Long-acting insulin analogues and the risk of breast cancer: a comparative effectiveness and methodological study

By: Jennifer W. Wu

A thesis submitted to McGill University in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Epidemiology

Department of Epidemiology, Biostatistics and Occupational Health
McGill University
Montreal, Canada

December 2016

Abstract

Basal insulins, intermediate-acting Neutral Protamine Hagedorn (NPH) insulin and long-acting insulin analogues glargine and detemir, are used to treat patients with advanced stages of type 2 diabetes. In the 1990s, long-acting insulin analogues, which are structurally modified human insulins, were developed as an alternative to NPH insulin to mitigate adverse events such as nocturnal hypoglycaemia. However, safety concerns were raised with experimental and observational studies suggesting that insulin glargine increases the risk of cancer, particularly breast. Several editorials and narrative reviews have criticised previous observational studies on long-acting insulin analogues and the risk of cancer for having methodological issues including prevalent user bias, lacked the use of lag periods to account for cancer latency, residual confounding, or short duration of follow-up. In contrast, randomised controlled trials (RCTs) showed null effects between long-acting insulin analogues and cancer but had small sample sizes and short duration of follow-up. Given the current evidence, uncertainty remains regarding the use of insulin glargine on the risk of breast cancer. Furthermore, no study to date has evaluated the use of the newer insulin detemir and the risk of breast cancer. The overall goal of my thesis was to evaluate the relationship between long-acting insulin analogues and risk of breast cancer while addressing the methodological issues encountered by the previous studies.

In the first manuscript, we conducted a systematic review of the observational studies of long-acting insulin analogues and the risk of any and site-specific cancers, with a particular focus on methodological strengths and limitations. We identified 16 cohort and 3 case-control studies in our systematic review. Generally, no association was observed between long-acting insulin analogues and the risk of any cancer and colorectal or prostate cancer whereas four of 13 studies reported an increased risk of breast cancer with insulin glargine use. Similarly, no association was observed between insulin detemir and any cancers. Moreover, no studies evaluated site-specific cancer risk with insulin detemir use. Importantly, in the study quality assessment, we identified limitations due to prevalent user bias, lack of inclusion of lag periods, time-related biases, and short duration of follow-up. Consequently, the relationship between long-acting insulin analogues and the risk of cancer, especially breast, remains uncertain due to the methodological limitations in the existing observational studies.

In the second manuscript, we used data from United Kingdom's Clinical Practice Research Datalink to evaluate whether the use of long-acting insulin analogues, when compared

with NPH insulin, is associated with an increased risk of breast cancer in women with type 2 diabetes using insulin therapies. Insulin glargine compared with NPH was associated with an overall increased risk of breast cancer (hazard ratio [HR]: 1.45, 95% confidence intervals [CI]: 1.12-1.87), which was particularly elevated after five years of use (HR: 2.25, 95% CI: 1.33-3.80), and after more than 30 insulin glargine prescriptions (HR: 2.28, 95% CI: 1.26-4.15). In contrast, the use of insulin detemir was not associated with an increased risk of breast cancer neither overall nor was a duration- or dose-response observed. We concluded that insulin glargine was associated with an increased long-term risk of breast cancer. In contrast, the risk with insulin detemir remains uncertain because of its more recent introduction to the United Kingdom (UK) market resulting in a smaller number of women and shorter duration of use.

In the third manuscript, we explored two different study designs and data analytical techniques in a comparative effectiveness study of insulin glargine and breast cancer risk to account for all types of patients – new and prior insulin users – in the real world. The time-dependent approach used the study cohort of women with type 2 diabetes using insulin glargine or NPH and classified women as unexposed until the first glargine prescription and exposed subsequently. The pseudo-matching approach identified the cohort from the initiation of insulin glargine and randomly selected a cohort entry date among the comparator (NPH insulin) to be similar to the glargine exposed cohort in terms of the prior history of insulin treatment. The HR for breast cancer with insulin glargine use compared with NPH users was 1.46 (95% CI: 1.16-1.84) and 1.44 (95% CI: 1.11-1.85) using the time-dependent and pseudo-matching approaches, respectively. The results suggested both approaches can be used to assess the safety profile of all types of patients but some complexities in the design and data analyses need to be considered when choosing one approach over the other.

My thesis furthers our knowledge of the relationship between long-acting insulin analogues and the risk of cancer, particularly breast. More specifically, the results from the first two manuscripts can be used to inform drug regulatory agencies and other relevant stakeholders of the safety profile of long-acting insulin analogues. In addition, my thesis contributes to the advancement of alternative study designs and data analytical approaches that could be used in comparative effectiveness observational studies to assess the safety of a new treatment in all patients in the real world setting. Future research and continual surveillance are still needed to assess the relationship of long-acting insulin analogues and the risk of breast cancer.

Résumé

Les insulines basales, l'insuline à action intermédiaire Hagedorn protamine neutre (NPH) et les analogues de l'insuline à action prolongée, l'insuline glargine et l'insuline détémir, sont utilisées pour traiter les patients à un stade avancé de diabète de type 2. Pendant les années 90, des analogues de l'insuline à action prolongée, qui sont des insulines humaines modifiées structurellement, ont été conçus comme une alternative à l'insuline NPH pour limiter les événements indésirables comme l'hypoglycémie nocturne. Cependant, des problèmes d'innocuité ont été soulevés à la suite d'études expérimentales et d'observation qui suggéraient que l'insuline glargine augmentait le risque de cancer, en particulier du cancer du sein. Plusieurs éditoriaux et revues narratives ont critiqué les études d'observation portant sur le risque de cancer associé aux analogues de l'insuline à action prolongée parce qu'elles présentaient des lacunes méthodologiques, y compris des biais liés à des échantillons de patients prévalents plutôt qu'incidents, l'absence de considération de périodes de latence entre l'exposition et la survenue du cancer, les biais de confusion résiduels ou une courte durée de suivi. En revanche, les essais cliniques aléatoires ne montraient pas d'association entre les analogues de l'insuline à action prolongée et le cancer, mais comportaient des échantillons de petite taille et des suivis de courte durée. Compte tenu des preuves actuelles, l'incertitude demeure quant à l'utilisation de l'insuline glargine et le risque de cancer du sein. Par ailleurs, aucune étude à ce jour n'a évalué l'utilisation de l'insuline détémir, une insuline plus récente, et le risque de cancer du sein. L'objectif général de ma thèse était d'évaluer la relation entre les analogues de l'insuline à action prolongée et le risque de cancer du sein, tout en traitant des lacunes méthodologiques rencontrées dans les études précédentes.

Dans le premier manuscrit, nous avons effectué une revue systématique des études d'observation portant sur le risque de tout type de cancer, et de certains types particuliers de cancers, associés aux analogues de l'insuline à action prolongée, l'accent étant mis sur les forces et les limites méthodologiques. Nous avons identifié 16 études de cohorte et trois études cas-témoins dans notre revue systématique. En général, aucun lien n'a été observé entre les analogues de l'insuline à action prolongée et le risque de toute de forme de cancer et de cancer colorectal ou de la prostate, alors que quatre des 13 études ont rapporté une augmentation du risque de cancer du sein avec l'utilisation de l'insuline glargine. De même, aucun lien n'a été observé entre l'insuline détémir et tout type de cancer. De plus, aucune étude n'a évalué le risque

de développer certains types particuliers de cancer suite à l'utilisation de l'insuline détémir. Fait à noter, dans l'évaluation de la qualité de ces études, nous avons relevé des limites associées à des biais liés à des échantillons de patients d'utilisateurs prévalents, l'absence d'inclusion de périodes de latence, des biais liés au temps et la courte durée du suivi. Par conséquent, le lien entre les analogues de l'insuline à action prolongée et le risque de cancer, et du cancer du sein en particulier, demeure incertain en raison des limites méthodologiques des études d'observation existantes.

Dans le deuxième manuscrit, nous avons utilisé les données provenant du Clinical Practice Research Datalink du Royaume-Uni pour déterminer si l'utilisation d'analogues de l'insuline à action prolongée, comparativement à l'insuline NPH, était associée à un risque accru de cancer du sein chez les femmes atteintes de diabète de type 2 traitées avec de l'insuline. L'insuline glargine, comparativement à l'insuline NPH, était associée à un risque accru de cancer du sein (risque relatif [RR] : 1,45; intervalle de confiance à 95 % [IC] : 1,12-1,87) qui était particulièrement élevé après cinq ans d'utilisation (RR : 2,25; IC de 95 % : 1,33-3,80) et après plus de 30 ordonnances d'insuline glargine (RR : 2,28; IC de 95 % : 1,26-4,15). En revanche, l'utilisation de l'insuline détémir n'était pas associée à un risque accru de cancer du sein en général et non plus en tenant compte de la durée ou de la dose du traitement. Nous avons conclu que l'insuline glargine était associée à un risque accru à long terme de cancer du sein. En revanche, le risque avec l'insuline détémir demeure incertain en raison de sa plus récente introduction sur le marché britannique, d'où un plus petit nombre de femmes et une plus courte durée d'utilisation.

Dans le troisième manuscrit, nous avons examiné deux plans d'étude et de méthodes d'analyse des données différents dans une étude d'efficacité comparative portant sur l'insuline glargine et le risque de cancer du sein pour tenir compte de tous les types de patients — anciens et nouveaux utilisateurs d'insuline — dans le monde réel. L'approche traitant l'exposition comme dépendante du temps inclut la cohorte de femmes atteintes de diabète de type 2 qui utilisent l'insuline glargine ou NPH. Les femmes étaient catégorisées comme non exposées avant la première ordonnance de glargine, et exposées par la suite. L'approche de pseudo-appariement a permis d'identifier la cohorte à compter du début de l'utilisation de l'insuline glargine et de choisir au hasard une date d'entrée dans la cohorte pour le groupe de comparaison (insuline NPH) afin qu'elle soit semblable à celle de la cohorte exposée à la glargine en ce qui a trait aux

antécédents de traitement à l'insuline. Le risque relatif de cancer du sein avec l'utilisation de l'insuline glargine comparativement aux utilisateurs de NPH était de 1,46 (IC à 95 % : 1,16-1,84) et de 1,44 (IC à 95 % : 1,11-1,85) selon l'approche liée au temps et l'approche de pseudo-appariement respectivement. Les résultats ont révélé que les deux approches pouvaient être utilisées pour évaluer le profil d'innocuité de tous les types de patients, mais certaines complexités dans le plan et l'analyse des données devaient être prises en considération lors du choix d'une approche par rapport à l'autre.

Ma thèse contribue à faire avancer notre savoir à propos du lien qui existe entre les analogues de l'insuline à action prolongée et le risque de cancer, notamment du cancer du sein. Plus précisément, les résultats des deux premiers manuscrits peuvent être utilisés pour informer les organismes de contrôle des médicaments et d'autres intervenants pertinents du profil d'innocuité des analogues de l'insuline à action prolongée. De plus, ma thèse contribue à l'avancement d'autres plans d'étude et méthodes d'analyse des données qui pourraient être utilisés dans les études d'observation de l'efficacité comparative pour évaluer l'innocuité d'un nouveau traitement chez tous les patients dans le monde réel. D'autres recherches et une surveillance continue sont toujours nécessaires pour évaluer le lien entre les analogues de l'insuline à action prolongée et le risque de cancer du sein.

Acknowledgements

I would like to take this moment to express my gratitude for all the individuals who have helped me throughout my doctorate training and scholarly journey.

First and foremost, I would like to thank my PhD supervisor, Dr. Samy Suissa, who has been a supportive and great mentor in all my endeavours. He has shared with me over the years his tremendous knowledge in pharmacoepidemiology and methods in epidemiology and trained me to become a better epidemiologist. I have also benefited greatly from the career advice he has given me as a blossoming epidemiologist.

Second, I am grateful for my thesis committee members – Drs. Laurent Azoulay, Jean-François Boivin, Agneiszka Majdan, and Michael Pollak – because this thesis would not have been possible without their expertise and contributions. Dr. Boivin – thank you for teaching me epidemiology by allowing me to teach epidemiology in your class. Dr. Pollak – thank you for helping me navigate through the web of biological mechanisms. Dr. Majdan – thank you for your clinical expertise and teaching me how patients with diabetes are managed. A special thanks to Dr. Azoulay for teaching me pharmacoepidemiology in my office every day and for always suggesting more sensitivity analyses making my research closer to perfection. I would also like to thank my unofficial committee member, Dr. Kristian Filion, for sharing your expertise in systematic reviews and meta-analyses, perfectionism, and career advices. I will miss the random hallway meetings!

Third, I would like to thank my PhD cohort at McGill for being so supportive throughout this journey of ours. In particular, thank you to Maggie Doll (and Whiskey), Miriam Kinkaid and Kathryn Morrison for your companionship, friendship dinners, stimulating debates, puppy and baby photos, and golden nuggets of wisdom about life.

Fourth, my colleagues and professors at McGill University and Jewish General Hospital/Lady Davis Institute (JGH/LDI) also need to be thanked for their time and sharing their knowledge and experiences in epidemiology and beyond. Another special thanks to the analysts at JGH/LDI, Janie Coulombe, Sophie Dell'Aneillo, and Hui Yin, without whom, none of the results in this

thesis would have come to fruition. Thank you for teaching me a lifetime and more of SAS knowledge.

Fifth, I also need to thank the administrative staff at McGill University and JGH/LDI for their help in the past few years. A special thanks to Marisa Mancini, Diane Gaudreau and Astrid Tchakmakchian at JGH/LDI for working with me so closely in the last couple of years. From providing me filing folders to answering my incessant questions about McGill forms, funding and travel reimbursement, thank you a million times.

Sixth, I would not have made it to this point without other professors and mentors from various points in my life. Thank you to Drs. Andrew Freedman, Linda Levesque, Victoria Kirsh, Xiaomei Ma, Yikyung Park, and Rashmi Sinha for all your career and life advice and encouraging me to take this scholarly road.

Seventh, I would like to thank my family and friends from all around the world for believing in me and providing me emotional support from afar. Particularly, thank you to Jessica Thom, for being my family, friend, personal chef and doctor in Montreal for the last 5 years…we made it!

Eighth, I would like to especially thank my partner Justin Chan for supporting me and literally sitting next to me while I worked on this thesis. Thank you for not minding having lost so many weekends together. You are simply the best.

Last but not least, my parents, Amy and Francis, have been an unwavering and continual source of financial and emotional support. Thank you for allowing me to take the road less travelled. Words cannot truly express how grateful I am for you two. This thesis is dedicated to both of you.

Contribution of Authors

Manuscript 1: The effect of long-acting insulin analogues on the risk of cancer: a systematic review of observational studies. Published in Diabetes Care. 2016;39(3):486-94.

I developed the objective of this systematic review with assistance from Drs. Filion and Suissa. I was primarily responsible for the systematic review protocol, systematic search, extracted the data, assessed the quality of the included studies, and the manuscript. My colleague Margaret Doll served as my second reviewer in the searching of the literature and data extraction and quality assessment process. Drs. Filion and Suissa served as adjudicators for disagreements in the inclusion of studies and quality assessment. Ms. Doll and Drs. Azoulay, Filion, and Suissa reviewed the manuscript for important intellectual content.

Manuscript 2: Long-term effect of long-acting insulin analogues on breast cancer incidence in women with type 2 diabetes. Submitted to The British Medical Journal.

The objective of this study was developed by Dr. Samy Suissa and me. I drafted the study protocol, submitted data requests from the UK, and obtained ethics approval from McGill University, Jewish General Hospital and UK's Independent Scientific Advisory Committee. I was responsible for all aspects of the study including preparing the data for analysis, performing all of the data analyses, interpreting the results, and drafting of the manuscript. Drs. Azoulay, Boivin, Majdan, Pollak and Suissa were also involved in the discussions of the study design and methodologies used and interpretation of the findings. All coauthors critically reviewed the manuscript for important intellectual content.

Manuscript 3: Alternative methods to account for patients with a prior history of treatment: an illustration with insulin glargine and breast cancer. To be submitted to Pharmacoepidemiology and Drug Safety.

The idea and objective of this study were conceptualised by Dr. Samy Suissa and me. I drafted the study protocol, submitted data requests from the UK, and obtained ethics approval from McGill University, Jewish General Hospital and UK's Independent Scientific Advisory Committee. Drs. Azoulay and Suissa reviewed and revised the study protocol. I was responsible

for all aspects of the project including preparing the data for analysis, performing all of the data analyses, interpreting the results, and drafting of the manuscript. Drs. Azoulay and Suissa were also involved in the discussions of the study designs and methodologies used and interpretations of the findings. Drs. Azoulay and Suissa critically reviewed the manuscript for important intellectual content.

Statement of Originality

The research presented here in this thesis constitutes original contributions which will further our knowledge about the effects of long-acting insulin analogues on the risk of cancer and methods in pharmacoepidemiology. In the first manuscript, we are the first to systematically evaluate observational studies on long-acting insulin analogues and its effects on cancer incidence with a particular focus on the methodological strengths and limitations. Although there have been several studies on long-acting insulin analogues and its effects on breast cancer incidence, prior studies have short follow-up and other methodological limitations identified in the first manuscript. Therefore, in our second manuscript, we estimate the effects of long-acting insulin analogues on the risk of breast cancer with longer durations of follow-up and addresses some of the methodological shortcomings of previously published studies. Finally, in our last manuscript, we provide alternative study designs and data analytical techniques to account for all types of patient in the real world setting for comparative effectiveness research in the context of insulin glargine and breast cancer. Moreover, we discuss the strengths and limitations to each alternative approach.

I declare that while I received guidance from my supervisors, thesis committee members and other co-authors on the substantive and methodological aspects of my thesis, the conception, execution, and drafting of the work in my thesis were entirely my own.

Statement of Financial Support

I am thankful for the various institutions that provided financial support for my research and studies at McGill University. The first year of my PhD was jointly funded by the Department of Epidemiology, Biostatistics and Occupational Health at McGill University and my supervisor Dr. Samy Suissa. The second to fourth year of my PhD was sponsored by a Doctoral Research Award from the Canadian Institute of Health Research's Priority Announcement Drug Safety and Effectiveness. In the final year of my PhD training, my supervisor has again kindly provided me with a stipend.

