ESTIMATING THE IMPACT OF MOLECULAR PROFILES AND PRESCRIPTION DRUGS ON BREAST CANCER OUTCOMES

PhD thesis

by

THOMAS PATRICK AHERN

Faculty of Health Sciences, Aarhus University
Department of Clinical Epidemiology, Aarhus University Hospital
Boston University School of Public Health

Supervisors

Timothy L. Lash, DSc, MPH
Associate Professor of Epidemiology

Henrik Toft Sørensen, DMSc, PhD
Professor of Clinical Epidemiology

Lars Pedersen, PhD
Associate Professor of Clinical Biostatistics

Stephen Hamilton-Dutoit, MD, FRCPath
Professor of Pathology

DEDICATION

For my parents,

Richard Charles Ahern & Shirley LaVigne Ahern

ACKNOWLEDGEMENTS

The following individuals have, in one way or another, made this dissertation possible. This accomplishment is as much theirs as it is mine.

My parents, Richard and Shirley Ahern, and my older brother, James Ahern. As I grew up and as I made my way through this program they gave love, support, and countless examples of the value of hard work.

My wife, Mariah McNamara, who first suggested that I consider a graduate degree in public health--and then supported me in every sense as I took her advice.

My sons, Liam Robert Ahern and Miles Cordell Ahern, who gave me vital daily doses of perspective and unconditional love. May they one day put their old man to shame.

My first scientific mentor, the late Dr. Cordell Gross. Dr. Gross was a vibrant character with an unquenchable scientific curiosity, which he focused with great success on the alleviation of human suffering. I continue to measure myself against his example.

My colleagues in the Department of Clinical Epidemiology at Aarhus University. Most notably Dr. Deirdre Cronin-Fenton, who helped get these studies airborne and gave helpful scientific insight throughout their conduct. While visiting Aarhus to conduct analyses for these studies, I received valuable advice and assistance from Morten Schmidt, Claus Sværke, Dóra Körmendiné Farkas, and Hanne Kjeldahl Schlosser.

The laboratory scientists from the Institute of Pathology at Aarhus University Hospital: Dr. Mariann Christensen, Dr. Ylva Hellberg, Kristina Lauridsen, Jesper

Bertelsen, Inge Krohn, and Ditte Mikkelsen. Without their hard work and attention to detail, the final study of this dissertation would not have been possible.

All my fellow doctoral students, for the community they have provided during this experience. Special acknowledgements go to Ryan Ferguson and Brenda Heaton (with whom studying for the qualifying exams was akin to a prolonged comedy sketch), and to Jaimie Gradus.

The faculty in the Departments of Epidemiology and Biostatistics at the Boston University School of Public Health, a cadre of talented educators who taught me things I never thought I could understand.

Dr. Rebecca A. Silliman, a gracious mentor whose expert perspectives on breast cancer epidemiology and career development were essential to this dissertation and to my professional evolution.

The reviewers of this dissertation, Dr. Kenneth Rothman and Dr. Carol Rosenberg. Their careful evaluation and insightful comments greatly improved the final quality and scope of the research herein.

Finally, I am most indebted to the members of my dissertation committee: Dr. Timothy L. Lash (Chair), Dr. Henrik Toft Sørensen, Dr. Lars Pedersen, and Dr. Stephen Hamilton-Dutoit. It has been an exceptionally rewarding experience to carry out these studies under their expert guidance. I am particularly grateful to Dr. Lash for the remarkable opportunities he gave me to teach and apply epidemiologic methods throughout my doctoral education. It is difficult indeed to imagine another mentor who could surpass Dr. Lash in knowledge, patience, and generosity.

Abstract

It is estimated that, in the year 2010, more than 200,000 American women will be diagnosed with breast cancer and that approximately 40,000 more will die from breast cancer-related causes. In Denmark, yearly breast cancer incidence between 2004 and 2008 equaled about 4,100 cases, with approximately 1,200 related deaths. The objective of this dissertation is to estimate associations between treatment with two classes of prescription drug (cardiac glycosides and vitamin K antagonists) and the incidence of invasive breast cancer, and to explore whether single-nucleotide polymorphisms in metabolic enzymes change the effectiveness of the drug tamoxifen in preventing breast cancer recurrence.

The first study enrolled 5,565 incident cases of invasive breast carcinoma and 55,650 matched controls from the female populations of two Danish counties. We characterized past exposure to cardiac glycoside drugs (*i.e.*, digoxin) using the counties' automated, electronic pharmacy databases. Women ever treated with digoxin had a 30% increased rate of developing breast cancer compared with women never treated with digoxin (OR: 1.30; 95% CI: 1.14, 1.48). The association magnitude increased directly with the length of digoxin exposure, and was robust to adjustment for age, co-medications, and medical indications similar to those for digoxin treatment.

The second study employed heart valve replacement as an instrumental variable to estimate the associations between vitamin K antagonist treatment and the incidence of breast and 23 other site-specific cancers. According to the instrumental variable estimator, women treated with a vitamin K antagonist did not have an elevated breast cancer incidence rate, compared with women never treated with a vitamin K antagonist (IRR: 1.1; 95% CI: 0.88, 1.3). Associations with most

other cancer sites were similarly null, and the overall pattern of site-specific cancer associations was consistent with an underlying null-centered distribution.

The third study evaluated the association between functional polymorphisms in the UDP-glucuronosyltransferase enzymes responsible for the conjugation and elimination of active tamoxifen metabolites and the rate of breast cancer recurrence among tamoxifen-treated women. We observed no association between breast cancer recurrence and having 2 variant alleles at the *UGT2B15**2 locus (OR: 0.68 95% CI: 0.45, 1.0), nor having 2 variant alleles at the *UGT2B7**2 locus (OR: 0.85; 95% CI: 0.54, 1.3).

CONTENTS

Page

ABBREVIATIONS

4HT	4-hydroxy-tamoxifen
ASA	Acetylsalicylic acid
BMI	Body mass index
CG	Cardiac glycoside
CI	Confidence interval
CPR	Danish Civil Personal Registry
CYP	Cytochrome P450
DBCG	Danish Breast Cancer Cooperative Group
DCR	Danish Cancer Registry
E2	Estradiol
ER	Estrogen receptor
HRT	Hormone replacement therapy
IRR	Incidence rate ratio
IV	Instrumental variable
LD	Linkage disequilibrium
MAF	Minor allele frequency
NRP	Danish National Registry of Patients
NSAID	Non-steroidal anti-inflammatory drug
OR	Odds ratio
PCR	Polymerase chain reaction
RMPS	Danish Register of Medicinal Product Statistics
SDS	Sodium dodecyl sulfate
SERM	Selective estrogen receptor modulator
SI	Simulation interval
SNP	Single nucleotide polymorphism
UGT	UDP-glucuronosyl transferase
UICC	International Union Against Cancer
VKA	Vitamin K antagonist

INTRODUCTION

The American Cancer Society estimates that, in the year 2010, over 200,000 new cases of invasive female breast cancer will be diagnosed, and that approximately 40,000 women will die from breast cancer-related causes.[1] The Danish Cancer Society reports that approximately 4,100 new breast cancer diagnoses and 1,200 related deaths occurred each year in the period from 2004 to 2008.[2] Breast cancer is a disease with well-characterized hereditary and reproductive risk factors, but research efforts to date have uncovered few strong associations between breast cancer and mutable exposures.

The major goal of this dissertation research is to add to current knowledge on modifiable risk factors for breast cancer incidence and recurrence through the following specific aims:

Study 1: Cardiac glycoside treatment and breast cancer incidence.

Specific Aim: To evaluate whether treatment with a cardiac glycoside (*i.e.*, digoxin or digitoxin), compared with no such treatment, affects the risk of incident breast cancer among Danish women.

Study 2: Vitamin K antagonist therapy and the incidence of site-specific cancers.

Specific Aim: To evaluate whether anticoagulation treatment with a vitamin K antagonist (*e.g.*, warfarin), compared with no such treatment, affects the incidence rate of breast or other site-specific cancers in the Danish population.

Study 3: Polymorphisms in UDP-glucuronosyltransferases and breast cancer recurrence in tamoxifen-treated women.

Specific Aim: To evaluate the associations between functional gene polymorphisms in two enzymes responsible for the systemic elimination of active tamoxifen metabolites and the rate of breast cancer recurrence among women treated with tamoxifen.

The three studies comprising this dissertation made use of various population-based registries in Denmark. The Danish government has systematically recorded information about its approximately 5.5 million citizens since as early as the 18th Century.[3] More than 200 specialized databases have been constructed since then, tracking vital status, health, employment, residence, and a host of other aspects of Danes' daily lives. In contrast with most other collections of such data, the Danish registries are not independent of one another: they can be linked together in any combination using the civil personal registration (CPR) number—a unique ten-digit identifier assigned to all Danish residents upon birth or immigration.

The studies in this dissertation employ (1) pharmacy databases (which operate at both regional and national levels) to ascertain drug exposures (*i.e.*, cardiac glycosides and vitamin K antagonists); (2) a specialized breast cancer database administered by the Danish Breast Cancer Cooperative Group (DBCG), which since 1977 has enrolled and tracked almost all women under 70 years of age diagnosed with invasive breast cancer; (3) the Danish Cancer Registry, which has recorded data on cancer diagnoses in the Danish population since 1943; (4) the Danish National Registry of Patients and county-based hospital discharge registries, which record all procedures and diagnoses arising from inpatient admission to any

non-psychiatric hospital in the entire nation; and (5) the Danish Civil Registry, which tracks residential address and vital status and is updated daily.

Using CPR numbers, we were able to link these registries at the level of the individual so that, for example, incident breast cancer cases identified in hospital discharge registries could be paired with population-based controls identified from the civil registry. Pharmaceutical exposure histories of these cases and controls were then characterized by CPR linkage to the prescription databases.

These registries enable the retrospective conduct (from the investigator's viewpoint) of epidemiologic studies with high-quality data collected prospectively on a very large number of individuals. It is certain that these dissertation studies would have been incredibly difficult, expensive, and time-consuming to conduct without access to the trove of resources offered by the Danish population-based registries.

CARDIAC GLYCOSIDE TREATMENT AND BREAST CANCER INCIDENCE

BACKGROUND:

Cardiac Glycosides

Cardiac glycosides (CGs) are natural steroid toxins derived from several botanical sources.[4] There are two classes of cardiac glycoside, cardenolides and bufadienolides.[5] Compounds from both of these classes have been used widely since the 18th century to treat congestive heart failure (CHF) and atrial fibrillation (AF).[6] All CGs share a similar structural motif: a 3- or 4-member glycone moiety linked to a steroid core with either a 5-member lactone ring (as in the cardenolides) or a 6-member pyrone ring (as in the bufadienolides).[5] Specific compounds are distinguished by their glycone composition and various substitutions on the steroid core. The clinically most prevalent CGs are the *Digitalis*-derived cardenolides (*e.g.*, digitoxin, digoxin, and oleandrin).

The primary physiologic effect of CGs results from inhibition of the sodium-potassium ATPase. The sodium-potassium ATPase is a transmembrane protein that maintains cellular ion homeostasis by ATP-dependent removal of sodium ions and importation of potassium ions in a 3:2 ratio. The ATPase consists of three distinct protein subunits. The alpha (α) subunit has four known variants and is the catalytically active member of the holoenzyme; it is the binding site for ATP, Na^+, K^+, CGs and a Mg^{2+} cofactor.[7] The beta (β) subunit, with three known variants, primarily regulates the structure and function of the ATPase, but is also involved in the regulation of gap junction proteins in cellular adhesion processes.[7] The gamma (γ,

or FXYD) subunit has seven known variants, and also regulates the holoenzyme.[7] Inhibition of the ATPase causes a rise in intracellular calcium ion (Ca^{2+}) concentration *via* the sodium-calcium exchanger. The high intracellular Ca^{2+} concentration enriches the calcium stores in the sarcoplasmic reticula of cardiac myocytes, increasing contractile force and cardiac output.

Despite their narrow therapeutic index, the popularity of CGs for treating CHF and AF has persisted worldwide, likely owing to their relative economy compared with newer alternative drugs. In addition to the well-established cardiotonic properties of these compounds, research spanning the past four decades has explored anti-neoplastic potential of CGs.

Epidemiologic Studies of Cardiac Glycosides and Breast Cancer

In 1979, Stenkvist *et al.* reported an unusual finding in a small cohort of breast cancer patients (n=142). Women in the cohort who were taking CGs (73% digoxin, 12% digitoxin, 15% other) at the time of their breast cancer diagnosis had tumors with less aggressive phenotypes than breast tumors from women not taking CGs.[8] Later, they reported a dramatically higher rate of breast cancer recurrence among the women not taking CGs after five[9] and approximately twenty-two[10] years of follow-up. The investigators noted structural similarity between digitalis glycosides and estradiol, namely their steroid nature, and posited that competitive inhibition of the estrogen receptor (ER) by CGs might have influenced tumor phenotype. In recent years this ER-interference hypothesis has been largely supplanted by laboratory studies implicating CGs in the signaling pathways mediated by the Na^+/K^+ ATPase.[7, 11]

Subsequent studies of the association between CG use and breast cancer incidence gave conflicting results. Haux *et al.* compared site-specific cancer

incidence rates among digitalis-treated Norwegian patients with expected rates in the general population.[12] Several cancers, including female breast cancer, occurred at higher rates among those treated with digitalis compared with the general population.[12] Also, Friedman reported no association between CG prescription history and breast cancer in a Kaiser-Permanente registry study.[13]

Given the continued importance of CG medicines to treat heart disease and the inconsistent results from earlier studies of the association between this therapy and breast cancer occurrence, we examined the association between digoxin treatment and breast cancer incidence in a population-based prospective case-control study of post-menopausal Danish women.

METHODS:

This study was approved by the Boston University Medical Campus Institutional Review Board and the Danish Registry Board.

Study Population

This study was nested within the female population of North Jutland and Aarhus Counties, Denmark.[14] We used county hospital registries to ascertain all cases of incident invasive breast cancer diagnosed in women age 55 or older. Ascertainment began on 1 January 1991 in North Jutland County and 1 January 1998 in Aarhus County, and continued until 31 December 2007.[15] The hospital registries contain data on patients' CPR number, date(s) of admission, date(s) of discharge, and up to twenty discharge diagnoses and medical procedures per discharge or outpatient visit. Diagnoses are assigned by the discharging physician,

and are coded according to the International Classification of Diseases, 8th revision (ICD-8, until 1995) and 10th revision (ICD-10, 1995 onward).

Controls were identified in the Danish Civil Registration System, which has tracked residential addresses, vital status, and date of emigration for the entire Danish population since 1968.[3] Controls were selected for each case by risk-set sampling, matching controls to cases on year of birth and county of residence. Within strata of the matching factors, we selected ten controls at random among those who were alive and without a history of breast cancer on the date of the matched case's diagnosis. This date was the case's and her matched controls' index date.

To ensure an adequate prescription data history for all subjects, we excluded cases and potential controls who had lived in the study counties less than two years between the start of the prescription registries and their index date.

Data Collection

We used each subject's unique CPR number to link the case-control roster to county prescription databases,[16, 17] which have recorded all prescriptions filled since 1989 in North Jutland County and 1996 in Aarhus County. (These counties are now part of the North and Mid regions of Denmark, respectively.) The databases encode drugs by the Anatomical Therapeutic Chemical (ATC) classification system[18] and record dates of all prescription fills along with the patient's CPR number. These systems report prescription data to the databases, as well as to the Danish National Health Service, which refunds a portion of medication costs. Prescriptions are logged in the registries after patients present to a pharmacy and pay their share of the prescription cost.

We ascertained medical history for cases and controls by extracting major diagnoses preceding index dates from the hospital registries. We also used these registries to identify all pre-diagnosis mammography procedures for cases and controls since 2001, the year mammography data began to be systematically recorded, to evaluate the possibility of detection bias.

Definitions of Analytic Variables

We identified cases of incident breast cancer in the hospital registries using ICD-8 and ICD-10 codes appropriate to the date ranges of the databases. ICD codes were also used to ascertain comorbid conditions for cases and controls (see full listing of ICD codes in Table 5).

We ascertained CG prescriptions by extracting all records from the prescription databases with ATC codes beginning with 'C01A'. CGs are available only by prescription in Denmark, and are dispensed at pharmacies equipped with automated electronic reporting systems described earlier. This strategy captured all CG prescriptions in the counties over the study period, which were for digoxin exclusively. Digoxin prescriptions were only considered if they occurred at least one year before the index date. Digoxin exposure was considered in broad terms as ever exposed (≥ 1 digoxin prescription at least one year before the index date) or never exposed (no record of digoxin prescription at least one year before the index date), and in finer terms according to the length of time between a woman's first digoxin prescription and her index date.

Candidate confounders were selected for adjustment by use of the directed acyclic graph (DAG) shown in Figure 1. A DAG encodes hypothesized relations between variables, which can aid in identifying confounders of a given exposure-disease association. Confounders in a DAG are variables along a causal path with

arrows pointing into both the exposure and disease, or an ancestor of such a variable.[19]

Figure 1 depicts the hypothesized relationships among the variables that influence digoxin prescription and breast cancer incidence. Using the back-door test described by Greenland *et al.*,[19] control for age, BMI, hormone replacement therapy (HRT) exposure, anticoagulant exposure, NSAID exposure, and aspirin exposure, were deemed minimally sufficient to address confounding, presuming the causal diagram faithfully depicts the causal relations among the variables.

Age was initially controlled by matching controls to cases on year of birth. We also calculated each subject's exact age on her index date to adjust for residual confounding by age. We additionally considered confounding by co-prescription of HRT, anticoagulants, NSAIDs, and aspirin. Anticoagulants are frequently prescribed for AF, and were associated with lower risk of urogenital cancers.[20] NSAID use has been associated with increased risk of CHF,[21] and these drugs have shown protective associations with breast cancer in some studies,[22] although not in Denmark.[23] Aspirin use, which may be more prevalent among digoxin users, has been associated with reduced breast cancer risk,[24] though data are conflicting.[22] We also evaluated HRT as a confounder because of its contribution to cumulative hormonal exposure and its association with breast cancer risk.[25]

Prescriptions for hormone replacement therapy were identified by ATC codes (estrogens: codes starting with either 'G03C' or 'L02AA'; progestin: codes beginning with 'G03D'; combination therapy: codes beginning with either 'G03F' or 'G03H'). Exposure to any of these drugs before the index date was classified as 'ever exposed to HRT' while exposure to none of them was classified as 'never exposed to HRT'. Similarly, we characterized ever/never exposure to anticoagulants, NSAIDs and aspirin by searching for ATC codes beginning with 'B01A,' 'M01A,' and

'B01AC06,' respectively.

We evaluated confounding by the medical indications for digoxin therapy by defining an alternative reference group of women who were never exposed to digoxin and who had a history of cardiovascular disease (excluding CHF or AF). These reference subjects should be more similar to the digoxin-treated women with regard to cumulative hormonal exposures and lifestyle factors that may modify risk for both heart disease and breast cancer. This reference group also facilitated evaluation of detection bias by allowing comparison of digoxin-exposed women with women with other medical histories who would likely have similar medical usage patterns.

We further evaluated detection bias by comparing mammography usage rates between cases and controls. Dates of all mammography procedures among cases and controls were identified in hospital registries using appropriate Danish medical procedure codes. We analyzed mammography usage among women with index dates from 1 January 2006 onward, the period of our study when opportunistic screening mammography would have been most common in Denmark. There was no formal mammography screening in this region of Denmark until after the study period.[26-28] For each subject who had undergone mammography before her index date, we identified her most recent procedure and calculated the time elapsed between that procedure and the index date.

Statistical Analysis

We characterized the names, doses, and prescribing frequencies of the various digoxin products used over the study period. We computed the frequency and proportion of cases and controls by digoxin exposure status, prevalent medical

conditions, use of other prescription drugs (HRT, anticoagulants, NSAIDs and aspirin), and age on index date.

We calculated associations between ever/never digoxin treatment and breast cancer incidence within joint strata of age group and past use of HRT. The odds ratio and 95% confidence interval summarized across all strata was computed by the Mantel-Haenszel pooling method.

We calculated the unadjusted odds ratios (OR) and 95% confidence intervals (CI) associating digoxin exposure categories with incident breast cancer and used conditional logistic regression to account for the matching factors and to adjust for any confounders that changed the estimated log odds by at least 10% (candidates confounders included exact age, and past use of HRT, anticoagulants, NSAIDs, and aspirin. Due to the risk-set sampling design, the odds ratio approximates the incidence rate ratio associating digoxin exposure with incident breast cancer.[29] All analyses were performed with SAS version 9 (SAS Institute, Cary, NC).

RESULTS:

Characteristics of Cases and Controls

We identified 5,565 cases and 55,650 matched population controls. Among the cases, 324 (5.8%) women had ever had a digoxin prescription at least one year before their diagnosis date. Among the controls, 2,546 (4.6%) women had ever had a digoxin prescription at least one year before their index date. The distributions of cases and controls according to age, mammography usage, comorbidity and relevant prescription drug usage are shown in Table 2. By virtue of the matching, cases and controls were identical with respect to age distribution. Cases were

somewhat more likely to have CHF, AF, chronic pulmonary disease, or diabetes, and were less likely to have a history of myocardial infarction, than controls. Cases also had more exposure to HRT, anticoagulants and NSAIDs than controls. As expected, mammography usage was substantially higher for cases than for controls in the year preceding the index date (81% vs. 1.6%, respectively; Table 2). However, usage was similar for cases and controls in time periods greater than one year from the index date.

Digoxin Treatment and Incident Breast Cancer

Table 1 shows all of the cardiac glycoside products recorded in the county prescription registries during the study period. We noted that digoxin was the sole CG used during this period. Approximately 97% of all digoxin prescriptions were for 62.5 μg tablets, indicating very little product heterogeneity among the digoxin-exposed subjects.

We observed a higher rate of breast cancer among ever-users of digoxin, relative to never users, in unadjusted, stratified, and regression analyses (adjusted OR: 1.30; 95% CI: 1.14, 1.48; Table 4). This association persisted in categories of drug exposure duration (for 1 to 3 years, adjusted OR: 1.25, 95% CI: 1.03, 1.52; for 4 to 6 years, adjusted OR: 1.30, 95% CI: 1.05, 1.61; and for 7 to 18 years, OR: 1.39, 95% CI: 1.10, 1.74), with a suggested positive trend in the ORs with increasing length of digoxin therapy. The association was strongest among women aged 85 or older who had a history of HRT exposure (OR: 2.4, 95% CI: 1.4, 4.4; Table 3). When we compared digoxin-exposed women with the alternative reference group of unexposed women with cardiovascular medical histories other than CHF or AF, we

continued to observe an association between digoxin exposure and incident breast cancer (adjusted OR: 1.42, 95% CI: 1.14, 1.77; Table 4).

DISCUSSION:

Strengths and Limitations

The main strengths of this study are its large size, use of high-validity registry data to ascertain diagnoses, use of prospectively-recorded exposure information, and lack of selection in enumerating cases and controls.

Our study design minimized the threat of selection bias, which can create the illusion of an exposure-disease association when, in fact, none exists.[30] We had only one subject exclusion criterion, and controls were selected completely at random within strata of the matching factors. Since no subject was required to give consent to participate, no self-selection mechanism could have influenced our results.

Our results are potentially subject to distortion by residual confounding and misclassification of exposure and outcome. We took measures to address confounding by age, past exposure to other prescription drugs, and the medical indications for digoxin prescription. We saw little change in the unadjusted association after accounting for these factors. Digoxin is ordinarily prescribed at an age when most women no longer bear children, so it is unlikely that digoxin exposure is strongly associated with the well-characterized reproductive factors that affect breast cancer risk.[31] It is also unlikely that use of other prescription drugs could confound our results since antibiotics, antihypertensives, statins, and antidepressants do not appear to modify breast cancer risk.[22]

Use of the alternative reference group resulted in a modest increase in the estimated odds ratio, which implies that confounding by indication actually served to attenuate the original association. Furthermore, detection bias is not likely to account for this observed association, since women with other cardiovascular diseases would have medical system usage similar to women treated with CGs. In the whole study population, we saw no material difference in mammography usage rates between cases and controls in time periods greater than one year from index dates, which further argues against detection bias.

