium channel blockers and renal cell carcinoma and that this was, at least partially, explaining the observed association. Future studies could add to our knowledge by using high-dimensional confounding adjustment methods e.g., propensity scores or disease risk scores and by utilizing data sources with serial blood pressure measurements.

Study IV

This was a hypothesis-free screening study aimed at uncovering potentially unknown carcinogenic effects of drugs. We included 456,828 individuals with incident cancer matched to 4,568,262 population controls. We examined 33 outcomes defined by cancer site and 85 outcomes defined by histologic subtype. A total of 13,577 associations were examined for individual drugs and 8996 associations for drug classes. From the initial 8373 associations between individual drugs and histologic subtypes of cancer, we highlighted 460 associations where the lower limit of the 95% CI for long-term use was above 1.25. Of these, 274 associations showed evidence of a cumulative dose-response pattern and were manually reviewed. The associations were classified in one of the following groups: *i)* likely explained by bias, *ii)* implausible to affect cancer risk due to pharmacological properties of the drug, and *iii)* not readily explained by bias. Of the drugs classified in the latter category, some were already established as carcinogens by the International Agency for Research on Cancers and others deserved further scrutiny. We compared our results to two recent screening-studies from Norway and Scotland to further qualify associations that could be worth further consideration [69, 70].

In conclusion, we conducted a hypothesis-free screening study intended to identify novel drug-cancer associations for further examination in tailored studies and for explorative and comparative purposes in drug-cancer research. The results should be considered hypothesis generating only, i.e., a positive association should not be interpreted causally, and the online results are not intended nor useful for clinicians and patients to inform decisions regarding drug use.

Study V

With the aim of developing a numerical summary score for comorbidity, frailty, and the general physical condition of an individual, we developed a 50-item index based on hospital diagnoses and prescriptions named the Nordic Comorbidity Index (NCI). The predictors in the NCI included among others high-ceiling diuretics, drugs for constipation, pneumonia, dementia, and metastatic cancer. The included predictors were consistent with the stereotypic frail patient seen in clinical practice and/or reflected specific conditions with a poor prognosis such as metastatic cancer. Each of the 50 items was assigned a weight that were summed to provide the final score for each individual. We validated the NCI in a temporal validation cohort separated in time as well as in six cohorts of new users of drugs. In all cohorts the performance of the NCI exceeded that of the typically used alternatives, the Charlson end Elixhauser indices.

We used a combination of variable selection based on subject matter knowledge and automatic variable selection to derive the predictors and their weights in the NCI. There are many ways of automatic variable selection all of whom have caveats, especially in small samples. Automatic variable selection techniques have been criticized because variables are dropped which invariably leads to loss of information [71, 72]. Automatic variable selection has also been criticized because of the risk of overfitting. Overfitting applies both to the coefficients of the included variables and the selection of variables to be included in the model [71, 73]. However, in our case with 38,301 events and an event to variable ratio of 255:1 it is less likely that overfitting has reduced the performance of the reduced model substantially compared to the full model. This is also evident from the good performance of the NCI in the validation cohorts. It has been argued that, with large samples, the assessment of model performance should focus on variation in performance

between settings, populations, and patient subgroups [74]. We assessed the performance of the NCI in several different settings and populations. The temporal validation cohort served as a type of external validation in that the patients were separated in time from the cohort where the model was developed, and the validation thus serves as a marker of the transportability of the NCI over calendar time. The validation in the new user cohorts showed that the NCI was transportable to specific patient populations with different baseline characteristics and mortality rates. Since most drugs in Denmark are prescribed by general practitioners, pharmacoepidemiologic studies are typically conducted in non-hospitalized populations. The NCI was developed in a general population as opposed to hospitalized patient populations used to develop the Charlson and Elixhauser indices. Even though the Nordic health registries are structurally similar, there may be different coding practices, drug use patterns and population demographics between countries. Hence, validation of the performance of the NCI in other Nordic countries is an area for future research.

To summarize, we concluded that the NCI had high predictive performance exceeding that of the Charlson and Elixhauser comorbidity scores and that the NCI could be used to adjust for confounding and describe the level of comorbidity in studies using Nordic registry data.

Discussion

In the five studies included in this thesis, we examined selected drug-cancer associations, conducted a hypothesis-free drug-cancer screening study, and developed a comorbidity index for use in pharmacoepidemiologic studies. In the first study, we summarized the conflicting evidence regarding use of VKAs and prostate cancer. The available evidence does not allow us to draw firm conclusions as to whether VKAs reduce the risk of prostate cancer, however, it seems justified to conclude that a clinically relevant effect of VKAs is unlikely. In the second study, we showed that even though most antiepileptic drugs are photosensitizing, their use is not linked to increased risk of skin cancer except for the association between squamous cell carcinoma and lamotrigine and carbamazepine. Due the nature of our study, these findings are to be considered hypothesis-generating and, accordingly, there are no direct clinical implications currently. In the third study, we illustrated methods to address confounding by indication and showed that the association between calcium channel blockers and renal cell carcinoma was at least partly explained by this bias. In the fourth study, we conducted a hypothesis-free screening of drug-cancer associations to identify unknown carcinogenic effects of drugs. We identified a number of drug-cancer associations that deserved further study. In the fifth study, we developed a comorbidity index to be used for descriptive purposes or confounding adjustment by comorbidity and frailty in Nordic registry-based studies.