Table of Contents

List of Tables

List of Figures

List of Acronyms and Abbreviations

BMI	Body mass index
CI	Confidence intervals
CNODES	Canadian Network for Observational Drug Effect Studies
CPRD	Clinical Practice Research Datalink
DPP-4i	Dipeptidyl-peptidase-4 inhibitors
EMA	European Medicine Agency
FDA	Food and Drug Administration
GLP-1	Glucagon-like peptide-1 analogues
GP	General practitioner
HbA1c	Glycated haemoglobin
HR	Hazard ratio
ICD	International Classification of Diseases
IGF	Insulin-like growth factor
IGFBP	Insulin-like growth factor binding protein
ISAC	Independent Scientific Advisory Committee
ITT	Intent-to-treat
MAPK	Mitogen-activated protein kinase
NCDR	National Cancer Data Repository
NOS	Newcastle Ottawa Scale
NPH	Neutral Protamine Hagedorn
NSAID	Non-steroidal anti-inflammatory drugs
ORIGIN	Outcomes Reduction with Insulin Glargine Intervention
RCT	Randomised controlled trials
ROBINS-I	Risk of Bias in Non-Randomised Studies – of Interventions
RR	Relative risks
SGLT2i	Sodium-glucose cotransporter 2 inhibitors
TZD	Thiazolidinediones
UK	United Kingdom
UTS	Up to standard

Chapter 1. Introduction

1.1. Overview

Type 2 diabetes, the most common type of diabetes, and cancer are global health burdens. According to the World Health Organization the current worldwide prevalence of diabetes is at 422 million, while 14 million new cancer cases have been reported in 2012 and expected to double in 2035.[1-3] The relationship between type 2 diabetes and cancer is believed to develop through biological mechanisms such as hyperinsulinemia defined as excess endogenous insulin levels circulating in the blood.[4 5] Experimental studies have shown that elevated levels of endogenous insulin and insulin-like growth factors (IGF) bind to the insulin receptor family resulting in the activation of mitogenic cellular pathways that can cause cancer cell proliferation, anti-apoptotic effects, and metastasis.[4] Consequently, it is possible that antidiabetic medications that can increase insulin levels may have implications for cancer risk.

When patients with type 2 diabetes have advanced to later stages of the disease, basal insulins such as intermediate-acting NPH insulin and long-acting insulin analogues (glargine and detemir) are usually prescribed.[6] While long-acting insulin analogues reduce the occurrence of nocturnal hypoglycaemia and have longer duration of treatment effects than NPH, there are concerns that their use may increase the risk of cancer, particularly breast. Indeed, some experimental studies conducted in breast cancer cells have shown that, due to the altered binding kinetics, long-acting insulin analogues bind more strongly to the insulin receptor family.[7] This results in long-acting insulin analogues initiating the same mitogenic signalling pathways as endogenous insulin and IGF.

In contrast to the experimental studies, analyses of Randomised controlled trials (RCT) of long-acting insulin analogues have not found an increased risk of overall and site-specific cancers.[8-11] However, these studies had small sample sizes and follow-up of fewer than seven years, which prevents the detection of a potential long-term effect on cancer outcomes. Since 2009, several observational studies, using cohort and case-control study designs, have been conducted on the use of insulin glargine and the risk of breast cancer and reported conflicting results.[12-25] Methodological shortcomings of a few of these observational studies, including reverse causation, lack of use of lag periods, prevalent user bias, and concerns about the data analysis and short duration of follow-up, have been previously highlighted.[26-31] Given the current

evidence, uncertainty remains regarding the use of insulin glargine on the risk of breast cancer, particularly given these methodological issues. Moreover, to date, no study has evaluated the use of the newer insulin detemir and the risk of breast cancer.

1.2. Research objectives

The primary goal of this Doctoral thesis dissertation was to evaluate the association between long-acting insulin analogues and risk of breast cancer while addressing the methodological issues encountered by the previous studies. I proposed three objectives to accomplish this:

1. Conduct a systematic review of observational studies examining the association of long-acting insulin analogues on the risk of cancer, with a particular focus on methodological strengths and weaknesses of these studies

2. Assess whether the use of long-acting insulin analogues, compared with the use of NPH insulin, is associated with an increased risk of breast cancer in women with type 2 diabetes using insulin therapy

3. Explore alternative study designs and data analytical techniques in a comparative effectiveness study of insulin glargine on the risk of breast cancer to account for all types of patients

1.3. Organisation of thesis

This is a manuscript-based thesis, which includes three manuscripts each with its research objective. We begin in chapter 2 by presenting the background on the link between type 2 diabetes, insulin therapies and cancer risk. In chapter 3, we provide an overview of the United Kingdom (UK)'s Clinical Practice Research Datalink (CPRD), which is the primary data source of this thesis. In chapter 4, we systematically review the literature for observational studies on long-acting insulin analogues and its effects on the risk of cancer. In chapter 5, we assess the effects of long-acting insulin analogues on the risk of breast cancer in a population-based cohort study. In chapter 6, we explore two alternative methods to account for all types of patients in the

real world setting. Finally, in chapter 7, we provide a summary of conclusions from the three research manuscripts, with a focus on key findings and implications for public health and future research. Appendices can be found at the end of each manuscript. References for each manuscript are placed at the end of the thesis.

Chapter 2. Background

2.1. Type 2 diabetes and cancer

2.1.1. Epidemiology and characteristics of type 2 diabetes and cancer

Together, diabetes and cancer is a global health burden with both the prevalence of diabetes and incidence of cancers doubling in 2035.[3] [32] Type 1 and 2 diabetes are the two primary subtypes of diabetes, with the latter accounting for 90-95% of all patients with diabetes.[4] [33] Type 1 diabetes is a result of autoimmune-mediated destruction of the β-cells (cells that produce and secrete insulin) in the pancreas whereas type 2 diabetes is characterized by hyperglycaemia (elevated glucose levels) leading to hyperinsulinemia (excess endogenous insulin levels) as a result of insulin resistance and damaged cells in the pancreas reducing insulin production.[5] [33] [34] Due to the destruction of β-cells in the pancreas, patients with type 1 diabetes are insulin dependent but these insulin administrations cannot artificially mimic normal physiological secretions and levels of insulin.[33] In contrast, patients with type 2 diabetes have higher than normal physiological levels of insulin in attempts to restore body fuel (e.g. glucose) to homoeostasis, which could play a role in cancer initiation.[33] Consequently, the differences in the pathophysiology of these two diabetes subtypes could explain the difference in effects on cancer risk.

Many systematic reviews and meta-analyses of observational studies have demonstrated that type 2 diabetes is associated with an increased risk of many site-specific cancers (bladder, blood, breast, colorectal, endometrial, gallbladder, gastrointestinal, kidney, liver, lung, oral, ovarian, pancreas, and thyroid) with summary relative risks (RR) ranging from 1.10-2.31.[35-44] However, a recent meta-analysis illustrated that many of these studies are biased, with respect to reporting of significant associations resulting in false positives and inflated estimates, except for common cancer sites such as breast cancer.[41] Despite this fact, in 2010, the joint consensus report published by the American Diabetes Association and American Cancer Society stated that diabetes, specifically type 2 diabetes, is a risk factor for cancer. [4] On the contrary, the evidence between type 1 diabetes and the risk of any and site-specific cancer (blood, cervix, endometrium, kidney, liver, lung, oral, pancreatic, skin, stomach) was limited.[45] Importantly, given type 1 diabetes is rare, the sample sizes were small in these previous studies. Moreover, the definition

of patients with type 1 diabetes was highly variable between studies and included patients with type 2 diabetes.

Given hyperinsulinemia conditions can persist for years during the natural progression from prediabetes to type 2 diabetes; it appears to provide an environment that will favour cancer initiation. Therefore, we will focus on the relationship between type 2 diabetes, treatments used for management of type 2 diabetes, and cancer risk for the remainder of this thesis.

2.1.2. The link between type 2 diabetes and cancer

In figure 2.1, the causal and non-causal relationships of type 2 diabetes and cancer are depicted. Many studies, including the consensus report, have suggested type 2 diabetes affects the neoplastic process through several biological mechanisms including hyperinsulinemia, hyperglycaemia, and chronic inflammation.[4 5 33 46] In addition to these biological mechanisms, common risk factors, such as obesity, may also explain the association between type 2 diabetes and cancer.[4 46-48] However, some observational studies have shown type 2 diabetes increases the risk of certain site-specific cancers despite adjusting on or stratifying by body mass index (BMI) levels.[33 49 50] In contrast, an experimental study, which developed a non-obese mice model with type 2 diabetes, showed enhanced breast cancer tumour growth.[46] Together, the evidence suggests type 2 diabetes, acting through any of the three biological mechanisms, may also be an independent risk factor for cancer. Although the other biological mechanisms are important to consider, hyperinulinemia will be the focus here because both excessive endogenous insulin and antidiabetic medications that alter insulin levels would have implications for cancer risk.[5]

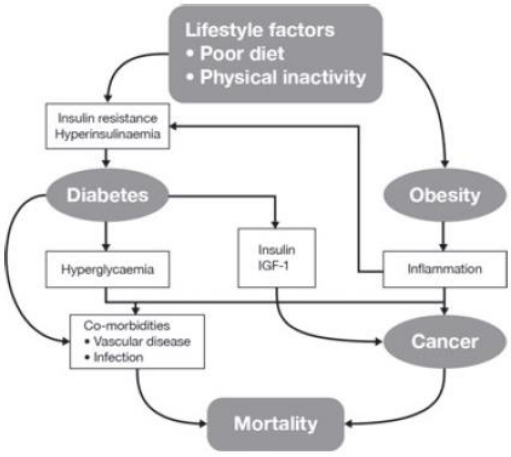


Figure 2.1 Interrelationship between type 2 diabetes and cancer [48]

Aforementioned, patients with type 2 diabetes are characterized by hyperglycaemia, potentially induced by excessive caloric intake, which leads to hyperinsulinemia.[5 33] Patients with type 2 diabetes can have these two conditions persist during the natural progression of this chronic disease. Moreover, excessive insulin levels can elevate other growth hormones such as the IGF and both insulin and IGF can upregulate cellular activities in peripheral tissues and ultimately initiate and promote cancerous tumour growth.

2.1.3. The insulin and insulin-like growth factor axis

Insulin and IGFs can mediate cellular activities by binding to the insulin receptor family, which includes insulin, IGFs, or a hybrid of insulin and IGF receptors. This is known as the insulin and IGF axis. The binding of insulin and IGFs to the insulin receptor family can activate both the metabolic and mitogenic cellular pathways.[46 51] More specifically, several experimental studies have suggested that insulin and IGFs can bind to the insulin receptor family to activate the mitogen-activate protein kinase (MAPK), which is also known as RAS/RAF/MEK/ERK, pathways to initiate cellular proliferation and anti-apoptotic effects.[4 46 51-54] Simultaneously, it has been shown that hyperinsulinemia can induce carcinogenesis through IGFs.[4 48] Insulin can reduce the production of IGF binding proteins (IGFBP) in the liver resulting in higher circulating levels of IGFs.[48 52] Consequently, this growth hormone IGF can bind to IGF receptors and predominantly stimulate the same mitogenic cellular activities as insulin. In addition, IGFBPs have been reported to have a pro-apoptotic effect, which could potentially reduce the risk of cancer.[54] Therefore, prolonged hyperinsulinemia conditions can provide an environment for cells to grow and potentially mutate, leading to cell transformation and eventually develop into clinically detectable cancer.[52]

The insulin receptor family are not only expressed in normal tissues but cancer cells as well and can potentially further promote the growth of tumours. Cancer cells have tended to overexpress insulin receptors.[51 55-59] Moreover, cancer cells can uptake glucose independently of insulin binding to its receptors.[4 46] Therefore, it suggests the role of insulin and binding to the insulin receptor family is to primarily activate the mitogenic pathways for cancer cell growth and survival. Although there are several cancer subtypes that overexpress insulin receptors,[51] there have been more experimental studies that suggested both insulin and IGF receptors tended to be

overexpressed in breast cancer cells.[55-59] Furthermore, there have also been experimental studies that have demonstrated insulin and IGF can stimulate the growth of breast tissues and breast cancer cell lines.[60 61]

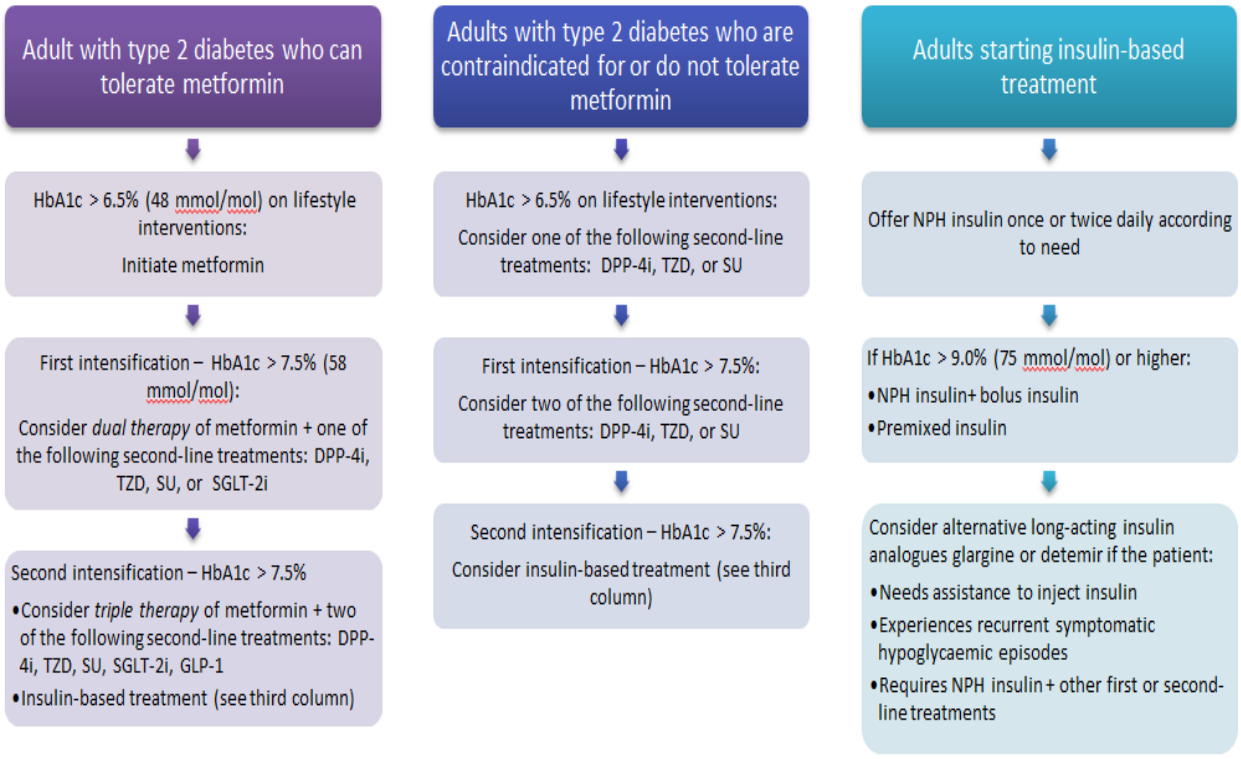
2.1.4. Observational studies on endogenous insulin, IGF, and IGFBP levels on the risk of cancer

Similar to type 2 diabetes and cancer epidemiologic research, there has been an abundance of observational studies that found a positive association between elevated levels of endogenous insulin and IGF levels and risk of several site-specific cancers whereas null associations were reported with IGFBPs. More specifically, systematic review of observational studies showed elevated levels of connecting or C-peptides, a biological marker for endogenous levels of insulin, were associated with an increased risk of common cancer sites such as breast and colorectal cancers with summary RRs ranging from 1.26-1.40.[62-65] However, other observational studies have also demonstrated null associations between C-peptide levels and the risk of breast and prostate cancer.[66-68] Similarly, IGFs, primarily IGF-1, have been associated with an increased risk of many sites-specific cancers such as breast, colorectal, and prostate, with RRs ranging from 1.14-1.52.[69-75] Other systematic reviews of observational studies on IGFs and site-specific cancer risk have also reported null associations for breast, lung, prostate, and ovarian cancers.[76-81] On the contrary, higher compared with lower levels of IGFBPs, primarily IGFBP-3, have reported null associations.[69-73 75 78-81] The inconsistencies observed in these studies may be due to differences in study populations and designs, cancer sites and assays of blood samples from patients.[73]

Together, the experimental and observational studies suggested, albeit inconsistently, excess endogenous insulin may play a role in cancer development. Therefore, antidiabetic medications used for the management of type 2 diabetes and can alter the levels of insulin, such as insulin therapies, could further affect the risk of cancer. In the next few sections, we discuss the types of insulin therapy that are currently available, insulin therapies on the risk of cancer and, in particular, the safety concerns associated with long-acting insulin analogues.

2.2. Insulin therapies and its implication for cancer risk

Current clinical guidelines for adult patients with newly diagnosed early stage type 2 diabetes is to commence with lifestyle modifications through diet or physical activity and then patients may progress to using antidiabetic medications to gain better control of glycated haemoglobin (HbA1c) levels.[6] [82-84] As depicted in figure 2.2, patients typically begin on metformin, unless contraindicated or not well tolerated, and then treatment can intensify to one or two add-ons of second line treatments if HbA1c targets are not met. Finally, insulin therapy is the last line of treatment for patients with advanced type 2 diabetes and up to 80% of patients with type 2 diabetes eventually advance to using insulin therapies.[4 5]



Abbreviations: DPP-4i – dipeptidyl-peptidase-4 inhibitor; GLP-1 – glucagon-like peptide-1 analogues; SGLT-2i – sodium-glucose cotransporter 2 inhibitors; SU – sulfonylurea; TZD – thiazolidinedione (only pioglitazone is recommended in the UK)

Figure 2.2 Brief overview of the treatment algorithm, adapted from the UK clinical guidelines, for adult patients with type 2 diabetes[6]

2.2.1. Types of insulin therapies

Presently, there are two major groups of insulin – prandial or bolus and basal. Bolus insulins are used after meal times to manage the sudden surge of glucose while basal insulins are

used to manage the baseline levels of insulin throughout the day.[85] Premixed insulins, which are a viable alternative, are a percentage of bolus insulins and an intermediate-acting NPH insulin (a type of basal insulin) and used when the patient needs a simple insulin therapy regimen.[86] We summarised in table 2.1 the types of insulin that are currently on the market and the characteristics of each type of insulin. Inhaled and animal insulins were not described in the table as these insulins are no longer available and not generally recommended.[87] Within each group of insulin, there are human insulins and insulin analogues, which are structurally modified human insulins via amino acid substitutions.[88] Generally, if patients with type 2 diabetes fail to achieve a specific HbA1c target, despite intensified therapy, clinical guidelines recommend patients starting with one of the three basal insulins (intermediate-acting NPH insulin, long-acting insulin analogue glargine or detemir) with or without bolus insulin or premixed insulins.[6 82-84] Starting in 2013, a new ultra long-acting insulin analogue, degludec, was introduced into the market worldwide but it has not been adopted by current clinical guidelines.[89-91] Although other non-insulin antidiabetic medications have been associated with a decreased and increased risk of cancer,[4 5] we focus on insulin therapy and its effects on cancer risk because subcutaneous injections of insulin yield higher dosages of insulin, compared with normal physiological levels, which could further amplify the risk of cancer.