We were not able to adjust directly for body mass index (BMI), which is positively associated with both cardiovascular disease (CVD) and breast cancer incidence.[32] Because of these relationships, the effect of BMI as a confounder of the digoxin/breast cancer association would be to positively displace the observed odds ratio from the true odds ratio, which could account for our positive observation. However, subjects in our alternative reference group likely had a BMI distribution similar to digoxin-exposed subjects due to the association between BMI and CVD.[33] Evaluation of the association using this reference group moved our estimated odds ratio upward, which is not the displacement we would expect upon adjustment for BMI. This displacement could be the result of control for factors other than BMI that may have differed between the initial and alternative reference groups, or it could have arisen from random mechanisms owing to the smaller sample size imposed by the alternative reference group.

Our characterization of digoxin exposure was informed only by the number and strength of prescriptions filled by study participants; the prescription registry data did not permit calculation of actual daily doses taken by exposed subjects. Because prescription records were generated automatically before breast cancer

diagnoses, and were logged in a registry separate from that containing the cancer diagnoses, we expect any exposure classification error to be non-differential and independent with respect to outcome. We are not aware of published validation data on the classification of incident breast cancer in the hospital discharge registries. However, breast cancer diagnoses were recorded without express knowledge of exposure, so outcome misclassification is also expected to be non-differential. Since non-differential and independent classification errors of dichotomous variables are expected to attenuate results, exposure and outcome misclassification cannot plausibly account for the positive association we observed.[30]

Much of the investigation into the digoxin-breast cancer association appears to be motivated by the original epidemiologic results reported by Stenkvist *et al*.[8-10, 34] Aside from a truly protective effect of CGs on breast cancer, a plausible alternative explanation for Stenkvist's observations is that detection bias gave rise to the more favorable tumor phenotypes seen among digoxin-treated women. The women in the cohort who were under cardiac glycoside treatment at the time of their breast cancer diagnosis must have been under regular medical care for CHF, AF, or both. The severity of these cardiac conditions would warrant relatively frequent medical office visits, and thus a higher probability of early breast cancer detection relative to healthier women with less medical contact. These baseline differences in breast cancer severity at diagnosis could also explain the favorable recurrence findings upon long-term follow-up. The possibility of distortion by medical surveillance bias was acknowledged and explored by Stenkvist *et al.*, and the authors concluded that no such distortion existed.[34] Similarly, our analysis of mammography usage among our cases and controls showed no evidence of differential surveillance. This surveillance information should be interpreted

conservatively because collection of mammography data is a recent practice in Denmark, and the data have not yet been well validated.

The CG protective effects on cancer incidence reported by Goldin *et al.*[35] are likely biased by the investigators' subject selection criteria. In particular, they reported a lower proportion of digitalis users among all cancer deaths compared with the proportion of users among all non-cancer deaths in their hospitals. The comparison set of non-cancer deaths would logically contain a large proportion of cardiac deaths, so the prevalence of digitalis usage in this reference group would naturally be higher than among the cancer deaths. This flaw was first noted in a letter by Friedman shortly after the Goldin study was published.[13] In the same letter, Friedman reported results from a Kaiser Permanente cohort study of the carcinogenic effects of prescription drugs, which showed no protective effect of digitalis compounds on breast cancer incidence. In fact, the standardized incidence ratio (SIR) for breast cancer among digitalis users, compared with non-users, was 1.2—very similar to the results of our study. No confidence interval was given for this measure of association.

Oddly enough, results from a supplemental case-control study by Stenkvist *et al.* agree with our present findings. The investigators compared the CG exposure history of the breast cancer cases from their original report[8] with the exposure history of an equal number of age-matched controls from the general population.[34] The authors cross-tabulated yes/no CG exposure with case/control status, and concluded that cardiac glycosides had no influence on breast cancer incidence, owing to a non-significant chi-square test for independence (P=0.25). The data from the cross-tabulation correspond to an odds ratio of 1.39, with a 95% confidence interval from 0.79 to 2.45. Figure 2 displays the p-value functions calculated from

our data and from the data reported by Stenkvist.[36, 37] Both studies arrived at approximately the same point estimate of the OR associating cardiac glycoside therapy with breast cancer incidence, but this study's estimate was measured with greater precision (resulting in a narrower p-value function). The majority of the area of the Stenkvist p-value function lies in the domain of positive associations, to the right of the null, making it likely that increased precision in their study would have alerted them to a positive association by the chosen chi-square test.

Other epidemiologic studies also appear consistent with our results. Haux and colleagues tabulated the number of incident cancers at several anatomic sites in a cohort of digitoxin-exposed Norwegian men and women. They then compared these counts with the expected number of cases based on age- and sex-standardized rates in the Norwegian population. No protective effect was observed at any of the anatomic sites; users of digitoxin showed a slightly elevated breast cancer incidence (SIR: 1.25; 95% CI: 0.95, 1.62).[12] A case-control study conducted by Ewertz *et al* found a positive association between digoxin usage and incident male breast cancer (OR for ≥5 years of digoxin use: 2.0; 95% CI: 0.9, 4.4)[38] which comports with earlier associations drawn between digitalis glycoside therapy and gynecomastia in men treated for heart disease.[39, 40] Together, these results argue against estrogen receptor antagonism by digoxin, and perhaps cardenolides in general.

Current laboratory findings implicate the sodium-potassium ATPase in a variety of signal transduction pathways. Many *in vitro* studies point toward a downstream anti-proliferative effect of CGs bound to the ATPase, but others leave open the possibility of a carcinogenic endpoint. For instance, the sodium-potassium ATPase appears to be a key modulator of cellular adhesion, which is impaired when

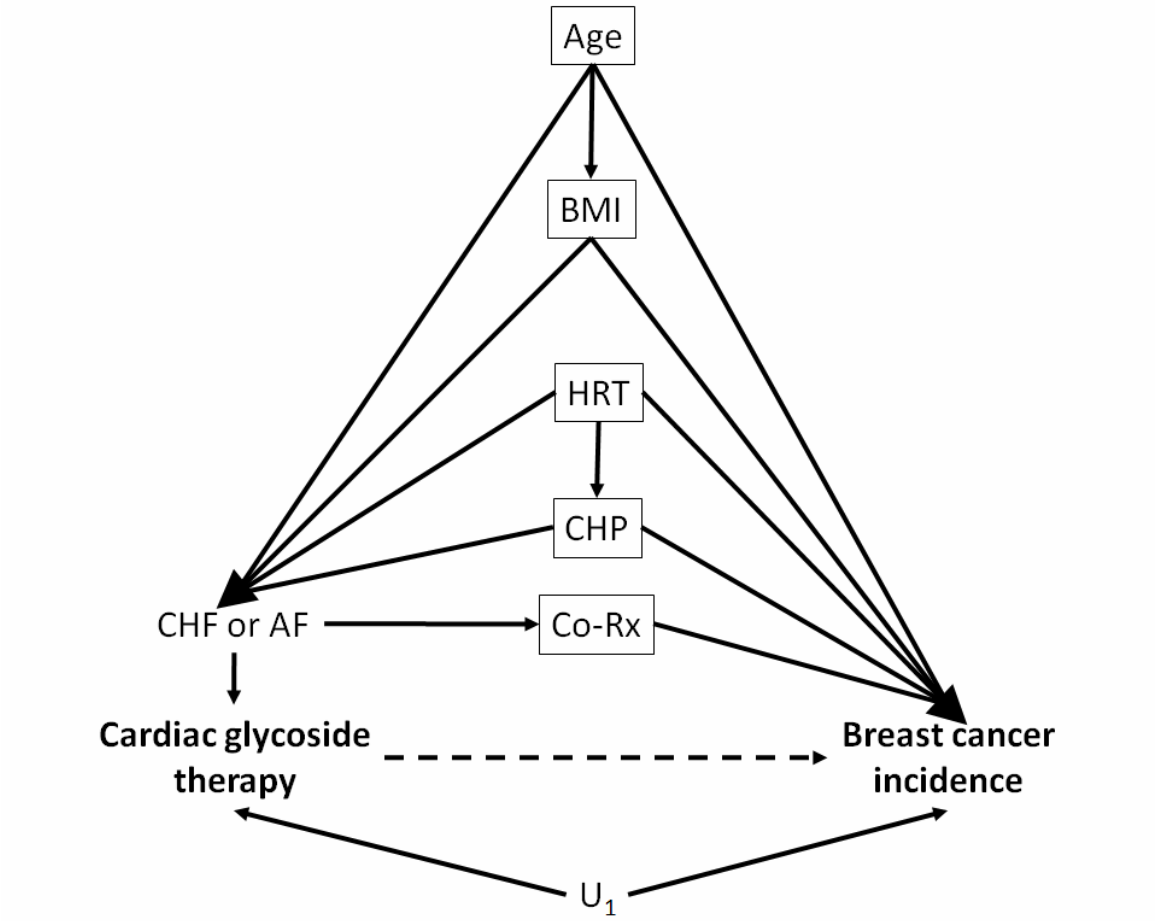
CGs are bound to the receptor *in vitro*.[11] Detachment from adjacent cells and the basement membrane and contact-independent growth are both hallmarks of the carcinogenic process.[41]

There is evidence that both the interaction of cardiac glycosides with the sodium-potassium ATPase and the consequential effects are highly dependent on the specific cardiac glycoside compound and the subunit makeup of the receiving ATPase.[7] Differences in the protein subunit composition of the ATPase are thought to regulate interspecies and inter-tissue variations in response to CG binding.

Summary

We observed a modestly increased rate of breast cancer among post-menopausal women with any history of digoxin use, compared with women with no such use, after adjustment for age, use of other prescription drugs, and cardiovascular indications. The associations persisted in long-term exposure categories. While a number of laboratory studies of cardiac glycosides and female breast cancer have suggested protective effects, our results suggest that one specific cardiac glycoside, digoxin, moderately increases the incidence rate of breast cancer. This finding agrees with results from past studies; [12, 13, 34] the importance of which were likely masked by relatively large standard errors of the association measures, in combination with an emphasis on statistical significance testing in causal inference.

Figure 1: Directed Acyclic Graph (DAG) depicting hypothesized relationships between analytic variables for the measurement of the association between digoxin therapy and breast cancer incidence.



CHF: congestive heart failure; AF: atrial fibrillation; BMI: body mass index; HRT: hormone replacement therapy; CHP: cumulative hormonal profile; Co-Rx: use of anticoagulants, NSAIDs, or aspirin concurrently with a cardiac glycoside. Boxes denote confounders identified by the back-door test (see text for details).

Table 1: All Cardiac Glycoside Products Prescribed to Study Subjects.

Product	Dose	Fill quantity	No. of prescriptions (% of total)
Digoxin	62.5 μg/tablet	100 tablets	83,094 (66)
	62.5 μg/tablet	200 tablets	38,188 (31)
	250 μg/tablet	100 tablets	4,047 (3.2)
	50 μg/mL	30 mL	28 (0.02)

Table 2: Characteristics of the Cases and Controls.

Variable	Cases (n=5,565)		Controls (n=55,650)	
Age on index date (years), n (%)				
55 – 64	2,116	(38)	21,160	(38)
65 – 74	1,800	(32)	18,000	(32)
75 – 84	1,356	(24)	13,560	(24)
≥ 85	293	(5.3)	2,930	(5.3)
Medical history, n (%)				
Congestive heart failure	160	(2.9)	1,337	(2.4)
Atrial fibrillation/ flutter	224	(4.0)	1,819	(3.3)
Pre-diagnosis mammography[†]				
< 1 year	417	(81)	84	(1.6)
1 to < 2 years	3	(0.6)	84	(1.6)
2 to < 3 years	9	(1.7)	84	(1.6)
≥ 3 years	17	(3.3)	130	(2.5)
Myocardial infarction	123	(2.2)	1,492	(2.7)
Chronic pulmonary disease	334	(6.0)	3,125	(5.6)
Peripheral vascular disease	167	(3.0)	1,563	(2.8)
Cerebrovascular disease	275	(4.9)	2,842	(5.1)
Lymphoma	12	(0.2)	155	(0.3)
Other solid tumor	0		0	
Liver disease	44	(0.8)	403	(0.7)
Diabetes (Type I or II)	215	(3.9)	1,706	(3.1)
Diabetes with complication	85	(1.5)	591	(1.1)
Renal disease	35	(0.6)	446	(0.8)
Other drug exposures, n (%)				
Hormone replacement therapy	2,062	(37)	17,582	(32)
Anticoagulants	231	(4.2)	2,109	(3.8)
NSAIDs	3,106	(56)	29,964	(54)
Aspirin, low-dose (<150 mg)	205	(3.7)	2,004	(3.6)
Aspirin, high-dose (≥150 mg)	505	(9.1)	4,878	(8.8)

[†]Restricted to cases and controls with index dates after 1 January 2006, when screening mammography data would have been most complete in Denmark. Categories reflect time elapsed between most recent mammogram and index date; proportion denominators are the total number of cases (n=516) or controls (n=5,160) in the restricted data set.

Table 3: Stratified analysis of the association between digoxin therapy and breast cancer incidence, by age category and use of hormone replacement therapy (HRT).

	Age 55 to 64		Age 65 to 74	
	HRT	No HRT	HRT	No HRT
	Cases/ Controls	Cases/ Controls	Cases/ Controls	Cases/ Controls
Exposed	15/ 81	13/ 101	27/ 181	58/ 421
Unexposed	991/ 8544	1097/ 12434	649/ 5328	1066/ 12070
OR (95% CI):	1.6 (0.92, 2.8)	1.5 (0.82, 2.6)	1.2 (0.81, 1.8)	1.6 (1.2, 2.1)

	Age 75 to 84		Age ≥ 85	
	HRT	No HRT	HRT	No HRT
	Cases/ Controls	Cases/ Controls	Cases/ Controls	Cases/ Controls
Exposed	39/ 253	116/ 1040	19/ 76	37/ 393
Unexposed	278/ 2690	923/ 9577	44/ 429	193/ 2032
OR (95% CI):	1.5 (1.0, 2.1)	1.2 (0.94, 1.4)	2.4 (1.4, 4.4)	1.0 (0.69, 1.4)

OR_{MH} (95% CI):	1.31 (1.16, 1.48)

Table 4: Associations Between Digoxin Use and Incident Breast Cancer.

Exposure categories	Cases (n=5,565)	Controls (n=55,650)	Unadjusted OR (95% CI)	Adjusted* OR (95% CI)
Ever/never prescribed				
Ever	324	2546	1.29 (1.14, 1.45)	1.30 (1.14, 1.48)
Never	5241	53104	1.0 (Ref)	1.0 (Ref)
Duration of digoxin therapy†				
7-18 y	93	694	1.35 (1.10, 1.69)	1.39 (1.10, 1.74)
4-6 y	103	811	1.29 (1.05, 1.58)	1.30 (1.05, 1.61)
1-3 y	128	1041	1.25 (1.03, 1.50)	1.25 (1.03, 1.52)
Never user	5241	53104	1.0 (Ref)	1.0 (Ref)
Ever/never prescribed (alternate reference group)	**(n=732)**	**(n=7,086)**		
Ever	324	2546	1.42 (1.21, 1.65)	1.42 (1.14, 1.77)
Never ‡	408	4540	1.0 (Ref)	1.0 (Ref)

*Adjusted for age (continuous), county of residence (categorical), and past receipt of hormone replacement therapy (dichotomous).

†Years elapsed between first digoxin prescription and index date.

‡The alternate reference group is additionally defined by a medical history positive for either myocardial infarction, peripheral vascular disease, cerebrovascular disease, or any combination thereof. Women with these histories should exhibit a hormonal profile similar to women with digoxin-indicating medical conditions.

Figure 2: P-value functions comparing association data reported by Stenkvist with the data reported in the present study (Ahern). The gray horizontal reference line intersects each p-value function at the lower and upper 95% frequentist confidence limits.

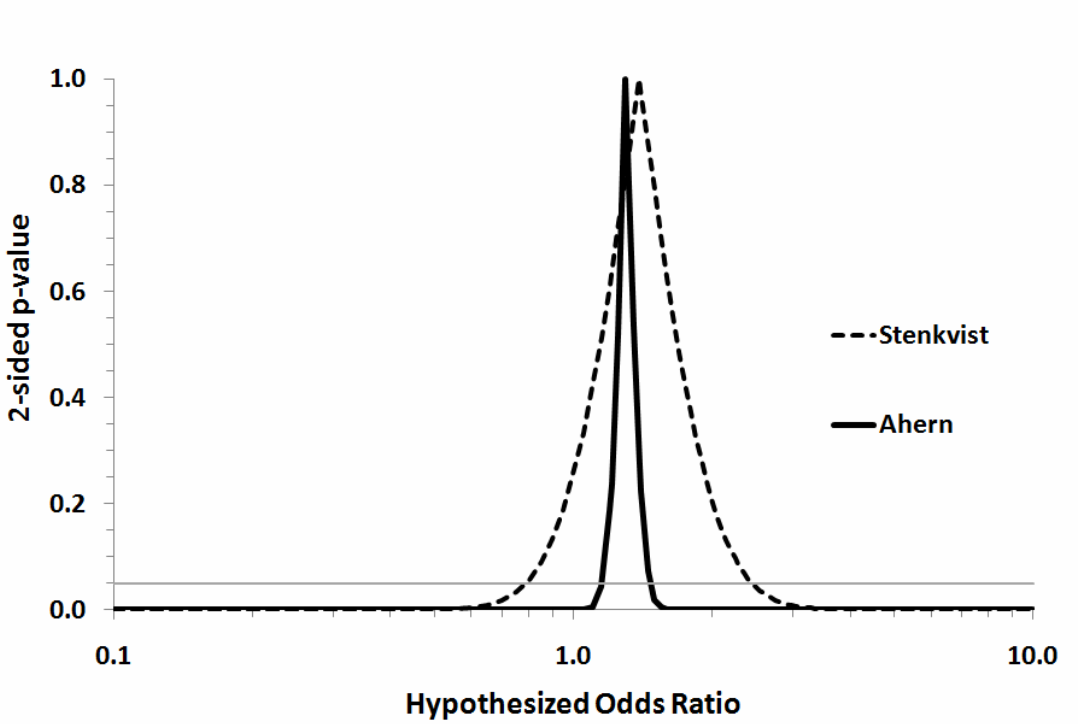


Table 5: ICD-8 and ICD-10 Codes Used to Ascertain Key Diagnoses.

Diagnosis	ICD-8	ICD-10
Invasive breast cancer	174.00-174.02; 174.08; 174.09;	C50.0-C50.6; C50.8; C50.9
Congestive heart failure	427.09; 427.10; 427.11; 427.19; 428.99; 782.49	I50; I11.0; I13.0; I13.2
Atrial fibrillation/ flutter†	427.93; 427.94	I48
Myocardial infarction	410	I21; I22; I23
Peripheral vascular disease	440; 441; 442; 443; 444; 445	I70; I71; I72; I73; I74; I77
Cerebrovascular disease	430-438	I60-I69; G45; G46
Chronic pulmonary disease	490-493; 515-518	J40-J47; J60-J67; J68.4; J70.1;J70.3; J84.1; J92.0; J96.1; J98.2; J98.3
Mild liver disease	571; 573.01; 573.04	B18; K70.0-K70.3; K70.9; K71; K73; K74; K76.0
Moderate to severe liver disease	070.00; 070.02; 070.04; 070.06; 070.08; 573.00; 456.00-456.09	B15.0; B16.0; B16.2; B19.0; K70.4; K72; K76.6; I85
Diabetes type1 Diabetes type2	249.00; 249.06; 249.07; 249.09 250.00; 250.06; 250.07; 250.09	E10.0, E10.1; E10.9 E11.0; E11.1; E11.9
Moderate to severe renal disease	403; 404; 580-583; 584; 590.09; 593.19; 753.10-753.19; 792	I12; I13; N00-N05; N07; N11; N14; N17-N19; Q61
Diabetes with end organ damage (types 1 and 2)	249.01-249.05; 249.08; 250.01-250.05; 250.08	E10.2-E10.8; E11.2-E11.8
Solid tumor	140-194	C00-C75
Lymphoma	200-203; 275.59	C81-C85; C88; C90; C96

†ICD-8 contained separate codes for atrial fibrillation (427.93) and flutter (427.94). These two diagnoses were combined into a single code in ICD-10 (I48).

VITAMIN K ANTAGONIST THERAPY AND THE INCIDENCE OF SITE-SPECIFIC CANCERS

BACKGROUND:

Vitamin K Antagonists: Background, Pharmacology, and Indications

Vitamin K antagonists (VKAs)—which include warfarin, phenprocoumon, and dicoumarol—are a class of oral anticoagulant drugs that have been used therapeutically in the United States since the 1950s.[42] Dicoumarol (bishydroxycoumarin) is a natural compound, discovered when hemorrhagic complications occurred in a herd of cattle that had fed on mold-contaminated sweet clover silage.[43] Warfarin, a synthetic congener of dicoumarol with higher potency, was first produced in the late 1940s and marketed as a rodenticide. Later, after validation of its safety profile in humans, warfarin became a standard treatment for the prevention of thromboembolism.[42]

All VKAs exert the same pharmacologic effect: the inhibition of vitamin K epoxide reductase. Vitamin K is a necessary cofactor for the activation of clotting factors II, IX, X, and proteins S and C—all critical to the coagulation pathway—*via* carboxylation of gamma glutamic acid residues. During the gamma carboxylation reaction, the vitamin K cofactor is oxidized to an inactive epoxide, which must be recycled to an active hydroquinone form *via* vitamin K epoxide reductase. Inhibition of vitamin K epoxide reductase activity therefore causes the one-way conversion of active vitamin K cofactor to an inactive form, which eventually halts the coagulation cascade.[42]

Oral anticoagulation therapy with a vitamin K antagonist is indicated for preventing recurrence of acute venous thrombosis and acute pulmonary

embolism.[42] VKA therapy is also used for the primary prevention of thromboembolism in patients undergoing surgery, and patients with acute myocardial infarction, chronic atrial fibrillation, or replacement heart valves.[42, 44]

The combination of efficacy and economy has made VKAs an enduring mainstay of oral anticoagulation therapy worldwide.

Vitamin K Antagonists and Cancer Risk

Epidemiologic evidence for an association between VKA therapy and cancer outcomes first appeared in the early 1980s with a pair of reports by Zacharski and colleagues of increased overall and recurrence-free survival time among patients with small cell lung carcinoma who were randomized to receive standard therapy plus warfarin, compared with patients who received standard therapy alone.[45, 46] No survival advantage among those randomized to warfarin was apparent for patients with non-small cell lung cancer, nor with cancers of the colon, head and neck, or prostate.[46] A 2001 systematic review of randomized studies of VKAs and overall cancer survival noted only a suggested protective association among patients with small cell lung carcinoma (1-year mortality OR: 0.72, 95% CI: 0.44, 1.16) but not among patients with other cancers.[47] Only one study so far has examined VKA therapy and breast cancer survival, and warfarin treatment had no detectable association with overall survival among patients diagnosed with metastatic breast cancer (mortality risk ratio: 0.90, p=0.55, derived 95% CI: 0.64, 1.3).[48]

Many years after publication of the Zacharski studies, Schulman and Lindmarker compared site-specific cancer incidence among clinical trial participants with a first incident venous thromboembolism (VTE) who were randomized to either short-duration (6 weeks) or long-duration (6 months) oral anticoagulation with warfarin.[20] The 854 participants with no history of malignancy at randomization were

followed from the time of their VTE diagnosis until the first incident cancer. The median follow-up time for the cohort was 8.1 years from the time of the VTE event. The short-duration warfarin group (n=419) experienced 66 cancer diagnoses; the long-duration warfarin group (n=435) experienced 45 cancer diagnoses. Comparing the two groups, the investigators observed an increased risk of urogenital cancer in the short-duration warfarin arm, compared with the long-duration arm (for a combined outcome of kidney, urinary bladder, prostate, ovary or uterine cancer: OR = 2.5; 95% CI = 1.3, 5.0). The authors asserted that their results strongly supported the notion that VKA drugs had anti-neoplastic activity, but their study received criticism for a number of unaddressed limitations including questionable sensitivity of outcome classification, lack of data on potentially confounding co-medications such as NSAIDs, selection bias arising from differential losses to follow-up, possible differential cancer surveillance in the two warfarin groups, and lack of a plausible biological mechanism for VKAs to specifically modify risk for the studied urogenital cancers.[49-53] No study has yet replicated the original findings of Schulman and Lindmarker. Most subsequent studies suffer severe limitations in precision owing to small numbers of incident site-specific cancers over relatively short follow-up times in randomized studies. Taliani showed a null association between 12 vs. 3 months of VKA therapy and the incidence of any cancer (19 exposed cancer events; RR=0.71; 95% CI: 0.36, 1.41).[54] Breast cancer incidence data from the Taliani study permit calculation of an imprecise cumulative incidence ratio comparing 12 month VKA therapy with 3 month VKA therapy that indicates the possibility of a positive association (RR=4.8; 95% CI: 0.6, 41).[54] Regarding urogenital cancers, Blumentals reported a null association between warfarin therapy and the incidence of bladder cancer in a Veterans' Administration case-control study (OR=1.27; 95% CI: 0.85,

1.89), while Tagalakis reported a reduced rate of prostate cancer among warfarin users compared with non-users in a Canadian population-based case-control study (OR=0.80; 95% CI: 0.65, 0.99).[55, 56] Tagalakis and colleagues observed null associations between VKA therapy and the incidence of bladder, kidney, uterine, and ovarian cancers.[56] A recent overview of this topic concluded that the evidence for an association between VKA therapy and cancer incidence is inconclusive, and that new randomized trials are necessary to confirm the observed utility of VKA drugs as adjuvant therapy for certain cancer subtypes.[57]

Rationale for the Current Design

To address important limitations in the earlier literature on this topic—most notably the limited number of VKA-exposed site-specific cancer diagnoses accrued by most studies and the narrow focus on urogenital cancers—we conducted a Danish population-based cohort study with a long follow-up period.