Compared to other disciplines within pharmacoepidemiology, drug-cancer studies seem to be less well represented and, as a field, still in its infancy. The reasons for this may simply include the lack of suitable data. Until recently, many registries did not provide sufficient follow-up to study cancer outcomes occurring e.g., 10 to 15 years after the first exposure. Another explanation is the difficulty in studying

outcomes that occur many years after the initial exposure, the fact that the effect of a drug depends on the cumulative exposure to that drug, or the fact that it is hard to know when the cancer developed – when did the outcome occur and what is the corresponding relevant exposure window? Another concern is that researchers may worry that their research does more harm than good. Cancer is a term with many connotations and a word that attract headlines and media attention. This puts a lot of responsibility on the researchers in communicating and disseminating their findings responsibly. The discussion section will evolve around some of these issues in drug-cancer research including methodological considerations, the dissemination of research findings to the public, and the future perspectives for hypothesis-free screening studies.

Cancer as outcome

It is common to see studies that use the composite endpoint `all cancers´. This outcome is heavily influenced by the most common cancers (i.e., lung, breast, prostate, and colorectal cancer) and, more importantly, the etiologies of the individual cancers that are combined in this endpoint are not the same. Some carcinogens are associated with several cancers, most notably tobacco smoking that is listed as a cause of cancer for 17 sites by the International Agency for Research on Cancer [1]. However, even a strong carcinogenic exposure such as tobacco smoking does not increase the risk of all cancers [75]. Similarly, no drugs have been shown to increase the risk of all cancers and newer drugs are not likely to be strong carcinogens for cancers in general given the premarketing safety assessment including in vitro and animal studies. With this line of reasoning, it is difficult to provide a meaningful interpretation of a study showing an association between a drug and `all cancers´ and the outcome `all cancers´ should probably be avoided in

drug-cancer research. Preferably, cancer outcomes are defined by the affected site or organ, e.g., breast cancer. Even this level of granularity may not suffice since the histologic subtypes of a given cancer may vary with regards to etiology. For example, the most common types of esophageal cancer; squamous cell carcinoma and adenocarcinoma have different etiologies with marked historical and geographical variation in incidence [76]. In the 20th century, squamous cell carcinoma was the most common type of esophageal cancer and this still holds true for Asia. In the western countries, however, adenocarcinomas have become dominant owing to a dramatic increase in incidence during the last 40 years [77]. In study IV, the histological differentiation of cancers enabled us to discriminate between different subtypes of a cancer that could differ with regards to etiology and also to study rare subtypes, e.g., colorectal neuroendocrine carcinomas that would otherwise be diluted among colorectal adenocarcinomas constituting more that 95% of all colorectal cancers [78]. With the rapid development of molecular pathology, it is likely that subtyping of cancer based on molecular profiles will enable even more detailed classification of cancer outcomes and provide more clues to the etiology of specific cancers [79, 80].

Modeling drug exposure

It is often easier to study an effect of a drug when it occurs shortly after initiation of the drug. When the effect is separated by many years in time from its cause, it becomes more difficult to establish a link between the two. In drug-cancer research, the latter is most often the case. Considerations regarding the latency period is relevant in drug-cancer studies because exposure that is not measured within the relevant time window leads to exposure misclassification [36]. For example, in a case-control study of smoking and lung cancer, the effect of smoking on lung cancer

risk would be diluted if smoking habits of study subjects were measured at the index date ignoring the latency period of lung cancer of maybe 10 to 20 years [36]. Likewise, current use of a drug at the time of cancer diagnosis is not a relevant exposure because the cancer in question has likely developed silently over ten or 15 years before being diagnosed. On the other hand, patients may be at risk of the cancer even years after having discontinued the drug. In general, there is limited biological evidence to guide assumptions about the latency period and relevant exposure time window, so the robustness of the assumptions regarding exposure windows are often examined in sensitivity analyses. In the studies in this thesis, a lag-time was applied by disregarding exposure in a period leading up to the index date, e.g., disregarding all prescriptions filled in the two years before the index date. This served several purposes. One was to limit exposure misclassification since recent exposure is unlikely to cause the cancer that probably developed years ago. Secondly, applying a lag-time reduced the risk of reverse causation and detection bias. Use of drugs increases markedly 6 months prior to a cancer diagnosis which may reflect increased health-care seeking behavior due to symptoms of the underlying cancer [81]. If we fail to take this into account, reverse causation bias may occur. For example, patients with gastric cancer may experience unspecific symptoms consistent with heartburn and be prescribed proton pump inhibitors inducing an association between proton pump inhibitors and gastric cancer. Detection bias may be introduced when users of a drug undergo more frequent screening or diagnostic work-up compared to non-users. For example, lab tests are often ordered before initiating chronic therapy with e.g., antihypertensives, antidiabetics, or cholesterol-lowering drugs. These lab tests often include urinalysis and blood tests where abnormal findings can lead to further diagnostic testing.