2.2.2. Observational studies on all insulin therapies and cancer risk

Many observational studies have evaluated the effect of insulin use on the risk of any and site-specific cancer among patients with diabetes.[7 92-101] In 2004, the first observational study on insulin therapy and colorectal cancer risk reported a 2-fold increase (HR: 2.1, 95% CI: 1.2-3.4).[102] Subsequently, albeit inconsistently, results from systematic reviews and meta-analyses of the published observational studies have largely observed insulin therapy use compared with no insulin therapy use or use of non-insulin antidiabetic medications increased the risk any and site-specific cancers, primarily colorectal cancer, with summary RRs ranging from 1.37-2.61.[94 98-100] However, it is important to note many of these observational studies have combined all types of insulins into one exposure definition which may not be clinically meaningful. As shown in table 2.1, each of the insulin therapies has different duration of effect in the human body, and it is possible that structurally modified human insulins such as long-acting insulin analogues may pose more of a risk for cancer compared with other insulins due to its altered pharmacokinetics.

Table 2.1 Characteristics of available insulin therapies

Insulin type (trade name)	Onset	Peak	Duration
Bolus (prandial) insulins			
Short-acting insulins			
Humulin R/S Novolin R	30 min	2-3 h	6.5 h
Rapid-acting insulin analogues			
Insulin aspart (NovoRapid)	10-15 min	1-1.5 h	3-5 h
Insulin lispro (Humalog)	10-15 min	1-2 h	3.5-4.75 h
Insulin glulisine (Apidra)	10-15 min	1-1.5 h	3-5 h
Basal insulins			
Intermediate-acting (NPH)			
Humulin N/I Novolin N	1-3 h	5-8 h	Up to 18 h
Long-acting insulin analogues			
Glargine (Lantus)			24 h
Detemir (Levemir)	30-90 min	None	16-24 h
Degludec (Tresiba)			> 24 h
Premixed insulins			
Regular insulin			
Humulin M2, M3, M5 Novolin 30/70, 40/60, 50/50	A single vial or cartridge contains a fixed ratio of insulin (% of short-acting or rapid-acting insulin to % intermediate-acting insulin or analogue and protamine)		
Insulin analogues			
Insulin aspart/aspart protamine (Novomix 30) Insulin lispro/lispro protamine (Humalog Mix25/Mix50)			

Adapted and modified from the Canadian Diabetes Association 2015 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada in Table 1.[87 103]
Onset: time to reach bloodstream and begins lowering glucose levels; Peak: time to reach maximum strength; Duration: length of time insulin continues to lower blood glucose levels; I=Isophane; M2=Mixed 20/80; M3=Mixed 30/70; M5=Mixed 50/50; N=Neutral Protamine Hagedorn or NPH; R=regular; S=short-acting

2.3. The rising safety concern for long-acting insulin analogues

2.3.1. The biological plausibility

Long-acting insulin analogues, structurally modified human insulins, were developed in the 1990s to help patients gain better glucose (HbA1c) control.[88] Long-acting insulin analogues are advantageous compared with the existing NPH insulin regarding reduced frequency of administration and adverse reactions to nocturnal hypoglycemia.[104 105] Nocturnal hypoglycaemia is defined as low blood glucose levels (≤ 63 mg/dL) occurring at night and has been reported to lead to fatal outcomes such as death.[105] Despite these advantages, safety concerns of long-acting insulin analogues have been raised. In the early 2000s, experimental studies suggested long-

acting insulin analogues can promote tumour growth due its altered pharmacokinetics. More specifically, there are two biological rationales as to how long-acting insulin analogues can cause an increased risk of cancer, especially in the breast.

First, the longer duration of effect of the long-acting insulin analogues in the human body means longer exposure to the insulin receptor family on cancer cells, allowing for further promotion of the tumour to clinically detectable levels. As further illustrated in table 2.1 and figure 2.3, long-acting insulin analogues have the longest duration of effect (up to 24 hours), as a result of the slower absorption, compared with endogenous insulin and even other insulin therapies such as NPH insulin.[87] Consequently, this longer duration of effect, resulting in longer exposure to peripheral tissues, could augment the risk of cancer.

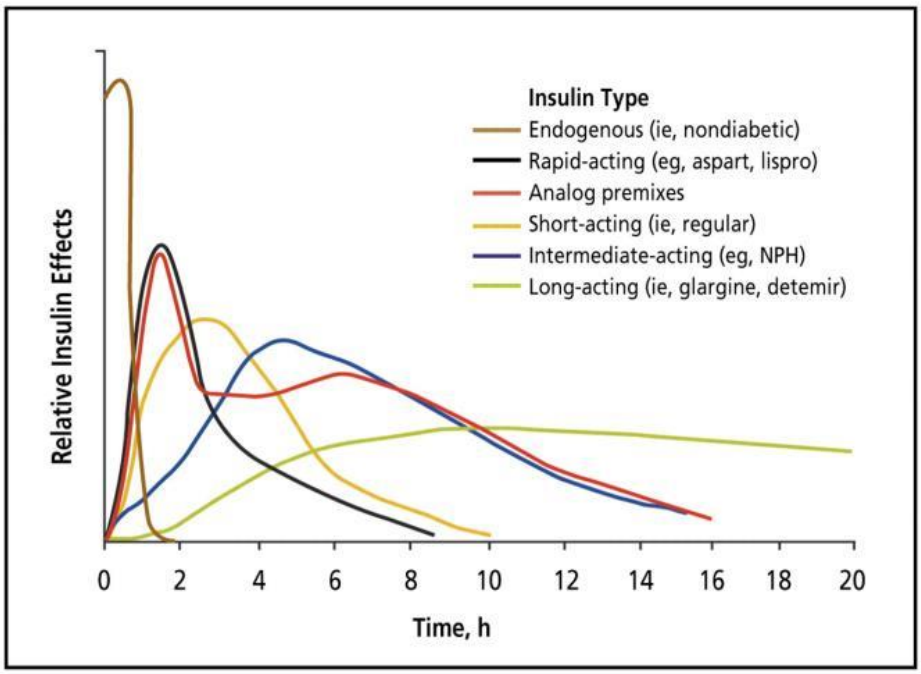


Figure 2.3 Duration of effect of each insulin therapy and endogenous insulin [106]

Second, the altered pharmacokinetics of the long-acting insulin analogues could allow for stronger binding to and prolonged residency on the IGF and insulin receptors, respectively, and thus increasing the mitogenic potential (i.e. increased cancer cell proliferation and anti-apoptotic effects). Many experimental studies have assessed the binding kinetics to both insulin and IGF receptors of long-acting insulin analogues in several cancer cell lines with some inconsistent findings.[88 107-113] In the beginning, insulin glargine but not detemir compared with human insulin was observed to have a 6 to 8-fold increase in binding affinity to IGF receptors.[113] In contrast,

another study of insulin glargine and detemir demonstrated that both were involved in tumour progression through IGF receptors.[109] It has been reported that insulin detemir binds to albumin, a blood protein, after injection and therefore it could reduce the amount of free insulin detemir available to target IGF receptors.[88] The amount of the binding of insulin detemir to albumin has not been thoroughly assessed in experimental studies. However, over the years, experimental studies have largely observed both long-acting insulin analogues bind to IGF receptors with greater affinity to activating the same MAPK pathways to initiate cancer cell proliferation and anti-apoptotic effects and primarily in breast cancer cell lines.[7] Some of the inconsistent results in the literature with insulin glargine could be because it gets immediately biotransformed in the body, at the site of the injection, to intermediate metabolites (M1 or M2 metabolites) and it has been shown to bind similarly as human insulin to the insulin receptor family.[108 110 114 115] However, it is uncertain the quantity of insulin glargine that gets biotransformed in the body after injection.[88] The dissociation rate of long-acting insulin analogues has not been studied extensively, but there have been a few studies suggesting this could also contribute to the mitogenic potential. The slower disassociation of the long-acting insulin analogues (1.5-3 times longer from insulin receptors) allows for the long-acting insulin analogues to continually upregulate the MAPK pathways.[88 116] Together, the stronger binding affinities to IGF receptors and prolonged residency on insulin receptors could enhance the mitogenic potential resulting in tumour growth, particularly breast.

2.3.2. Observational studies

The safety concerns of long-acting insulin analogues amplified when four cohort studies were published in Diabetologia reporting an increased risk of cancer, particularly breast, with insulin glargine use compared with other insulins in 2009.[14 15 19 117] Although these observational studies raised important safety considerations of the long-acting insulin analogues, these studies were criticised for their methodological shortcomings, which included reverse causation, did not account for cancer latency, inclusion of prevalent insulin users, and concerns about the data analysis.[26-30] Since 2009, a total of 21 cohort and case-control studies have been published on long-acting insulin analogues on the risk of any and common site-specific cancers (breast, colorectal, and prostate).[12-25 117-123] Of the 21 observational studies, only four have assessed the effect of insulin detemir on any cancer incidence.[16 118 121 122] The current evidence has shown some

inconsistencies between long-acting insulin analogues, specifically glargine, and any and site-specific cancer risk. However, the more recent observational studies may also have methodological limitations that need important consideration and be systematically assessed, particularly for biases that are prevalent in pharmacoepidemiology such as time-related biases (immortal time, time-window, and time-lag bias). Due to the potential for methodological shortcomings of the currently available studies, the evidence for the effect of long-acting insulin analogues on the risk of any and site-specific cancers remain unclear.

Despite the previously discussed methodological and analytical limitations of the existing literature on studies of long-acting insulin analogues and any and site-specific cancer incidence, six meta-analyses have been conducted in this area and focused on insulin glargine only.[7 92 93 95 124 125] More importantly, quality assessment of the observational studies in these meta-analyses were not conducted at all or adequately. The Newcastle Ottawa Scale (NOS) was the quality assessment tool that was primarily used in these previous meta-analyses. Although it assesses for the more typical epidemiological biases such as selection bias, information bias, and confounding it cannot identify biases in pharmacoepidemiology such as the aforementioned time-related biases.[126] Recently, Cochrane developed the Risk of Bias in Non-Randomised Studies – of Interventions (ROBINS-I), which is a more comprehensive quality assessment tool but it still does not address all the pharmacoepidemiology biases (e.g. immortal time, time-lag or time-window biases).[127] Pharmacoepidemiology biases are not necessarily different types of biases, but they are framed differently. For example, immortal time bias is a form of selection or information bias depending if person-time, in a cohort study using survival analysis, is excluded or misclassified, respectively, by design of the study and data analysis chosen.[128] Hence, current quality assessment tools may not be designed to allow the user (i.e. assessor of the study quality) to identify these subtle yet important variations in these biases. Consequently, most of these meta-analyses reported that the quality of these studies was moderate to high quality using these current quality assessment tools, including the previous highly criticised four cohort studies.[92 93 95] By restricting quality assessment to NOS,[92 93 95] or other quality assessment tools (e.g. Grading of Recommendations Assessment, Development and Evaluation),[7 124 125] previous systematic reviews and meta-analyses have not thoroughly assessed the quality of the existing literature due to many biases that are specific to pharmacoepidemiology studies are not captured in the quality assessment tools. Therefore, the full extent of the bias in the current observational studies of

long-acting insulin analogues and the risk of any and site-specific cancers may not have been evaluated.

2.3.3. Randomised controlled trials

In contrast to observational studies, RCTs assessing long-acting insulin analogues did not observe adverse events of any cancer.[9-11 129 130] However, there were some methodological shortcomings present in these RCTs. In several individual and pooled RCTs of either insulin glargine or detemir, these trials tended to have short follow-up (< 1 year) or small sample size (≤10,000 patients) reducing the ability to detect an effect.[8-10] In general, these trials were primarily designed to assess immediate changes to glucose levels in the blood (e.g. fasting glucose) and not aimed at long-term outcomes such as cancer. The most notable trial, Outcomes Reduction with Insulin Glargine Intervention (ORIGIN), had 12,537 patients and a median follow-up of 6.2 years with a large number of events with 476 cancers reported; an updated analysis with 906 incident or recurrent cancers still found no effect.[11 129] The ORIGIN trial had strengths including longer duration of follow-up compared with previous RCTs and adjudication of cancer events; however, the trial included both incident and recurrent cancer events and may not have been sufficiently powered for studying cancer-specific events. RCTs are considered the highest level of scientific evidence for estimating causal effects in epidemiology but it cannot be concluded from the RCTs presented thus far that long-acting insulin analogues have no effect on all and cancer-specific incidence given the methodological limitations. Furthermore, RCTs may not be an appropriate study design because of the rarity of cancer events and long latency period between first exposures to the antidiabetic medication such as long-acting insulin analogues to the development of cancer.

2.3.4. Drug regulatory agencies worldwide

Given the potential methodological issues present in many of the existing observational studies and RCTs on long-acting insulin analogues and cancer risk, drug regulatory agencies such as the U.S. Food and Drug Administration (FDA) and Health Canada, have not been able to conclude that long-acting insulin analogues do not increase the risk of cancer.[131-134] In contrast, the European Medicine Agency (EMA), has found the evidence inconclusive to suggest insulin glargine increases the risk of cancer.[135 136] All agencies have not suggested changing clinical

guidelines and practices, but FDA and Health Canada are requesting for more epidemiological evidence and continually reviewing the literature on this safety issue.[131-134]

2.4. Comparative effectiveness research

Comparative effectiveness research aims to evaluate the benefits and risks associated with new treatments compared with an active comparator or older treatments.[137] Comparative effectiveness research uses many types of study designs to assess new treatments including observational studies.[138] Comparative effectiveness observational studies may be better at assessing the benefits and risks associated with a new treatment in the real world setting because it is comprised of many types of patients including new and prior treatment users (i.e. patients without and with a history of treatment use). Guidelines for comparative effectiveness observational studies have been proposed and study design techniques such as a new user study design and matching on prior history of treatment have been recommended.[139-142] Both these study design techniques have been used in comparative effectiveness observational studies to assess the effect of insulin glargine on the risk of breast cancer.[12 13 15 16 23-25] However, new user study designs may be too restrictive and prevents the assessment of the safety profile of a new treatment in all patients in the real world setting, which is contrary to the goal of the comparative effectiveness research. Alternatively, study designs that match patients on prior history of treatment can include both new and prior treatment users,[142] but this method becomes challenging when there are more patients on the newer compared with older treatments over the calendar period. There is an increasing trend to prescribe long-acting insulin analogues due to its reduced adverse events of nocturnal hypoglycaemia compared with NPH insulin.[104 105] Future research to explore alternative approaches is needed to mitigate these issues with current study designs for comparative effectiveness research.

2.5. Summary

Elevated endogenous insulin and IGF levels bind to insulin and IGF receptors on breast cancer cell tissues predominantly, which can activate mitogenic cellular pathways leading to cellular proliferation and anti-apoptotic effects. Long-acting insulin analogues have altered pharmacokinetics, with respect to the binding affinity to IGF receptors, compared with human

insulins triggering similar mitogenic cellular activities primarily in breast cancer cells. Although there have been several cohort and case-control studies on long-acting insulin analogues and the risk of any and site-specific cancers, there are potential methodological shortcomings. Given the biological plausibility and current evidence, the relationship between long-acting insulin analogues and the risk of cancer, particularly breast, needs to be further elucidated. Lastly, comparative effectiveness observational studies tend to use study design techniques that do not allow the assessment of the safety profile of a drug in the entire patient population but only in subpopulations. Alternative study designs and data analytical techniques are needed in the context of long-acting insulin analogues and breast cancer.

Chapter 3. Data source

3.1. Overview of United Kingdom's Clinical Practice Research Datalink

The UK's CPRD, formerly known as the General Practice Research Database, was established in 1987.[143-145] The CPRD is a primary care electronic medical records database, which routinely collects data from consented general practices in England, Scotland, and Northern Ireland (figure 3.1).[143] As of 2013, there were 674 general practices comprising a total of 11.3 million patients, with 4.4 million whom are active patients (alive and currently registered) and have an average duration of follow-up of 9.4 years.[143] The CPRD represents 6.9% of the UK general population and is similar with respect to age, sex, and ethnicity.[143 146] From the beginning, CPRD was implemented in a general practitioner's (GP) office for both administrative and research purposes.[145] Presently, the CPRD is used by researchers worldwide on a variety of health-related topics such as pharmacoepidemiology.

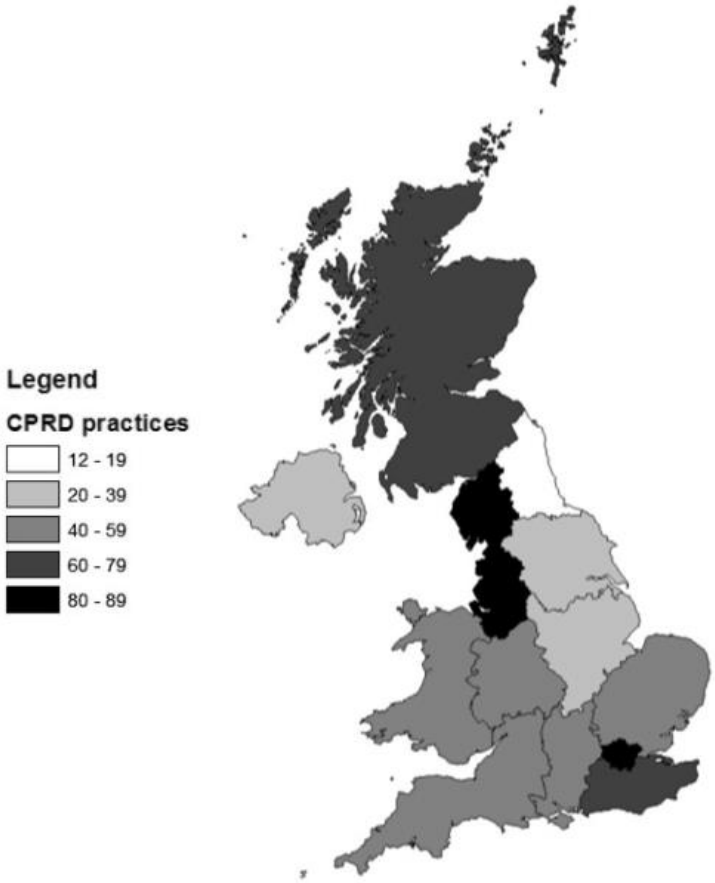


Figure 3.1 Geographical locations of the general practices in the CPRD in England, Scotland and Northern Ireland [143]

3.1.1. Data quality standards

To ensure that the data within CPRD are of high quality and complete, the CPRD provides GPs monetary incentives per patient per year.[144] Notably, the CPRD undergoes rigorous quality checks at the patient and practice level.