The Danish National Registry of Patients (NRP) has electronically recorded all inpatient diagnoses made at non-psychiatric Danish hospitals since 1977. The Danish Cancer Registry (DCR) is a consolidated, automated registry of cancer diagnoses made since 1943, with mandatory reporting beginning in 1987. Prescription drug data are available from the Danish Register of Medicinal Product Statistics (RMPS), but are only available nationwide from 1995 onward. Therefore, if we were to limit our study to the time period for which actual prescription data were available, our follow-up time—and therefore our cancer event frequencies—would not likely represent an improvement over earlier studies. To enable a longer follow-up period, we capitalized on the longer period of coverage of the NRP, from which we ascertained heart valve replacement procedures as a proxy for exposure to a vitamin K antagonist. As we argue below, heart valve replacement appears to

satisfy the requirements of an instrumental variable for the associations between VKA therapy and the incidence of site-specific cancers. Consequently, use of this variable not only lengthens the available study period, but also permits the estimation of unconfounded associations between VKA therapy and the cancer outcomes, presuming that the instrumental variable assumptions are valid.

METHODS:

Study Population and Data Collection

This study was nested within the entire population of Denmark during the period January 1, 1989 to December 31, 2005. We used the Danish National Registry of Patients to identify all instances of heart valve replacement that occurred during the study period. Heart valve replacement patients are expected to receive VKA therapy immediately after surgery.[58] To allow a reasonable time to pass for the VKA to be a plausible etiologic agent in the development (or prevention) of cancer, we began follow-up one year after the index (surgical) date. Replacements were either for the mitral, aortic, tricuspid, or pulmonic valve, and could be either biological (*i.e.*, porcine) or artificial. We defined each subject's index date as the date of their heart valve replacement procedure, and excluded all subjects with a diagnosis of any cancer either before—or within the year after—the index date. The resulting roster comprised our cohort of exposed subjects.

We sampled up to 10 unexposed subjects for each exposed subject from the general population, without replacement, matching on year of birth and sex. Matched unexposed subjects were required to have no history of heart valve replacement or cancer diagnosis on their matched exposed subject's index date, or within the year after that date.

We searched the Danish Cancer Registry for diagnoses of the 49 site-specific cancers listed in Table 7 among our cohort members. Follow-up of each subject began one year after their index date and continued until the first of the following events occurred: first incident cancer, death, emigration (loss to follow-up), or the end of the study period. We ensured that no subject appeared in the cohort more than once; as a result, a matched unexposed person enrolled in the cohort could not go on to receive a heart valve replacement. We searched the NRP to enumerate prevalent comorbidities on the index date according to the Charlson Comorbidity Index, as well as to ascertain past history of two other major indications for VKA therapy: venous thromboembolism (VTE) and atrial fibrillation (AF).

To evaluate the performance of heart valve replacement status as a dichotomous proxy variable for exposure to a vitamin K antagonist drug, we conducted a nested validation study in the subset of cohort members whose areas of residence and index dates fell within the geographic and temporal coverage of four Danish county-specific automated prescription registries (Aarhus County, data since 1996; North Jutland County, data since 1989; Viborg and Ringkøbing Counties, data since 1998). (These are now in the North, Mid, and South regions of Denmark, respectively.) Vitamin K antagonist drugs are available only by prescription in Denmark. Patients must present their prescription to a pharmacy and pay the entire cost of the medication. Once paid, the medication is dispensed and the transaction is automatically logged into the county registry and the Danish government reimburses the patient for a portion of the drug cost. All county prescription registries encode dispensed drugs according to the Anatomic Therapeutic Chemical (ATC) classification system,[18] and additionally record dispense dates, fill quantities, and the receiving patient's CPR number. Using these

registries, we determined whether or not each member of the validation subset received a prescription for one of the vitamin K antagonists available in Denmark.

Definitions of Analytic Variables

We identified heart valve replacement surgeries in the NRP using the procedure codes listed in Table 6. For heart valve recipients, the index date was defined as the date of their surgical procedure. The same index date for a given heart valve recipient was assigned to all of his or her matched unexposed subjects.

We defined each subject's age as that on their index date by calculating the number of years elapsed between their birth date and the index date. For stratified analyses, age was categorized as follows: less than 18 years, 18 to 24 years, 25 to 34 years, 35 to 44 years, 45 to 54 years, 55 to 64 years, 65 to 74 years, 75 to 84 years, and greater than or equal to 85 years.

We ascertained cancer incidence by linking our cohort roster to the Danish Cancer Registry using each subject's CPR number. By design, all exposed and unexposed subjects were without a prior cancer diagnosis one year after their index date. We identified the first diagnosis of any of the cancers listed in Table 7 using ICD-10 codes. The cancer registry has translated earlier entries under past ICD conventions into ICD version 10 to create uniform ascertainment criteria. We considered as separate outcomes all cancers with a total of 5 or more identified cases.

For subjects without a cancer diagnosis, we determined the end of follow-up by CPR linkage with the Danish Civil Registry, which updates address and vital status for all Danish residents on a daily basis. Each subject contributed person-time at risk for a first cancer diagnosis from one year after their index date until the

first occurrence of (a) a cancer diagnosis, (b) emigration from Denmark, (c) death from any cause, or (d) December 31, 2006, the end of available follow-up.

For descriptive purposes, we calculated the Charlson Comorbidity Index (CCI) for all subjects based on appropriate diagnoses prevalent on the index date. The CCI summarizes the extent of one's medical history for major diagnoses into a convenient ordinal score for use in stratified analyses and regression models. Beyond the Charlson comorbid conditions, we also assessed history of atrial fibrillation, pulmonary embolism, portal vein thrombosis, and unspecified venous thrombosis. All such diagnoses were ascertained by CPR linkage to the NRP; Table 8 lists the ICD-8 and ICD-10 codes for these diagnoses. We calculated the weighted Charlson comorbidity score and summarized the ordinal index according to the published method.[59] We defined a positive history of venous thromboembolism (VTE) as having been diagnosed with either pulmonary embolism, portal vein thrombosis, unspecified venous thrombosis, or any combination thereof, before the index date.

For the cohort members in the validation subset, we searched appropriate county prescription registries for all records of vitamin K antagonist (VKA) prescription after their index dates. We ascertained VKA prescriptions by searching for ATC data fields containing the character string "B01AA," which identified all specific drugs in the VKA class of anticoagulants. We recorded the specific VKAs and the total number of prescriptions of each recorded for members of the validation subset. Since heart valve replacement patients are almost always placed on life-long VKA therapy after surgery,[60] determination of 'ever exposure' to a VKA after the index date is expected to indicate enduring use of the drug.

Statistical Analysis

Within heart valve replacement exposure groups, we calculated the frequency of subjects and the sum and proportion of person-time according to incidence of site-specific cancers, age category, sex, Charlson Comorbidity Index, history of VTE, and history of AF.

Associations Estimated in the Validation Subset

Within the validation subset, for site-specific cancers with at least 5 identified cases, we calculated incidence rate ratios (IRR) and 95% confidence intervals comparing the VKA-exposed group with the VKA-unexposed group. The point estimates and 95% confidence intervals for these IRRs were plotted according to rank to provide a visual depiction of the overall association pattern. Since idiopathic VTE motivates intensive screening for asymptomatic cancer,[61, 62] we also re-calculated the cancer associations in the validation subset further restricted to subjects with no history of VTE on their index date.

Estimation of Validation Parameters for Heart Valve Replacement as a Proxy Variable for VKA Therapy

We stratified the validation subset by disease status (any cancer versus no cancer diagnosed during follow-up) and cross-tabulated heart valve replacement status with VKA exposure status. Because of the 10:1 matched design, only predictive values—not sensitivity and specificity—could be calculated from our validation data.[63] Positive predictive values (PPV) quantified the performance of a positive heart valve replacement history as a proxy for receiving VKA therapy during follow-up, and were calculated as the number of subjects positive for both valve replacement and VKA therapy, divided by the total number of heart valve recipients.

Negative predictive values (NPV) quantified the performance of a negative heart valve replacement history as a proxy for never receiving VKA therapy during follow-up, and were calculated as the number of subjects negative for both heart valve replacement and VKA therapy, divided by the total number of subjects with no heart valve replacement.

We examined stability of the PPV and NPV point estimates over time by stratifying the validation data by subjects' index year and calculating PPV and NPV within each interval.

Probabilistic Bias Analysis

Estimation of the PPV and NPV enabled us to adjust for misclassification of exposure in the non-validated subset of the cohort, yielding VKA/cancer associations rooted in a larger body of data. We employed probabilistic bias analysis methods described by Lash, Fox, and Fink to carry out these corrections.[64]

To begin, we allowed for uncertainty in the PPV and NPV estimates themselves by empirically characterizing separate beta distributions for these parameters according to cancer diagnosis status (any cancer diagnosed vs. no cancer diagnosed). A beta distribution is defined by two positive shape parameters, alpha (α) and beta (β), and is restricted to the range (0,1).[64] This range restriction makes the beta distribution particularly suitable for describing proportions, as in the case of PPV and NPV. The α and β parameters can be defined in a variety of ways to yield distributions that are either nearly uniform, bimodal, symmetric, or skewed, depending on the desired distribution of the modeled proportion.[64] Using validation data according to cancer diagnosis status (any vs. no cancer diagnosed over follow-up, denoted by the subscripts 1 and 0, respectively), beta distributions for PPV_1 and PPV_0 were defined by setting alpha parameters equal to the number of heart valve

recipients who also received VKA therapy plus one, and beta parameters equal to the number of heart valve recipients who did not receive VKA therapy plus one.[64] Similarly, beta distributions for NPV_1 and NPV_0 were defined by setting alpha parameters equal to the number of subjects with no valve replacement who did not receive VKA therapy plus one, and beta parameters equal to the number of subjects with no valve replacement who did receive VKA therapy plus one.[64] Plots of the resulting beta distributions appear in Figure 3. The VKA exposure data in the prescription validation subset of the cohort were considered to be correctly classified, and tables for the site-specific cancer associations were arranged according to the example in Table 9. Let a_{val} represent the number of site-specific cancer cases among those with a documented VKA prescription, and b_{val} represent the number of cases among those with no documented VKA prescription. Let $PT1_{val}$ and $PT0_{val}$ represent the total person-time at risk contributed by those exposed to VKA and those unexposed to VKA, respectively. The subscript 'val' denotes that these observed cell frequencies arise from the validation subset with correctly classified VKA exposure status.

Site-specific cancer association data in the non-validation subset were arranged according to the example in Table 10. Let a_{mis} represent the number of site-specific cancer cases among cohort members who received a heart valve replacement and let b_{mis} represent the number of cases among non-recipients. Similarly, let $PT1_{mis}$ represent the total person-years at risk for a first incident cancer accumulated by heart valve recipients and let $PT0_{mis}$ represent the total person-years at risk accumulated by non-recipients. The subscript 'mis' denotes that these observed cell frequencies arise from the non-validation subset and are subject to misclassification of VKA exposure by the heart valve replacement proxy variable.

We performed 100,000 iterations of table-level misclassification adjustment[65] according to the algorithm depicted in Figure 4. Each iteration began by drawing a random value from the standard uniform probability distribution and a second random value from the standard uniform that was correlated with the first (r=0.8); this was done separately by cancer status strata. These were in turn used to draw values of PPV and NPV from the stratum-appropriate beta distributions. Within the non-validated subset, we adjusted the cell frequencies in each of the site-specific cancer 2x2 tables using the selected PPV and NPV values according to the equations in

Table 11. The observed cell frequencies from the validated subset were then added to the re-classified cell frequencies to yield a combined table for each site-specific cancer association.

For each site-specific cancer, the incidence rate ratio (IRR) comparing those classified as VKA-exposed to those classified as VKA-unexposed was calculated, along with the conventional standard error. To re-incorporate random error into the adjusted estimates, we subtracted the product of the conventional standard error and a random standard normal deviate from the natural logarithm of the IRR. Exponentiating this difference yielded an IRR estimate adjusted for exposure misclassification, incorporating uncertainty due to random error as well as the variation in validation parameters. IRR values from each iteration were stored, and the bias analysis routine was re-iterated until 100,000 estimates were accumulated. Thus, for each site-specific cancer outcome, we generated a distribution of 100,000 misclassification-corrected estimates.

We reported the median of each distribution as the point estimate for each association. For each group of corrected IRRs (one for each site-specific cancer),

we calculated univariate statistics to characterize the 2.5th and 97.5th percentiles of the distribution of estimates. These percentile values were reported as the 95% simulation interval, a measure of uncertainty in the misclassification adjustment as well as sampling error.

Empirical Bayes Analysis of the Associations Estimated from the Validation Subset and from the Probabilistic Bias Analysis

Because we are estimating associations between VKA therapy and the incidence of 24 separate cancers, the potential generation of false-positive associations (that is, either spuriously 'causal' or spuriously 'protective') by random mechanisms is a limitation of our study. To address this possibility, and to remove emphasis from any high magnitude associations that are measured with considerable imprecision, we subjected the vectors of VKA/cancer associations from both the validation subset and the probabilistic bias analysis to empirical Bayes shrinkage.[66, 67]

The empirical Bayes (EB) method is applied to groups of associations from a common population—here, the individual associations are all first incident cancers among the cohort subjects—and 'shrinks' each association toward the overall mean, in proportion to the ratio of its variance to the true population variance.[68] Associations with large variances relative to the true population variance are displaced further toward the overall mean than associations with smaller relative variances. The prime objective of empirical Bayes shrinkage is to de-emphasize estimates that have impressive magnitudes but were measured with poor precision, such that time and resources are less likely to be spent in pursuit of false-positive associations.

We performed the EB analysis according to the method published by Steenland *et al.*,[68] adapting their calculations to the SAS/IML programming environment (SAS version 9.1). We incorporated no prior information about the associations and assumed that the cancer outcomes were independent of one another.

Instrumental Variable Analysis

In our age- and sex-matched cohort, receipt of a heart valve replacement appears to satisfy the necessary criteria to serve as an instrumental variable (IV) for the association between VKA therapy and the incidence of site-specific cancers.[69, 70] Figure 6 is a directed acyclic graph (DAG) depicting hypothesized relationships between heart valve replacement (Z; the instrumental variable), VKA prescription (X; the target exposure), and cancer incidence (Y; the outcome). In order for valve replacement to be a valid IV for the VKA/cancer associations, it must have no direct causal effect on the incidence of the site-specific cancers we studied (*i.e.*, no plausible arrow 'b' in Figure 6), its effect on cancer incidence must be mediated by VKA exposure (presence of arrow 'a' in Figure 6), and there must be no open backdoor path[19] between valve replacement and cancer incidence (that is, no common cause of valve replacement and cancer incidence (absence of, or adequate conditioning on, node 'U_2' in Figure 6)). If these assumptions hold, then bias due to residual confounding of the VKA/cancer associations (node 'U_1' in Figure 6) is negated, albeit at the expense of non-differential exposure misclassification by the instrument.[69, 70]

We believe the IV assumptions hold because (1) there is no literature supporting a causal relationship between surgical installation of a heart valve and the incidence of any cancer, nor is there a biological pathway that implies such an

association (*i.e.*, no arrow 'b'), (2) heart valve replacement essentially mandates lifelong VKA therapy,[60] while the prevalence of VKA treatment is relatively low in the general population (*i.e.*, there exists a strong causal association depicted by arrow 'a'), and (3) after matching valve recipients to non-recipients on age and sex, we expect no important uncontrolled common causes of valve replacement and cancer incidence (*i.e.*, no other variables that satisfy the 'U_2' node for any of the site-specific cancers examined).

Instrumental variable analyses are most commonly performed with two-stage linear regression. In the first stage, the association between the instrument and the outcome (Z-Y) is estimated, which—conditional on the validity of the instrument—yields an estimate of the exposure-outcome relationship that is unconfounded but inherently misclassified.[71, 72] The strength of the association between the instrument and the target exposure (Z-X) informs the severity of this misclassification.[70] In the second stage of a traditional IV analysis, the instrument-outcome association (Z-Y) is scaled by the instrument-exposure association (Z-X), yielding an unconfounded estimate of the exposure-outcome risk difference (X-Y) that has been adjusted for misclassification of the target exposure by the instrument.[71] This approach requires that each subject in a study has data for both the instrument and the target exposure. In our study, all subjects have data on the instrument (heart valve replacement), but only ~25% of those subjects have data on the target exposure (VKA prescription). Under a traditional two-stage IV analysis framework, we could use our data to estimate theoretically unconfounded instrument-outcome associations but, because not all subjects have recorded VKA prescription status, we would not be able to carry out the second stage regression to correct the inherent exposure misclassification. Our solution to this limitation was to multiply

impute the VKA prescription status of the ~75% of subjects with missing data.[73] We used a logistic regression method[74] to impute binary VKA prescription status in subjects with missing VKA data, using the heart valve replacement instrument as the predictor. The regression procedure imputes the missing VKA exposure based on coefficients drawn from a posterior distribution characterized by the regression of observed VKA exposure on the heart valve replacement instrument within the validation subset.[73, 74] We performed a total of five imputations[73] using the MI procedure in SAS version 9.1.

Using the imputed data sets, we estimated IRRs associating VKA therapy with site-specific cancer incidence by using separate Poisson regression models for each of the 24 site-specific cancers in each of the five imputed data sets. VKA status was the sole independent variable in each of these models, and the logarithm of person-years at risk served as the offset variable. We used generalized estimating equations (GEE) to calculate variances empirically. We clustered by subject and specified an exchangeable covariance matrix. The GEE parameter estimates and standard errors for each cancer site were used to calculate summary IRRs and 95% confidence intervals reflective of the uncertainty in imputed exposure values (PROC MIANALYZE, SAS version 9.1).[75]

To evaluate whether the pattern of IV-estimated associations was consistent with a null-centered distribution of associations, the estimates were ranked by magnitude and plotted against the inverse-normal of rank percentile (INRP). This plot was overlaid with predicted log IRR values calculated from the inverse-variance weighted linear regression of the observed log IRR estimates on their INRP values.[76] Associations drawn from a null-centered Gaussian distribution are

expected to lie along the predicted log IRR line, which would intersect the plot coordinate where INRP and the log IRR both equal 0.[76]

RESULTS:

Baseline Characteristics of the Cohort

We identified 9,727 subjects who underwent a heart valve replacement procedure during the study period. To these subjects we matched 95,481 subjects from the general population with no history of heart valve replacement (average matching ratio = 9.8 unexposed : 1 exposed). Heart valve recipients contributed a total of 52,510 person-years of follow-up, and matched non-recipients contributed a total of 556,021 person-years of follow-up.

A comparison of baseline characteristics between heart valve recipients and non-recipients appears in Table 13. Approximately equal proportions of person-time were contributed across exposure groups with respect to sex and age on the index date, indicating successful matching on these variables. Heart valve recipients had a higher comorbidity burden than the matched unexposed (43% vs. 15% of person-time contributed by subjects with CCI ≥ 1). As expected, heart valve recipients were about nine times more likely to have a history of atrial fibrillation/flutter than non-recipients (8.6% vs. 1.0% of person-time positive, respectively), and were approximately three times more likely to have a history of venous thromboembolism (1.0% vs. 0.3% of person-time positive, respectively). The imbalance in positive history of VTE between exposure groups motivated re-calculation of cancer incidence rate ratios in the subset of cohort members with no history of VTE, since primary VTE (*i.e.*, VTE not linked to pregnancy, traumatic injury, or surgery) increases suspicion of (and screening for) previously undetected cancer.[77] The

higher baseline prevalence of VTE among the exposed subjects could lead to artificially higher cancer incidence rates in this group, yielding an upward bias in IRR values. Analyses restricted to the subset with no VTE history did not differ from analyses on the entire cohort (Table 16, compared with Table 15), so results presented henceforth are based on the entire cohort.

Validation of Vitamin K Antagonist Exposure Classification by the Heart Valve Replacement Proxy Variable

Of the entire cohort, 24,647 (23%) subjects had index dates within the joint temporal and geographic coverage of available prescription drug registries. After stratifying this subset on cancer diagnosis status, there were 2,087 subjects diagnosed with any cancer over follow-up and 22,560 subjects with no cancer diagnosed over follow-up. Table 14 shows the cross-tabulation of heart valve replacement status with VKA prescription status in these two strata. Among subjects with any cancer diagnosed over follow-up, classification of VKA prescription by the heart valve replacement proxy resulted in a positive predictive value (PPV) of 0.956 and a negative predictive value (NPV) of 0.880. Among subjects with no cancer diagnosed over follow-up, classification of VKA prescription by the heart valve replacement proxy resulted in a PPV of 0.972 and a NPV of 0.916. These statistics imply that more than 96% of heart valve recipients were placed on VKA therapy after their surgery, and that approximately 90% of subjects without a heart valve replacement were not placed on VKA therapy after the index date. Figure 3 shows the empirically characterized beta probability distributions for PPV and NPV in the cancer-positive and cancer-negative strata.

Figure 7 and Figure 8 display the trend in PPV and NPV values across years of the validation period. These plots indicate moderate variability of the PPV and

NPV estimates with respect to estimates over the entire period (shown in each plot as a dashed gray line), with no consistent increase or decrease over the validation period. Thus the beta distributions, parameterized with respect to the validation data, should adequately reflect uncertainty in PPV and NPV values over the entire temporal domain of the study, including the earlier time period without prescription data to inform the PPV and NPV estimates.

Associations between Vitamin K Antagonist Therapy and Site-Specific Cancer Incidence in the Validation Subset

Known drug exposure status among subjects in the validation subset permitted direct calculation of associations between VKA therapy and site-specific cancer incidence in this smaller body of data. These estimates, though less precise, are subject to exposure misclassification only to the degree of imperfection in the sensitivity and specificity of the prescription registries for classifying actual internalized VKA exposure. Cancer IRRs comparing VKA-exposed with VKA-unexposed in the validation subset are listed in Table 15; the point estimates and 95% confidence intervals are depicted graphically in Figure 9. There were no cases of liver cancer among the VKA-exposed in the validation subset, so this site was omitted from this analysis.

Of the 23 remaining cancer sites that were analyzed, IRR point estimates ranged from 0.46 to 4.6, and precision of the estimates—defined as the ratio of the upper and lower 95% confidence limits—ranged from 1.8 to 80. Most of the 23 sites showed point estimates near unity with narrow, null-centered 95% confidence intervals. Five of the sites showed modestly elevated incidence rates among the VKA-exposed, compared with the VKA-unexposed, counter to the *a priori* hypothesis of protective effects based on Schulman's observations.[20] Four of these

associations were measured with good precision (prostate cancer IRR: 1.3, 95% CI: 1.0, 1.7; basal cell skin cancer IRR: 1.4, 95% CI: 1.0, 1.8; bladder cancer IRR: 1.7, 95% CI: 1.2, 2.5; colon cancer IRR: 1.7, 95% CI: 1.2, 2.4), and one was measured with poor precision (myelodysplastic syndromes IRR: 4.6, 95% CI: 1.3, 16).