In drug-cancer studies, cumulative exposure is of main interest. Often, drugs are used over many years, i.e., the exposure is chronic. It may take years before the drug has had its effect in the sequence of events that eventually result in the development of cancer. Since cancer development is a stochastic, multistage process dependent on multiple failures [35], the continuing exposure to a carcinogenic drug may increase cancer risk in a fashion that is dependent on the cumulative exposure to the drug. We defined a minimum threshold for a cumulative dose to plausibly affect cancer development as the main exposure of interest. This threshold was, admittedly, arbitrarily defined, but justified by the above considerations regarding exposure misclassification. The assessment of cumulative dose-response was carried out by including cumulative dose as a categorical exposure variable. The levels for the categories were guided by pharmacologic reasoning or assumptions related hereto. A common practice is to assert dose-response patterns using the observed distribution of doses in the study population; however, this approach may fail to define biologically plausible thresholds of when a cumulative dose is likely to increase cancer risk. Another option is to avoid categorization altogether and retain the cumulative dose as continuous variable, e.g., using restricted cubic splines or fractional polynomials as we did in study III [82].

Case-control studies and cohort studies

The drug-cancer studies in this thesis were case-control studies. We chose this design because of the need to study effects of cumulative exposure which is more efficient in case-control studies compared to cohort studies. This was particularly important in the screening study. A limitation of the case-control study is that absolute risk estimates are not readily available. Relative risk estimates are frequently used to describe the association between an exposure and disease risk in

epidemiological studies and are arguably relevant when we are interested in whether a drug increases the risk of cancer in the first place. However, if a drug is established as a possible carcinogen, absolute risks are essential to guide public health interventions and inform patients and physicians. It is difficult to interpret the magnitude of harm from relative risks and people are prone to read a high relative risk as equal to a highly clinically relevant risk [83]. Take as example the contraceptive pill scare in UK in 1995. The UK committee of Safety of Medicines issued a statement in 1995 that third-generation contraceptives increased the risk of thromboembolic disease by 100% (relative risk of 2). This was reported in the media and led to a surge of unintended pregnancies and abortions in the UK [84]. We may imagine that many women would have interpreted the corresponding absolute risk increase differently. The absolute risks was 1 in 7000 women using second generation contraceptives compared to 2 in 7000 women using third generation contraceptives [83]. The same applies to drug-cancer research where the absolute risks differences are often even lower. We did provide a measure related to absolute risks in study II by providing the `exposure needed for one additional patient to be harmed´ [85]. For carbamazepine we estimated that 6335 persons years of long-term use of carbamazepine were required for one additional squamous cell carcinoma to occur.

Dissemination of findings

During the thesis, I have increasingly recognized that public dissemination of drug-cancer research is an important and integrated part of conducting research in this field. Cancer is a disease that inspires fear. For example, 59% of all individuals were more afraid of cancer than any other disease in a study from the UK [86]. In a systematic review of cancer fears in the general population, cancer fear was linked

to seeing cancer as a hostile, volatile, and malicious personality, and not just a disease much aligned with the common "war on cancer" theme in the public and the media [87]. With a word and disease that is so emotionally loaded, drug-cancer research may tend to be misrepresented and misinterpreted in the media. If research findings are exaggerated and sensationalized, this could lead to unnecessary worry and even do harm. We may fear that patient seek to non-working or harmful alternatives, opt for less effective treatments, or simply discontinue their drugs. As researchers, we may engage with the established media as well as social media to increase the likelihood that our research is communicated responsibly. As a member of the Danish Society for Pharmacoepidemiology I have been involved in the formulation of a position paper outlining a set of principles and concrete guidance on how to talk to the media [88]. Among others, we underlined the need to simplify messages by the researcher. The media cannot convey the many uncertainties and assumptions in research, so the conclusions will be simplified, and the researcher is most likely the one who is able to simplify the message in the most responsible way. We also stressed the importance of being specific. If the research revolves around a specific drug, it is preferrable to use the specific name of the drug, e.g., by using acetaminophen instead of referring to the drug class weak analgesics. In general, it is of value to be realistic about the importance and implications of the specific study and recognize the uncertainties related to the findings including the inherent sources of bias and play of chance. In one of the studies included in our systematic review on VKAs and prostate cancer risk, the authors found a reduced risk of the outcome `all cancers´ associated with use of warfarin as well as breast cancer, lung cancer, prostate cancer, gynecological cancers, bladder cancer, brain cancer, stomach cancer, head and neck cancer, and cancer of the endocrine glands [54]. The authors concluded that, "*Warfarin use may have broad anticancer potential in a large, population-based*

cohort of persons older than 50 years. This finding could have important implications for the selection of medications for patients needing anticoagulation"[54] and in the discussion section it was stated that, "*The well-known challenges of warfarin dosing that necessitate regular monitoring have fueled a transition to new oral anticoagulants. An unintended consequence of this switch to new oral anticoagulants may be an increased incidence of cancer, which is an important consideration for public health.*"[54]. The study was criticized for introducing immortal time bias since the exposure definition seemed to depend on future events [55, 56], a bias that has been described in several studies reporting a beneficial effect of various drugs [89]. Unlike our meta-analysis on VKA use and prostate cancer, this study was widely reported in the media with headlines such as "*A common blood thinner may protect against some cancers*"[90] and "*Warfarin shows broad anticancer potential*"[91]. These headlines align well with the above citations from the discussion and conclusions section of the paper, however, there are no clinical or public health implications of the study considering the (lack of) validity of the study findings and lack of replication hereof in other populations.