At the patient level, a patient could be considered 'acceptable' or not for research purposes (recorded as a binary variable in the patient file of the CPRD data).[143] A patient would be excluded or deemed unacceptable for research if the patient does not have continuous follow-up or poor data recording. More specifically, if the patient meets any of following exclusion criteria,[143] the patient would be considered unacceptable and not recommended for research:

- An empty or invalid first or current registration date
- No record for year of birth
- A first or current registration date before their birth year
- A transferred out reason but no transferred out date and vice versa
- A transferred out date before first or current registration date
- A current registration data before their first registration date
- A gender other than female/male/indeterminate
- Age > 115 years at the end of follow-up
- Recorded health care episodes in years before their birth year
- All recorded health care episodes have empty or invalid event dates
- Registration status of temporary patients

At the practice level, the practice is considered as 'up to standard' (UTS) if it has continuous high-quality data.[143] To be considered as a UTS practice, the practice is evaluated on the continuity of data based on two components: gap analysis and death recordings.[143 147] For the gap analysis, the purpose is to identify significant gaps between event recordings. A significant gap is defined as if there are five consecutive seven day windows where the recording of an event in each window is below 30% of the median number of events for the practice.[147] The earliest date after which no significant gaps are identified is defined as the gap date. For the death recordings, the CPRD practices are expected to have a similar number of deaths recorded over time. The gaps of death recordings are evaluated with a maximum allowable time of seven times the expected gaps based on the UK death rates with consideration of geographical and

seasonal variations, size of the practice, and a grace period of 31 days.[147] The earliest date after which no significant gaps in death are observed would be considered as the removal date. The UTS date (the date when practice is determined to be up to standard) is based on the gap or removal date, whichever comes first.[147]

3.1.2. Medical and product codes

In the CPRD, the data is primarily coded using the Read code classification system version 2, developed by Dr. James Read in 1982, with over 96,000 codes presently.[143 148 149] The Read or medical codes are used by GPs, other healthcare professionals, and administrators to describe a patient's condition, which includes, for example, symptoms and history, laboratory procedures, medical diagnoses, and so forth.[143 149] Read terms are organised by numerical and alphabetical chapters.[149] From chapters 0-9, the medical codes include history, examination, procedure and administration. From chapters A-Z, the medical codes include conditions, diagnoses, and injuries. Chapters A-Z are linked to the International Classification of Diseases (ICD).[148] For benign and malignant neoplasms specifically, chapter B is mapped to ICD for Oncology (ICD-O-3). Read codes contain five characters and are hierarchical with the first character being more general and every character or number added subsequently gets more specific. When GPs prescribe medication, prescriptions are automatically recorded with a product name, product code and British National Formulary code with dosage instructions (but not specific recommendations to each patient) and quantity.

3.1.3. Strengths and limitations

The strengths and limitations of the CPRD for research purposes have been highlighted previously.[143-145] In summary, the strength of the CPRD is that it is population-based, representative, has a large number of patients and follow-up, undergoes quality controls and checks, and has a unique set of variables and extensive data. First, aforementioned, the CPRD is a sample of the GP offices across three different countries, and it has been demonstrated to have similar age, sex, and ethnicity distributions as the UK general population-based on census data. Second, the CPRD currently includes over 11.3 million patients from nearly 700 practices and has up to 11.1 years of follow-up. This allows for research questions to evaluate rare exposures and outcomes (e.g. cancer). Third, the CPRD implements quality controls and checks at three

levels – physician, patient, and practice. Lastly, one unique aspects of CPRD is the availability of lifestyle or behaviour factors and laboratory values that are not typically available in health administration databases. Moreover, the CPRD can link with other data sources such as the Hospital Episode Statistics for hospitalisation data or the National Cancer Data Repository (NCDR).[143]

Despite the strengths, the CPRD does have a few limitations. First, not all data are captured in CPRD such as over-the-counter prescriptions and adherence. However, if the main exposure and other medications require a prescription and are a necessity for the management of a chronic disease (e.g. insulin therapy for patients with type 2 diabetes), this is not a major limitation. In addition, certain subpopulations such as patients in elderly homes are not included in the database. Second, medical reports on diagnoses from specialists (e.g. endocrinologists) must be manually inputted into the database. Consequently, certain information such as tests performed may be missing. Lastly, the data are not complete across all patients. For example, certain key variables such as smoking, blood pressure, BMI, alcohol intake, total cholesterol, and ethnicity were poorly recorded. However, starting in 2004, the Quality and Outcomes Framework was an initiative created to incentivize GPs to record information on common chronic diseases (e.g. diabetes, cancer), public health concerns (e.g. smoking status), and preventative measures (e.g. blood pressure).[150] Since the implementation of Quality and Outcomes Framework, the proportion of key variables have been on the rise (figure 3.2). Moreover, key variables recorded tended to be high for subpopulations. For example, among patients with type 2 diabetes, 97% had a recent recording for BMI.[151] Therefore missing data may not be too problematic in our studies since our cohorts aimed to include patients with type 2 diabetes using insulin therapy only in manuscripts 2-3.

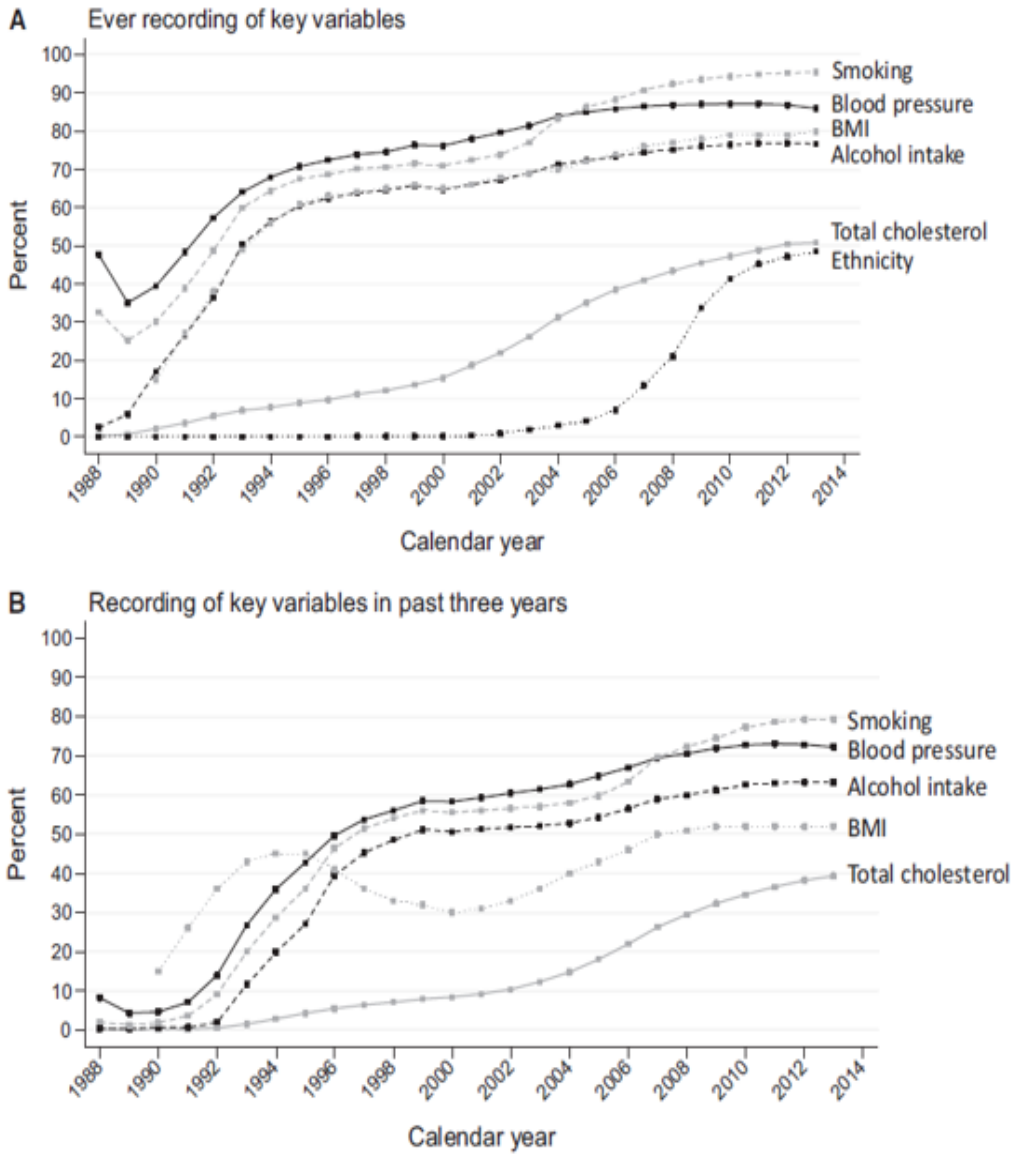


Figure 3.2 Proportion of recording for key variables since 1988 (A) and in the last three years (B) among acceptable patients ≥ 18 years or older [143]

3.1.4. Validation studies

Medical codes in the CPRD have been highly validated for various health-related behaviours (e.g. smoking) and health outcomes (e.g. cancer and acute myocardial infarction).[152-155] In a systematic review of validation studies of diagnostic codes in the CPRD, a variety of methods, including diagnostic algorithms, manual review of free text, sensitivity analyses, questionnaires sent to GP, medical record review, and comparison of rates of CPRD and non-

CPRD practices, were used to assess the validity of the diagnostic codes and determined a majority of the codes had approximately 90% or more confirmed diagnoses.[152-154] More specifically, diagnostic codes of many site-specific cancers among patients 40 years or older has been validated against the UK's NCDR, the gold standard cancer registry, and confirmed many cancer diagnostic codes (for cancer or cancer therapy) in the CPRD.[155] For breast cancer cases in the CPRD, ≥90% of cases were confirmed in the NCDR and over 63% of the cases were reported within the first month of the CPRD and NCDR databases.[155] Although prescriptions have not been validated in the CPRD, they are reported to be well documented by GPs and considered to be nearly complete with the exception for prescriptions provided in a secondary care setting (e.g. from specialists) or over-the-counter medications.[153]

3.2. Cohort formation and variable definitions

3.2.1. Base cohort formation

We discuss the base and study cohort study definitions and exclusion criteria in further detail in manuscripts 2-3. In this section, we describe in detail the start and end dates used to form the base cohort. For manuscripts 2-3, we created a base cohort of women 40 years or older with type 2 diabetes using any insulin therapy. To enter the base cohort, a woman must have at least one insulin prescription between the maximum start date of either the start date of CPRD (1 January 1988), date of first or current registration date plus one year, UTS date plus one year, or year of birth plus 40 years (i.e. women must be 40 years or older) and the minimum end date of either the last collection date (last date where data was collected from the practice), transferred out date (date where women left a CPRD practice), date of death, or end of recruitment period (31 December 2012). In addition, the patient must be considered acceptable. For the registration date, we used first registration date if the registration gap was ≤ 30 days or current registration date if registration gap was > 30 days. We illustrate in figure 3.3 the time period based on the maximum start and minimum end dates in which insulin prescriptions for one hypothetical patient would eligible for the base cohort study.

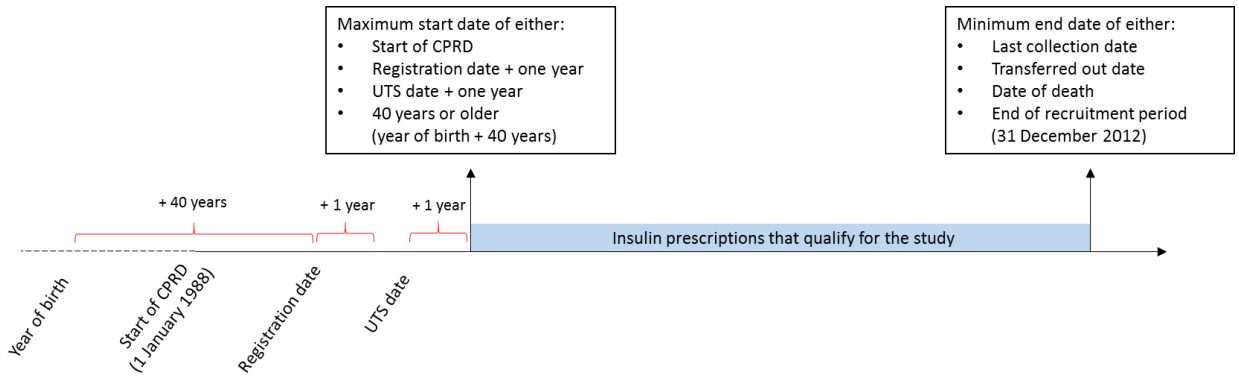


Figure 3.3 Hypothetical illustration of the time period insulin prescriptions for one patient would be eligible for the base cohort study

3.2.2. Exposure and comparator and outcome definitions

In table 3.1, we provided a list of product codes for both the exposure (long-acting insulin analogue glargine and detemir) and comparator (NPH insulin) definitions used in manuscripts 2-3. For insulin glargine, Abasaglar developed by Eli Lilly and Company Ltd and Toujeo developed by Sanofi (contains a higher dosage of glargine: 300 IU/mL) were recently approved by UK.[156 157] However, these types of insulin glargine were not available in the CPRD until after November 2015. Hence, these specific product codes (64354, 64723, 64987) were not included in the study as there were very few patients (< 300 therapy events) using this new brand or dosage of insulin glargine.

In table 3.2, we included a list of malignant breast cancer diagnostic codes from the CPRD for outcome definition of the first, primary malignant breast cancer in manuscripts 2-3. Aforementioned, the medical codes for incident breast cancers in CPRD have been previously validated with the NCDR, UK's cancer registry, with ≥90% of the cases confirmed.[155] We used 'B' medical codes only because previous studies that used a breast cancer algorithm, which used cancer diagnoses, procedures (e.g. radiotherapy and chemotherapy), and medications (e.g. tamoxifen) to identify breast cancer found over 95% of breast cancer cases had 'B' medical code for malignant breast cancer.[25 158]

3.2.3. Covariate definitions

We summarised the definition, type of variable, and lookback periods for each of the confounders in table 3.3. In this section, we describe in further detail how potential confounders

were defined for multivariable models in manuscripts 2-3. We used variable lookback windows instead of a time-fixed window to enhance the sensitivity and specificity of the confounder and subsequently reduce the potential for residual confounding.[159] [160] Previous studies have shown that using a longer (or all-available) lookback window enhances sensitivity and reduces bias.[161] [162] Although we did not have a gold standard (e.g. comparing CPRD medical codes to medical records from physicians) to compare the sensitivity and specificity of the confounders, we assessed the proportion of women that would be included by looking at different lookback periods (tables 3.4 and 3.5). We also discuss in the following sections the rationale for the definition of certain variables in further detail. For details on the formation of the study cohort and study cohort entry definition, please see sections 5.5 and 6.5 of this thesis.

Excessive alcohol use

For excessive alcohol use definition, we used the most recent medical code prior to study cohort entry for alcohol-related diseases (alcoholic fatty liver, hepatitis, cirrhosis, pancreatitis), symptoms, questionnaire for alcoholism, and stages of alcoholism (e.g. declining he or she is an alcoholism). These codes may not be diagnosed sequentially, and not all codes may be captured; therefore, it is important to use the entire lookback period to determine if women use excessive amounts of alcohol. Furthermore, there were more women included over the entire lookback period.

Smoking status

For smoking status, we defined women as an ever smoker if there was at least one code prior to study cohort entry in the entire lookback period for medical code for smoker, past smoker, occasional smoker, and smoking cessation medications (nicotine or bupropion hydrochloride) prior to cohort entry and a never smoker was the most recent code for non-smoker and no code for ever smoker. We used the entire lookback period because we captured more women that smoked earlier in life. Moreover, smoking and cancer has a long latency period and thus it is important to consider all-available lookback period prior to cohort entry.

Body mass index

For BMI, we used the average height because recent weights (defined as last weight recorded prior to study cohort entry) could not be used since the recent height (recorded on the same day) was excluded because it was not within acceptable ranges of 1-3 metres or 4-7 feet. However, other previously recorded heights were within acceptable ranges. Therefore, we averaged all previously recorded heights within the acceptable range and used the most recent weight and averaged heights to calculate the most recent BMI. We considered using obesity medical codes as well to determine if the woman was normal, overweight or obese based on predefined medical codes. However, the use of these medical codes would not have provided an exact BMI value. In addition, the number of additional women that it would have been included was 1.8% (n=55). We used a lookback period of five years because the proportion of women included was over 90%.

Glycated haemoglobin

For HbA1c, we used the most recent laboratory tests, defined as entity type 275 in the CPRD, in the last two years prior to study cohort entry and percentage unit (%) as the main unit in manuscripts 2-3. In October 2011, the UK converted to using the new international HbA1c unit, mmol per mol (mmol/mol), as proposed by the International Federation of Clinical Chemistry and Laboratory Medicine.[163 164] Therefore, we converted all laboratory tests in the new international unit to the percentage unit since the recruitment period was between 2002-2012 when the percentage unit was used primarily. Each laboratory tests in the CPRD have an assigned specimen unit of measure code in the test files. The following unit codes were accepted and included in our study: 1, 96, 97, 205, and 215. Specifically, the unit codes 96, 97, and 205 were in mmol/mol and converted to a percent using the following equation: hba1c_percent = (hba1c_mmolmol / 10.929) + 2.15.[163] We used a lookback period of two years because a recent HbA1c may be more likely to determine the type of basal insulin women would be prescribed.

Diabetes duration

For diabetes duration, we used either of the first of non-insulin antidiabetic medication prescription, any insulin prescription, type 2 diabetes diagnosis, and HbA1c test with a value of ≥ 6.5%, whichever occurred first. The decision to use the earliest of any of these diagnostic,

laboratory, and prescription dates was because women with type 2 diabetes may be seen at endocrinologists. For example, a woman with a high HbA1c value could precede the diagnosis of type 2 diabetes (possibly diagnosed at the referred endocrinologists) and, therefore, the HbA1c test could represent the start of type 2 diabetes. Moreover, it is also possible the diagnostic code for type 2 diabetes may not be retroactively recorded by the GP in the CPRD.[143] We evaluated in our second manuscript the proportion of women with a diagnosis, prescription or laboratory tests as the start of type 2 diabetes. We observed 21.0% had non-insulin antidiabetic medication prescription, 17.9% had an insulin prescription, 47.2% had a type 2 diabetes diagnosis, and 14.0% had an HbA1c test to define the start date of type 2 diabetes.

Medication use

For all medications, we used the most recent prescription prior to cohort entry to define exposure status to the medication. Moreover, a lookback period of two years was used because medications used just prior to study cohort entry may be a good proxy for disease severity, particularly for non-insulin antidiabetic medications.

Deyo's Charlson comorbidity score

We evaluated three lookback periods only: one year, five years and all-available. We decided to use the one year prior to cohort entry because Deyo's Charlson comorbidity score also used a one year lookback period.[165] Furthermore, there was a greater proportion of women with CCS compared with using a long lookback period (table 3.5). We modified the disease risk score to exclude previous history of any cancers as this was an exclusion criterion in manuscripts 2-3.