Probabilistic Bias Analysis: Estimated VKA and Cancer Associations in the Entire Cohort

Table 17 shows the median IRRs and 95% simulation intervals for the site-specific cancer associations after probabilistic misclassification adjustment. This analysis included all 23 cancer sites from the validation subset, plus liver cancer. The effect of misclassification adjustment was to positively displace the distribution of IRR estimates, and to modestly widen the intervals about these estimates. Of the 5 cancer sites that appeared to be positively associated with VKA exposure in the validation subset (prostate, basal cell skin, bladder, colon, and myelodysplastic syndromes), 4 also had a positive association suggested by point estimates and intervals calculated in the probabilistic bias analysis (prostate cancer median IRR: 1.3, 95% SI 1.0, 1.6; basal cell skin cancer median IRR: 1.3, 95% SI 1.1, 1.6; colon cancer median IRR: 1.3, 95% SI 1.1, 1.7; bladder cancer median IRR: 1.4, 95% SI 1.1, 1.9).

Empirical Bayes Analysis of the Associations Estimated from the Validation Subset and from the Probabilistic Bias Analysis

Associations from both the validation subset and the probabilistic bias analysis were not compatible with the empirical Bayes estimation algorithm. The algorithm first calculates the observed variance of the log IRRs (Var_{obs}) and the mean of the individual IRR variances across the entire vector of associations (Var_{mean}).[68] The (unknown) true variance, Var_{true}, is initially estimated as the

difference between the observed and mean variances.[67] Since this estimated true variance must be positive, the empirical Bayes method requires that Var_{obs} be greater than Var_{mean}.[68] The estimate of Var_{true} is updated iteratively until a solution is converged upon. The final estimate of Var_{true} is compared with the variance of the individual associations, and this comparison informs the magnitude of the new estimate's displacement toward the mean.

In both the validation subset and probabilistic bias analysis data sets, the calculated mean variance was greater than the calculated observed variance, so there were no solutions for Var_{true}. Despite this limitation to our analysis, it is notable that the ordinary objective of an empirical Bayes analysis is to remove emphasis from spuriously non-null associations. In this study, the totality of the evidence from the three estimation strategies did not suggest any meaningful associations to pursue with further study. It therefore stands to reason that no adjustment for the multiple comparisons is necessary to qualify our inferences.

Instrumental Variable Analysis: Associations between Heart Valve Replacement Receipt and Incidence of Site-Specific Cancers in the Entire Cohort

Table 18 summarizes results from the instrumental variable analysis. The instrumental variable analysis included all 24 cancer sites. IRR point estimates ranged from 0.83 to 1.6. The individual point estimates and confidence intervals under the IV analysis all suggested null associations between vitamin K antagonist therapy and incidence of the specific cancers. In support of this interpretation is the INRP analysis presented in Figure 10. In this plot, the individual IV estimates fall almost perfectly along the line of IRR values predicted under a Gaussian distribution model, and this line approximately intersects the point defined by a null IRR and the

center of the INRP scale. Together, these indicate that the vector of IV associations is consistent with associations drawn at random from a null-centered normal distribution.[76]

DISCUSSION:

Considering our results in aggregate, there appear to be null associations between VKA therapy and the incidence of all 24 site-specific cancers we were able to evaluate.

In the validation subset and the misclassification-adjusted analyses, four cancer sites consistently showed mildly elevated incidence rates in the VKA-exposed compared with the VKA-unexposed, and the IRRs were measured with good precision (prostate, bladder, colon, and basal cell skin cancers). Nevertheless, under the IV analysis, these four measurements—and all other cancer association measurements—appeared null individually and conformed altogether with the expected distribution of values under a null model, as seen in the INRP analysis in Figure 10. This finding reduces the credibility that the elevated associations represent causal effects, and we focus on residual sources of bias as an explanation.

Residual Confounding

Residual confounding may explain the positive associations observed between VKA therapy and prostate, bladder, colon, and basal cell skin cancers. The risk factors for valvular disease (the indication for heart valve replacement) include the general risk factors for coronary heart disease (*i.e.*, age, sex, smoking, body mass index, hypertension, and hypercholesterolemia) and history of endocarditis, congenital heart disease, or rheumatic heart disease.[78, 79] These risk factors must

also be associated with one or more of the 4 site-specific cancers with elevated IRRs in order for confounding to explain the observed increase.

Age, sex, and smoking are expected to be the strongest confounders of all VKA/cancer associations. Age and sex were balanced between exposure groups in the design of our study by matching. The table of baseline characteristics (Table 13) shows very similar proportions of person-time between exposure groups in the age and sex categories. This balance indicates successful matching and yields no expectation of residual confounding by age or sex.

Smoking is a well-established risk factor for cardiovascular disease in general, though its specific association with the incidence and progression of valvular disease has not been well studied. It seems that this association would be positive and of modest magnitude. Smoking has shown modest associations with the risks of bladder,[80, 81] prostate,[82-84] and colon cancer.[85] In our study, no elevation was observed for lung cancer under any of the estimation strategies. Since the cancer-smoking association should be strongest for the lung site, but no positive displacement of the lung cancer IRR was observed, residual confounding by smoking is an unlikely explanation for the observed positive IRRs.

Data on BMI were not available in the Danish registries. Obesity is positively associated with both valvular disease[78] and the incidence of colorectal, bladder and prostate cancer.[85, 86] While a positive association with both the exposure and the outcome could explain the observed non-null IRRs for these cancers (assuming a truly null association), other cancers known to be related to obesity—particularly breast cancer—did not show positive associations. This incongruity, as well as the low prevalence of obesity in Denmark, makes residual confounding by BMI an unlikely explanation for the observed positive associations.

While no evidence has associated hypertension and hypercholesterolemia with cancer incidence, these conditions are sometimes treated by long-term prescription drug therapy that may modify cancer risk (*e.g.*, aspirin, statins, ACE inhibitors, and calcium channel blockers).

No class of antihypertensive drug showed an association with incidence of any cancer in a meta-analysis of 27 randomized trials.[87] One study showed a protective association between use of beta blockers and incidence of prostate cancer.[88] Regardless, a null or protective association between antihypertensive therapy and incidence of cancer would not give rise to the positive confounding that would be necessary to account for our observations.

Hypercholesterolemia is frequently managed with long-term statin therapy. Statin drugs, among other pleiotropic effects, are hypothesized to reduce the incidence of some types of cancer. While some individual epidemiology studies support this hypothesis, meta-analyses indicate null associations.[22, 89, 90] As with antihypertensives, null or protective associations between statins and cancer incidence could not explain the positive associations observed in our study.

Two other drugs—the prescription pharmaceutical cabergoline and the recreational drug 3,4-methylenedioxymethamphetamine (MDMA)—are positively associated with valvular disease,[91] but neither has been associated with cancer incidence.

The remaining risk factors for valve disease (history of endocarditis, congenital heart disease or rheumatic heart disease) also have no demonstrated association with cancer incidence, and were not considered candidate confounders of the estimated VKA/cancer associations.

Selection Bias

This study employed population-based, prospective, and automatically updated medical and civil registries for ascertainment of subjects and all exposure, outcome, and follow-up data. There was no active tracking of the research subjects, so no participation choices could affect loss to follow-up. Losses to follow-up due to emigration from Denmark are expected to occur at random; that is, with no joint dependence on heart valve replacement (or VKA prescription) status and cancer incidence.

However, exposed subjects had a higher mortality rate compared with unexposed subjects, which was initially suggested by the total person-time contributed by each group (52,510 person-years contributed by heart valve recipients and 556,021 person-years contributed by 10:1 matched unexposed) and then confirmed by Kaplan-Meier survival analysis (Figure 11). To examine whether the differential mortality rate generated informative censoring, we created a secondary data set in which unexposed subjects were censored on the last follow-up date of matched exposed subjects who were censored. We compared associations between heart valve replacement and incidence of the site-specific caners in the original and re-censored data sets (Figure 12), and noted similar results under both approaches. Selection bias from differential loss to follow-up is thus an implausible source of inaccuracy in our association estimates.

Information Bias

The primary exposure in our study was receipt of a heart valve replacement, and was ascertained from the Danish National Registry of Patients. The nature of this exposure requires admission to a hospital and receipt of a surgical procedure,

leaving little room for errors in its classification. Our target exposure, however, was vitamin K antagonist therapy. Our probabilistic bias analysis, which addressed misclassification of VKA therapy by the heart valve proxy variable, was only as valid as the classification performance measures (PPV and NPV) that informed it. Imperfect classification of VKA exposure by the prescription registries would then lead to flawed inference about the ability of our proxy variable to capture the actual exposure of interest. The prescription registries log a patient's pharmacy transaction only after they have paid the full price of the drug, after which the patient is reimbursed for a portion of the drug cost by the Danish healthcare system. Thus there is a strong expectation that an entry in the registry means that the patient took possession of the drug. Though possession does not guarantee actual use of the drug, the consequences are potentially dire if a heart valve recipient does not take their anticoagulant, so we expect high compliance among valve recipients with a VKA prescription. Poor prescription compliance among non-recipients of heart valves would only strengthen the contrast between our index and reference conditions.

Cancer incidence should be classified with high specificity in the registries; the most plausible source of classification error comes from imperfect sensitivity. If this were the case, the classification error would almost certainly be non-differential and independent since the prescription data were recorded prospectively (before the occurrence of cancer in any subject) and were contained in a wholly independent register than that used to ascertain the cancer diagnoses. Misclassification of the outcome with near perfect specificity and imperfect yet non-differential and independent sensitivity is not expected to bias our estimated incident rate ratios appreciably, since our cancer outcomes are quite rare.[92]

Summary

We estimated associations between VKA therapy and the incidence of 24 different site-specific cancers using three estimation strategies. The first strategy examined direct associations between VKA prescription and incidence of the cancers in a validation subset. The second strategy used receipt of a heart valve replacement as a proxy variable for VKA exposure in an age- and sex-matched cohort, with probabilistic adjustment for misclassification of VKA prescription by the heart valve proxy. The third strategy recognized heart valve replacement as a potential instrumental variable for the association between VKA therapy and the incidence of any cancer.

Treatment with a vitamin K antagonist showed positive associations with the incidence of prostate, bladder, colon, and basal cell skin cancers in both the validation subset and probabilistic bias analyses. No cancer site showed an association (positive or negative) under the instrumental variable analysis. Of the three analytic approaches, the validation subset analysis likely best characterizes true exposure to a VKA; however age and sex are the only two candidate confounders for which the estimated VKA/cancer associations were controlled. The probabilistic bias analysis allowed for a longer study period and accrual of more cancer events, thus decreasing random error, but likely harbors residual exposure misclassification in addition to controlling only for the candidate confounders age and sex. Our first two analyses are therefore threatened by residual confounding, and each suffers from exposure misclassification in its own way. In the validation subset, prescription for a VKA does not necessarily mean that an individual was compliant and took their drug. In the probabilistic bias analysis, the misclassification adjustment brings the misclassified heart valve proxy as close as possible to

representing actual VKA prescription status, but this adjustment is likely incomplete and, even under the best of circumstances, would still require the qualification that having a prescription does not necessarily equal an internalized exposure. The instrumental variable analysis, however, is not only adjusted for confounding by age and sex, but by other potential confounders—measured and unmeasured—if the IV requirements are indeed satisfied. Use of the validation data to multiply impute VKA status from the heart valve replacement instrumental variable also adjusted for misclassification of the target exposure. Therefore, the IV analysis consolidates adjustment for known and unknown confounders with adjustment for exposure misclassification by the instrument—with the same qualification regarding prescription and actual exposure to a drug. For these reasons we consider the IV analysis to be the most valid of our three estimation strategies, and this analysis clearly depicts null associations between vitamin K antagonist therapy and the incidence of the 24 site-specific cancers we were able to examine (Table 18 and Figure 10).

It should be noted that, even if the observed non-null observations from the validation subset and probabilistic bias analyses were not the result of residual bias or random error, their direction is opposite to that observed by Schulman and Lindmarker (causal, not protective), and do not support the proffered hypothesis that VKA therapy protects against cancer incidence.

Table 6: Procedure codes used to ascertain biological and artificial heart valve replacement procedures in the Danish National Registry of Patients.

Surgical procedure	Procedure codes
Aortic valve replacement	KFMD, 31268, 31269
Mitral valve replacement	KFKD, 31129, 31130
Pulmonic valve replacement	KFJF, 30959
Tricuspid valve replacement	KFGE, 30729

Table 7: ICD-10 codes used to ascertain incident site-specific cancer diagnoses in the Danish Cancer Registry.

Cancer site	ICD-10 code(s)
Lip	C00.x
Tongue	C01.x – C02.x
Oral cavity	C02.x – C06.x
Salivary gland	C07.x – C08.x
Tonsils and mouth	C09.x – C10.x
Nasal	C11.x
Esophagus	C15.x
Stomach	C16.x
Small intestine	C17.x
Colon	C18.x – C19.x
Rectum	C20.x – C21.x
Liver	C22.x
Gallbladder & bile ducts	C23.x – C24.x
Pancreas	C25.x
Larynx	C32.x
Lungs, bronchi & trachea	C33.x – C34.x
Thymus	C37.x
Cardiac/thoracic	C38.1 – C38.3, C38.8
Pleural	C38.4, C45.0
Bones & articular cartilage	C40.x – C41.x
Mesothelioma	C45.1 – C45.9
Kaposi sarcoma	C46.x, B21.0
Peripheral nerves	C47.x
Peritoneum	C48.x
Breast	C50.x
Cervix	C53.x
Uterus	C54.x – C55.x
Ovary	C56.x, C57.0 – C57.4
Placenta	C58.x
Penis	C60.x
Prostate	C61.x
Testicle	C62.x
Kidney	C64.x
Renal pelvis	C65.x, D30.1, D41.1
Urinary bladder	C67.x, D09.0, D30.3, D41.4
Eye	C69.x
Brain	C71.x, D33.x, D35.2 – D35.4, D43.x, D44.3 – D44.5
Spinal cord & cranial nerves	C72.x
Thyroid	C73.x
Adrenal gland	C74.x

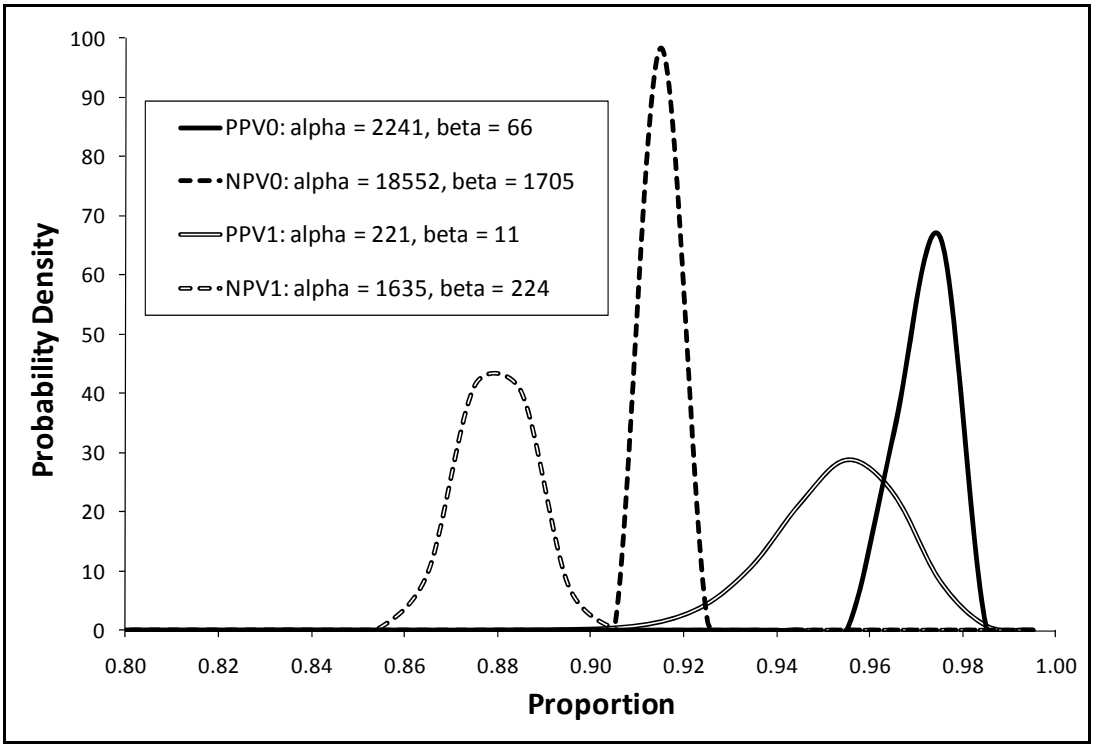
Cancer site	ICD-10 code(s)
Hodgkin's malignant lymphoma Non-Hodgkin's malignant lymphoma Malignant myeloproliferative disease	C81.x C82.x – C85.x, C90.x C88.x
Leukemia (lymphocytic) Leukemia (myeloid)	C91.x C92.x
Basal cell skin cancer Non-basal cell skin cancer	D44.x (Various site codes with morphology code 809)
Polycythemia vera Myelodysplastic syndromes	D45.x D46.x

Table 8: ICD-8 and ICD-10 codes used to ascertain past medical history variables from the Danish National Registry of Patients.

Diagnosis	ICD-8 code(s)	ICD-10 code(s)
Charlson comorbidities		
Myocardial infarction	410.x	I21 – I23
Congestive heart failure	427.09 – 427.11, 427.19, 428.99, 782.49	I50.x, I11.0, I13.0, I13.2
Peripheral vascular disease	440 – 445	I70.x – I74.x, I77.x
Cerebrovascular disease	430 – 438	I60.x – I69.x, G45.x – G46.x
Dementia	290.09 – 290.19, 293.09	F00.x – F03.x, F05.1, G30.x
Chronic pulmonary disease	490 – 493, 515 – 518	J40.x – J47.x, J60.x – J67.x, J68.4, J70.1, J70.3, J84.1, J92.0, J96.1, J98.2, J98.3
Connective tissue disease	712, 716, 734, 446, 135.99	M05.x, M06.x, M08.x, M09.x, M30.x – M36.x, D86.x
Ulcer disease	530.91, 530.98, 531 - 534	K22.1, K25.x – K28.x
Mild liver disease	571, 573.01, 573.04	B18.x K70.0 – K70.4, K70.9, K71.x, K73.x, K74.x, K76.0
Diabetes (Type I)	249.00, 249.06, 249.07, 249.09	E10.0, E10.1, E10.9
Diabetes (Type II)	250.00, 250.06, 250.07, 250.09	E11.0, E11.1, E11.9
Hemiplegia	344	G81.x – G82.x
Moderate/severe renal disease	403, 404, 580 – 584, 590.09, 593.19, 753.10 – 753.19, 792	I12.x, I13.x, N00.x – N05.x, N07.x, N11.x, N14.x, N17.x – N19.x, Q61.x
Diabetes with complications	249.01 – 249.05, 249.08	E10.2 – E10.8, E11.2 – E11.8

Any tumor	140 - 194	C00.x – C75.x
Leukemia	204 - 207	C91.x – C95.x
Lymphoma	200 – 203, 275.59	C81.x – C85.x, C88.x, C90.x, C96.x
Moderate/severe liver disease	070.00, 070.03, 070.04, 070.06, 070.08, 573.00, 456.00 – 456.09	B15.0, B16.0, B16.2, B19.0, K70.4, K72.x, K76.6, I85.x
Metastatic solid tumor	195 – 198, 199.x	C76.x – C80.x
AIDS	079.83	B21.x – B24.x
Venous thromboembolic events (VTE)		
Pulmonary embolism	450.x	I26.x
Portal vein thrombosis	452.x	I81.x
Other VTE	453.x	I82.x
Atrial fibrillation	427.4, 427.9	I48.x

Figure 3: Beta probability density functions defined for PPV and NPV for VKA exposure proxy by heart valve receipt, by cancer diagnosis status.



PPV0, NPV0: Predictive values among subjects with no incident cancer during follow-up. **PPV1, NPV1:** Predictive values among subjects with any incident cancer during follow-up.

Figure 4: Algorithm for probabilistic bias analysis using positive and negative predictive values to adjust associations between VKA exposure and site-specific cancer incidence for misclassification by the heart valve replacement proxy variable.

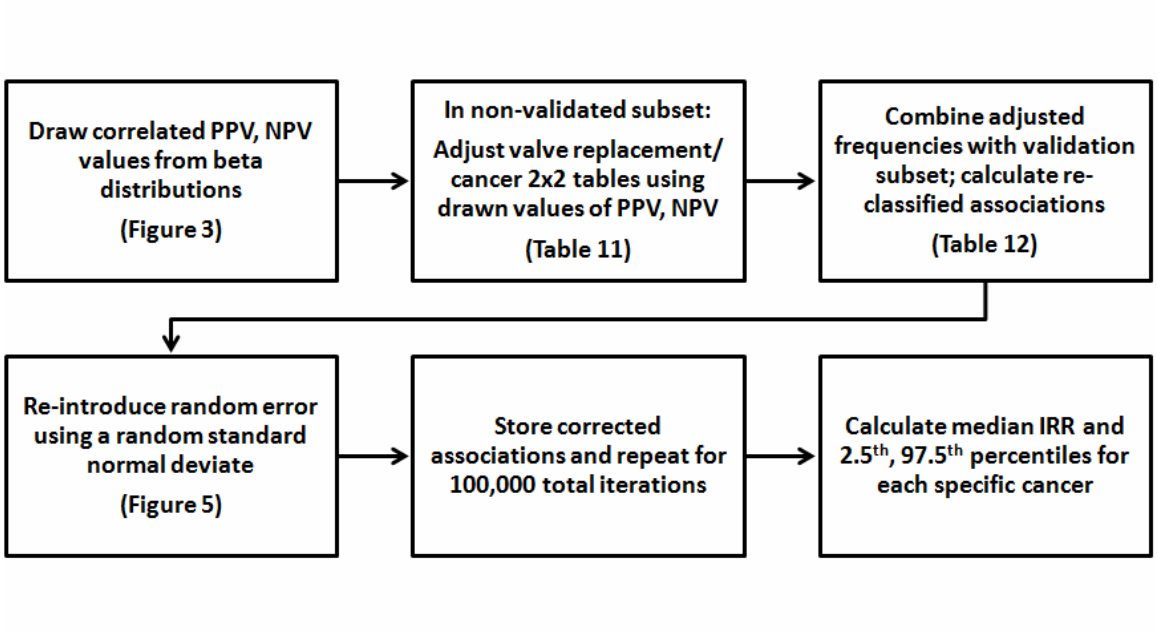


Table 9: Data arrangement for the j^{th} site-specific cancer association with known VKA exposure data in the prescription validation subset.

(Validated)	VKA +	VKA -
Cancer cases	$a_{val,j}$	$b_{val,j}$
Person-years	$PT1_{val,j}$	$PT0_{val,j}$

Table 10: Data arrangement for the j^{th} site-specific cancer association subject to VKA exposure misclassification under the heart valve replacement proxy.

(Non-validated subset)	Heart valve +	Heart valve -
Cancer cases	$a_{mis,j}$	$b_{mis,j}$
Person-years	$PT1_{mis,j}$	$PT0_{mis,j}$

Table 11: Calculation of adjusted cell values for the j^{th} site-specific cancer association in the i^{th} iteration of the probabilistic bias analysis routine, using observed values in the non-validated subset (Table 10) and randomly drawn values of PPV and NPV from empirically characterized beta distributions.

(Adjusted)	Classified VKA +	Classified VKA -
Cancer cases	$a_{adj,ij} = a_{mis,j}(PPV1_i) + b_{mis,j}(1-NPV1_i)$	$b_{adj,ij} = b_{mis,j}(NPV1_i) + a_{mis,j}(1-PPV1_i)$
Person-years	$PT1_{adj,ij} = PT1_{mis,j}(PPV0_i) + PT0_{mis,j}(1-NPV0_i)$	$PT0_{adj,ij} = PT0_{mis,j}(NPV0_i) + PT1_{mis,j}(1-PPV0_i)$

Table 12: Combination of observed data from the validation subset and re-classified data from the i^{th} iteration of the probabilistic bias analysis routine for the j^{th} site-specific cancer.

(Combined)	Exposed	Unexposed
Cancer cases	a_{ij} = $a_{val,j} + a_{adj,ij}$	b_{ij} = $b_{val,j} + b_{adj,ij}$
Person-years	$PT1_{ij}$ = $PT1_{val,j} + PT1_{adj,ij}$	$PT0_{ij}$ = $PT0_{val,j} + PT0_{adj,ij}$

Figure 5: Equation for calculation of the cancer incidence rate ratio comparing those classified as VKA-exposed with those classified as VKA-unexposed, with re-incorporation of random error, for the j^{th} site-specific cancer at the i^{th} iteration of the probabilistic bias analysis routine (Z_{ij} is a random standard normal deviate).

$$IRR_{ij} = e^{\ln\left(\frac{a_{ij}/PT1_{ij}}{b_{ij}/PT0_{ij}}\right) - \left(Z_{ij} \times \sqrt{1/a_{ij} + 1/b_{ij}}\right)}$$

Figure 6: Directed acyclic graph (DAG) depicting necessary conditions for heart valve replacement to serve as an instrumental variable (IV) for exposure to a vitamin K antagonist.