Screening studies and future of pharmacovigilance

Traditional pharmacovigilance is based on spontaneous reporting of suspected adverse events, however this method has several disadvantages including being influenced by media attention, incapability of detection common adverse effects (for example cardiovascular events with the so-called coxibs) and delayed adverse effects such as cancer [92, 93]. Therefore, there is a need to utilize other methods in pharmacovigilance, and study IV is an example hereof. By adapting traditional epidemiological principles towards hypothesis-generation and utilizing the increasingly available health registry data worldwide, screening studies are potentially going to change pharmacovigilance practices. However, there are still limitations

that must be addressed. When the aim moves from examining a single hypothesis related to a specific drug and cancer outcome towards an agnostic approach, several challenges arise. A main challenge is how to process the output of the screening studies. How do we prioritize between the thousands of examined associations? It is not feasible to manually review all associations, applying a significance threshold based on multiple testing will lead us to oversee important adverse effects, and applying thresholds related to e.g., the strength of association and dose-response assessment involves arbitrary decisions on where to place these thresholds. In the drug-cancer screening study, we combined a filtration on strength of association, dose-response assessment, and manual review. Even after keeping only those associations strongly associated with the drug and evidence of a dose-response pattern, most associations were readily attributable to bias by confounding. The associations we were not able to explain were compared to drug-cancer screening studies from Norway and Scotland and the minority of associations were replicated in all three countries.

Another potential issue is how to evaluate drug-cancer associations of potential interest. Can we follow-up on an association from the screening study using the same data as the screening study is based on? If the same hypothesis is essentially tested re-using the same data as the screening study, the follow-up study should not be regarded as a confirmation of the results of the screening study but would still hold value in refining the very crude estimates of the screening study [94, 95]. The follow-up study would be tailored to the specific drug-cancer association of interest and potential confounders, mediators and colliders would be identified and guide the analyses for that specific association and cumulative dose-response associations could be examined in more detail. Confirmatory results from the same data source can be obtained if it is possible to define orthogonal hypotheses, i.e., hypotheses

that are not merely a repetition of the initial finding in the screening study [96]. A hypothesis is orthogonal if evolves around the same biological phenomena but is statistically independent from the initial hypothesis [97]. This can be achieved by using a different exposure, a different outcome, or a different population than the study where the initial hypothesis was born. For example, an association between hydrochlorothiazide and lip cancer could be qualified further by testing the orthogonal hypothesis that hydrochlorothiazide is associated with skin cancers other than lip cancer [96]. Replication in other populations, however, remain important and there is considerable value in replicating study findings across different populations. Such replication studies can be facilitated using common data models and currently, such a setup is underway for Danish, Norwegian and Dutch data allowing for replication of drug-cancer associations with limited delay. As another means of identifying relevant associations, associations with a known pathway through which the drug could exert a carcinogenic effect could be identified. A potential database to assess this could be the Kyoto Encyclopedia of Genes and Genomes pathway database where pathways for specific cancers are listed [98].

Concluding remarks

The work presented here on drug-cancer associations have aimed at increasing our understanding of drugs as possible causes of cancer. While we can analyze associations of drug use and cancers in observational studies based on large health registries such as the Danish registries, there is still work to be done to improve our understanding beyond merely identifying and describing associations. In the end, we aim to understand the causal mechanisms at play for drugs to induce or promote cancer. Further, although hypothesis-free screening studies hold promise in pharmacovigilance, advances must be made with regards to signal identification and establishing collaborations where drug-cancer associations can be replicated in different populations within a short time frame.

References

1. International Agency for Research on Cancer, World Health Organization (2021) List of Classifications – IARC Monographs on the Identification of Carcinogenic Hazards to Humans. https://monographs.iarc.who.int/list-of-classifications. Accessed 30 Nov 2021
2. Coggon D, Rose G, Barker DJP (2003) Epidemiology for the Uninitiated, 5th ed. BMJ Books, London, UK
3. Melicow MM (1975) Percivall Pott (1713–1788) 200th Anniversary of First Report of Occupation-Induced Cancer of Scrotum in Chimney Sweepers (1775). Urology 6(6):745–749. https://doi.org/10.1016/0090-4295(75)90812-2
4. Pott P (1775) Chirurgical observations relative to the cataract, the polypus of the nose, the cancer of the scrotum, the different kinds of ruptures, and the mortification of the toes and feet. T.J. Carnegy for L. Hawes, W. Clarke & R. Collins, London
5. Morabia A (2014) History of Epidemiological Methods and Concepts. In: Ahrens W, Pigeot I (eds) Handbook of Epidemiology. Springer, New York, NY, pp 43–74
6. Doll R, Hill AB (1950) Smoking and Carcinoma of the Lung. Br Med J 2(4682):739–748
7. Wynder EL, Graham EA (1950) Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma; a study of 684 proved cases. J Am Med Assoc 143(4):329–336. https://doi.org/10.1001/jama.1950.02910390001001
8. White C (1990) Research on smoking and lung cancer: a landmark in the history of chronic disease epidemiology. Yale J Biol Med 63(1):29–46
9. IARC Working Group on the Evaluation of Carcinogenic Risk to Humans (2012) Personal Habits and Indoor Combustions, IARC monograph volume 100E. International Agency for Research on Cancer, Lyon, France
10. Hill AB (1965) The environment and disease: association or causation? Proc R Soc Med 58:295–300