3.3. Additional information on exposure, outcome, and covariate definitions

Table 3.1 Product codes for basal insulins

Basal insulin	Description	Product codes
Long-acting insulin analogues (exposures of interest)		
Glargine/Lantus	Developed by Sanofi; Strength: 100IU/mL; each cartridge/vial contains 3mL of solution or 300 IU of glargine	5953, 6057, 7237, 7266, 7393, 7400, 7402, 10225, 10259, 36853, 49831, 50633, 56495
Detemir/Levemir	Developed by Novo Nordisk; 100IU/mL; each cartridge/vial contains 3mL of solution or 300 IU of detemir	6958, 6965, 10184, 14301, 14330, 35260, 55618
Intermediate-acting NPH insulin (active comparator)		
Contains NPH only	Humulin I: Developed by Eli Lily and Company Ltd; 100 IU/mL; each cartridge/vial contains 3 or 10 mL of solution	1839, 4760, 10229, 14357, 14918, 43950
	Insulatard: Developed by Novo Nordisk; 100 IU/mL; each cartridge/vial contains 3 or 10 mL of solution	1593, 1595, 1886, 5891, 9737, 10208, 14290, 14928, 24866, 33966, 52748
	Insuman Basal: Developed by Sanofi; 100 IU/mL; each cartridge/vial contains 3 or 10 mL of solution	5501, 23992, 27461, 35468, 46001
	Other non-specific brands	1587, 1643, 1649, 1844, 4248, 7537, 7765, 7771, 7772, 7783, 8322, 8376, 8646, 9376, 10175, 10207, 10547, 10691, 11080, 13729, 14925, 15040, 15624, 15961, 18461, 18645, 18931, 19829, 20671, 22496, 22823, 23003, 25812, 26784, 30861, 34097, 55517, 59500
Premixed	Humulin M: Developed by Eli Lily and Company Ltd; NPH: 70 IU/mL, Human: 30 IU/mL; each cartridge/vial contains 3 or 10 mL of solution	4093, 4198, 4199, 8841, 10277, 10545, 10546, 10566, 11107, 16160, 19513, 22094, 22161, 43991, 57620, 60933
	Insuman comb: Developed by Sanofi; NPH: 85/75/50 IU/mL, Human: 15/25/50 IU/mL; each cartridge/vial contains 3 or 10 mL of solution	15199, 20422, 21554, 24002, 24993, 25133, 30819, 31205, 35253, 44480, 45158
	Other non-specific brands NPH: 90/85/75/50 IU/mL, Human: 10/15/25/50 IU/mL	1805, 1806, 2220, 2221, 2454, 2455, 2456, 2812, 2929, 3396, 3439, 3550, 3551, 4790, 5255, 5845, 5933, 7231, 7300, 7319, 7959, 8203, 8322, 8354, 9341, 10244, 10245, 10484, 10887, 11055, 11056, 12818, 13277, 13416, 13550, 13837, 14644, 14649, 16152, 17731, 18461, 19878, 21110, 21232, 21347, 21374, 21395, 21422, 22058, 22697, 24722, 25735, 25736, 26403, 26784, 27614, 28096, 28978, 29837, 31267, 33167, 33232, 36194, 41120, 42954, 44378, 50691, 52722, 54462, 56857, 60938

Table 3.2 Diagnostic codes for malignant breast cancer

Medical code	Read code	Read term
348	B34..11	Ca female breast
3968	B34..00	Malignant neoplasm of female breast
9470	B34z.00	Malignant neoplasm of female breast NOS
12499	Byu6.00	[X]Malignant neoplasm of breast
20685	B346.00	Malignant neoplasm of axillary tail of female breast
23380	B340000	Malignant neoplasm of nipple of female breast
23399	B344.00	Malignant neoplasm of upper-outer quadrant of female breast
26853	B340.00	Malignant neoplasm of nipple and areola of female breast
29826	B342.00	Malignant neoplasm of upper-inner quadrant of female breast
31546	B341.00	Malignant neoplasm of central part of female breast
38475	B34yz00	Malignant neoplasm of other site of female breast NOS
42070	B345.00	Malignant neoplasm of lower-outer quadrant of female breast
45222	B343.00	Malignant neoplasm of lower-inner quadrant of female breast
49148	B347.00	Malignant neoplasm, overlapping lesion of breast
56715	B34y.00	Malignant neoplasm of other site of female breast
59831	B340z00	Malignant neoplasm of nipple or areola of female breast NOS
64686	B340100	Malignant neoplasm of areola of female breast
95057	B34y000	Malignant neoplasm of ectopic site of female breast

Abbreviations: Ca – cancer; NOS – not otherwise specified; [X] – terms that have been added to the Read Codes to ensure that every ICD-10 code is cross-mapped to a Read Code.

Table 3.3 Definition and lookback periods used for potential confounders in multivariable models

Covariate	Definition	Lookback period*	Type of variable†
Age	Year of study cohort entry minus year of birth	Defined at study cohort entry	Continuous; unit: years
Calendar year	Year of study cohort entry	Defined at study cohort entry	Continuous; unit: years
Excessive alcohol use	Most recent medical code for alcohol-related diseases (cirrhosis, fatty liver, hepatitis, pancreatitis), symptoms, questionnaire for alcoholism, or stages of alcoholism	Anytime until 1/1/1988	Binary: yes/no
Smoking status	Ever user: at least one medical codes for smoker, past smoker, and smoking cessation (nicotine or bupropion hydrochloride) prior to cohort entry; Never smoker: most recent code for non-smoker and no code for ever smoker	Anytime until 1/1/1988	Categorical: never (referent), ever, unknown
BMI	Weight: Most recent prior to cohort entry weight within acceptable range ≥ 40 kg [166] Height: Average of all heights prior to cohort entry recorded within acceptable range 1-3 m (4-7 ft) [166]; BMI=weight/(height)^2	5 years	Categorical: 18-25 (referent), 25-30, ≥ 30, unknown; unit: kg/m^2
HbA1c	Most recent prior to cohort entry entity type (laboratory test; type=275) for HbA1c values HbA1c values with mmol/mol units were converted to %: hba1c_per=(hba1c_mmol/10.929)+2.15	2 years	Categorical: <6.5 (referent), 6.5-8.0, >8.0; unit: %
Diabetes duration	Study cohort entry date minus date of first of either of the following events: non-insulin antidiabetic medication prescription, any insulin prescription, type 2 diabetes diagnosis, HbA1c clinical values	Anytime until 1/1/1988	Continuous; unit: years
Prior insulin use duration	Study cohort entry date minus date of first-ever insulin prescription	Anytime until 1/1/1988	Continuous; unit: years
Metformin	Presence of most recent metformin prescription	2 years	Binary: yes/no
Sulfonylureas	Presence of most recent sulfonylurea prescription	2 years	Binary: yes/no
TZD	Presence of most recent TZD prescription	2 years	Binary: yes/no
Other	Presence of most recent other non-insulin antidiabetic medication prescription prior to study cohort entry included: meglitinides, DPP-4i, GLP-1, alpha-glucosidase inhibitors, guar gum, and SGLT-2i	2 years	Binary: yes/no
Charlson comorbidity score	Deyo's method but modified to exclude any cancers [165]	1 year	Categorical: ≤ 1 comorbidities (referent), 2-3, >3
Statins	Presence of most recent statins prescription	2 years	Binary: yes/no
Aspirin	Presence of most recent aspirin prescription	2 years	Binary: yes/no
NSAID	Presence of most recent NSAID prescription	2 years	Binary: yes/no

*Lookback period starts from the date of study cohort entry. †Binary variables used 'no' as the referent.
Abbreviations: BMI – body mass index; DPP-4i – dipeptidyl-peptidase-4 inhibitor; GLP-1 – glucagon-like peptide-1 analogues; HbA1c – glycated haemoglobin; NSAID – non-steroidal anti-inflammatory drug; SGLT-2i – sodium-glucose cotransporter 2 inhibitors; TZD – thiazolidinedione

Table 3.4 Proportion of women with type 2 diabetes using basal insulin with the covariate based on different lookback periods

Covariates	At study cohort entry	1 year	2 years	3 years	4 years	5 years	All-available*
Alcohol-related disease	0.5	2.1	2.7	3.0	3.2	3.4	4.1
Smoking status†	45.9	82.6	88.2	90.9	92.4	93.5	97.9
BMI†	53.7	84.3	89.0	90.9	91.9	92.7	95.3
HbA1c†	65.7	90.6	92.6	93.3	93.7	93.9	94.3
Metformin	27.2	33.9	36.1	37.7	39.0	40.0	43.5
Sulfonylureas	19.3	23.8	25.9	27.6	29.6	31.2	39.0
TZD	6.5	9.7	11.3	12.3	13.0	13.4	14.1
Other ADM	4.7	6.2	6.8	7.3	7.9	8.4	10.8
Statins	59.7	68.6	70.3	71.1	71.6	71.9	72.5
Aspirin	41.0	49.5	51.9	53.5	54.6	55.5	58.1
NSAID	16.5	29.9	37.7	43.5	48.6	52.3	66.3

*All-available lookback period is up until 1 January 1988; †Missing values accounted for ~5% of all women; NA=not available; Abbreviations: ADM – antidiabetic medications; BMI – body mass index; HbA1c – glycated haemoglobin; NSAID – nonsteroidal anti-inflammatory drugs; TZD – thiazolidinedione

Table 3.5 Distribution of Charlson comorbidity score (CCS) based on different lookback periods starting from study cohort entry

CCS	1-year		5-years		All-available*	
	N	%	N	%	N	%
1	11,571	51.7	7551	33.7	5927	26.5
2	8,032	35.9	8823	39.4	8634	38.6
3	2,792	12.5	6021	26.9	7834	35.0

*All-available lookback period is up until 1 January 1988

Chapter 4. Manuscript 1 – The effect of long-acting insulin analogues on the risk of cancer: a systematic review of observational studies

4.1. Preamble

In the background chapter, we highlighted many observational studies on long-acting insulin analogues on the risk of any and site-specific cancers have been conducted; however, the extent of the methodological issues of these studies have not been fully assessed especially for pharmacoepidemiology biases. Therefore, the primary motivation for this following systematic review was to evaluate the association between long-acting insulin analogues and the risk of any and site-specific cancers with a particular focus on methodological strengths and limitations using our quality assessment criteria. After reviewing the current quality assessment tools, including the Cochrane's quality assessment tool for observational studies on interventions (ROBINS-I), we determined the existing tools do not consider all the biases in pharmacoepidemiology. Therefore, to fully assess the quality of the observational studies on long-acting insulin analogues and cancer risk in our systematic review, we developed our own criteria focusing on several biases that are prevalent in pharmacoepidemiology including time-related biases (immortal time, time-lag, and time-window), prevalent user bias, inclusion of lag periods, and length of follow-up between treatment initiation and outcome of interest.

In 2015, this manuscript from my thesis, entitled "The effect of long-acting insulin analogues on the risk of cancer: a systematic review of observational studies", was published in the Diabetes Care journal.[31] After the publication of my systematic review, two additional observational studies were published on long-acting insulin analogues and the risk of cancer.[12 123] Similar to the findings of our systematic review, these studies also had methodological limitations including short follow-up and had inappropriate comparators.

The findings from this systematic review would demonstrate that there is still uncertainty regarding whether insulin glargine increases the risk of site-specific cancers, particularly with the breast. Furthermore, it would identify that currently there are no studies on insulin detemir and the risk of breast cancer. In addition, it highlights methodological issues identified in this area of research that still needs to be addressed.

4.2. Title page

Title: The effect of long-acting insulin analogues on the risk of cancer: a systematic review of observational studies

Authors: Jennifer W. Wu MPH[1, 2], Kristian B. Filion PhD[1,2,3], Laurent Azoulay PhD[1,2,4], Margaret K. Doll MPH[1], and Samy Suissa PhD[1,2,3]

Affiliations:

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Quebec, Canada

[2] Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, Montreal, Quebec, Canada

[3] Division of Clinical Epidemiology, Department of Medicine, McGill University, Montreal, Quebec, Canada

[4] Department of Oncology, McGill University, Montreal, Quebec, Canada

Corresponding author:

Dr. Samy Suissa
Centre for Clinical Epidemiology
Lady Davis Institute, Jewish General Hospital
3755 Cote Ste-Catherine, H-461
Montreal, Quebec, Canada, H3T 1E2
Tel: (514) 340-7593
Fax: (514) 340-7564
Email: samy.suissa@mcgill.ca

4.3. Abstract

Background: Observational studies examining the association between long-acting insulin analogues and cancer incidence have produced inconsistent results. We conducted a systematic review of these studies, focusing on their methodological strengths and weaknesses.

Research design and methods: We systematically searched MEDLINE and EMBASE from 2000 to 2014 to identify all observational studies evaluating the relationship between the long-acting insulin analogues and the risk of any and site-specific cancers (breast, colorectal, prostate). We included cohort and case-control studies published in English on insulin glargine and detemir and any cancer incidence among patients with type 1 or 2 diabetes. The methodological assessment involved the inclusion of prevalent users, inclusion of lag periods, time-related biases, and duration of follow-up between insulin initiation and cancer incidence.

Results: A total of 16 cohort and 3 case-control studies met our inclusion criteria. All studies evaluated insulin glargine, while four studies were also examining insulin detemir. Follow-up ranged 0.9 to 7.0 years. Thirteen out of fifteen studies reported no association between insulin glargine and detemir and any cancer. Four out of thirteen studies reported an increased risk of breast cancer with insulin glargine. In the quality assessment, 7 studies included prevalent users, 11 did not consider a lag period, 6 had time-related biases, and 16 had short (< 5 years) follow-up.

Conclusion: The observational studies examining the risk of cancer associated with long-acting insulin analogues have important methodological shortcomings that limit the conclusions that can be drawn. Thus, uncertainty remains, particularly for breast cancer risk.

4.4. Introduction

Neutral Protamine Hagedorn (NPH) insulin has been the mainstay treatment for type 1 diabetes and advanced type 2 diabetes since the 1950s. However, this insulin is associated with an increased risk of nocturnal hypoglycemia, and its relatively short half-life requires frequent administration.[6 82] Consequently, structurally modified insulins, known as long-acting insulin analogues (glargine and detemir) were developed in the 1990s to circumvent these limitations. However, there are concerns that long-acting insulin analogues may be associated with an increased risk of cancer. Indeed, some laboratory studies showed, long-acting insulin analogues were associated with cancer cell proliferation and protected against apoptosis via their higher binding affinity to IGF-I receptors.[109 113]

In 2009, four observational studies associated the use of the insulin glargine with an increased risk of cancer.[14 15 19 117] These studies raised important concerns, but were also criticised for important methodological shortcomings.[26-28 30 167] Since then, several observational studies assessing the association between long-acting insulin analogues and cancer have been published but yielded inconsistent findings.[13 16-18 20-25 118-122] Such discrepancies may be due to methodological limitations, including inadequate durations of follow-up between insulin initiation and cancer incidence, protopathic bias, detection bias, the inclusion of prevalent users, and time-related biases such as immortal time bias, time-window bias, and time-lag bias.[126]

Randomised controlled trials (RCT) have reported the effects of long-acting insulin analogues on the risk of any cancers,[8-10] but most of these RCTs were designed to study efficacy (e.g., fasting plasma glucose level) and not designed to assess cancer. The most notable RCT, the Outcomes Reduction with Insulin Glargine Intervention (ORIGIN) trial, did not observe an effect of insulin glargine on the composite outcome of any cancer.[11] The ORIGIN trial had several strengths, including the power to detect a clinically important effect of insulin glargine on any cancer, adjudication of cancer outcomes, and follow-up was longer compared with previous RCTs (median: 6.2 years). However, it was not powered to detect site-specific cancers, combined both new and recurrent cancers, and follow-up was still considerably short given the long latency of cancer (follow-up duration up to 7 years only).

Several meta-analyses of observational studies have investigated the association between insulin glargine and cancer risk.[7 93 124 125] These previous meta-analyses assessed the quality of

included studies, but the methodological issues particular to pharmacoepidemiologic research were not fully considered. In addition, given the presence of important heterogeneity in this literature, the appropriateness of pooling the results of these studies remains unclear. We, therefore, conducted a systematic review of observational studies examining the association between long-acting insulin analogues and cancer incidence, with a particular focus on methodological strengths and weaknesses of these studies.

4.5. Methods

This systematic review was conducted following a prespecified protocol and reported following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses guidelines.[168]

Search strategy

We systematically searched MEDLINE and EMBASE via Ovid from 1 January 2000 to 8 October 2014 for observational studies examining the association between long-acting insulin analogues and cancer incidence. The detailed search strategy is reported in Supplementary Table 1. Briefly, the search included MeSH terms, Emtree terms, and keywords for diabetes, long-acting insulin analogues, neoplasm, and observational studies. The publication type search terms used in this search strategy were adopted from the Scottish Intercollegiate Guidelines Network group.[169] The search was limited to articles published from 2000 onwards because long-acting insulin analogues were not available globally until after 2000. Our search was also limited to studies published in English. We hand-searched relevant systematic reviews and meta-analyses to identify additional articles that were not identified in our electronic literature search.

Inclusion and exclusion criteria

Cohort, case-control, and case-cohort studies evaluating the association of long-acting insulin analogues (glargine and detemir) and cancer incidence among patients with type 1 or 2 diabetes were eligible for inclusion. Inclusion was restricted to studies reporting any incident cancer or site-specific cancers as primary or secondary outcomes. Studies that did not exclude

prevalent cancer cases were eligible for inclusion. We excluded studies that did not meet these inclusion criteria.

The literature search was conducted independently by two reviewers (JWW and MKD), who assessed the titles and/or abstracts of identified publications. The full-text of any publication deemed potentially relevant by either reviewer at this stage was then retrieved for detailed review. Discrepancies in determining whether the study met our inclusion criteria during full-text review were resolved by consensus or, when necessary, a third reviewer (KBF).

Data extraction and quality assessment

We developed a data extraction form, which was pilot-tested on six included studies. Two independent reviewers (JWW and MKD) extracted data, with disagreements resolved by consensus or a third reviewer (KBF or SS). Disagreements could have occurred when extracting individual data points (e.g., study characteristics and measures of association) or when evaluating the quality of the studies.

Extracted information included: 1) study characteristics (source population, country, sample size, study design, type of database used to ascertain information about exposure and outcome); 2) patient characteristics (age); 3) exposure and comparator definitions (ever vs. never use, duration of use, dose, use of time-independent or dependent approach); 4) incidence of any and/or site-specific cancers; 5) odds ratios, risk ratios, rate ratios, or hazard ratios (HR) with corresponding 95% CIs; 6) methods of adjustment for confounders (matching, regression-based adjustments, propensity scores, disease risk scores) and list of potential confounders; and 7) quality of the studies. We extracted any site-specific cancer but did not report on relative risks (RR) for sites that were not commonly reported in the included studies.

No available quality assessment tool adequately captures the methodological issues and biases that are particular to pharmacoepidemiology. Therefore, we assessed the quality of studies for key components, including time-related biases (immortal time, time-lag, and time-window), inclusion of prevalent users, inclusion of lag periods, and length of follow-up between insulin initiation and cancer incidence.