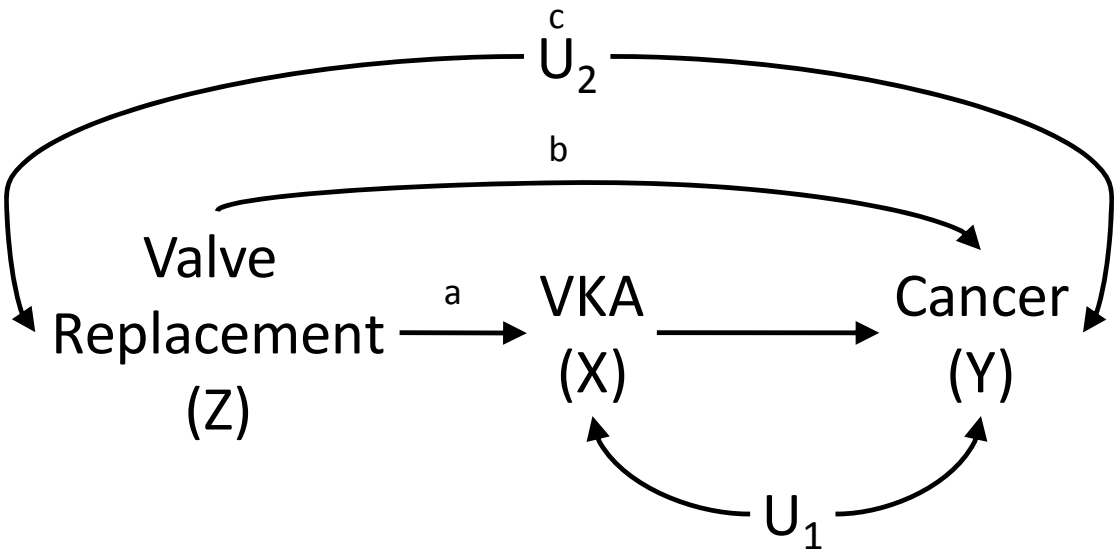


Table 13: Baseline characteristics of the matched cohort.

Variable	Valve recipients (n=9,727)		Matched unexposed (n=95,481)	
	n	Person-years (%)	n	Person-years (%)
Sex				
Male	6,024	32,293 (61)	58,953	337,992 (61)
Female	3,703	20,217 (39)	36,528	218,029 (39)
Age on index date				
< 18	176	1,186 (2.3)	1,727	11,845 (2.1)
18-24	88	742 (1.4)	930	8,108 (1.5)
25-34	260	2,022 (3.9)	2,577	20,859 (3.8)
35-44	520	3,698 (7.0)	5,221	40,615 (7.3)
45-54	1,142	7,747 (15)	11,409	84,893 (15)
55-64	2,247	13,140 (25)	22,475	144,861 (26)
65-74	3,317	16,763 (32)	32,449	177,318 (32)
75-84	1,885	6,956 (13)	17,819	65,551 (12)
≥ 85	92	257 (0.5)	874	1,972 (0.4)
Charlson Comorbidity Index [a]				
0	4,932	29,817 (57)	75,171	471,516 (85)
1	4,010	19,842 (38)	17,825	77,264 (14)
2	664	2,540 (4.8)	2,136	6,461 (1.2)
3	121	313 (0.6)	349	780 (0.1)
History of atrial fibrillation/flutter				
Positive	1,183	4,513 (8.6)	2,070	5,378 (1.0)
Negative	8,544	47,998 (91)	93,411	550,642 (99)
History of VTE [b]				
Positive	101	508 (1.0)	487	1,917 (0.3)
Negative	9,626	52,003 (99)	94,994	554,103 (99.7)

[a] Charlson Comorbidity Index is classified by the sum of weighted prevalent medical conditions. The diagnoses, weights, and ordinal index classification are described in detail elsewhere.[59, 93]

[b] VTE: Venous thromboembolism.

Table 14: Validation of the heart valve replacement proxy for VKA exposure in the subset of subjects for whom both prescription and hospital discharge data were available (n = 24,647).

	Any cancer diagnosis		No cancer diagnosis	
Valve replacement history:	**VKA+**	**VKA-**	**VKA+**	**VKA-**
Positive	220	10	2,240	65
Negative	223	1,634	1,704	18,551
Positive predictive value (PPV):	220/230 = 0.956		2,240/2,305 = 0.972	
Negative predictive value (NPV):	1,634/ 1,857 = 0.880		18,551/20,255 = 0.916	

Figure 7: Trend in positive predictive value (PPV) estimates by year of the prescription validation period.

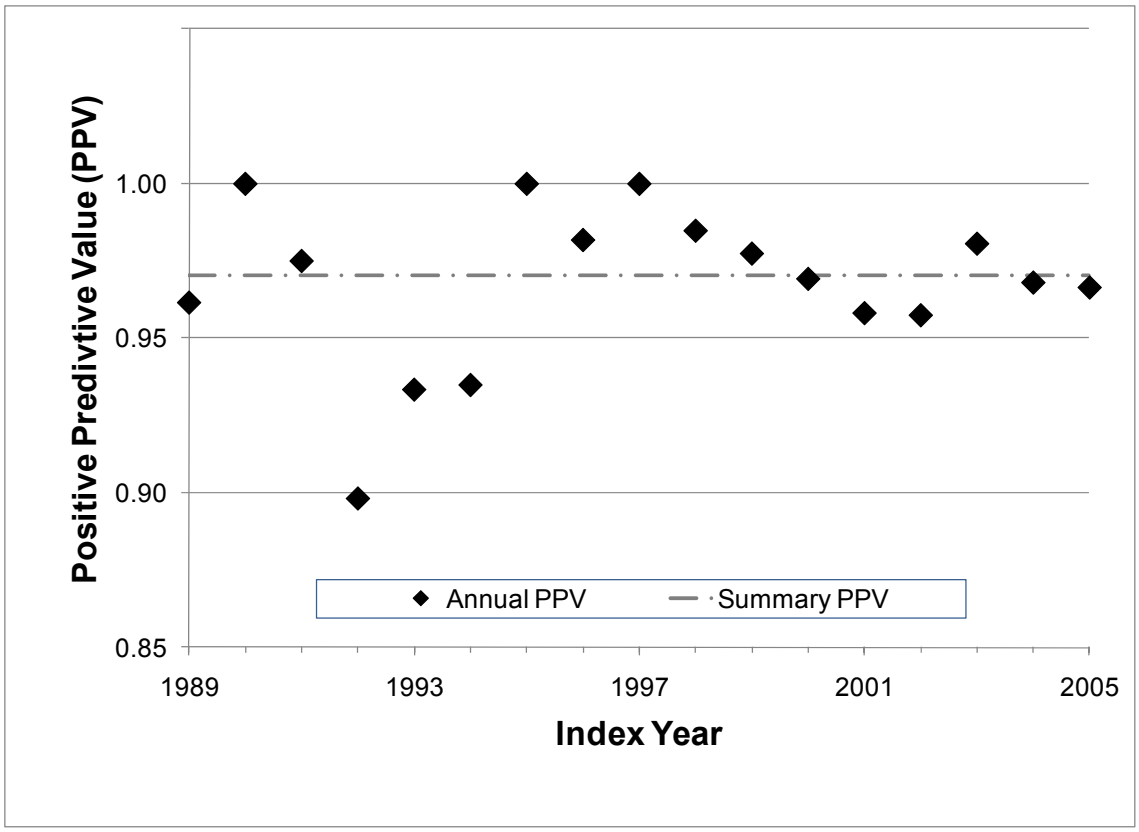


Figure 8: Trend in negative predictive value (NPV) estimates by year of the prescription validation period.

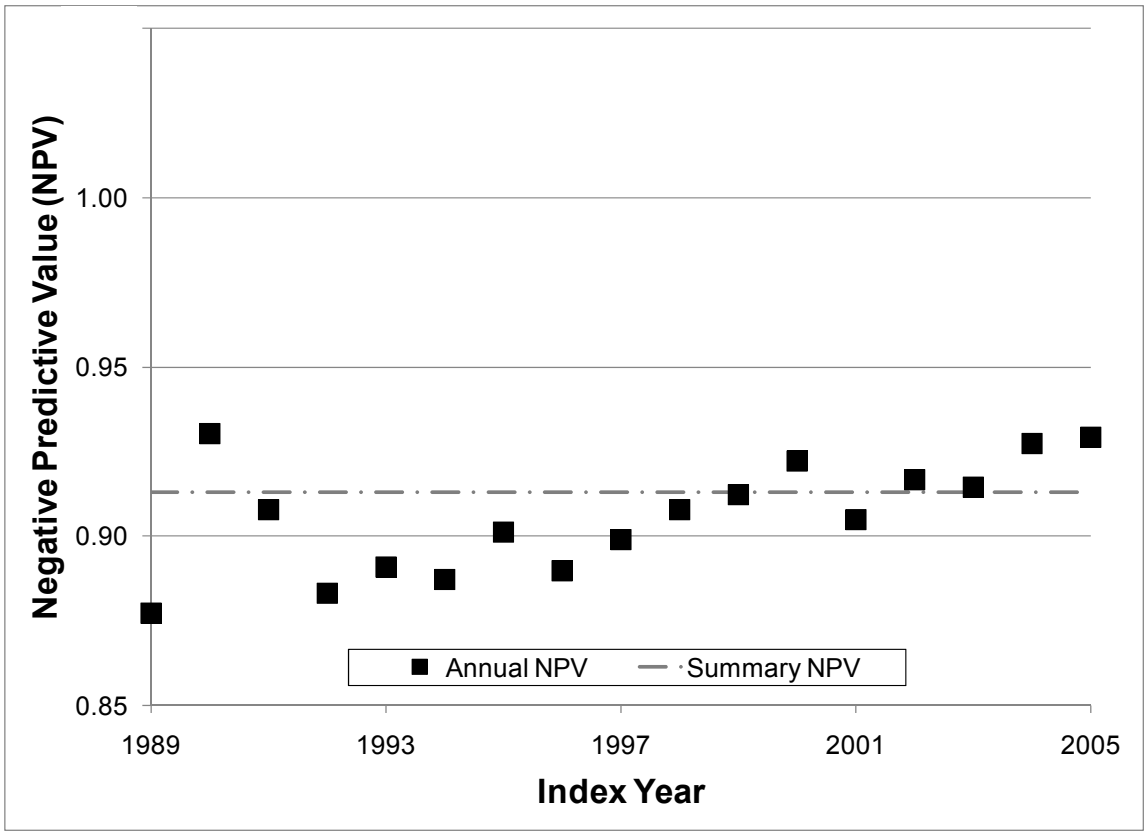


Table 15: Associations between VKA therapy and site-specific cancer incidence in the validation subset (n = 24,647), with rows ordered by rank of IRR point estimate.

Cancer site	VKA Exposed: Cases/ Person-years	VKA Unexposed: Cases/ Person-years	IRR (95% CI)	Precision [a]
Esophagus	2/ 18526.2	20/ 85465.1	0.46 (0.11, 2.0)	18
Brain	4/ 18526.8	25/ 85473.5	0.74 (0.26, 2.1)	8.3
Skin, non-basal cell	9/ 18583.2	52/ 85582.3	0.80 (0.39, 1.6)	4.1
Non-Hodgkin's lymphoma	10/ 18543.4	53/ 85549.8	0.87 (0.44, 1.7)	2.7
Breast (female)	18/ 7053.5	93/ 33619.1	0.92 (0.56, 1.5)	3.9
Lungs	50/ 18735.7	211/ 86181.9	1.1 (0.80, 1.5)	1.9
Rectum	20/ 18584.8	84/ 85683.3	1.1 (0.67, 1.8)	3.7
Pancreas	11/ 18562.8	45/ 85515.1	1.1 (0.58, 2.2)	2.7
Gallbladder	1/ 18518.6	4/ 85385.5	1.2 (0.13, 10)	80
Leukemia, myeloid	6/ 18529.9	23/ 85465.1	1.2 (0.49, 3.0)	6.0
Prostate	57/ 11769.0	198/ 52989.7	1.3 (1.0, 1.7)	1.8
Ovary	7/ 7003.4	25/ 33398.0	1.3 (0.58, 3.1)	5.4
Skin, basal cell	63/ 18737.6	210/ 86141.3	1.4 (1.0, 1.8)	1.8
Uterus	7/ 7012.7	24/ 33370.5	1.4 (0.60, 3.2)	5.4
Kidney	7/ 18537.7	23/ 85460.0	1.4 (0.60, 3.3)	5.4
Larynx	4/ 18529.6	13/ 85425.7	1.4 (0.46, 4.4)	9.4
Oral cavity	4/ 18532.4	12/ 85440.4	1.5 (0.50, 4.8)	9.6
Stomach	10/ 18565.3	29/ 85505.9	1.6 (0.77, 3.3)	4.2
Colon	44/ 18684.3	119/ 85943.9	1.7 (1.2, 2.4)	2.0

Cancer site	VKA Exposed: Cases/ Person-years	VKA Unexposed: Cases/ Person-years	IRR (95% CI)	Precision [a]
Bladder	40/ 18672.0	106/ 85717.5	1.7 (1.2, 2.5)	2.1
Leukemia, lymphoid	6/ 18532.7	15/ 85432.9	1.8 (0.72, 4.8)	6.6
Cervix	4/ 6982.8	8/ 33304.5	2.4 (0.72, 7.9)	11
Myelodysplastic syndromes	5/ 18552.1	5/ 85391.3	4.6 (1.3, 16)	12

[a] Precision is calculated as the ratio of the upper and lower 95% confidence limits.

Table 16: Associations between VKA therapy and site-specific cancer incidence in the validation subset, restricted to subjects with no history of venous thromboembolism.

Cancer site	VKA Exposed: Cases/ Person-years	VKA Unexposed: Cases/ Person-years	IRR (95% CI)
Esophagus	2/ 18336.8	20/ 85271.7	0.47 (0.11, 2.0)
Brain	4/ 18337.4	25/ 85280.1	0.74 (0.26, 2.1)
Skin, non-basal cell	9/ 18393.9	52/ 85388.9	0.80 (0.40, 1.6)
Breast (female)	18/ 6986.8	92/ 33490.1	0.94 (0.57, 1.6)
Non-Hodgkin's lymphoma	10/ 18354.0	53/ 85356.4	0.88 (0.45, 1.7)
Lungs	48/ 18524.0	211/ 85988.5	1.1 (0.77, 1.4)
Pancreas	11/ 18373.4	45/ 85321.7	1.1 (0.59, 2.2)
Rectum	20/ 18395.4	84/ 85489.9	1.1 (0.68, 1.8)
Gallbladder	1/ 18329.2	4/ 85192.1	1.2 (0.13, 10)
Leukemia, myeloid	6/ 18340.6	23/ 85271.7	1.2 (0.49, 3.0)
Ovary	7/ 6936.7	25/ 33269.2	1.3 (0.58, 3.1)
Prostate	56/ 11646.2	198/ 52925.2	1.3 (1.0, 1.7)
Kidney	7/ 18348.3	23/ 85266.6	1.4 (0.61, 3.3)
Larynx	4/ 18340.2	13/ 85232.3	1.4 (0.47, 4.4)
Skin, basal cell	62/ 18548.0	208/ 85944.5	1.4 (1.0, 1.8)
Uterus	7/ 6946.0	24/ 33241.6	1.4 (0.60, 3.2)
Oral cavity	4/ 18343.0	12/ 85247.0	1.6 (0.50, 4.8)
Stomach	10/ 18375.9	29/ 85312.5	1.6 (0.78, 3.3)
Bladder	40/ 18482.7	106/ 85524.1	1.8 (1.2, 2.5)

Cancer site	VKA Exposed: Cases/ Person-years	VKA Unexposed: Cases/ Person-years	IRR (95% CI)
Colon	43/ 18492.8	118/ 85737.2	1.7 (1.2, 2.4)
Leukemia, lymphoid	6/ 18343.3	15/ 85239.5	1.9 (0.72, 4.8)
Cervix	4/ 6916.1	8/ 33175.7	2.4 (0.72, 8.0)
Myelodysplastic syndromes	5/ 18362.7	5/ 85197.9	4.6 (1.3, 16)

Figure 9: Plot of the associations between VKA therapy and site-specific cancer incidence in the validation subset, ordered left to right according to rank.

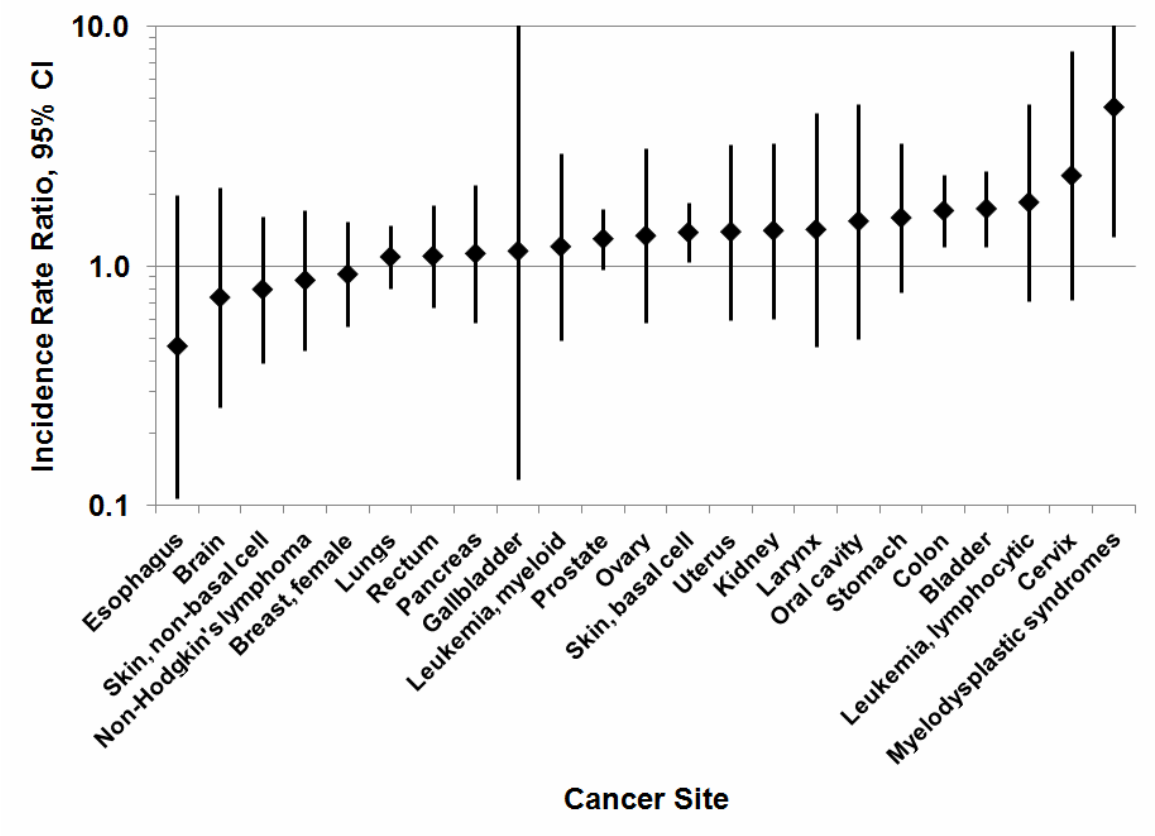


Table 17: Site-specific cancer associations following probabilistic bias analysis to correct VKA exposure misclassification by the heart valve proxy: median incidence rate ratios and 95% simulation intervals from distributions of 100,000 adjusted estimates. Listed by rank.

Cancer Site	Median IRR	95% SI
Pancreas	1.0	0.57, 1.9
Leukemia, lymphocytic	1.1	0.39, 2.9
Brain	1.1	0.57, 2.1
Ovary	1.2	0.52, 2.7
Larynx	1.2	0.43, 3.2
Lungs	1.2	0.96, 1.5
Gallbladder	1.2	0.44, 3.4
Rectum	1.2	0.86, 1.8
Uterus	1.2	0.65, 2.4
Esophagus	1.3	0.74, 2.1
Prostate	1.3	1.0, 1.6
Leukemia, myeloid	1.3	0.51, 3.2
Breast	1.3	0.99, 1.7
Non-Hodgkin's lymphoma	1.3	0.91, 1.9
Kidney	1.3	0.75, 2.4
Skin, basal cell	1.3	1.1, 1.6
Colon	1.3	1.1, 1.7
Skin, non-basal cell	1.4	1.0, 2.0
Bladder	1.4	1.1, 1.9
Stomach	1.5	0.90, 2.4
Cervix	1.6	0.62, 3.9
Liver	1.6	0.82, 3.0
Myelodysplastic syndromes	1.6	0.50, 5.2
Oral cavity	1.7	0.70, 4.0

Table 18: Instrumental variable analysis: Associations between VKA therapy and site-specific cancer incidence, estimated by using receipt of a heart valve replacement as an instrumental variable.

Cancer Site	IRR (95% CI)	Inverse Variance	Rank	INRP	Modeled IRR [a]
Leukemia, lymphocytic	0.83 (0.48, 1.4)	13	1	-1.95	0.82
Pancreas	0.84 (0.55, 1.3)	22	2	-1.50	0.88
Brain	0.91 (0.51, 1.6)	13	3	-1.24	0.91
Larynx	0.93 (0.47, 1.8)	9	4	-1.04	0.94
Ovary	0.95 (0.51, 1.8)	11	5	-0.88	0.97
Gallbladder	0.99 (0.44, 2.3)	6	6	-0.73	0.99
Rectum	1.0 (0.75, 1.4)	42	7	-0.60	1.0
Lung	1.0 (0.87, 1.3)	121	8	-0.48	1.0
Uterus	1.1 (0.68, 1.7)	19	9	-0.37	1.1
Breast	1.1 (0.88, 1.3)	87	10	-0.26	1.1
Prostate	1.1 (0.92, 1.3)	142	11	-0.16	1.1
Liver	1.1 (0.51, 2.4)	7	12	-0.05	1.1
Esophagus	1.1 (0.66, 1.9)	16	13	0.05	1.1
Kidney	1.1 (0.70, 1.8)	18	14	0.16	1.1
Colon	1.2 (0.90, 1.5)	73	15	0.26	1.2
Skin, basal	1.2 (1.0, 1.4)	183	16	0.37	1.2
Non-Hodgkin's lymphoma	1.2 (0.83, 1.7)	32	17	0.48	1.2
Leukemia, myeloid	1.2 (0.65, 2.3)	10	18	0.60	1.2
Stomach	1.3 (0.86, 1.9)	27	19	0.73	1.3

Cancer Site	IRR (95% CI)	Inverse Variance	Rank	INRP	Modeled IRR [a]
Bladder	1.3 (1.0, 1.6)	80	20	0.88	1.3
Skin, non-basal	1.3 (0.96, 1.8)	44	21	1.04	1.3
Cervix	1.3 (0.68, 2.6)	9	22	1.24	1.4
Oral cavity	1.5 (0.75, 3.0)	8	23	1.50	1.4
Myelodysplastic syndromes	1.6 (0.60, 4.3)	5	24	1.95	1.5

[a] Modeled IRR was calculated with the following equation, estimated from the inverse-variance weighted regression of log observed IRRs on the inverse normal of rank percentile: $IRR_{null} = EXP(0.107 + 0.159*INRP)$

Figure 10: Plot of the associations between VKA therapy and the incidence of site-specific cancers according to inverse normal of rank percentile: heart valve replacement as an instrumental variable for exposure to a vitamin K antagonist. Cancer sites are ordered left to right according to the list in Table 18.