11. Pottegård A, Schmidt SAJ, Wallach-Kildemoes H, Sørensen HT, Hallas J, Schmidt M (2017) Data Resource Profile: The Danish National Prescription Registry. Int J Epidemiol 46(3):798–798f. https://doi.org/10.1093/ije/dyw213

12. Schmidt M, Pedersen L, Sorensen HT (2014) The Danish Civil Registration System as a tool in epidemiology. Eur J Epidemiol 29(8):541–9. https://doi.org/10.1007/s10654-014-9930-3

13. Schmidt M, Schmidt SAJ, Adelborg K, et al (2019) The Danish health care system and epidemiological research: from health care contacts to database records. Clin Epidemiol Volume 11:563–591. https://doi.org/10.2147/CLEP.S179083

14. Schmidt M, Schmidt SAJ, Sandegaard JL, Ehrenstein V, Pedersen L, Sørensen HT (2015) The Danish National Patient Registry: a review of content, data quality, and research potential. Clin Epidemiol 2015:7:449—490. https://doi.org/10.2147/CLEP.S91125

15. Gjerstorff ML (2011) The Danish Cancer Registry. Scand J Public Health 39(7 Suppl):42–5. https://doi.org/10.1177/1403494810393562

16. Pukkala E, Engholm G, Højsgaard Schmidt LK, et al (2018) Nordic Cancer Registries – an overview of their procedures and data comparability. Acta Oncol 57(4):440–455. https://doi.org/10.1080/0284186X.2017.1407039

17. Friis S, Jørgensen T, Mellemkjær L, Olsen JH (2012) Validation of The Danish Cancer Registry and selected Clinical Cancer Databases. The Danish Cancer Society and Statens Serum Institut. https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-de-nationale-sundhedsregistre/sygedomme-laegemidler-og-behandlinger/cancerregisteret. Accessed 30 Nov 2021

18. The Danish Health Data Authority (2019) Nye kræfttilfælde i Danmark 2018. https://sundhedsdatastyrelsen.dk/da/tal-og-analyser/analyser-og-rapporter/sygdomme-og-behandlinger/kraeft/kraeft-_-nye-tilfaelde. Accessed 30 Nov 2021

19. Bjerregaard B, Larsen OB (2011) The Danish Pathology Register. Scand J Public Health 39(7_suppl):72–74. https://doi.org/10.1177/1403494810393563

20. The Danish Health Authority (2009) Det moderniserede Cancerregister – metode og kvalitet. https://sundhedsdatastyrelsen.dk/da/registre-og-services/om-de-nationale-sundhedsregistre/sygedomme-laegemidler-og-behandlinger/cancerregisteret. Accessed 30 Nov 2021

21. Sobin LH, Gospodarowicz MK, Wittekind C, International Union against Cancer (2010) TNM classification of malignant tumours, 7th ed. Wiley-Blackwell, Chichester, West Sussex, UK

22. Soegaard M, Olsen M (2012) Quality of cancer registry data: completeness of TNM staging and potential implications. Clin Epidemiol 4 (Suppl 2):1–3. https://doi.org/10.2147/CLEP.S33873

23. Støvring H, Pottegård A, Hallas J (2017) Refining estimates of prescription durations by using observed covariates in pharmacoepidemiological databases: an application of the reverse waiting time distribution: WTD estimation with covariates. Pharmacoepidemiol Drug Saf 26(8):900–908. https://doi.org/10.1002/pds.4216

24. Pottegård A, Olesen M, Christensen B, Christensen MB, Hallas J, Rasmussen L (2021) Who prescribes drugs to patients: A Danish register-based study. Br J Clin Pharmacol 87(7):2982–2987. https://doi.org/10.1111/bcp.14691

25. Herrett E, Gallagher AM, Bhaskaran K, et al (2015) Data Resource Profile: Clinical Practice Research Datalink (CPRD). Int J Epidemiol 44(3):827–836. https://doi.org/10.1093/ije/dyv098

26. Pottegård A, Christensen R dePont, Houji A, et al (2014) Primary non-adherence in general practice: a Danish register study. Eur J Clin Pharmacol 70(6):757–763. https://doi.org/10.1007/s00228-014-1677-y

27. Furu K, Wettermark B, Andersen M, Martikainen JE, Almarsdottir AB, Sørensen HT (2010) The Nordic Countries as a Cohort for Pharmacoepidemiological Research. Basic Clin Pharmacol Toxicol 106(2):86–94. https://doi.org/10.1111/j.1742-7843.2009.00494.x

28. The Danish Health Data Authority: SKS-browser, vers 4.06. https://medinfo.dk/sks/brows.php. Accessed 30 Nov 2021

29. World Health Organization: International Statistical Classification of Diseases and Related Health Problems 10th Revision. https://icd.who.int/browse10/2019/en. Accessed 5 Nov 2021

30. Fritz A, Constance P, Jack A, et al (2000) International classification of diseases for oncology (ICD-O), 3rd ed. World Health Organization, Geneva

31. WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health (2019) Guidelines for ATC classification and DDD assignment 2020. WHO Collaborating Centre for Drug Statistics Methodology, Oslo, Norway