Immortal time bias is defined by a period of unexposed person-time that is misclassified as exposed person-time or excluded, resulting in the exposure of interest appearing more favourable.[128 170] Time-lag bias occurs when treatments used later in the disease management

process are compared with those used earlier for less advanced stages of the disease. Such comparisons can result in confounding by disease duration or severity of disease if duration and severity of disease are not adequately considered in the design or analysis of the study.[126] This is particularly true for chronic disease with dynamic treatment processes such as type 2 diabetes. Currently, American and European clinical guidelines suggest using basal insulin (e.g., NPH, glargine and detemir) as the last line of treatment if glycated haemoglobin (HbA1c) targets are not achieved with other antidiabetic medications.[83] Therefore, studies that compare long-acting insulin analogues to non-basal insulin may introduce confounding by disease duration. Time-window bias occurs when the opportunity for exposure differs between cases and controls.[126 171]

Prevalent user bias occurs when new and prevalent treatment (e.g., insulin) users are combined and prior history of treatment use is not accounted for in the study design or analysis.[28 141] Including prevalent users without accounting for the prior history of treatment use in a study can result in under ascertaining the outcome of interest.[141] In a cohort study, new treatment users are defined as a patient without a history of treatment use for a specified time period (e.g., one year) prior to cohort entry whereas prevalent treatment users are patients with a history of treatment use within the specified time period. In a case-control study, history of treatment use would be assessed prior to the exposure time window.[141]

The importance of considering a lag period is necessary for latency considerations (i.e., a minimum time between treatment initiation and the development of cancer) and to minimise protopathic and detection bias. Protopathic bias, or reverse causation, is present when a medication (exposure) is prescribed for early symptoms related to the outcome of interest, which can lead to an overestimation of the association. Lagging the exposure by a predefined time window in cohort studies or excluding exposures in a predefined time period prior to the event in case-control studies is a means of minimising this bias.[172] Detection bias is present when the exposure leads to higher detection of the outcome of interest due to the increased frequency of clinic visits (e.g., newly diagnosed patients with type 2 diabetes or new users of another antidiabetic medication), which also results in an overestimation of risk.[173] Thus, by including a lag period, such as starting follow-up after one year of the initiation of a drug, simultaneously considers a latency period while also minimising protopathic and detection bias.

We also assessed the studies for traditional epidemiological biases such as selection bias, information bias, and confounding. For confounding, we considered three potential sources: 1)

imbalances between measured baseline covariates that were not addressed in the study design or analytically, 2) residual confounding due to unmeasured confounders, and 3) lack of adjustment for time-dependent confounders.

This assessment focused on the discussion of key components of design and analysis rather than the creation of an aggregate score, as has been suggested elsewhere.[174] We used the primary analysis of each included study for the qualitative assessment, but if the issue or bias was addressed in an appropriate sensitivity analysis, we considered this in the qualitative assessment.

Data analysis

Given the methodological focus of this review and heterogeneity among published studies, we conducted a systematic review without meta-analysis. Nonetheless, forest plots were constructed with Stata version 13 (StataCorp, College Station, TX, USA) to present the available data graphically. Although each observational study presented multiple effect estimates, in each of the forest plots of long-acting insulin analogues and any and commonly reported site-specific cancer incidence, we displayed one effect estimate from each observational study that yielded minimal biases.

4.6. Results

Study selection

Our search of MEDLINE and EMBASE yielded a total of 4,417 potentially relevant articles (Supplemental Fig. 1). Following our inclusion criteria, 16 cohort and 3 case-control studies were included in this systematic review.[13-25 117-122] All studies evaluated insulin glargine, with four studies also investigating insulin detemir.[16 118 121 122]

Study characteristics and effect estimates

The study populations ranged from 1,340 to 275,164 patients (table 1). The mean or median durations of follow-up and age ranged from 0.9 to 7.0 years and from 52.3 to 77.4 years, respectively. Thirteen studies examined ever use of long-acting insulin analogues, which was defined as at least one prescription, compared to non-use, other, human, or NPH insulin.[13-15 17-19]

[21 22 24 25 117 119 121] One study examined the duration of time since starting long-acting insulin analogues, and one examined mean daily dose.[118 120] Four studies used time-dependent exposure definitions.[16 20 23 122] All included studies evaluated cancer incidence as a primary outcome.

Of the 16 studies that evaluated the relationship between long-acting insulin analogues and any, colorectal, and/or prostate cancer, 13 reported no associations (Figure 1 and Supplemental Figure 2).[13 14 16 18-22 24 118 119 121 122] Four out of thirteen studies reported an association between insulin glargine and breast cancer in the overall study population or among new insulin users only.[18 19 21 23] Moreover, two of the four studies found an association among new insulin users. [18 23] In two other studies, insulin glargine was also observed to be associated with an increased risk of breast cancer among the entire study population (i.e. combining both new and prevalent insulin users) (RR: 3.65, 95% CI: 1.05-12.68) or among prevalent insulin users only (RR: 2.70, 95% CI: 1.10-6.50).[14 25] The additional effect estimates from these studies were not displayed in the forest plots because we opted to present the effect estimate from each study that tended to reduce the prevalent user bias (e.g., study employed the new user study design).

Quality assessment

The different key components of the quality assessment are summarised in table 2 and discussed in detail below.

Immortal time bias

Of the 19 studies in this review, immortal time bias may have been introduced in one study based on the time-independent exposure and cohort entry definitions that were used in this cohort study.[119] For the exclusive user definition, patients needed to have either insulin glargine or human insulin only between the first and last prescription to be considered exposed to that one insulin only. However, the follow-up started from the first insulin prescription and as a result, the time before the last insulin prescription was misclassified as exposed when it should be classified as unexposed. Similarly, for the predominant user definition, the patient needed to have at least 12 prescriptions of insulin and be exposed 80% of the follow-up time to be considered exposed but the time before the twelfth prescription and meeting the 80% exposure time should be

considered unexposed (as depicted in Supplemental Fig. 3). As a result, the adjusted HRs for any cancer were ~0.60, although the results were not statistically significant.

Time-lag bias

Time-lag bias may have occurred in four studies that compared insulin glargine to human or other (non-basal) insulin or highest-to-lowest duration of insulin use without adjusting or matching on diabetes duration.[22 117-119] The presence of time-lag bias is well illustrated in a cohort study in which individuals who received human insulin or any insulin analogue for the first time were included in the cohort.[117] Such individuals could be at earlier stages of the disease than those who received insulin glargine (as depicted in Supplemental Figure 4). Unfortunately, diabetes duration was not reported. This study observed an association between insulin glargine and cancer was observed in this study (HR: 1.19; 95% CI: 1.09-1.29), but it is possible that more cases of cancer occurred in the insulin glargine group due to the longer diabetes duration rather than due to exposure to insulin glargine.

A variation of time-lag bias was observed in a cohort study of new insulin users.[118] For the exposure definition, highest duration since the start of insulin use was compared to the lowest. It is expected that the risk of cancer would increase with longer duration of insulin use; however, the opposite was reported (with RRs ranging from 0.50 to 0.90). The protective association observed could be due to competing risks (e.g., death from cardiovascular-related events).[175 176] Patients with diabetes have a higher risk of cardiovascular-related deaths compared to patients with no diabetes.[177 178] Therefore, patients with diabetes who die from cardiovascular-related events do not have the opportunity to develop cancer, resulting in an underestimation the risk of cancer.

Time-window bias

Time-window bias was observed in two studies.[17 120] In one of two studies, despite matching on calendar time, time-window bias was potentially present because cases and controls were not matched on diabetes duration (as depicted in Supplemental Figure 5).[17] Consequently, the opportunity for exposure differed between the cases and controls due to the varying diabetes durations (a mean of 14.5 years among cases and 13.2 among controls). Although one would expect an increased risk due to the presence of time-window bias, a null effect was observed.

This suggests that other biases, such as selection bias resulting from the selection of cases and controls from different study bases, may also be present.

Residual confounding

We evaluated the patient characteristics in each of the 19 studies and observed that the measured covariates (e.g., age, sex, HbA1c, diabetes duration, comorbidities, prior medication use, smoking status, and/or alcohol use) were generally balanced between groups (either exposed versus comparator or cases versus controls, depending on the study design). Residual confounding may have been present in all 19 studies given the data sources used and methods to deal with unmeasured confounding such as instrumental variables were not used in any of these.[179] More specifically, residual confounding may have resulted due to the presence of unmeasured confounders. HbA1c and diabetes duration were not accounted for in 15 out of 19 studies, resulting in likely residual confounding.[13 16 17 19-24 117-122] In addition, residual confounding may have occurred in nine studies using a time-dependent exposure because none of these studies adjusted for time-dependent covariates such as the addition of short-acting insulins or other antidiabetic medication (e.g., metformin) at all or appropriately (i.e., used a marginal structural model and inverse probability weighting to adjust for a time-dependent confounders on the causal pathway).[13 14 16 18 20 23 117 121 122]

Other methodological issues

Seven studies included prevalent insulin users,[17 19-22 121 122] which is problematic because of the corresponding depletion of susceptibles in other insulin groups (i.e. patients may have developed cancer prior to the start of the study and these patients will be subsequently excluded from the analysis) compared with long-acting insulin analogues. Only one of the studies observed an association between insulin glargine and breast cancer after five years of use among matched prevalent users on prior history of insulin use (HR: 2.7, 95% CI: 1.1-6.5), which may highlight the importance of including new and prevalent insulin users while stratifying on these two types of users.[25 180] Protopathic or detection bias could have resulted in 11 of the 19 studies because a lag period was not incorporated in the study design.[13 15 17 18 20-22 117-119 122] Furthermore, given the cancer latency and the time required to observe all the cancers that will occur in patients in these studies, short follow-up (defined here as < 5 years) was an issue in 16 studies,

whose follow-up time (reported as mean, median, or maximum duration of follow-up) ranged 0.9 to 4.5 years.[13-16 18 19 21-25 117-119 121 122]

4.7. Discussion

Summary

Our systematic review identified 16 cohort and 3 case-control studies on long-acting insulin analogues and cancer risk. We have shown that 7 studies included prevalent users, 11 did not incorporate a lag period, 6 were subject to time-related biases (4 of which had time-lag bias), and 16 had a short follow-up (< 5 years). The RR reported in the existing literature on long-acting insulin analogues and cancer suggests there is no increased risk for any, colorectal, or prostate cancers but four studies observed an increased risk with breast cancer when comparing insulin glargine with other insulins. However, due to the methodological issues, the conclusions that can be drawn from observational studies on long-acting insulin analogues and cancer are limited due to the methodological issues.

Implications and solutions to the methodological issues

Given the methodological issues present in many of the existing studies, the currently available evidence is insufficient to draw definitive conclusions regarding the association between long-acting insulin analogues and cancer. The U.S. Food and Drug Administration arrived at similar conclusions.[132-134] In contrast, the European Medicine Agency concluded that insulin glargine does not increase the risk of cancer.[135] Given the limitations of the existing literature, there remains a need for methodologically rigorous studies conducted with longer follow-up to clearly evaluate the relationship between long-acting insulin analogues and site-specific cancers. Such studies must use study designs and analytical approaches that consider the biases and issues that were discussed in detail above and summarised in Supplemental Table 2.

Previous observational studies, reviews, and RCTs on antidiabetic medications and cancer

To the best of our knowledge, this is the first systematic review of the methodological strengths and limitations of existing studies on long-acting insulin analogues and cancer. However, earlier editorials and narrative reviews have criticised the four cohort studies on

insulin glargine and cancer for their methodological shortcomings, which included reverse causation, lack of lag periods, inclusion of prevalent insulin users, and concerns about the data analysis.[26-30] In our systematic review, we also identified the lack of lag periods used and the inclusion of prevalent users as additional limitations in a few studies.

One of the insulin glargine and breast cancer studies only observed an association among prevalent insulin users, matched on prior history of insulin use, but not among new insulin users.[25] This study suggests that duration of insulin use could be an effect measure modifier of the insulin glargine and breast cancer relationship and studies with shorter follow-up may not be sufficient to observe these effects. Moreover, it also highlights the importance of separating new or first-time users from patients who are switchers from one type of insulin to another as the risk may not be uniform across user types. Along with using more appropriate comparators, one of the strengths of a recent study by Habel and colleagues was the separation of new users and switchers.[18] Studies only considering new users may not provide adequate evidence for decision making in a real world setting as the risk of cancer may differ among patients who switch from other insulins to long-acting insulin analogues.[24]

The methodological limitations, particularly time-related biases, of previous studies of antidiabetic medications and cancer incidence were previously discussed in a review of observational studies of metformin and cancer.[126] Compared with the literature examining the association between metformin and cancer incidence, we observed a smaller prevalence of time-related biases. However, we identified the presence of other methodological issues not addressed in this previous work. Importantly, unlike the previous review, the present methodological assessment was conducted in the context of a systematic review.

Similar to observational studies, RCTs assessing long-acting insulin analogues among patients with diabetes did not observe an increased risk of cancer,[8-10] but these RCTs were designed to study efficacy (e.g., improvements in fasting plasma glucose level) and not cancer outcomes. The most notable RCT was the ORIGIN trial, which had 12,537 patients in whom 953 new or recurrent cancers occurred during of the median follow-up of 6.2 years.[11] This secondary analysis of the ORIGIN trial had 90% power to detect a 20% increased risk of cancer with use of insulin glargine, cancer outcomes were adjudicated by a blinded (to treatment assignment) assessor, and longer follow-up compared with previous RCTs. Despite these strengths, the study was insufficiently powered to assess site-specific cancers conclusively. Furthermore, given the

long latency of cancer, the duration of follow-up of ORIGIN (up to seven years only) was likely insufficient to assess cancer risk conclusively.

Strength and limitations

Our study has several important strengths. First, to our knowledge, it is the first systematic review to methodologically assess the literature on long-acting insulin analogues and their effects on cancer in patients with type 1 or 2 diabetes. This includes the assessment of biases and methodological issues that are particularly prevalent in pharmacoepidemiologic research. Second, a prespecified protocol was used to conduct the systematic review. Finally, our systematic search was conducted in duplicate, ensuring the inclusion of all relevant studies in the present systematic review.

There are also some potential limitations. First, we did not search the grey literature, contact other experts in the field, or attempt to obtain unpublished work. Second, the search was restricted to studies published in English. However, this restriction did not result in the exclusion of many studies (see supplemental table 1). Third, the presence of residual confounding due to unmeasured confounders was evaluated based on confounders (i.e., HbA1c and diabetes duration) previously identified in the literature and conclusions could vary based on the assessment of other potential confounders. Fourth, this systematic review focused on the association between long-acting insulin analogues and cancer incidence. Consequently, it did not assess the literature in which the cancer risk of any insulin compared with that of no insulin, an area that warrants further investigation, particularly given the emergence of new medications for patients with type 2 diabetes. Finally, as is true with any systematic review, there is the potential for publication bias. However, given our focus on the methodological aspects of the literature on this topic and a large number of included studies with null results, the impact of publication bias on the present study was likely minimal.

4.8. Conclusions

We identified several methodological issues in observational studies on long-acting insulin analogues and cancer incidence, including the inclusion of prevalent users, lack of lag periods, and time-lag bias. In addition to these three prevalent methodological issues, most

studies had a short follow-up, which could prevent the observation of a relationship given the long latency of cancer. Therefore, the conclusions that can be drawn from existing observational studies of long-acting insulin analogues and cancer are limited. Future studies addressing these issues must use appropriate study designs and analytical approaches that address these limitations to conclusively address the potential association between long-acting insulin analogues and cancer incidence.

4.9. Tables

Table 4.1 Characteristics of observational studies examining the association between long-acting insulin analogues and cancer incidence

Source	Study design	Follow-up (y)*	Age (y)*	Study sample size	Type of database		Exposure vs. Comparator†	Type of cancer(s)‡
					Exposure	Outcome		
Colhoun [14]	Cohort	~3.0	54.7	36,524 patients with DM	Health research database	Disease registry	Insulin glargine vs other insulin§	Any, breast, CRC, lung, pancreas, prostate
Currie [15]	Cohort	2.4	62.0	62,809 patients with DM	Health research database	Health research database	Insulin glargine vs other insulin§	Breast
Hemkens [117]	Cohort	1.6	68.0	127,031 insulin users	Health administrative database	Health administrative database	Insulin glargine vs. human insulin	Any
Jonasson [19]	Cohort	~2.0	n/a	114,841insulin users	Pharmacy dispensing records	Disease registry	Insulin glargine vs other insulin§	Any, breast, GI, prostate
Mannucci [120]	Nested CC	6.3	63.1	1,340 patients with DM Cases: 112; Controls: 370	Medical records	Health administrative database, disease registry	Mean daily dose of insulin glargine (≥ 0.3 vs. < 0.3 IU/kg/day)	Any
Buchs [122]	Cohort	4.5	60.0	36,342 patients with DM	Pharmacy dispensing records	Disease registry	Total purchases of insulin glargine/detemir	Any
Chang [13]	Cohort	1.7	61.4	59,443 new insulin users	Health administrative database	Disease registry	Insulin glargine vs intermediate/long-acting human insulin	Any, bladder or kidney, breast, CRC, liver, lung, pancreas, prostate, skin, stomach
Ljung [21]	Cohort	~3.0	n/a	114,838 insulin users	Pharmacy dispensing records	Disease registry	Insulin glargine vs other insulin§	Any, breast, CRC, GI, pancreas, prostate
Morden [22]	Cohort	1.9	77.4	81,681 patients with DM	Health administrative database	Health administrative database	Insulin glargine vs other insulin§	Any, breast, colon, pancreas, prostate
Suissa [25]	Cohort	~4.0	65.0	15,227 female, insulin users	Health research database	Health research database	Insulin glargine vs other insulin	Breast
Blin [119]	Cohort	1.4	68.9	6,649 insulin users	Health administrative database	Health administrative database	Insulin glargine vs human insulin (≥ 2 prescriptions)	Any

Continued on following page

Table 4.1–*Continued*

Source	Study design	Follow-up (y)*	Age (y)*	Study sample size	Type of database		Exposure vs. Comparator†	Type of cancer(s)‡
					Exposure	Outcome		
Lind [20]	Cohort	7.0	52.3	7,942 (breast) and 11,613 (prostate) patients with DM	Health research database	Disease registry	Current use of insulin glargine	Breast, prostate
Ruiter [23]	Cohort	3.1	63.3	19,337 patients with DM	Pharmacy dispensing records	Medical records	Cumulative duration of insulin glargine	Any, bladder, breast, colon, endometrial, pancreas, prostate, respiratory
van Staa [118]	Cohort	4.0	65.0	23,005 insulin users	Health research database	Health research database, disease registry	6-24, 24-60, or >60 vs 0-6 months since starting insulin glargine/detemir	Any
Fagot [16]	Cohort	2.8	63.2	70,027 insulin users	Health administrative database	Health administrative database	Cumulative dose of insulin glargine/detemir	Any, bladder, breast, CRC, head and neck, kidney, liver, lung, prostate
Habel [18]	Cohort	3.3	n/a	115,514 patients with DM	Pharmacy dispensing records	Disease registry	Insulin glargine vs. NPH insulin (≥ 2 prescriptions)	Any, breast, CRC, prostate
Simo [121]	Nested CC	~2.0	72	275,164 patients with T2DM Cases: 764; Controls: 2,292	Pharmacy dispensing records	Disease registry	Insulin glargine/detemir vs non-use‖	Any
Sturmer [24]	Cohort	0.9	60.1	52,453 patients with DM	Health administrative database	Health administrative database	Insulin glargine vs. NPH insulin (≥ 2 prescriptions)	Any, breast, colon, prostate
Grimaldi-Bensouda [17]	CC	n/a	66.4	Cases: 775 Controls: 3,050	Questionnaire	Medical records, health research database	Insulin glargine vs other insulin§	Breast

~ indicates that it was estimated from the start and end of the study or total person-time; CC, case-control; CRC, colorectal cancer; DM, diabetes mellitus; GI, gastrointestinal cancer. *Reported as means, medians or maximum range. †Ever versus never exposure definitions were reported unless the study only reported other exposure definitions. ‡All cancers that were reported in the study are presented but only the RRs of the four most common cancer sites among the included studies were reported. §Other insulin comparator definition in the study can include rapid-acting, short-acting, other basal (NPH, detemir), premixed, inhaled, and animal insulin. ‖Non-use can include non-insulin antidiabetic medication and other insulins as listed above.