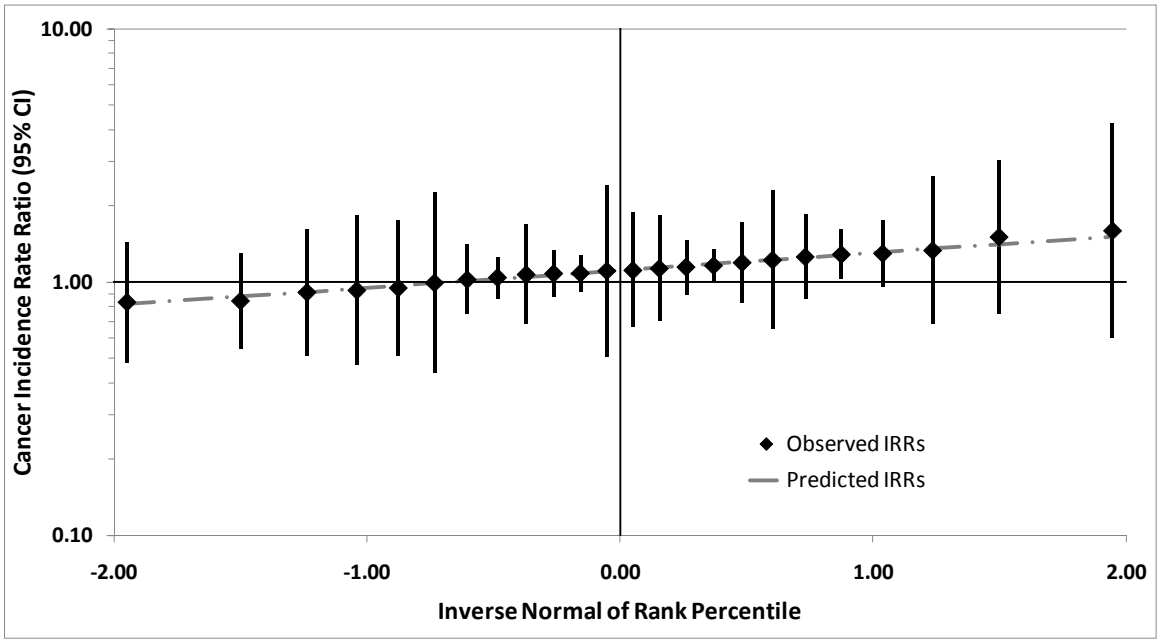


Figure 11: Kaplan-Meier survival curves for total mortality, stratified by heart valve replacement status.

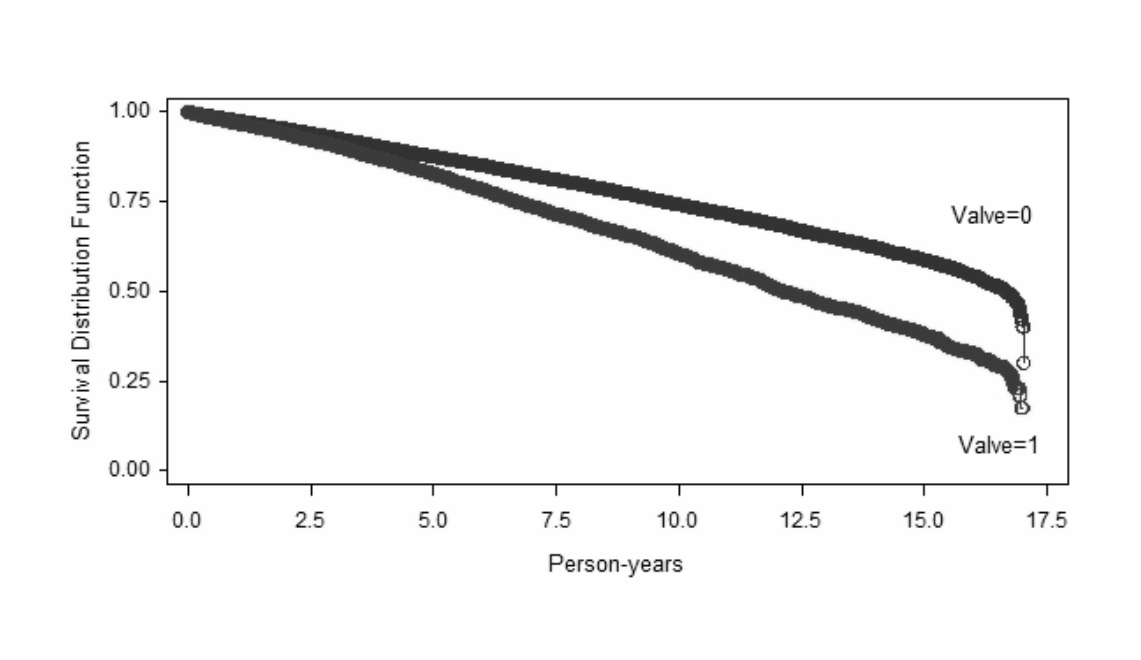
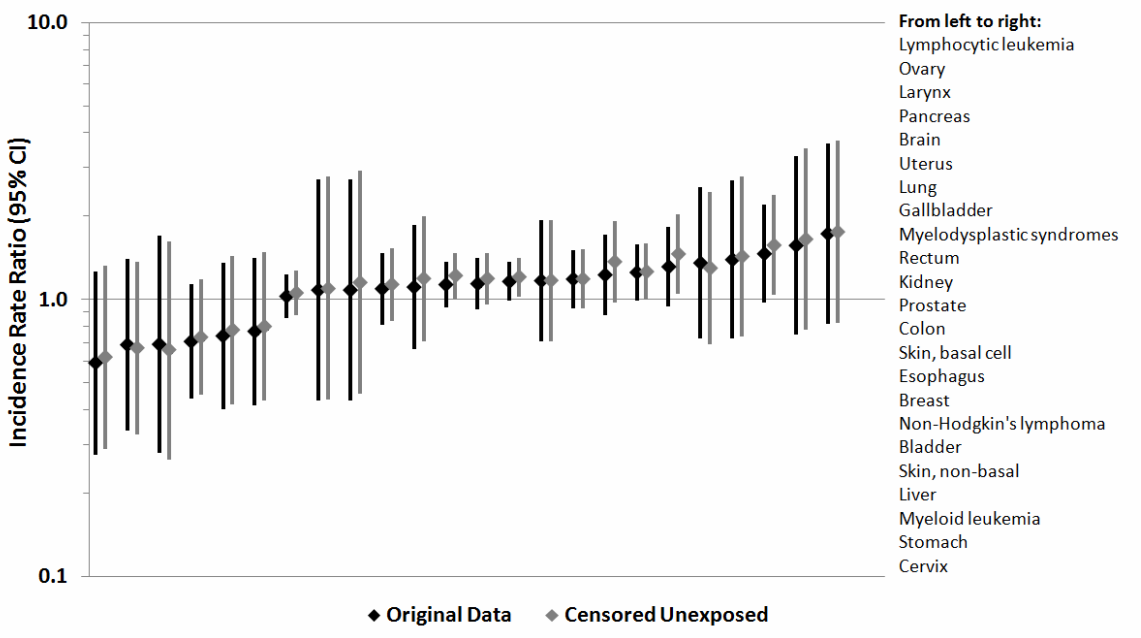


Figure 12: Comparison of the associations between heart valve replacement and site-specific cancer incidence in the original data, and in a secondary data set with re-censored unexposed subjects whose matched exposed subjects were lost to follow-up.



POLYMORPHISMS IN THE UDP-GLUCURONOSYL TRANSFERASES AND BREAST CANCER RECURRENCE IN TAMOXIFEN-TREATED WOMEN

BACKGROUND:

Tamoxifen (TAM) is a selective estrogen receptor modulator (SERM) that binds the estrogen receptor (ER) and inhibits breast cancer cell growth stimulation by estradiol (E2).[94] Cytochrome P450 enzymes (CYPs) oxidize tamoxifen and N-desmethyl tamoxifen to 4-OH-tamoxifen (4HT) and 4-OH-N-desmethyl tamoxifen (endoxifen), respectively, which bind the ER with approximately 100-fold higher affinity than tamoxifen itself. Tamoxifen's intended therapeutic effect is potentiated by cytochrome P450 (CYP)-mediated conversion to its more active metabolites, 4-OH-tamoxifen (4HT) and 4-OH-N-desmethyl-tamoxifen (endoxifen) which exert most of the anti-estrogenic effect.[95] Phase II metabolic enzymes catalyze the addition of polar moieties to the active tamoxifen metabolites, enhancing their water solubility and promoting their excretion. One such set of enzymes is the UDP-glucuronosyltransferase family (UGTs)—hepatic enzymes that convert 4-OH-tamoxifen and endoxifen into tamoxifen glucuronides.[96] Glucuronidation is the chief mechanism by which active tamoxifen metabolites are eliminated from the body,[97] but a minor elimination pathway proceeds *via* sulfonation reactions catalyzed by *SULT1A1*, a polymorphic enzyme in the sulfotransferase family.[98] The main phase I and II reactions of tamoxifen metabolism are summarized in Figure 13.

Four members of the UGT family—*UGT2B15*, *UGT1A8*, *UGT2B7* and *UGT1A10*—are chiefly responsible for the glucuronidation of tamoxifen.[96] Three of these UGTs (*UGT2B15*, *UGT1A8* and *UGT2B7*) harbor single nucleotide polymorphisms (SNPs), with variant alleles yielding phenotypes that have altered

enzyme functionality. *UGT2B15* has wild-type (*1) and variant (*2) alleles, with the variant encoding an enzyme that has a roughly two-fold higher rate of catalysis than the wild-type.[99] *UGT1A8* has wild-type (*1) and two variant (*2 and *3) alleles. The *UGT1A8**2 allele produces an enzyme with catalytic activity similar to that of the wild-type, and the *3 allele—though it completely eradicates glucuronidation activity—is very uncommon in the population.[100, 101] *UGT2B7* has wild-type (*1) and variant (*2) alleles, and the variant allele encodes an enzyme with reduced activity.[101]

Consequently, of the genes involved in the conjugation and elimination of active tamoxifen metabolites, polymorphisms in *UGT2B15* and *UGT2B7* have the greatest potential to alter elimination rates of 4HT and endoxifen, either prolonging or reducing the *in vivo* availability of these metabolites and potentially modifying the effectiveness of tamoxifen therapy.

METHODS:

This study was approved by the Regional Committee on Biomedical Research Ethics of Aarhus County, Denmark, and by the Boston University Medical Campus Institutional Review Board.

Study Population

We studied the association between UGT variants and breast cancer recurrence in tamoxifen-treated women using a population-based case-control design. The source population for cases and controls was all women aged 35 to 69 years who were diagnosed with UICC Stage I, II or III primary breast cancer between 1985 and 2001 in one of seven Danish counties (Aarhus, North Jutland, Viborg, Ringkøbing, South Jutland, Vejle, or Ribe) and who were reported to the

Danish Breast Cancer Cooperative Group (DBCG) registry.[102] (The included counties are now in the North, Mid, and South regions of Denmark.) The study population was divided into two groups; (1) women whose primary tumors expressed the estrogen receptor and who were treated with tamoxifen for at least one year (ER+/Tam+), and (2) women whose primary tumors did not express the estrogen receptor, were not treated with tamoxifen, and who survived at least one year after breast cancer diagnosis (ER-/Tam-). The latter group was included in the study to estimate the direct effect (*i.e.*, not mediated through the tamoxifen pathway), if any, of the UGT variants on breast cancer recurrence risk. Women not meeting either of the group definitions were excluded from the study.

Cases were those women from the source population who experienced a breast cancer recurrence after at least one year of tamoxifen treatment (if in the ER+/Tam+ group) or who survived at least one year after breast cancer diagnosis (if in the ER-/Tam- group).

Eligible controls were those women from the source population who had not had a breast cancer recurrence after the same amount of follow-up time as each index case.

Registry Data Collection

Cases of breast cancer recurrence were identified in the DBCG registry. Risk sets were enumerated for each case, conditional on the following matching factors: (a) group membership (ER+/Tam+ or ER-/Tam-), (b) menopausal status at diagnosis (premenopausal or postmenopausal), (c) date of breast cancer surgery (caliper matched ± 12 months), (d) county of residence at diagnosis, and (e) UICC stage at diagnosis (stage I, II, or III). One control was sampled at random from the

risk set of each case. The risk-set sampling of controls results in a case-control odds ratio that estimates the breast cancer recurrence rate ratio.[103]

The recurrence risk period began after at least one year of tamoxifen treatment (for cases and controls in ER+/Tam+) or after survival for one year after breast cancer diagnosis (for cases and controls in ER-/Tam-). Follow up ended with the first of breast cancer recurrence, death from any cause, emigration from Denmark, or 1 September 2006.

Breast cancer and demographic data (age, menopausal status, hospital of diagnosis, UICC stage, histologic grade, estrogen receptor expression, primary surgery type, receipt of radiation therapy, receipt of chemotherapy, and receipt of tamoxifen therapy) were ascertained from the DBCG database for all cases and controls.

Tissue Processing

We used the CPR numbers of cases and controls to identify and retrieve archived tumors that were resected during primary therapy and stored in the pathology archives of treating hospitals. Hematoxylin and eosin stained sections or written pathological descriptions of the tumor blocks were reviewed by a pathologist to identify appropriate blocks for processing. Tissue blocks were manipulated in a laminar flow hood that had been sterilized before use with at least one hour of ultraviolet light exposure. Contaminating nucleases were removed by wiping all work surfaces and instruments with an aqueous solution of 1% (w/v) sodium dodecyl sulfate (SDS), followed by 99.9% ethanol. Several sections were cut from each block and discarded to ensure a clean starting surface for final sample procurement. For DNA extractions, three to six 10 μm pieces were cut from each tumor block and placed in a sterile 1.5 mL centrifuge tube. The cutting knife was wiped clean with the

SDS/ethanol series after each block, and the blade was replaced after every two blocks. Each blade was of sufficient length to permit the cutting of two blocks without cross-contamination. Laboratory personnel changed gloves between blocks, and cut a pure paraffin control block after every 10 tumor blocks to serve as a checkpoint for contamination.

DNA Extraction

Following removal of paraffin by xylene treatment, DNA was extracted from tumor samples by two serial incubations with 99% ethanol at 60°C and 800 rpm for 20 minutes, followed by centrifugation at 12,200 rpm for 5 minutes. One drop of acetone was added to each resulting pellet, and samples were evaporated to dryness at 60°C. Tissues were then dissolved overnight in a shaking incubator at 55°C and 1,000 rpm after addition of 150 µL proteinase K solution (10 µg/µL). The following morning, proteinase K reactions were halted by a 20 minute incubation at 98°C. DNA was then separated from the reactants using a robotic magnet-assisted nucleic acid isolation instrument (MagnaPure, Roche Applied Science).

Amplification and Genotyping

From each tumor sample, 50 ng of extracted DNA were amplified in 25 µL PCR reactions with 50 denaturation cycles at 92°C for 15 seconds, followed by annealing and extension at 60°C for 90 seconds, using primers and reagents supplied with TaqMan genotyping kits (Applied Biosystems, Foster City, California, USA).

Genotyping of the *UGT2B15*2* allele was accomplished with a commercially available kit (TaqMan real-time PCR, Applied Biosystems, Inc. Product # C-27028164-10).

Genotyping the *UGT2B7**2 allele required development of a custom TaqMan assay (Applied Biosystems, Inc.). The nucleotide sequence flanking the SNP site is relatively rich in guanine and cytosine residues, which impeded successful binding of a custom-synthesized fluorescent probe for this locus. We therefore identified another *UGT2B7* SNP—with sparse guanine and cytosine in its flanking region—estimated to be in perfect linkage disequilibrium (LD) with the *2 variant in Caucasians, to use as a proxy SNP. Linkage disequilibrium measures the degree of non-random association between alleles at different loci, usually on the same chromosome.[104] An allele is said to be in perfect LD with another if the two correlate perfectly (*i.e.*, $r^2 = 1$, or D' = 1).[104] Therefore, measuring the genotype of one SNP that is in perfect LD with a second SNP is equivalent to measuring the genotype of the second SNP. Linkage disequilibrium parameters are estimated for all SNPs surveyed in the human genome by the International HapMap Project (HapMap). We used the Haploview software application (Broad Institute, Cambridge, Massachusetts, USA) to download and analyze public HapMap data (version 2, release 22) from the region of chromosome 4 containing the *UGT2B7* gene. We searched the 5 kilobase area flanking the target *UGT2B7**2 allele (rs7439366) for candidate proxy SNPs with high LD values. We identified a second *UGT2B7* SNP estimated to be in perfect LD with the target (proxy SNP ID: rs7434332; $r^2 = 1$; D' = 1, 95% CI: 0.95, 1). Based on these LD estimates, detecting a variant allele at the proxy SNP is essentially equivalent to detecting a variant allele at the *UGT2B7**2 locus. A custom TaqMan assay for the proxy SNP was successfully developed by contract with Applied Biosystems, Inc.

All samples were assayed in duplicate using the MX3000P real-time PCR system (Stratagene, Cedar Creek, Texas, USA). Positive controls for each variant

were identified by sequencing peripheral blood DNA from 30 healthy individuals and included with each batch of assays. Negative controls, with sterile water substituted for DNA, were also included in each batch.

Genotypes were classified as (a) homozygous wild-type, (b) heterozygous, or (c) homozygous variant, according to the auto-call feature of the analytic software (MXPro QPCR version 4.1, Stratagene).

Definitions of Analytic Variables

UGT2B7 and *UGT2B15* genotypes at the *2 loci were recorded as (a) 2 normal alleles (homozygous wild-type), (b) 1 variant allele (heterozygous), or (c) 2 variant alleles (homozygous variant).

The *UGT2B7**2 variant confers a phenotype of *reduced* glucuronidation rate, while the *UGT2B15**2 variant confers a phenotype of *increased* glucuronidation rate. Women with variant alleles in both UGTs will therefore have a complicated glucuronidation phenotype. Table 21 shows how women with different genotype combinations for the two UGTs were classified according to overall phenotype. This predicted phenotype served as an alternative exposure definition.

Statistical Analysis

Within analytic groups (ER+/Tam+ and ER-/Tam-) we calculated the frequency and proportion of cases and controls according to UGT genotype, predicted glucuronidation phenotype, and tumor, treatment and socio-demographic characteristics.

We used UGT genotype data to calculate the observed minor allele frequency (MAF) for each UGT, and to test whether the genotype distributions were in Hardy-Weinberg equilibrium among the controls.[105]

Logistic regression was used to estimate the odds ratio (by design approximating the rate ratio) associating UGT polymorphisms with breast cancer recurrence among women taking tamoxifen. The analyses were repeated in the group of women who did not take tamoxifen to evaluate non-tamoxifen mediated effects of UGT genotype on recurrence rate. Logistic regression models were constructed in two ways: (1) conditioned on matched factors, thus controlling for any selection bias induced by matching controls to cases, and (2) adjusting for the matching factors stage and menopausal status at diagnosis as independent variables in the model.

We used 4 different approaches to characterizing exposure to variant UGTs. (1) We characterized UGT genotypes as exposure categories in models unadjusted and adjusted for any substantial confounders (tumor, treatment, and demographic variables)—using separate models for *UGT2B15* and *UGT2B7* (Model 1). (2) We characterized UGT genotypes as exposure categories with mutual adjustment (that is, evaluating independent effects when both genes are modeled simultaneously), with and without adjustment for any substantial confounders (Model 2). (3) We built restricted models, in which we estimated associations for one UGT within the stratum of homozygote wild-types for the other UGT. (4) We classified UGT exposure by the predicted glucuronidation phenotype (detailed in Table 21).

We evaluated potential confounding by other covariates using the modified stepwise procedure recommended by Greenland (using a ≥ 10% change in the estimated log-odds ratio as an indicator of substantial confounding). No covariate was expected to confound the associations, since none could act causally on, or likely share common causal ancestors with, genotype.

RESULTS:

Baseline Characteristics of Cases and Controls

We identified 355 ER+/Tam+ cases of recurrent breast cancer, to which were matched 360 ER+/Tam+ controls. We identified 214 ER-/Tam- cases of recurrent breast cancer, to which were matched 206 ER-/Tam- controls.

The distribution of cases and controls according to UGT genotypes, predicted glucuronidation phenotypes, and key demographic, tumor, and treatment variables is shown in Table 20. The matching factors (diagnosis year, menopausal status at diagnosis, and tumor stage at diagnosis) were well-balanced between cases and controls in both the ER+/Tam+ and ER-/Tam-groups. The majority of cases and controls were age 55 or older when their breast cancer was diagnosed. In the ER+/Tam+ group, more than 90 percent of cases and controls were post-menopausal at breast cancer diagnosis, compared with only 60 percent in the ER-/Tam- group. The ER+/Tam+ group had fewer stage I tumors at diagnosis compared with the ER-Tam- group. Women in the ER-/Tam- group were somewhat more likely to undergo radiation therapy, and were much more likely to received adjuvant chemotherapy, compared with women in the ER+/Tam+ group. Similar proportions of cases and controls in both groups underwent breast-conserving surgery instead of radical mastectomy (approximately 15 percent), which is consistent with an earlier report on trends of breast-conserving surgery use in Denmark.[106] Within the ER+/Tam+ group, 91 percent of cases and 90 percent of controls were initially assigned to tamoxifen treatment durations of two years or more. Based on results from an internal medical record review, most tamoxifen-treated women placed on

one- or two-year treatment protocols ultimately stayed on the drug longer than initially assigned.

Critical comparison of our tamoxifen-treated subjects with the 'super-population' of tamoxifen-treated Danish breast cancer patients is prevented by the process through which tamoxifen status is recorded by the DBCG registry: only women with higher-risk breast cancers were enrolled in the tamoxifen treatment protocols established by the DBCG, and it is only these women whose tamoxifen status is reliably recorded in the registry. As a result, for example, the lower prevalence of stage I tumors observed in the tamoxifen-treated study subjects is an artifact of the registry reporting criteria, and likely does not reflect the true stage distribution among all Danish breast cancer patients who underwent tamoxifen therapy.

Table 24 shows the genotype and minor allele frequencies for the two UGT variants that were observed among controls. Genotype frequencies among controls were in Hardy-Weinberg equilibrium.[107] For *UGT2B15**2, we observed a minor allele frequency (MAF) of 0.49, which is almost equal to the benchmark MAF reported for Caucasians in the PharmGKB database (0.47; Table 19). For *UGT2B7**2, we observed a MAF of 0.55, which compares well with the benchmark of 0.50 reported for Caucasians in the NCBI SNP database (Table 19).

UGT Single-Nucleotide Polymorphisms and Breast Cancer Recurrence

No covariate changed the estimated log-odds by more than 10%. Therefore all models were adjusted only for the matched factors (menopausal status and UICC stage at diagnosis) either through estimation of conditional likelihoods or through inclusion as independent variables.

We observed no association between the UGT variants and breast cancer recurrence among women not treated with tamoxifen (the ER-/Tam- group). Thus, any association seen among the tamoxifen-treated women would reasonably be mediated through UGT variant effects on the bioavailability of tamoxifen metabolites.

Estimates across all modeling strategies showed null or near-null associations between variant alleles at *UGT2B15**2 and *UGT2B7**2 and the rate of recurrence among breast cancer patients treated with tamoxifen (Table 22 and Table 23). Under the Model 1 estimation strategy (genotype associations adjusted for the matching factors UICC stage and menopausal status at diagnosis), having 2 variant alleles at the *UGT2B15**2 locus was associated with a breast cancer recurrence rate ratio of 0.68 (95% CI: 0.45, 1.0); having 2 variant alleles at the *UGT2B7**2 locus was associated with a breast cancer recurrence rate ratio of 0.85 (95% CI: 0.54, 1.3); having a predicted phenotype of increased glucuronidation rate was associated with a breast cancer recurrence rate ratio of 1.0 (95% CI: 0.67, 1.5); and having a predicted phenotype of reduced glucuronidation rate was associated with a breast cancer recurrence rate ratio of 1.1 (95% CI: 0.73, 1.5).

DISCUSSION:

We observed no association between variant alleles of *UGT2B7* and *UGT2B15* and the rate of breast cancer recurrence among breast cancer patients treated with tamoxifen.

Variant alleles at *UGT2B15**2 appeared somewhat protective against recurrence in crude conditional logistic regression analyses (Table 22; Crude OR for 1 variant allele: 0.70, 95% CI: 0.47, 1.0; Crude OR for 2 variant alleles: 0.57, 95%

CI: 0.35, 0.91). These associations appeared null upon use of the more statistically efficient Model 1 (adjustment for stage and menopausal status instead of conditioning on matched strata) and also appeared null upon mutual adjustment for *UGT2B7**2 genotype. Since a protective association with variant *UGT2B15**2 alleles is not expected under the biological model, in which the variant *increases* the rate of elimination of active tamoxifen metabolites, we interpret the results of the multivariate models as indicating no association between this SNP and breast cancer recurrence.