32. Moriyama IM, Loy RM, Robb-Smith AHT, Rosenberg HM, Hoyert DL (2011) History of the statistical classification of diseases and causes of death. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, Hyattsville, Md

33. The WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health (2018) ATC/DDD methodology - History. https://www.whocc.no/atc_ddd_methodology/history/. Accessed 30 Nov 2021

34. The WHO Collaborating Centre for Drug Statistics Methodology, Norwegian Institute of Public Health (2018) DDD - Definition and general considerations. https://www.whocc.no/ddd/definition_and_general_considera/. Accessed 30 Nov 2021

35. Nunney L (2016) Commentary: The multistage model of carcinogenesis, Peto's paradox and evolution. Int J Epidemiol 45(3):649–653. https://doi.org/10.1093/ije/dyv201

36. Rothman KJ (1981) Induction and latent periods. Am J Epidemiol 114(2):253–259. https://doi.org/10.1093/oxfordjournals.aje.a113189

37. Nadler DL, Zurbenko IG (2014) Estimating Cancer Latency Times Using a Weibull Model. Adv Epidemiol 2014:1–8. https://doi.org/10.1155/2014/746769

38. Ozasa K, Grant EJ, Kodama K (2018) Japanese Legacy Cohorts: The Life Span Study Atomic Bomb Survivor Cohort and Survivors' Offspring. J Epidemiol 28(4):162–169. https://doi.org/10.2188/jea.JE20170321

39. Demoor-Goldschmidt C, de Vathaire F (2019) Review of risk factors of secondary cancers among cancer survivors. Br J Radiol 92(1093):20180390. https://doi.org/10.1259/bjr.20180390

40. Vandenbroucke JP, Pearce N (2012) Case–control studies: basic concepts. Int J Epidemiol 41(5):1480–1489. https://doi.org/10.1093/ije/dys147

41. Pottegård A, Friis S, Stürmer T, Hallas J, Bahmanyar S (2018) Considerations for Pharmacoepidemiological Studies of Drug-Cancer Associations. Basic Clin Pharmacol Toxicol. https://doi.org/10.1111/bcpt.12946

42. Kumar V, Abbas AK, Aster JC, Perkins JA (2018) Robbins Basic Pathology, 10th ed. Elsevier, Philadelphia, Pennsylvania

43. Hallas J, Hansen MR, Pottegård A, Støvring H (2021) Bottleneck analysis: Simple prediction of the precision of a planned case–control or cohort study based on healthcare registers. Pharmacoepidemiol Drug Saf 30(5):619–625. https://doi.org/10.1002/pds.5200

44. Rockhill B, Newman B, Weinberg C (1998) Use and misuse of population attributable fractions. Am J Public Health 88(1):15–19. https://doi.org/10.2105/AJPH.88.1.15

45. Groenwold RHH, Goeman JJ, Cessie SL, Dekkers OM (2021) Multiple testing: when is many too much? Eur J Endocrinol 184(3):E11–E14. https://doi.org/10.1530/EJE-20-1375

46. Charlson ME, Pompei P, Ales KL, MacKenzie CR (1987) A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. J Chronic Dis 40(5):373–83

47. Elixhauser A, Steiner C, Harris DR, Coffey RM (1998) Comorbidity measures for use with administrative data. Med Care 36(1):8–27

48. Desai RJ, Wang SV, Vaduganathan M, Evers T, Schneeweiss S (2020) Comparison of Machine Learning Methods With Traditional Models for Use of Administrative Claims With Electronic Medical Records to Predict Heart Failure Outcomes. JAMA Netw Open 3(1):e1918962. https://doi.org/10.1001/jamanetworkopen.2019.18962

49. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B (2019) A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol 110:12–22. https://doi.org/10.1016/j.jclinepi.2019.02.004

50. Steyerberg EW, Vickers AJ, Cook NR, et al (2010) Assessing the Performance of Prediction Models: A Framework for Traditional and Novel Measures. Epidemiology 21(1):128–138. https://doi.org/10.1097/EDE.0b013e3181c30fb2

51. Danckert B, Ferlay J, Engholm G, et al NORDCAN: Cancer Incidence, Mortality, Prevalence and Survival in the Nordic Countries, Version 8.2 (26.03.2019). Association of the Nordic Cancer Registries. Danish Cancer Society. https://www.ancr.nu/. Accessed 30 Nov 2021

52. Bray F, Kiemeney LA (2017) Epidemiology of Prostate Cancer in Europe: Patterns, Trends and Determinants. In: Bolla M, van Poppel H (eds) Management of Prostate Cancer. Springer International Publishing, pp 1–27

53. Schulman S, Lindmarker P (2000) Incidence of Cancer after Prophylaxis with Warfarin against Recurrent Venous Thromboembolism. N Engl J Med 342(26):1953–1958. https://doi.org/10.1056/NEJM200006293422604

54. Haaland GS, Falk RS, Straume O, Lorens JB (2017) Association of Warfarin Use With Lower Overall Cancer Incidence Among Patients Older Than 50 Years. JAMA Intern Med 177(12):1774–1780. https://doi.org/10.1001/jamainternmed.2017.5512