Table 4.2 Pharmacoepidemiology biases in studies examining the association between long-acting insulin analogues and cancer incidence

Study	Short follow-up*	Prevalent insulin users†	Lack of lag period	Residual confounding‡	Time-related biases			Main limitation§
					Immortal time	Time-lag	Time-window	
Colhoun [14]	●			●				Short follow-up
Currie [15]	●		●	●				Short follow-up
Hemkens [117]	●		●	●		●		Time-lag bias
Jonasson [19]	●	●		●				Inclusion of prevalent users
Mannucci [120]				●			●	Time-window bias
Buchs [122]	●	●	●	●				Inclusion of prevalent users
Chang [13]	●		●	●				Selection bias and lack of lag period
Ljung [21]	●	●	●	●				Inclusion of prevalent users
Morden [22]	●	●	●	●		●		Time-lag bias
Suissa [25]	●			●				Short follow-up
Blin [119]	●		●	●	●	●		Immortal time bias
Lind [20]		●	●	●				Inclusion of prevalent users
Ruiter [23]	●			●				Short follow-up
van Staa [118]	●		●	●		●		Time-lag bias
Fagot [16]	●			●				Short follow-up
Habel [18]	●		●	●				Lack of lag period
Simo [121]	●	●		●				Inappropriate comparator
Sturmer [24]	●			●				Short follow-up
Grimaldi-Bensouda [17]		●	●	●			●	Selection bias

● = indicates presence of the methodological issue or bias in the study

*Short follow-up is defined as < 5 years of follow-up.

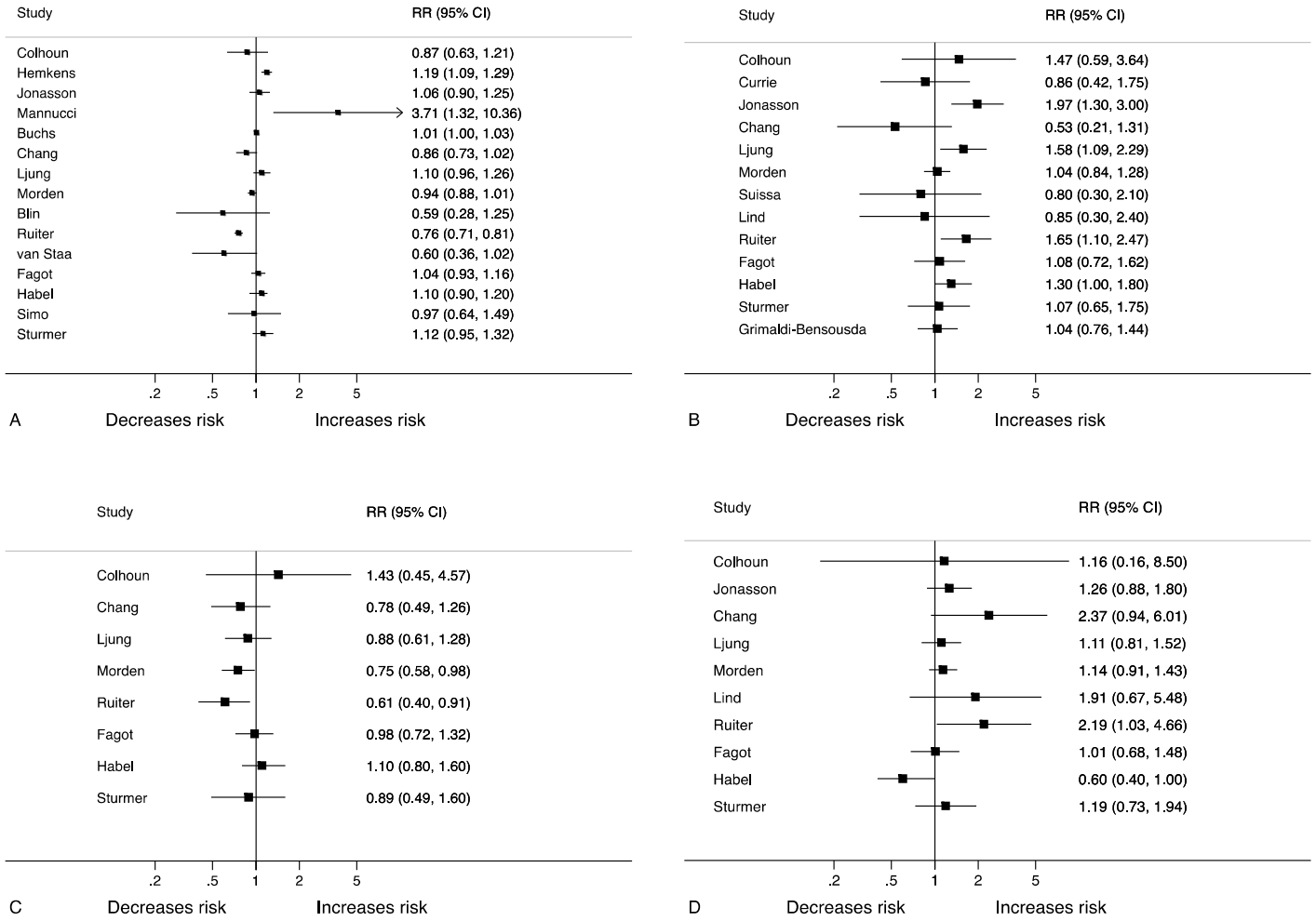
†Prevalent insulin users refer to the study not distinguishing between prevalent and new insulin users.

‡Residual confounding as a result of unmeasured confounders (e.g., HbA1c and diabetes duration) or lack of adjustments for time-dependent confounders.

§Main limitation refers to bias or methodological issue that changed the RR.

4.10. Figures

Figure 4.1 Forest plots of relative risks from studies on insulin glargine and any (A), breast (B), colorectal (C), and prostate (D) cancers



4.11. Appendices

Supplemental Table 1 Search strategy for observational studies on long-acting insulin analogues and cancer using MEDLINE and EMBASE databases from 1 January 2000 to 8 October 2014

MEDLINE

Description	Key terms	Records
Patients with type 1 or type 2 diabetes	exp Diabetes Mellitus, Type 1/ **OR** exp Diabetes Mellitus, Type 2/ **OR** diabet*.mp.	490,207
Insulin glargine or detemir	exp Insulins/ **OR** insulin*.mp. **OR** glargine.mp. **OR** Lantus.mp. **OR** detemir.mp. OR Levemir.mp.	336,783
Cancer	exp Neoplasms/ **OR** cancer*.mp. **OR** neoplasm*.mp. **OR** tumour*.mp. **OR** tumour*.mp. **OR** malignan*.mp.	3,363,037
Publication type: observational studies (cohorts and case-controls)	exp cohort studies/ **OR** exp case control studies/ **OR** observational study/ **OR** observational study as topic/ **OR** Epidemiologic studies/ **OR** Registries/ **OR** (cohort adj (study or studies)).tw. **OR** cohort*.tw. **OR** (case adj (control or cohort) adj (study or studies)).tw. **OR** (nested case control adj (study or studies)).tw. **OR** (case adj (control or controls or cohort or cohorts)).tw. **OR** (observational adj (study or studies)).tw. **OR** (epidemiologic$ adj (study or studies)).tw.	1,801,696
PICO and publication type		1,576
Limits	Publication year: 2000-current	1,379
	Language: English	1,303

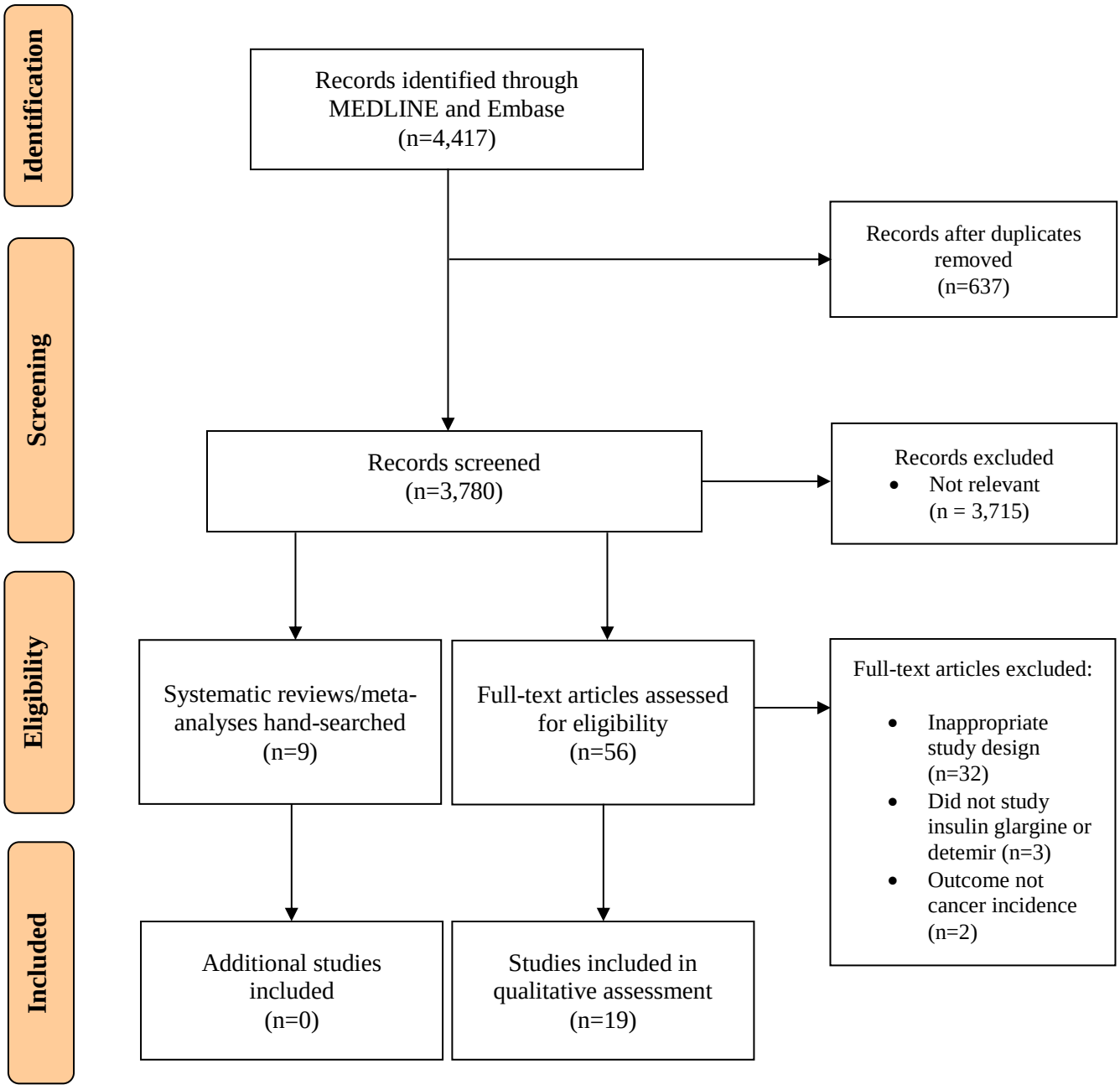
EMBASE

Description	Key terms	Records
Patients with type 1 or type 2 diabetes	exp insulin dependent diabetes mellitus/ **OR** exp non insulin dependent diabetes mellitus/ **OR** diabet*.mp.	762,737
Insulin glargine or detemir	exp insulin derivative/ **OR** insulin*.mp. **OR** glargine.mp. **OR** Lantus.mp. **OR** detemir.mp. OR Levemir.mp.	574,961
Cancer	exp neoplasms/ **OR** cancer*.mp. **OR** neoplasm*.mp. **OR** tumour*.mp. **OR** tumour*.mp. **OR** malignan*.mp.	4,412,637
Publication type: observational studies (cohorts and case-controls)	observational study/ **OR** cohort analysis/ **OR** retrospective study/ **OR** prospective study/ **OR** longitudinal study/ **OR** case control study/ **OR** register/ **OR** (cohort adj (study or studies)).tw. **OR** cohort*.tw. **OR** (case adj (control or cohort) adj (study or studies)).tw. **OR** (nested case control adj (study or studies)).tw. **OR** (case adj (control or controls or cohort or cohorts)).tw. **OR** (observational adj (study or studies)).tw. **OR** (epidemiologic$ adj (study or studies)).tw.	1,302,843
PICO and publication type		3,380
Limits	Publication year: 2000-current	3,254
	Language: English	3,114

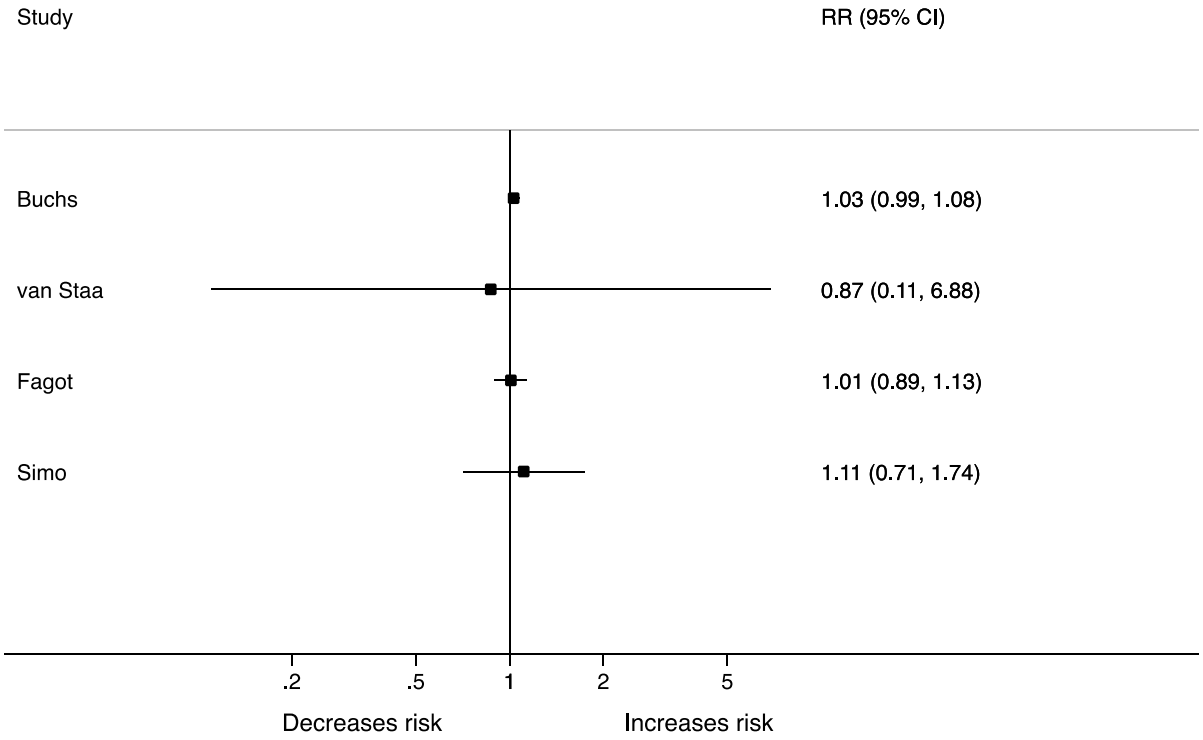
Supplemental Table 2 Solutions to methodological issues present in literature examining long-acting insulin analogues and cancer

Type of bias or methodological issues	Solutions		References
	Design	*Analysis*	
Immortal time bias	For exposure-based cohorts (e.g., enter the cohort based on *n*th prescription(s) of insulin), define cohort entry or time zero (t_0) as the *n*th prescription exposure and comparator. Alternatively, a nested case-control study can be used because of its inherent time-dependent characteristic.	Use a time-dependent analysis (e.g., Cox model with a time-dependent exposure) to classify unexposed person-time until meeting the exposure definition	128 170
Time-lag bias	Match on duration of disease for the exposed and comparator groups.		126
Time-window bias	Match cases and controls on duration of disease.		126 171
Residual confounding	Use an active comparator with similar indications which will help reduce confounding.	Instrumental variables, sensitivity analyses	179 181
Disease latency, protopathic bias, and detection bias	*Cohorts:* Lagging the exposure by *n* amount of years from cohort entry *Case-controls:* Lag the exposure by *n* amount of years prior to case or control index date.		126 172
Prevalent user bias	Matching on prior treatment history; restrict the cohort to a minimum period of non-use (wash-out) prior to t_0.	Adjust for prior treatment history and stratify by new and prevalent users	141 142 180

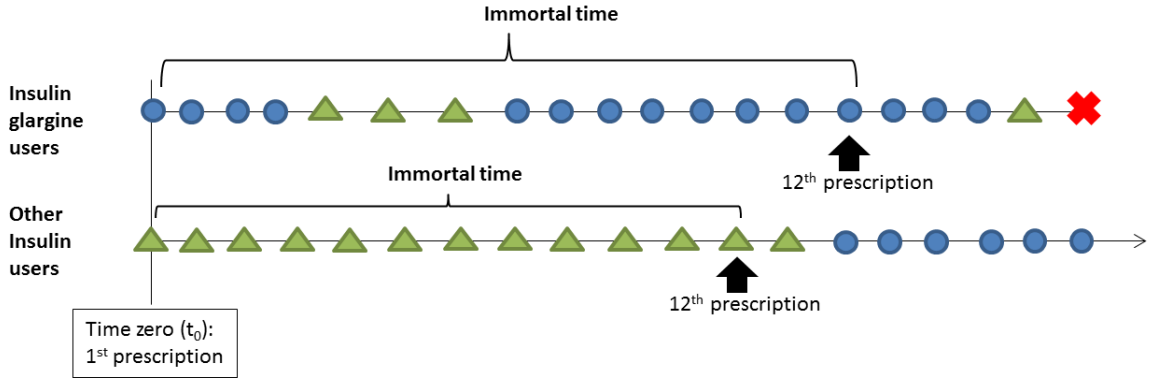
Supplemental Figure 1 Flow chart of included and excluded studies on long-acting insulin analogues and risk of cancer identified from the literature search from 1 January 2000 to 8 October, 2014



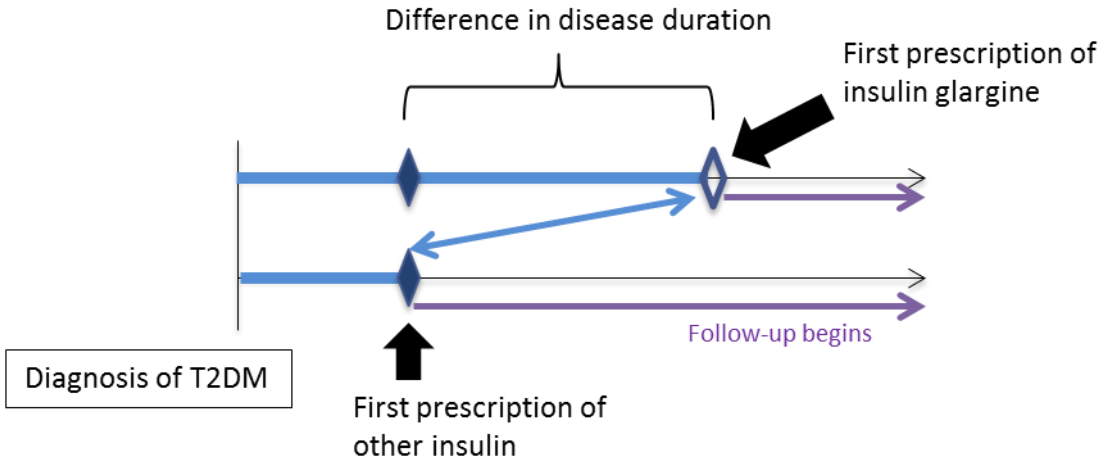
Supplemental Figure 2 Forest plots of relative risks (solid squares) and 95% CIs (solid horizontal lines) from studies on insulin detemir and any cancer. For exposure and comparator definitions in each study, please refer to Table 1.