Analyses using the predicted glucuronidation phenotype—estimated from the joint *UGT2B15**2 and *UGT2B7**2 genotypes (Table 21)—as an alternative exposure definition were similarly null.

Analyses of each UGT that were restricted to the stratum of homozygote wild-types (*i.e.*, 2 normal alleles) on the other UGT also indicated null associations (Table 23), though these models yielded imprecise estimates due to sparse data.

Limitations

A modest proportion (16%) of our subjects had missing genotype data at the *UGT2B7**2 locus due to a limited availability of laboratory personnel and resources. Because we expect these data to be missing at random, we do not anticipate that any systematic bias influenced the observed associations between *UGT2B7**2 genotypes and breast cancer recurrence. However, the impaired sample size reduced the precision with which the *UGT2B7* and predicted phenotype associations—and the *UGT2B15* associations adjusted for, or restricted to, *UGT2B7* status—could be estimated.

Due to assay development challenges, a polymorphism in the *SULT1A1* gene (the *2 variant) could not be measured and incorporated into an overall

'elimination phenotype' exposure definition. The *SULT1A1**2 variant increases the rate of sulfation, and as with *UGT2B15**2 would be expected to reduce tamoxifen effectiveness under the biological model. However, tamoxifen metabolites appear to be largely excreted in the bile as tamoxifen glucuronides,[97] and the sulfation pathway appears to be a minor contributor to elimination. Furthermore, Jin and colleagues found no important difference in plasma concentrations of tamoxifen and its metabolites in carriers of *SULT1A1**2 variants, compared with *SULT1A1* wild-types.[108] Therefore our inability to measure the *SULT1A1**2 variant likely did not substantially bias our observations.

Residual confounding of our observed associations is not likely because, as expected, none of the examined candidate confounders substantially altered the association between genotype and recurrence. Another variable that is as strongly related to recurrence as those already examined, and also strongly (and causally) related to genotype, is unlikely to exist.

Misclassification of genotype could attenuate associations in some exposure categories and give rise to our null observations. However, substantial misclassification would be expected to yield a departure from Hardy-Weinberg equilibrium proportions, which was not observed in our study. Additionally, genotyping data were only included if the assays reported correct results for positive and negative controls for each gene variant.

Outcome misclassification is not likely to explain our results because validation of the DBCG register data has shown that breast cancer recurrence is measured with high sensitivity and near-perfect specificity.[109] Selection bias is also highly unlikely because our case and control inclusion criteria are almost certainly unrelated to subjects' genotypes on the two UGTs we examined.

Summary

We observed no associations between functional polymorphisms in *UGT2B7* and *UGT2B15* and the rate of breast cancer recurrence, either among women with estrogen receptor-positive tumors who were treated with tamoxifen or women with estrogen receptor-negative disease who were not treated with tamoxifen. No systematic error plausibly affected our observations.

A key limitation of this study is that we have characterized functional polymorphisms in only two of the enzymes involved in phase II biotransformation of active tamoxifen metabolites, and have not accounted for polymorphisms in other (minor) phase II enzymes nor for polymorphisms in the phase I enzymes that catalyze formation of the active metabolites (*e.g.*, *CYP2D6*). For instance, if a tamoxifen-treated woman harbored a polymorphism in *CYP2D6* that substantially reduced the rate at which active metabolites were formed, then the ability to detect a clinical effect may also depend upon the rate at which those active metabolites are eliminated from the body. Thus, a comprehensive assessment of the major metabolic pathways affecting the bioavailability of active tamoxifen metabolites would be an advantage over the current design. Statistical methods have recently been developed that allow classification of overall phenotypes based upon laboratory investigation into the quantitative impact of different metabolic enzyme polymorphisms on the *in vivo* concentration of active drug metabolites.[110] Application of this method to the association between tamoxifen metabolic enzyme polymorphisms and breast cancer recurrence must await comprehensive and large-sample laboratory data on the quantitative effects of phase I and II enzyme polymorphisms on active tamoxifen metabolite concentrations.

It is also possible that tamoxifen and its high ER-affinity metabolites are present in such molar excess *in vivo* (with respect to the estrogen receptor) that the drug's effectiveness is invariant to the concentration fluctuations conferred by variant phenotypes of the key metabolic enzymes.[111]

Figure 13: Phase I and II biotransformation of tamoxifen.

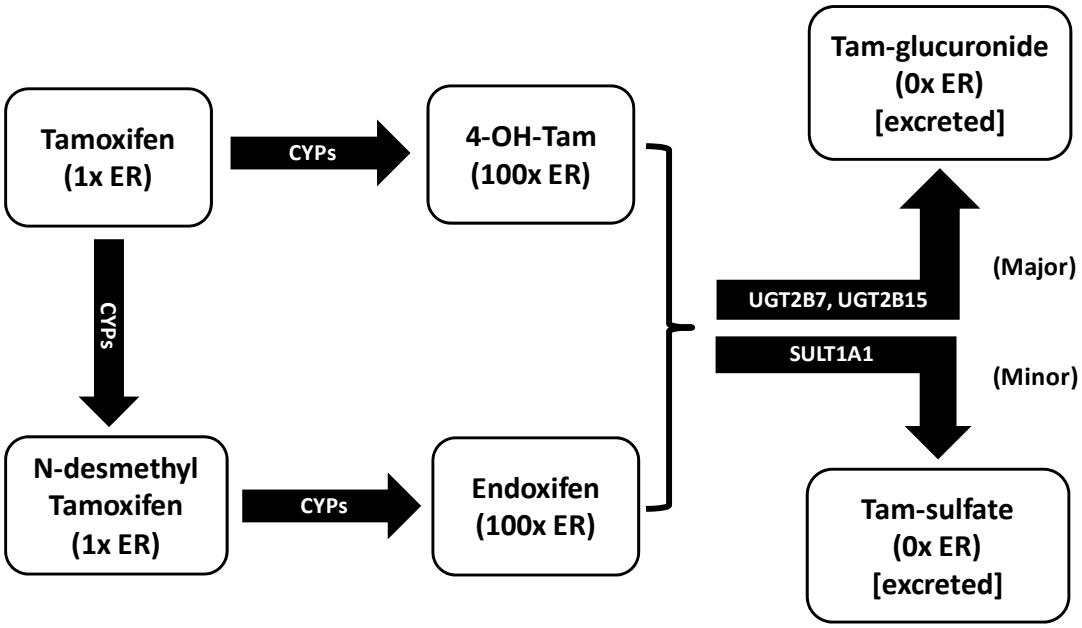


Table 19: Summary of single-nucleotide polymorphisms in key UDP-glucuronosyltransferase enzymes.

Gene	Variant allele/ SNP ID	Amino acid change	Minor allele frequency	Hypothesized enzyme function
UGT2B15	*2/ rs1902023	85 D>Y	0.47 [a]	Increased
UGT2B7	*2/ rs7439366	268 H>Y	0.50 [b]	Reduced

[a] Reported for Caucasian population PS206125 in the PharmGKB database: (http://www.pharmgkb.org/views/reports/loadFrequencyInSampleSets.action?varRptId=133585530&submissionId=PS206125).

[b] Reported for Caucasians in the NCBI SNP database: http://www.ncbi.nlm.nih.gov/projects/SNP/snp_ref.cgi?rs=7439366.

Table 20: Characteristics of breast cancer recurrence cases and matched controls according to estrogen receptor/tamoxifen treatment strata.

	ER+/Tam+ n (%)		ER-/Tam- n (%)	
	Cases	**Controls**	**Cases**	**Controls**
***UGT2B15*2* genotype**				
2 normal alleles	110 (31)	91 (25)	60 (28)	66 (32)
1 variant allele	163 (46)	171 (48)	101 (47)	90 (44)
2 variant alleles	80 (23)	97 (27)	52 (24)	48 (24)
(missing)	2	1	1	2
***UGT2B7*2* genotype**				
2 normal alleles	67 (22)	62 (21)	39 (21)	37 (21)
1 variant allele	143 (48)	142 (47)	90 (49)	86 (49)
2 variant alleles	88 (30)	96 (32)	56 (30)	51 (29)
(missing)	57	60	29	32
Predicted glucuronidation phenotype				
normal	118 (40)	121 (40)	80 (43)	69 (40)
increased	73 (25)	75 (25)	45 (24)	42 (24)
reduced	105 (35)	103 (34)	59 (32)	61 (35)
(missing)	59	61	30	34
Diagnosis year†				
1985–1993	166 (47)	162 (45)	73 (34)	65 (32)
1994–1996	68 (19)	69 (19)	56 (26)	55 (27)
1997–2001	121 (34)	129 (36)	85 (40)	86 (42)
(missing)	0	0	0	0
Age at diagnosis				
35–44	7 (2.0)	9 (2.5)	47 (22)	42 (20)
45–55	75 (21)	66 (18)	86 (40)	71 (34)
55–65	184 (52)	186 (52)	58 (27)	60 (29)
65–70	89 (25)	99 (28)	23 (11)	33 (16)
(missing)	0	0	0	0
Menopausal status at diagnosis†				
pre-menopausal	21 (5.9)	23 (6.4)	85 (40)	83 (40)
post-menopausal	334 (94)	337 (94)	129 (60)	123 (60)
(missing)	0	0	0	0
UICC tumor stage at diagnosis†				
stage I	6 (1.7)	5 (1.4)	17 (7.9)	16 (7.8)
stage II	169 (48)	172 (48)	107 (50)	110 (53)
stage III	180 (51)	183 (51)	90 (42)	80 (39)
(missing)	0	0	0	0

	ER+/Tam+ n (%)				ER-/Tam- n (%)			
	Cases		Controls		Cases		Controls	
Surgery type								
breast conserving	39	(11)	51	(14)	34	(16)	41	(20)
mastectomy	316	(89)	309	(86)	179	(84)	165	(80)
(missing)	0		0		1		0	
Radiation therapy								
yes	120	(34)	125	(35)	90	(43)	80	(45)
no	235	(66)	235	(65)	118	(57)	97	(55)
(missing)	0		0		6		29	
Tamoxifen protocol								
one year	183	(52)	176	(49)				
two years	57	(16)	61	(17)	N/A		N/A	
five years	115	(32)	123	(34)				
(missing)	0		0					
Systemic adjuvant chemotherapy								
yes	44	(12)	42	(12)	178	(83)	129	(63)
no	311	(88)	318	(88)	36	(17)	77	(37)
(missing)	0		0		0		0	
Estrogen receptor expression								
positive	324	(92)	342	(96)	50	(24)	49	(25)
negative	27	(7.7)	13	(3.7)	157	(76)	151	(76)
not available‡	4		5		7		5	
(missing)	0		0		0		1	

† Variable included in risk set sampling to match controls to cases.

‡ No malignant tissue available for assay or results indeterminate.

Table 21: Joint distribution of *UGT2B152 and *UGT2B7**2 genotypes [n, (%)], with assigned predicted glucuronidation phenotype.**

		***UGT2B15**2 (Variant increases elimination of active tamoxifen metabolites)**		
		2 normal alleles	**1 variant allele**	**2 variant alleles**
***UGT2B7**2 (Variant reduces elimination of active tamoxifen metabolites)**	**2 normal alleles**	normal 71 (7.5)	increased 105 (11)	increased 27 (2.8)
	1 variant allele	reduced 142 (15)	normal 214 (23)	increased 103 (11)
	2 variant alleles	reduced 63 (6.6)	reduced 123 (13)	normal 103 (11)

Table 22: Associations between single-nucleotide polymorphisms in UDP-glucuronosyltransferases, overall predicted glucuronidation phenotype, and breast cancer recurrence.

	Cases/ Controls	Crude OR (95% CI)*	Model 1 OR (95% CI)†	Model 2 OR (95% CI)‡
ER+/Tam+				
*UGT2B15*2*				
2 normal alleles	110/ 91	1. (reference)	1. (reference)	1. (reference)
1 variant allele	163/ 171	0.70 (0.47, 1.0)	0.79 (0.55, 1.1)	0.86 (0.58, 1.3)
2 variant alleles	80/ 97	0.57 (0.35, 0.91)	0.68 (0.45, 1.0)	0.78 (0.50, 1.2)
*UGT2B7*2*				
2 normal alleles	67/62	1. (reference)	1. (reference)	1. (reference)
1 variant allele	143/ 142	1.0 (0.63, 1.7)	0.94 (0.62, 1.4)	0.95 (0.62, 1.4)
2 variant alleles	88/ 96	0.84 (0.49, 1.4)	0.85 (0.54, 1.3)	0.87 (0.55, 1.4)
Predicted glucuronidation phenotype				
Normal	118/ 121	1. (reference)	1. (reference)	N/A
Increased	73/ 75	0.97 (0.60, 1.6)	1.0 (0.67, 1.5)	
Reduced	105/ 103	1.1 (0.70, 1.7)	1.1 (0.73, 1.5)	
ER-/Tam-				
*UGT2B15*2*				
2 normal alleles	60/ 66	1. (reference)	1. (reference)	1. (reference)
1 variant allele	101/ 90	1.5 (0.92, 2.5)	1.2 (0.79, 1.9)	1.3 (0.81, 2.1)
2 variant alleles	52/ 48	1.5 (0.84, 2.8)	1.2 (0.71, 2.0)	1.2 (0.66, 2.1)
*UGT2B7*2*				
2 normal alleles	39/ 37	1. (reference)	1. (reference)	1. (reference)
1 variant allele	90/ 86	1.1 (0.54, 2.2)	1.0 (0.58, 1.7)	1.0 (0.57, 1.7)
2 variant alleles	56/ 51	1.3 (0.63, 2.6)	1.1 (0.58, 1.9)	1.1 (0.59, 2.0)
Predicted glucuronidation phenotype				
Normal	80/ 69	1. (reference)	1. (reference)	N/A
Increased	45/ 42	1.0 (0.54, 2.0)	0.95 (0.56, 1.6)	
Reduced	59/ 61	0.82 (0.45, 1.5)	0.85 (0.52, 1.4)	

* Crude: Conditional logistic regression accounting for all matching factors, with no additional independent variables.

† Model 1: Adjusted for UICC stage and menopausal status at diagnosis.

‡ Model 2: Adjusted for UICC stage and menopausal status at diagnosis, and mutually for *UGT2B15**2 or *UGT2B7**2 genotype.

Table 23: UGT analyses restricted to strata of subjects with homozygous wild-type genotype (no variant allele) for complementary UGT.

	Cases/ Controls	OR (95% CI)[a]
ER+/Tam+		
*UGT2B15*2*		
2 normal alleles	25/ 17	1. (reference)
1 variant allele	33/ 35	0.67 (0.30, 1.5)
2 variant alleles	8/ 9	0.62 (0.19, 2.0)
*UGT2B7*2* [b]		
2 normal alleles	25/ 17	1. (reference)
1 variant allele	46/ 45	0.70 (0.33, 1.5)
2 variant alleles	18/ 16	0.77 (0.31, 1.9)
ER-/Tam-		
*UGT2B15*2*		
2 normal alleles	15/ 14	1. (reference)
1 variant allele	20/ 17	1.1 (0.39, 3.0)
2 variant alleles	4/ 6	0.74 (0.15, 3.8)
*UGT2B7*2*		
2 normal alleles	15/ 14	1. (reference)
1 variant allele	23/ 18	0.71 (0.27, 1.9)
2 variant alleles	14/ 15	0.81 (0.27, 2.4)

[a] *UGT2B15**2 model restricted to subjects with no variant allele at *UGT2B7**2, and vice-versa. Models were adjusted for UICC stage and menopausal status at diagnosis.

[b] Restricted model for *UGT2B7**2 in ER+/Tam+ group could not support adjustment for stage and menopausal status; results reflect estimates from an unconditional model without adjustment for these variables.

Table 24: Genotype and minor allele frequencies observed among controls for *UGT2B152 and *UGT2B7**2.**

Gene variant	Genotype frequencies, observed (expected) [a]			Minor allele frequency	Hardy-Weinberg P-value [b]
	2 normal alleles	1 variant allele	2 variant alleles		
*UGT2B15**2	157 (147)	261 (281)	145 (135)	0.49	0.09
*UGT2B7**2	99 (96)	228 (235)	147 (144)	0.55	0.54

[a] Expected values are genotype frequencies predicted under Hardy-Weinberg equilibrium.

[b] Two-sided p-value (alpha=0.05) from chi-squared test of H_0: observed genotypes are in Hardy-Weinberg equilibrium.

CONCLUSIONS

The studies comprising this dissertation have addressed three timely hypotheses in breast cancer epidemiology. The first study examined the association between treatment with the prescription drug digoxin and the incidence of invasive breast carcinoma, and was motivated by a growing body of literature that suggested a protective role of cardiac glycoside drugs in both cancer incidence and survival. We observed a 30% increase in the rate of breast cancer incidence among ever-users of digoxin compared with never users (OR: 1.30, 95% CI: 1.14, 1.48). The magnitude of this association increased directly, though modestly, with duration of digoxin exposure (for 1 to 3 years of digoxin exposure, OR: 1.25, 95% CI: 1.03, 1.52; for 4 to 6 years of digoxin exposure, OR: 1.30, 95% CI: 1.05, 1.61; for 7 to 18 years of exposure, OR: 1.39, 95% CI: 1.10, 1.74). While the positive association we observed was opposite to the association hypothesized by earlier investigators,[8-10] our result was actually corroborated by un-emphasized measurements in earlier papers.[12, 13, 34]

The second study used three different strategies to estimate the association between treatment with a vitamin K antagonist and the incidence of 24 site-specific cancers in a cohort of heart valve recipients with age- and sex-matched non-recipients. The first approach used a VKA prescription validation subset to estimate the drug-cancer associations under known prescription status. The second and third approaches expanded follow-up time (enabling collection of more rare cancer events to augment the precision of estimated associations) by using heart valve replacement history as a proxy variable for exposure to a vitamin K antagonist. The second approach treated the heart valve proxy as a misclassified VKA exposure

measurement, and we adjusted for misclassification with a probabilistic bias analysis parameterized by positive and negative predictive values estimated in the prescription validation subset. The third approach treated heart valve replacement as an instrumental variable for the VKA/cancer associations which, in addition to scaling estimates to adjust for misclassification by the instrument, has the added advantage of removing confounding due to known and unknown variables (presuming the instrumental variable assumptions are satisfied). All approaches showed null estimates between VKA exposure and the rate of breast cancer incidence (in the validation subset, breast cancer IRR: 0.9, 95% CI: 0.6, 1.5; in the probabilistic bias analysis, breast cancer IRR: 1.3, 95% CI: 1.0, 1.7; in the instrumental variable analysis, breast cancer IRR: 1.1, 95% CI: 0.8, 1.4). Associations with the other 23 cancer sites were similarly null. Earlier studies had indicated potential protective associations between VKA therapy and cancer incidence,[20, 56] but had also been criticized for having several unaddressed threats to their internal validity.[49-53]

The third study examined the association between functional single nucleotide polymorphisms in the UDP-glucuronosyl transferase enzymes chiefly responsible for the elimination of tamoxifen metabolites with the highest affinity for the estrogen receptor. We enrolled breast cancer recurrence cases and matched breast cancer controls in two groups: women with estrogen receptor positive disease who were treated with tamoxifen (ER+/Tam+), and women with estrogen receptor negative disease who were not treated with tamoxifen (ER-/Tam-). Variant alleles at the *UGT2B15**2 and *UGT2B7**2 loci showed null associations with breast cancer recurrence in the ER-/Tam- group, so any association seen in the ER+/Tam+ group would reasonably be mediated through effects on the tamoxifen pathway, and

not reflective of a direct effect of the gene polymorphisms on breast cancer recurrence. However, the same variant alleles also showed null associations in the ER+/Tam+ group (for 2 variant alleles at *UGT2B15**2, OR: 0.68, 95% CI: 0.45, 1.0; for 2 variant alleles at *UGT2B7**2, OR: 0.85, 95% CI: 0.54, 1.3). These results do not support the *a priori* biological hypothesis that enhanced (or reduced) elimination of tamoxifen metabolites increases (or decreases) the rate of breast cancer recurrence among tamoxifen-treated women. It may be that an effect of this sort can only be detected through comprehensive evaluation of genetic variation in tamoxifen metabolic pathways, which may be possible to execute with recently developed statistical methods,[110] but only after extensive new data on tamoxifen pharmacodynamics are added to the existing literature. Another possibility is that tamoxifen and its metabolites are present in such excess that their therapeutic effect is insensitive to the concentration fluctuations induced by variant metabolic enzymes.[111]

REFERENCES

1. Breast Cancer Overview. July 1, 2010; http://www.cancer.org/Cancer/BreastCancer/OverviewGuide/breast-cancer-overview-key-statistics. Accessed July 2, 2010, 2010.

2. NORDCAN: Cancer Incidence, Mortality, Prevalence and Prediction in the Nordic Countries. Association of the Nordic Cancer Registries. Danish Cancer Society.; 2010. http://www.ancr.nu. Accessed August 6, 2010.

3. Frank L. Epidemiology. When an entire country is a cohort. *Science.* 2000;287:2398-9.

4. Lewis WH, Elvin-Lewis MPF. *Medical botany : plants affecting man's health*. New York: Wiley; 1977.

5. Melero CP, Medarde M, San Feliciano A. A short review on cardiotonic steroids and their aminoguanidine analogues. *Molecules.* 2000;5:51-81.

6. Harrison TR, Fauci AS. Harrison's principles of internal medicine. 14th ed. New York: McGraw-Hill, Health Professions Division; 1998:1264-95.

7. Newman RA, Yang P, Pawlus AD, Block KI. Cardiac glycosides as novel cancer therapeutic agents. *Molecular Interventions.* 2008;8:36-49.

8. Stenkvist B, Bengtsson E, Eriksson O, Holmquist J, Nordin B, Westman-Naeser S. Cardiac glycosides and breast cancer. *Lancet.* 1979;1:563.

9. Stenkvist B, Pengtsson E, Dahlqvist B, Eriksson O, Jarkrans T, Nordin B. Cardiac glycosides and breast cancer, revisited. *New England Journal of Medicine.* 1982;306:484.

10. Stenkvist B. Is digitalis a therapy for breast carcinoma? *Oncology Reports.* 1999;6:493-6.

11. Chen JQ, Contreras RG, Wang R, et al. Sodium/potassium ATPase (Na+, K+-ATPase) and ouabain/related cardiac glycosides: A new paradigm for development of anti- breast cancer drugs? *Breast Cancer Research and Treatment.* 2006;96:1-15.

12. Haux J, Klepp O, Spigset O, Tretli S. Digitoxin medication and cancer; case control and internal dose-response studies. *BMC Cancer.* 2001;1:11.

13. Friedman GD. Digitalis and breast cancer. *Lancet.* 1984;2:875.

14. Sorensen HT, Christensen S, Mehnert F, et al. Use of bisphosphonates among women and risk of atrial fibrillation and flutter: population based case-control study. *British Medical Journal.* 2008;336:813-6.

15. Sorensen HT, Skriver MV, Friis S, McLaughlin JK, Blot WJ, Baron JA. Use of antibiotics and risk of breast cancer: a population-based case-control study. *British Journal of Cancer.* 2005;92:594-6.

16. Floyd A, Pedersen L, Nielsen GL, Thorlacius-Ussing O, Sorensen HT. Risk of acute pancreatitis in users of azathioprine: a population-based case-control study. *American Journal of Gastroenterology.* 2003;98:1305-8.

17. Lash TL, Pedersen L, Cronin-Fenton D, et al. Tamoxifen's protection against breast cancer recurrence is not reduced by concurrent use of the SSRI citalopram. *British Journal of Cancer.* 2008;99:616-21.

18. WHO Collaborating Center for Drug Statistics Methodology. [World Wide Web] http://www.whocc.no/atcddd/, 2008.

19. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10:37-48.

20. Schulman S, Lindmarker P. Incidence of cancer after prophylaxis with warfarin against recurrent venous thromboembolism. Duration of Anticoagulation Trial. *New England Journal of Medicine.* 2000;342:1953-8.

21. McGettigan P, Han P, Jones L, Whitaker D, Henry D. Selective COX-2 inhibitors, NSAIDs and congestive heart failure: differences between new and recurrent cases. *British Journal of Clinical Pharmacology.* 2008;65:927-34.

22. Moysich KB, Beehler GP, Zirpoli G, Choi JY, Baker JA. Use of common medications and breast cancer risk. *Cancer Epidemiology Biomarkers & Prevention.* 2008;17:1564-95.