55. Li X, Lund JL, Toh S (2018) Lower Cancer Incidence—Warfarin Effect or Immortal Time Bias? JAMA Intern Med 178(4):584. https://doi.org/10.1001/jamainternmed.2018.0367

56. Svendsen K, Karlstad Ø, Småbrekke L (2018) Lower Cancer Incidence—Warfarin Effect or Immortal Time Bias? JAMA Intern Med 178(4):585. https://doi.org/10.1001/jamainternmed.2018.0370

57. Ward MM (2020) Chronic oral anticoagluation and risk of prostate cancer: Evidence of detection bias. Int J Cancer 146(11):3022–3025. https://doi.org/10.1002/ijc.32712

58. Nguyen-Nielsen M, Frøslev T, Friis S, Harving, Borre, Soegaard M (2012) Completeness of prostate cancer staging in the Danish Cancer Registry, 2004-2009. Clin Epidemiol 4(Suppl 2):17–23. https://doi.org/10.2147/CLEP.S32004

59. Madan V, Lear JT, Szeimies R-M (2010) Non-melanoma skin cancer. The Lancet 375(9715):673–685. https://doi.org/10.1016/S0140-6736(09)61196-X

60. Kaae J, Boyd HA, Hansen AV, Wulf HC, Wohlfahrt J, Melbye M (2010) Photosensitizing Medication Use and Risk of Skin Cancer. Cancer Epidemiol Biomarkers Prev 19(11):2942–2949. https://doi.org/10.1158/1055-9965.EPI-10-0652

61. Kaae J, Carstensen L, Wohlfahrt J, Melbye M, Allison Boyd H (2014) Epilepsy, anti-epileptic medication use and risk of cancer: Epilepsy, Anti-Epileptic Medication Use and Risk of Cancer. Int J Cancer 134(4):932–938. https://doi.org/10.1002/ijc.28396

62. GlaxoSmithKline (2017) Lamictal: Full Prescribing Information. https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/020241s045s051lbl.pdf. Accessed 30 Nov 2021

63. Novartis Pharmaceuticals Corporation (2007) Tegretol: Full Prescribing Information. https://www.accessdata.fda.gov/drugsatfda_docs/label/2015/016608s097,018281s045,018927s038,020234s026lbl.pdf. Accessed 30 Nov 2021

64. Silva R, Machado A, Brandão M, Gonçalo S (1986) Patch test diagnosis in carbamazepine erythroderma. Contact Dermatitis 15(4):254–255

65. Bilski PJ, Wolak MA, Zhang V, Moore DE, Chignell CF (2009) Photochemical Reactions Involved in the Phototoxicity of the Anticonvulsant and Antidepressant Drug Lamotrigine (Lamictal®). Photochem Photobiol 85(6):1327–1335. https://doi.org/10.1111/j.1751-1097.2009.00590.x

66. Onoue S, Tsuda Y (2006) Analytical Studies on the Prediction of Photosensitive/Phototoxic Potential of Pharmaceutical Substances. Pharm Res 23(1):156–164. https://doi.org/10.1007/s11095-005-8497-9

67. Welch HG, Mazer BL, Adamson AS (2021) The Rapid Rise in Cutaneous Melanoma Diagnoses. N Engl J Med 384(1):72–79. https://doi.org/10.1056/NEJMsb2019760

68. Chow W-H, Dong LM, Devesa SS (2010) Epidemiology and risk factors for kidney cancer. Nat Rev Urol 7(5):245–257. https://doi.org/10.1038/nrurol.2010.46

69. Andreassen BK, Støer NC, Martinsen JI, et al (2019) Identification of potential carcinogenic and chemopreventive effects of prescription drugs: a protocol for a Norwegian registry-based study. BMJ Open 9(4):e028504. https://doi.org/10.1136/bmjopen-2018-028504

70. McDowell RD, Hughes C, Murchie P, Cardwell C (2021) A systematic assessment of the association between frequently prescribed medicines and the risk of common cancers: a series of nested case-control studies. BMC Med 19(1):22. https://doi.org/10.1186/s12916-020-01891-5

71. Heinze G, Wallisch C, Dunkler D (2018) Variable selection – A review and recommendations for the practicing statistician. Biom J 60(3):431–449. https://doi.org/10.1002/bimj.201700067

72. Sauerbrei W, Perperoglou A, Schmid M, et al (2020) State of the art in selection of variables and functional forms in multivariable analysis—outstanding issues. Diagn Progn Res 4(1):3. https://doi.org/10.1186/s41512-020-00074-3

73. Steyerberg EW (2019) Clinical Prediction Models: A Practical Approach to Development, Validation, and Updating. Springer International Publishing, Cham

74. Riley RD, Ensor J, Snell KIE, et al (2016) External validation of clinical prediction models using big datasets from e-health records or IPD meta-analysis: opportunities and challenges. BMJ 353:i3140. https://doi.org/10.1136/bmj.i3140

75. Cogliano VJ, Baan R, Straif K, et al (2011) Preventable Exposures Associated With Human Cancers. JNCI J Natl Cancer Inst 103(24):1827–1839. https://doi.org/10.1093/jnci/djr483

76. Yang CS, Chen X, Tu S (2016) Etiology and Prevention of Esophageal Cancer. Gastrointest Tumors 3(1):3–16. https://doi.org/10.1159/000443155

77. Pohl H, Sirovich B, Welch HG (2010) Esophageal adenocarcinoma incidence: are we reaching the peak? Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 19(6):1468–1470. https://doi.org/10.1158/1055-9965.EPI-10-0012

78. Stewart SL, Wike JM, Kato I, Lewis DR, Michaud F (2006) A population-based study of colorectal cancer histology in the United States, 1998–2001. Cancer 107(S5):1128–1141. https://doi.org/10.1002/cncr.22010

79. Zhao L, Lee VHF, Ng MK, Yan H, Bijlsma MF (2019) Molecular subtyping of cancer: current status and moving toward clinical applications. Brief Bioinform 20(2):572–584. https://doi.org/10.1093/bib/bby026

80. Brennan P, Smith GD (2021) Identifying Novel Causes of Cancers to Enhance Cancer Prevention: New Strategies are Needed. JNCI J Natl Cancer Inst. https://doi.org/10.1093/jnci/djab204

81. Pottegard A, Hallas J (2016) New use of prescription drugs prior to a cancer diagnosis. Pharmacoepidemiol Drug Saf 26(2):223–227. https://doi.org/10.1002/pds.4145

82. Harrell FE (2015) Regression Modeling Strategies: With Applications to Linear Models, Logistic and Ordinal Regression, and Survival Analysis, 2nd ed. Springer International Publishing, Cham, Switzerland

83. Gigerenzer G, Gaissmaier W, Kurz-Milcke E, Schwartz LM, Woloshin S (2007) Helping Doctors and Patients Make Sense of Health Statistics. Psychol Sci Public Interest 8(2):53–96. https://doi.org/10.1111/j.1539-6053.2008.00033.x

84. Furedi A (1999) Social consequences. The public health implications of the 1995 "pill scare." Hum Reprod Update 5(6):621–626. https://doi.org/10.1093/humupd/5.6.621

85. Hallas J, Christensen RD, Sturmer T, Pottegard A (2014) Measures of 'exposure needed for one additional patient to be harmed' in population-based case-control studies. Pharmacoepidemiol Drug Saf 23(8):868–74. https://doi.org/10.1002/pds.3635

86. Vrinten C, Waller J, von Wagner C, Wardle J (2015) Cancer fear: facilitator and deterrent to participation in colorectal cancer screening. Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol 24(2):400–405. https://doi.org/10.1158/1055-9965.EPI-14-0967

87. Vrinten C, McGregor LM, Heinrich M, et al (2017) What do people fear about cancer? A systematic review and meta-synthesis of cancer fears in the general population. Psychooncology 26(8):1070–1079. https://doi.org/10.1002/pon.4287

88. The Danish Society for Pharmacoepidemiology (2018) Responsible dissemination of research within the pharmaceutical field. http://www.dsfe.dk/responsible-dissemination-of-research/. Accessed 30 Nov 2021

89. Suissa S (2008) Immortal Time Bias in Pharmacoepidemiology. Am J Epidemiol 167(4):492–499. https://doi.org/10.1093/aje/kwm324

90. Searing L (2017) A common blood thinner may protect against some cancers. Washington Post. https://www.washingtonpost.com/national/health-science/a-common-blood-thinner-may-protect-against-some-cancers/2017/11/17/7de9b788-cafe-11e7-8321-481fd63f174d_story.html. Accessed 30 Nov 2021

91. Cowen L (2017) Warfarin shows broad anti-cancer potential. MedwireNews. https://www.medwirenews.com/oncology/haematology/warfarin-shows-broad-anti-cancer-potential-/15200676. Accessed 30 Nov 2021

92. Sharrar RG, Dieck GS (2013) Monitoring product safety in the postmarketing environment. Ther Adv Drug Saf 4(5):211–219. https://doi.org/10.1177/2042098613490780

93. De Bruin ML, van Puijenbroek EP, Egberts ACG, Hoes AW, Leufkens HGM (2002) Non-sedating antihistamine drugs and cardiac arrhythmias -- biased risk estimates from spontaneous reporting systems? Br J Clin Pharmacol 53(4):370–374. https://doi.org/10.1046/j.1365-2125.2002.01569.x

94. Hallas J, Wang SV, Gagne JJ, Schneeweiss S, Pratt N, Pottegård A (2018) Hypothesis-free screening of large administrative databases for unsuspected drug-outcome associations. Eur J Epidemiol 33(6):545–555. https://doi.org/10.1007/s10654-018-0386-8

95. Pottegård A, Hallas J, Wang SV, Gagne JJ (2018) Identifying signals of interest when screening for drug-outcome associations in health care data. Br J Clin Pharmacol 84(9):1865–1867. https://doi.org/10.1111/bcp.13634

96. Wang SV, Kulldorff M, Glynn RJ, et al (2018) Reuse of data sources to evaluate drug safety signals: When is it appropriate? Pharmacoepidemiol Drug Saf 27(6):567–569. https://doi.org/10.1002/pds.4442

97. Walker AM (2010) Orthogonal predictions: follow-up questions for suggestive data. Pharmacoepidemiol Drug Saf 19(5):529–532. https://doi.org/10.1002/pds.1929

98. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Res 45(D1):D353–D361. https://doi.org/10.1093/nar/gkw1092

Appendices