23. Cronin-Fenton DP, Pedersen L, Lash TL, Friis S, Baron JA, Sorensen HT. Prescriptions for selective cyclooxygenase-2 inhibitors, non-selective non-

steroidal anti-inflammatory drugs, and risk of breast cancer in a population-based case-control study. *Breast Cancer Research.*12:R15.

24. Terry MB, Gammon MD, Zhang FF, et al. Association of frequency and duration of aspirin use and hormone receptor status with breast cancer risk. *Journal of the American Medical Association.* 2004;291:2433-40.

25. Cuzick J. Hormone replacement therapy and the risk of breast cancer. *European Journal of Cancer.* 2008;44:2344-9.

26. Jensen A, Olsen AH, von Euler-Chelpin M, Helle Njor S, Vejborg I, Lynge E. Do nonattenders in mammography screening programmes seek mammography elsewhere? *International Journal of Cancer.* 2005;113:464-70.

27. Olsen AH, Jensen A, Njor SH, et al. Breast cancer incidence after the start of mammography screening in Denmark. *British Journal of Cancer.* 2003;88:362-5.

28. Svendsen AL, Olsen AH, von Euler-Chelpin M, Lynge E. Breast cancer incidence after the introduction of mammography screening: what should be expected? *Cancer.* 2006;106:1883-90.

29. Greenland S, Lash TL, Rothman KJ. Concepts of Interaction. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3 ed. Philadelphia: Lippincott Williams & Wilkins; 2008:71-83.

30. Rothman KJ, Greenland S, Lash TL. Introduction to stratified analysis. *Modern Epidemiology*. 3 ed. Philadelphia: Lippincott Williams & Wilkins; 2008:111-125.

31. Russo J, Moral R, Balogh GA, Mailo D, Russo IH. The protective role of pregnancy in breast cancer. *Breast Cancer Research.* 2005;7:131-42.

32. Hjartaker A, Langseth H, Weiderpass E. Obesity and diabetes epidemics: cancer repercussions. *Advances in Experimental Medicine and Biology.* 2008;630:72-93.

33. Zalesin KC, Franklin BA, Miller WM, Peterson ED, McCullough PA. Impact of obesity on cardiovascular disease. *Endocrinology & Metabolism Clinics of North America.* 2008;37:663-84.

34. Stenkvist B, Bengtsson E, Eklund G, et al. Evidence of a modifying influence of heart glucosides on the development of breast cancer. *Analytical and Quantitative Cytology.* 1980;2:49-54.

35. Goldin AG, Safa AR. Digitalis and cancer. *Lancet.* 1984;1:1134.

36. Poole C. Beyond the confidence interval. *American Journal of Public Health.* 1987;77:195-9.

37. Sullivan KM, Foster DA. Use of the confidence interval function. *Epidemiology.* 1990;1:39-42.

38. Ewertz M, Holmberg L, Tretli S, Pedersen BV, Kristensen A. Risk factors for male breast cancer--a case-control study from Scandinavia. *Acta Oncologica.* 2001;40:467-71.

39. Lewinn EB. Gynecomastia during digitalis therapy; report of eight additional cases with liver-function studies. *New England Journal of Medicine.* 1953;248:316-20.

40. Braunstein GD. Gynecomastia. *New England Journal of Medicine.* 1993;328:490-5.

41. Simpson CD, Anyiwe K, Schimmer AD. Anoikis resistance and tumor metastasis. *Cancer Letters.* 2008;272:177-85.

42. Goodman LS, Gilman A, Brunton LL, Lazo JS, Parker KL. *Goodman & Gilman's the pharmacological basis of therapeutics*. 11th ed. New York: McGraw-Hill; 2006.

43. Runciman DJ, Lee AM, Reed KF, Walsh JR. Dicoumarol toxicity in cattle associated with ingestion of silage containing sweet vernal grass (Anthoxanthum odoratum). *Australian Veterinary Journal.* 2002;80:28-32.

44. Katzung BG. *Basic & clinical pharmacology*. 9th ed. New York: Lange Medical Books/McGraw Hill; 2004.

45. Zacharski LR, Henderson WG, Rickles FR, et al. Effect of warfarin on survival in small cell carcinoma of the lung. Veterans Administration Study No. 75. *Journal of the American Medical Association.* 1981;245:831-5.

46. Zacharski LR, Henderson WG, Rickles FR, et al. Effect of warfarin anticoagulation on survival in carcinoma of the lung, colon, head and neck, and prostate. Final report of VA Cooperative Study #75. *Cancer.* 1984;53:2046-52.

47. Smorenburg SM, Vink R, Otten HM, Swaneveld F, Buller HR. The effects of vitamin K-antagonists on survival of patients with malignancy: A systematic analysis. *Thrombosis and Haemostasis.* 2001;86:1586-7.

48. Levine M, Hirsh J, Gent M, et al. Double-blind randomised trial of a very-low-dose warfarin for prevention of thromboembolism in stage IV breast cancer. *Lancet.* 1994;343:886-9.

49. Eikelboom JW, Mehta SR. Venous thromboembolism and cancer. *New England Journal of Medicine.* 2000;343:1337.

50. Hughes-Davies TH. Venous thromboembolism and cancer. *New England Journal of Medicine.* 2000;343:1337.

51. Roychowdhury D. Venous thromboembolism and cancer. *New England Journal of Medicine.* 2000;343:1337-8.

52. Zacharski LR, Ornstein DL. Venous thromboembolism and cancer. *New England Journal of Medicine.* 2000;343:1338.

53. Zielinski CC, Hejna M. Warfarin for cancer prevention. *New England Journal of Medicine.* 2000;342:1991-3.

54. Taliani MR, Agnelli G, Prandoni P, et al. Incidence of cancer after a first episode of idiopathic venous thromboembolism treated with 3 months or 1 year of oral anticoagulation. *Journal of Thrombosis and Haemostasis.* 2003;1:1730-3.

55. Blumentals WA, Foulis PR, Schwartz SW, Mason TJ. Does warfarin therapy influence the risk of bladder cancer? *Thrombosis and Haemostasis.* 2004;91:801-5.

56. Tagalakis V, Tamim H, Blostein M, Collet JP, Hanley JA, Kahn SR. Use of warfarin and risk of urogenital cancer: a population-based, nested case-control study. *Lancet Oncology.* 2007;8:395-402.

57. Pengo V, Denas G, Jose SP, Pengo MF. Cancer prevention and vitamin K antagonists: an overview. *Thrombosis Research.* 2010;125 Suppl 2:S103-5.

58. Bonow RO, Carabello BA, Kanu C, et al. ACC/AHA 2006 guidelines for the management of patients with valvular heart disease: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines (writing committee to revise the 1998 Guidelines for the Management of Patients With Valvular Heart Disease): developed in collaboration with the Society of Cardiovascular Anesthesiologists: endorsed by the Society for Cardiovascular Angiography and Interventions and the Society of Thoracic Surgeons. *Circulation.* 2006;114:e84-231.

59. Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *Journal of Chronic Diseases.* 1987;40:373-83.

60. Aurigemma GP, Gaasch WH. Antithrombotic therapy in patients with prosthetic heart valves. In: Otto CM, ed. *UpToDate*. Waltham, MA: UpToDate; 2010.

61. Baron JA, Gridley G, Weiderpass E, Nyren O, Linet M. Venous thromboembolism and cancer. *Lancet.* 1998;351:1077-80.

62. Sorensen HT, Mellemkjaer L, Olsen JH, Baron JA. Prognosis of cancers associated with venous thromboembolism. *New England Journal of Medicine.* 2000;343:1846-50.

63. Marshall RJ. Validation study methods for estimating exposure proportions and odds ratios with misclassified data. *Journal of Clinical Epidemiology.* 1990;43:941-7.

64. Lash TL, Fox MP, Fink AK. Probabilistic Bias Analysis. *Applying Quantitative Bias Analysis to Epidemiologic Data.* Dordrecht, NL: Springer; 2009:117-150.

65. Rothman KJ, Greenland S, Lash TL. Bias Analysis. *Modern Epidemiology*. 3rd ed. Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2008.

66. Greenland S, Robins JM. Empirical-Bayes adjustments for multiple comparisons are sometimes useful. *Epidemiology.* 1991;2:244-51.

67. Greenland S, Poole C. Empirical-Bayes and semi-Bayes approaches to occupational and environmental hazard surveillance. *Archives of Environmental Health.* 1994;49:9-16.

68. Steenland K, Bray I, Greenland S, Boffetta P. Empirical Bayes adjustments for multiple results in hypothesis-generating or surveillance studies. *Cancer Epidemiology Biomarkers & Prevention.* 2000;9:895-903.

69. Greenland S. An introduction to instrumental variables for epidemiologists. *International Journal of Epidemiology.* 2000;29:722-9.

70. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology.* 2006;17:360-72.

71. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Instrumental variable analysis of secondary pharmacoepidemiologic data. *Epidemiology.* 2006;17:373-4.

72. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. *Epidemiology.* 2006;17:268-75.

73. Rubin DB. *Multiple Imputation for Nonresponse in Surveys*. New York: John Wiley & Sons; 1987.

74. The MI Procedure. *SAS OnlineDoc 9.1.3* [June 16, 2010]; http://support.sas.com/onlinedoc/913/docMainpage.jsp.

75. The MIANALYZE Procedure. *SAS OnlineDoc 9.1.3* [World Wide Web] http://support.sas.com/onlinedoc/913/docMainpage.jsp. Accessed August 15, 2007.

76. Cunnane C. Unbiased plotting positions: a review. *Journal of Hydrology.* 1978;37:205-22.

77. Sorensen HT, Johnsen SP, Norgard B, Zacharski LR, Baron JA. Cancer and venous thromboembolism: a multidisciplinary approach. *Clinical Laboratory.* 2003;49:615-23.

78. Libby P. The Pathogenesis, Prevention, and Treatment of Atherosclerosis. In: Fauci AS, Braunwald E, Kasper DL, et al., eds. *Harrison's Principles of Internal Medicine*. 17e ed.

79. Maganti K, Rigolin VH, Sarano ME, Bonow RO. Valvular heart disease: diagnosis and management. *Mayo Clinic Proceedings.* 2010;85:483-500.

80. Brennan P, Bogillot O, Cordier S, et al. Cigarette smoking and bladder cancer in men: a pooled analysis of 11 case-control studies. *International Journal of Cancer.* 2000;86:289-94.

81. Morrison AS, Buring JE, Verhoek WG, et al. An international study of smoking and bladder cancer. *Journal of Urology.* Apr 1984;131(4):650-654.

82. Cerhan JR, Torner JC, Lynch CF, et al. Association of smoking, body mass, and physical activity with risk of prostate cancer in the Iowa 65+ Rural Health Study (United States). *Cancer Causes & Control.* 1997;8:229-38.

83. Hiatt RA, Armstrong MA, Klatsky AL, Sidney S. Alcohol consumption, smoking, and other risk factors and prostate cancer in a large health plan cohort in California (United States). *Cancer Causes & Control.* 1994;5:66-72.

84. Honda GD, Bernstein L, Ross RK, Greenland S, Gerkins V, Henderson BE. Vasectomy, cigarette smoking, and age at first sexual intercourse as risk factors for prostate cancer in middle-aged men. *British Journal of Cancer.* 1988;57:326-31.

85. Coyle YM. Lifestyle, genes, and cancer. *Methods in Molecular Biology.* 2009;472:25-56.

86. Koebnick C, Michaud D, Moore SC, et al. Body mass index, physical activity, and bladder cancer in a large prospective study. *Cancer Epidemiology Biomarkers & Prevention.* 2008;17:1214-21.

87. Coleman CI, Baker WL, Kluger J, White CM. Antihypertensive medication and their impact on cancer incidence: a mixed treatment comparison meta-analysis of randomized controlled trials. *Journal of Hypertension.* 2008;26:622-9.

88. Perron L, Bairati I, Harel F, Meyer F. Antihypertensive drug use and the risk of prostate cancer (Canada). *Cancer Causes & Control.* Aug 2004;15:535-41.

89. Dale KM, Coleman CI, Henyan NN, Kluger J, White CM. Statins and cancer risk: a meta-analysis. *Journal of the American Medical Association.* 2006;295:74-80.

90. Bonovas S, Filioussi K, Tsavaris N, Sitaras NM. Use of statins and breast cancer: a meta-analysis of seven randomized clinical trials and nine observational studies. *Journal of Clinical Oncology.* 2005;23:8606-12.

91. Bhattacharyya S, Schapira AH, Mikhailidis DP, Davar J. Drug-induced fibrotic valvular heart disease. *Lancet.* 2009;374:577-85.

92. Rothman KJ, Greenland S, Lash TL. Validity in Epidemiologic Studies. *Modern Epidemiology*. 3 ed. Philadelphia: Lippincott Williams & Wilkins; 2008:128-147.

93. Ahern TP, Lash TL, Thwin SS, Silliman RA. Impact of acquired comorbidities on all-cause mortality rates among older breast cancer survivors. *Medical Care.* 2009;47:73-9.

94. Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: an overview of the randomised trials. *Lancet.* 2005;365:1687-1717.

95. Desta Z, Ward BA, Soukhova NV, Flockhart DA. Comprehensive evaluation of tamoxifen sequential biotransformation by the human cytochrome P450 system in vitro: prominent roles for CYP3A and CYP2D6. *Journal of Pharmacology and Experimental Therapeutics.* 2004;310:1062-75.

96. Sun D, Sharma AK, Dellinger RW, et al. Glucuronidation of active tamoxifen metabolites by the human UDP glucuronosyltransferases. *Drug Metabolism & Disposition.* 2007;35:2006-14.

97. Lazarus P, Blevins-Primeau AS, Zheng Y, Sun D. Potential role of UGT pharmacogenetics in cancer treatment and prevention: focus on tamoxifen. *Annals of the New York Academy of Sciences.* 2009;1155:99-111.

98. Nowell S, Sweeney C, Winters M, et al. Association between sulfotransferase 1A1 genotype and survival of breast cancer patients receiving tamoxifen therapy. *Journal of the National Cancer Institute.* 2002;94:1635-40.

99. Levesque E, Beaulieu M, Green MD, Tephly TR, Belanger A, Hum DW. Isolation and characterization of UGT2B15(Y85): a UDP-glucuronosyltransferase encoded by a polymorphic gene. *Pharmacogenetics.* 1997;7:317-25.

100. Blevins-Primeau AS, Sun D, Chen G, et al. Functional significance of UDP-glucuronosyltransferase variants in the metabolism of active tamoxifen metabolites. *Cancer Research.* 2009;69:1892-1900.

101. Levesque E, Delage R, Benoit-Biancamano MO, et al. The impact of UGT1A8, UGT1A9, and UGT2B7 genetic polymorphisms on the pharmacokinetic profile of mycophenolic acid after a single oral dose in healthy volunteers. *Clinical Pharmacology & Therapeutics.* 2007;81:392-400.

102. Blichert-Toft M, Christiansen P, Mouridsen HT. Danish Breast Cancer Cooperative Group--DBCG: History, organization, and status of scientific achievements at 30-year anniversary. *Acta Oncologica.* 2008;47:497-505.

103. Rothman KJ, Greenland S, Lash TL. Case-Control Studies. *Modern Epidemiology*. 3 ed. Philadelphia: Lippincott Williams & Wilkins; 2008:111-127.

104. Martin ER. Linkage Disequilibrium and Association Analysis. In: Haines JL, Pericak-Vance M, eds. *Genetic Analysis of Complex Diseases*. 2 ed. Hoboken, New Jersey: John Wiley & Sons; 2006:329-353.

105. Weir BS. *Genetic data analysis II : methods for discrete population genetic data*. Sunderland, Mass.: Sinauer Associates; 1996.

106. Ahern TP, Larsson H, Garne JP, Cronin-Fenton DP, Sorensen HT, Lash TL. Trends in breast-conserving surgery in Denmark, 1982-2002. *European Journal of Epidemiology.* 2008;23:109-14.

107. Trikalinos TA, Salanti G, Khoury MJ, Ioannidis JP. Impact of violations and deviations in Hardy-Weinberg equilibrium on postulated gene-disease associations. *American Journal of Epidemiology.* 2006;163:300-9.

108. Jin Y, Desta Z, Stearns V, et al. CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. *Journal of the National Cancer Institute.* 2005;97:30-9.

109. Hansen PS, Andersen E, Andersen KW, Mouridsen HT. Quality control of end results in a Danish adjuvant breast cancer multi-center study. *Acta Oncologica.* 1997;36:711-4.

110. Li L, Borges S, Jason RD, Shen C, Desta Z, Flockhart D. A Penalized Mixture Model Approach in Genotype/Phenotype Association Analysis for Quantitative Phenotypes. *Cancer Informatics.* 2010;9:93-103.

111. Lash TL, Lien EA, Sorensen HT, Hamilton-Dutoit S. Genotype-guided tamoxifen therapy; time to pause for reflection? *Lancet Oncology.* 2009;10:825-33.

Reports/ PhD theses from Department of Clinical Epidemiology

1. Ane Marie Thulstrup: Mortality, infections and operative risk in patients with liver cirrhosis in Denmark. Clinical epidemiological studies. *2000*.

2. Nana Thrane: Prescription of systemic antibiotics for Danish children. *2000*.

3. Charlotte Søndergaard. Follow-up studies of prenatal, perinatal and postnatal risk factors in infantile colic. *2001*.

4. Charlotte Olesen: Use of the North Jutland Prescription Database in epidemiological studies of drug use and drug safety during pregnancy. *2001*.

5. Yuan Wei: The impact of fetal growth on the subsequent risk of infectious disease and asthma in childhood. *2001*.

6. Gitte Pedersen. Bacteremia: treatment and prognosis. *2001*.

7. Henrik Gregersen: The prognosis of Danish patients with monoclonal gammopathy of undertermined significance: register-based studies. *2002*.

8. Bente Nørgård: Colitis ulcerosa, coeliaki og graviditet; en oversigt med speciel reference til forløb og sikkerhed af medicinsk behandling. *2002*.

9. Søren Paaske Johnsen: Risk factors for stroke with special reference to diet, Chlamydia pneumoniae, infection, and use of non-steroidal anti-inflammatory drugs. *2002*.

10. Elise Snitker Jensen: Seasonal variation of meningococcal disease and factors associated with its outcome. *2003*.

11. Andrea Floyd: Drug-associated acute pancreatitis. Clinical epidemiological studies of selected drugs. *2004*.

12. Pia Wogelius: Aspects of dental health in children with asthma. Epidemiological studies of dental anxiety and caries among children in North Jutland County, Denmark. *2004*.

13. Kort-og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Nordjyllands, Viborg og Århus amter 1985-2003. *2004*.

14. Reimar W. Thomsen: Diabetes mellitus and community-acquired bacteremia: risk and prognosis. *2004*.

15. Kronisk obstruktiv lungesygdom i Nordjyllands, Viborg og Århus amter 1994-2004. Forekomst og prognose. Et pilotprojekt. *2005*.

16. Lungebetændelse i Nordjyllands, Viborg og Århus amter 1994-2004. Forekomst og prognose. Et pilotprojekt. *2005*.

17. Kort- og langtidsoverlevelse efter indlæggelse for nyre-, bugspytkirtel- og leverkræft i Nordjyllands, Viborg, Ringkøbing og Århus amter 1985-2004. *2005*.

18. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Nordjyllands, Viborg, Ringkøbing og Århus amter 1995-2005. *2005*.

19. Mette Nørgaard: Haematological malignancies: Risk and prognosis. *2006*.

20. Alma Becic Pedersen: Studies based on the Danish Hip Arthroplastry Registry. *2006*.

21. Særtryk: Klinisk Epidemiologisk Afdeling - De første 5 år. *2006*.

22. Blindtarmsbetændelse i Vejle, Ringkjøbing, Viborg, Nordjyllands og Århus Amter. *2006*.

23. Andre sygdommes betydning for overlevelse efter indlæggelse for seks kræftsygdomme i Nordjyllands, Viborg, Ringkjøbing og Århus amter 1995-2005. *2006*.

24. Ambulante besøg og indlæggelser for udvalgte kroniske sygdomme på somatiske hospitaler i Århus, Ringkjøbing, Viborg, og Nordjyllands amter. *2006*.

25. Ellen M Mikkelsen: Impact of genetic counseling for hereditary breast and ovarian cancer disposition on psychosocial outcomes and risk perception: A population-based follow-up study. *2006*.

26. Forbruget af lægemidler mod kroniske sygdomme i Århus, Viborg og Nordjyllands amter 2004-2005. *2006*.

27. Tilbagelægning af kolostomi og ileostomi i Vejle, Ringkjøbing, Viborg, Nordjyllands og Århus Amter. *2006*.

28. Rune Erichsen: Time trend in incidence and prognosis of primary liver cancer and liver cancer of unknown origin in a Danish region, 1985-2004. *2007*.

29. Vivian Langagergaard: Birth outcome in Danish women with breast cancer, cutaneous malignant melanoma, and Hodgkin's disease. *2007*.

30. Cynthia de Luise: The relationship between chronic obstructive pulmonary disease, comorbidity and mortality following hip fracture. *2007*.

31. Kirstine Kobberøe Søgaard: Risk of venous thromboembolism in patients with liver disease: A nationwide population-based case-control study. *2007*.

32. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Region Midtjylland og Region Nordjylland 1995-2006. *2007*.

33. Mette Skytte Tetsche: Prognosis for ovarian cancer in Denmark 1980-2005: Studies of use of hospital discharge data to monitor and study prognosis and impact of comorbidity and venous thromboembolism on survival. *2007*.

34. Estrid Muff Munk: Clinical epidemiological studies in patients with unexplained chest and/or epigastric pain. *2007*.

35. Sygehuskontakter og lægemiddelforbrug for udvalgte kroniske sygdomme i Region Nordjylland. *2007*.

36. Vera Ehrenstein: Association of Apgar score and postterm delivery with neurologic morbidity: Cohort studies using data from Danish population registries. *2007*.

37. Annette Østergaard Jensen: Chronic diseases and non-melanoma skin cancer. The impact on risk and prognosis. *2008*.

38. Use of medical databases in clinical epidemiology. *2008*.

39. Majken Karoline Jensen: Genetic variation related to high-density lipoprotein metabolism and risk of coronary heart disease. *2008*.

40. Blodprop i hjertet - forekomst og prognose. En undersøgelse af førstegangsindlæggelser i Region Nordjylland og Region Midtjylland. *2008*.

41. Asbestose og kræft i lungehinderne. Danmark 1977-2005. *2008*.

42. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme i Region Midtjylland og Region Nordjylland 1996-2007. *2008*.

43. Akutte indlæggelsesforløb og skadestuebesøg på hospiter i Region Midtjylland og Region Nordjylland 2003-2007. Et pilotprojekt. *2009*.

44. Peter Jepsen: Prognosis for Danish patients with liver cirrhosis. *2009*.

45. Lars Pedersen: Use of Danish health registries to study drug-induced birth defects – A review with special reference to methodological issues and maternal use of non-steroidal anti-inflammatory drugs and Loratadine. *2009*.

46. Steffen Christensen: Prognosis of Danish patients in intensive care. Clinical epidemiological studies on the impact of preadmission cardiovascular drug use on mortality. *2009*.

47. Morten Schmidt: Use of selective cyclooxygenase-2 inhibitors and nonselective nonsteroidal antiinflammatory drugs and risk of cardiovascular events and death after intracoronary stenting. *2009*.

48. Jette Bromman Kornum: Obesity, diabetes and hospitalization with pneumonia. *2009*.

49. Theis Thilemann: Medication use and risk of revision after primary total hip arthroplasty. *2009*.

50. Operativ fjernelse af galdeblæren. Region Midtjylland & Region Nordjylland. 1998-2008. *2009*.

51. Mette Søgaard: Diagnosis and prognosis of patients with community-acquired bacteremia. *2009*.

52. Marianne Tang Severinsen. Risk factors for venous thromboembolism: Smoking, anthropometry and genetic susceptibility. *2010*.

53. Henriette Thisted: Antidiabetic Treatments and ischemic cardiovascular disease in Denmark: Risk and outcome. *2010*.

54. Kort- og langtidsoverlevelse efter indlæggelse for udvalgte kræftsygdomme. Region Midtjylland og Region Nordjylland 1997-2008. *2010*.

55. Prognosen efter akut indlæggelse på Medicinsk Visitationsafsnit på Nørrebrogade, Århus Sygehus. *2010*.

56. Kaare Haurvig Palnum: Implementation of clinical guidelines regarding acute treatment and secondary medical prophylaxis among patients with acute stroke in Denmark. *2010*.