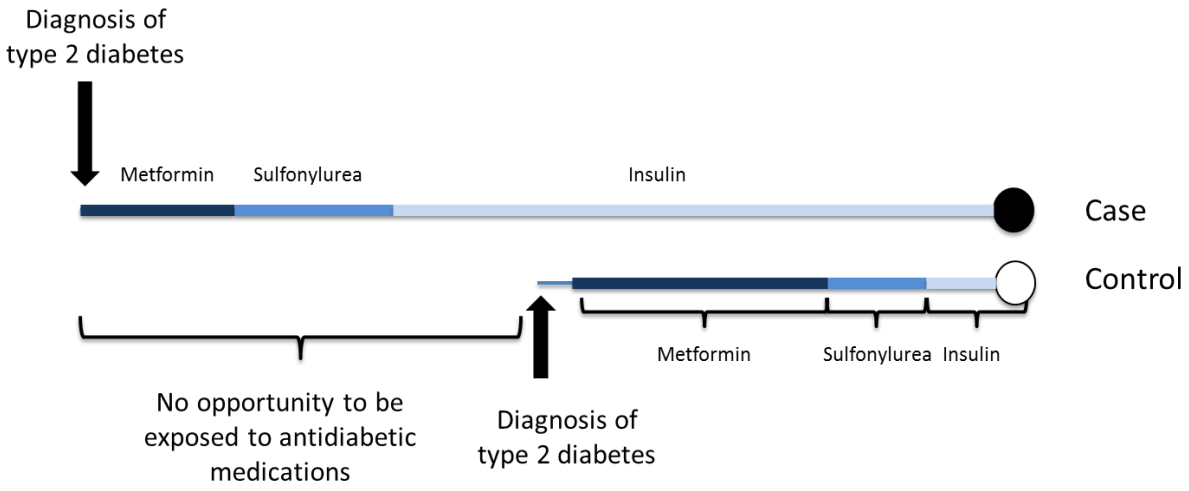
Supplemental Figure 3 Graphical representations of immortal time bias. Immortal time refers to a time period in the follow-up where the outcome of interest cannot occur as a result of the study design and analysis.[128 170] This immortal time occurs due to an arbitrary delay period that is imposed (e.g., using at least 12 prescriptions for the medication to define exposure status but with follow-up time beginning with the first prescription). Patients are considered "immortal" because he or she has to survive (i.e., be free of the outcome of interest) before satisfying the exposure definition. If the patient has the outcome before satisfying the exposure definition, then he or she will be classified as unexposed. This bias occurs when the immortal time period or person-time is misclassified as exposed or excluded from the analysis.[128] Immortal time bias may have been introduced based on the definitions of exposure and cohort entry that were used in one cohort study.[119] For the predominant user definition, the patient needed to have at least 12 prescriptions of insulin and be exposed 80% of the follow-up time to be considered exposed but the time before the twelfth prescription and meeting the 80% exposure time should be considered unexposed. As a result, the adjusted hazard ratios for any cancers were ~0.60, although the results were not statistically significant. The circles (●) indicate a prescription for insulin glargine and triangles (▲) indicate a prescription for other insulins.



Supplemental Figure 4 An illustration of time-lag bias when comparing insulin glargine to other insulins (e.g., human or short-acting insulin), which may be used in earlier stages of diabetes and thus introducing confounding by disease duration.[126] The arrow represents when the comparison begins.



Supplemental Figure 5 Depiction of potential time-window bias, that can occur when cases and controls are not matched on duration of diabetes resulting in differential exposure opportunity.[126 171] Time-window bias was observed in the population-based case-control study.[17] Despite matching on calendar time, time-window bias was potentially present because cases and controls were not matched on duration of diabetes. Consequently, the opportunity for exposure differed between the cases and controls due to the varying diabetes durations (a mean of 14.5 years among cases and 13.2 among controls). While one would expect an increased risk due to the presence of time-window bias, a null effect was observed. This suggests that other biases, such as selection bias resulting from the selection of cases and controls from different study bases, may also be present.



Chapter 5. Manuscript 2 – Long-term effect of long-acting insulin analogues on breast cancer incidence in women with type 2 diabetes

5.1. Preamble

The experimental studies, presented in the background chapter, suggested long-acting insulin analogues bound to the insulin receptor family (primarily IGF) with greater affinity and causes cancer cell proliferation and anti-apoptotic effects primarily in breast cancer cells.[7 55-61]

In the last chapter, we observed in our systematic review many observational studies had evaluated insulin glargine and the risk of any and common site-specific cancers including breast, colorectal, and prostate. There were no associations between insulin glargine and any, colorectal, or prostate cancers whereas the results for breast cancer were inconsistent. More importantly, many of these observational studies had methodological limitations including prevalent user bias, lacked the use of lag periods to account cancer latency, time-related biases and short duration of follow-up between insulin initiation and cancer incidence. Furthermore, no studies have evaluated the relationship between insulin detemir and breast cancer. Consequently, the evidence of long-acting insulin analogues on the risk of breast cancer remains uncertain.

Together with the current evidence from experimental and observational studies, our primary motivation for this manuscript was to assess the relationship of long-acting insulin analogues on the risk of breast cancer among women with type 2 diabetes using insulin therapy. The findings from this study would further elucidate the relationship between long-acting insulin analogues and breast cancer risk and provide additional evidence on the safety profile of long-acting insulin analogues to drug regulatory agencies.

This manuscript has been submitted to *Journal of Clinical Oncology*.

5.2. Title page

Title: Long-term effect of long-acting insulin analogues on breast cancer incidence in women with type 2 diabetes

Authors: Jennifer W. Wu[1,2], Laurent Azoulay[1,2,3,4], Agnieszka Majdan[5], Jean-François Boivin[1,2], Michael Pollak[4], and Samy Suissa[1,2,3]

Affiliations:

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada

[2] Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, Montreal, Canada

[3] Division of Clinical Epidemiology, Department of Medicine, McGill University, Montreal, Canada

[4] Gerald Bronfman Department of Oncology, McGill University, Montreal, Canada

[5] Division of Endocrinology, Jewish General Hospital, Montreal, Canada

Corresponding author:

Dr. Samy Suissa
Centre for Clinical Epidemiology
Lady Davis Institute, Jewish General Hospital
3755 Cote Ste-Catherine, H-461
Montreal, Quebec, Canada, H3T 1E2
Tel: (514) 340-7593
Fax: (514) 340-7564
Email: samy.suissa@mcgill.ca

5.3. Abstract

Objective: To assess whether the use of long-acting insulin analogues, when compared with Neutral Protamine Hagedorn (NPH) insulin, is associated with an increased risk of breast cancer.

Design: Population-based cohort study

Setting: United Kingdom Clinical Practice Research Datalink

Participants: 22,395 women aged 40 years or older, treated with insulin glargine (n=9,575), detemir (3,271) or NPH (n=9,549) between 2002 and 2012 and followed until 2015.

Main outcome measures: Cox proportional hazards models were used to estimate adjusted hazard ratios (HRs) and 95% confidence intervals (CIs) of incident malignant breast cancer, comparing long-acting insulin analogues with NPH insulin overall and by duration since first insulin prescription and cumulative dose.

Results: A total of 321 women were newly diagnosed with breast cancer during the follow-up period (crude incidence rate: 3.3 per 1000 person-years). Compared with NPH insulin, insulin glargine was associated with an overall increased risk of breast cancer (HR: 1.45, 95% CI: 1.12 to 1.87). This HR was elevated after five years since initiation (HR: 2.25, 95% CI: 1.33 to 3.80). A dose-response relationship was also demonstrated, where women with >30 insulin glargine prescriptions were at an increased risk of breast cancer (HR: 2.28, 95% CI: 1.26 to 4.15). In contrast, the use of insulin detemir was not associated with an increased risk of breast cancer (HR: 1.18, 95% CI: 0.78 to 1.78).

Conclusions: Insulin glargine was associated with an increased long-term risk of breast cancer in women with type 2 diabetes. The risk with insulin detemir remains uncertain because of the fewer women and shorter duration of use from its more recent introduction to the market.

5.4. Introduction

Basal insulins, which include Neutral Protamine Hagedorn (NPH) insulin and the long-acting insulin analogues glargine and detemir, are commonly used to treat patients with type 1 diabetes and advanced type 2 diabetes.[6 82] While long-acting insulin analogues reduce the occurrence of nocturnal hypoglycaemia and have longer treatment effects than NPH insulin,[6 82] there are concerns that their use may increase the risk of breast cancer. Indeed, several experimental studies have shown that long-acting insulin analogues have stronger binding affinities to the insulin receptor family; a proposed mechanism for the increased cellular proliferation and inhibition of apoptosis observed primarily with breast cancer cells.[7 88 182]

Several observational studies have been conducted to assess whether insulin glargine is associated with an increased risk of breast cancer incidence,[12-25] half these studies reported an association.[12 14 18 19 21 23 25] Overall, several of these studies had a number of methodological limitations, including prevalent user bias, no consideration of cancer latency, time-related biases, and short durations of follow-up.[31] Furthermore, none of these studies assessed the association between insulin detemir and breast cancer incidence. Finally, while several randomised controlled trials (RCTs) of long-acting insulin analogues did not find significant associations with cancer incidence, these RCTs were not designed to assess this safety endpoint and lacked statistical power and long duration of follow-up to investigate the association with breast cancer.[8-11] Thus, the relationship between long-acting insulin analogues and breast cancer incidence remains uncertain.

Therefore, with the longer market availability of these insulins, we assessed whether the use of long-acting insulin analogues over the long-term, compared with the use of NPH insulin, is associated with an increased risk of breast cancer in women with type 2 diabetes using insulin therapy.

5.5. Methods

Data source

The study was conducted using the United Kingdom (UK) Clinical Practice Research Datalink (CPRD).[143 144] The CPRD includes approximately 700 practices with over 14 million patients and has been shown to be representative of the UK general population.[143] The CPRD

uses the Read or medical code classification system to record medical diagnoses and procedures whereas prescriptions are based on the British National Formulary system. Furthermore, the CPRD collects anthropometric data, lifestyle indicators, and laboratory values. The CPRD data was previously validated and shown to be of high quality.[143 144 153 155] This study was approved by the International Scientific Advisory Committee of CPRD (protocol number 15_005R), Institutional Review Board of McGill University (A11-M114-14B), and Research Ethics Board of Jewish General Hospital in Montreal, Quebec.

Study population

We assembled a base cohort of women who received at least one prescription for any type of insulin (rapid-acting insulin analogues, short-acting insulin, intermediate-acting insulin or long-acting insulin analogues [basal insulins], premixed [a combination of rapid-acting or short-acting and intermediate-acting], and animal insulin) from 1 January 1988 to 31 December 2012. Since women with type 1 diabetes have a different pathophysiology than women with type 2 diabetes,[33] we restricted the cohort to women 40 years or older at the time of base cohort entry because type 2 diabetes is more likely to be diagnosed in the elderly population.[4 183] We also excluded women with less than one year of medical history in the CPRD, and those with a previous diagnosis of gestational diabetes at any time before base cohort entry.

From the base cohort, we created a study cohort of women with at least one prescription for basal insulin (glargine, detemir, or NPH) from 1 September 2002 (the year the first long-acting insulin analogue, glargine, entered the UK market) to 31 December 2012.

We further categorised the study cohort by new insulin users (first-time users or initiators of insulin) and prior insulin users (prevalent insulin users). Women were classified as new insulin users provided that they not receive any insulin at any time before study cohort entry. Conversely, women were classified as prior insulin users if they had received any insulin prescription at any time before study cohort entry.

Exposure definition

Among new insulin users, women were exposed to insulin glargine, detemir, or NPH, depending on which basal insulin prescription occurred first. Study cohort entry was defined based on the date of first basal insulin prescription. Among prior insulin users, we used a

hierarchical exposure definition. Women with at least one insulin glargine or detemir prescription were considered exposed to insulin glargine or detemir, respectively, whereas the remaining women in the study cohort were considered exposed to NPH insulin. More long-acting insulin analogue users could have entered the study cohort later while NPH insulin users could have entered earlier. To ensure that NPH insulin user had a similar starting time point as insulin glargine users, we randomly selected one NPH insulin prescription among NPH insulin users between 2002 and 2012 and used this prescription as the study cohort entry date. Hence, study cohort entry was defined based on the date of the first long-acting insulin analogue prescription or randomly selected NPH insulin prescription. After the random selection, we then excluded basal insulin users with a history of cancer any time prior to the study cohort entry.

For the primary exposure definition, we compared long-acting insulin analogues using an approach that is analogous to intent-to-treat (ITT). For our secondary exposure, we also evaluated the duration of use since first basal insulin prescription and was classified according to four predefined categories: <3, 3-5 and ≥ 5 years. In addition, we also assessed the cumulative dose defined as cumulative number of insulin glargine and NPH prescriptions and was classified according to three predefined categories: <10, 10-30, >30 prescriptions. We used these categories based on a priori knowledge of where the risk of breast cancer would change over the follow-up.[25]

Outcome definition and follow-up

Women were followed from study cohort entry to an incident diagnosis of primary malignant breast cancer (available upon request), death due to any causes, end of CPRD registration, or end of study date (28 February 2015), whichever came first. Breast cancer medical codes reported in the CPRD were shown to be highly concordant with breast cancer cases identified in the UK National Cancer Data Repository (NCDR) (>90%).[155]

Covariates

The models were adjusted for potential confounders that were known to be a risk factor for cancer prior to study cohort entry using variable lookback periods (see table 1).[4] We included age, year of study cohort entry, excessive alcohol use (medical codes for alcohol-related diseases or alcoholism screening), smoking status (never (reference), ever, missing), body mass index

(BMI; <25 (reference), 25-30, ≥ 30 kg/m^2, missing), glycated haemoglobin (HbA1c; (<6.5 (reference), 6.5-8.0, >8.0%, missing), diabetes duration (time between study cohort entry date and date of first non-insulin antidiabetic medication or insulin prescription, diagnosis of type 2 diabetes, or HbA1c value ≥6.5%), prior duration of insulin (time between study cohort entry date and date of first-ever of any type of insulin prescription), non-insulin antidiabetic medication use vs. no use (metformin, sulfonylureas, thiazolidinedione [TZD], and others), Deyo's Charlson comorbidity score (≤ 1 (reference), 2-3, >3; excludes previous cancers), [165] and other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications [NSAID]). Other antidiabetic medications included meglitinides, dipeptidyl-peptidase-4 inhibitors (DPP-4i), glucagon-like-peptide-1 analogues (GLP-1), alpha-glucosidase inhibitors, guar gum, and sodium/glucose cotransporter-2 inhibitors (SGLT2i). In the primary analysis, we included an indicator for missingness for smoking, BMI and HbA1c categorical variables because missingness was minimal (~5%).

Statistical analysis

The crude incidence rate of breast cancer and 95% confidence intervals (CIs) was estimated by cumulating the person-time over the follow-up using the Poisson distribution. We also constructed cumulative incidence of breast cancer curves for each exposure group.[184] For the primary analysis, we used a time-independent Cox model to estimate the unadjusted and adjusted hazard ratios (HRs) and 95% CIs of breast cancer incidence comparing long-acting insulin analogues with NPH insulin. For the duration- and dose-response analyses, we used a time-dependent Cox model to estimate the unadjusted and adjusted HRs and 95% CIs to update the risk set for each of the three duration and dose categories of insulin glargine or detemir compared with NPH insulin in the same category. We also conducted stratified analyses by new and prior insulin users. Restricted cubic splines were used to test the linearity assumptions of continuous variables and assess the HR as a function of duration since first long-acting insulin analogues with the SAS macro developed by Heinzl and Kaider (appendix 1).[185]

Sensitivity analysis

For more details on the methods of sensitivity analyses conducted, please see appendix 2. In brief, sensitivity analyses were conducted to assess the robustness of the estimates. First, in

the stratified analyses of prior insulin users, we further restricted these prior insulin users to switchers only (e.g. women who switched from NPH or other insulins to either long-acting insulin analogue). Second, to evaluate detection bias, we included a one to three years lag period to account for cancer latency and conducted age (<50, 50-70, and 70+ years) and prior mammography screening in the last three years and entire history (i.e. with or without) stratified analyses. Third, we performed multiple imputations to account for the missing data in the analysis. Fourth, we used the "array approach" to determine the strength and prevalence of unmeasured confounders that would be required to bring the adjusted HRs towards the null.[179] Lastly, we conducted a competing risk analysis for all-cause mortality based on the Fine and Gray Cox proportional hazard model to determine the subdistribution hazard ratios.[186 187]

All data analyses were performed using SAS 9.4 statistical software (SAS Institute, Cary, NC) and forest plots were constructed with the 'meta' package from R version 3.3.1 (R Development Core Team, Vienna, Austria).

5.6. Results

Study population characteristics

A total of 22,395 women were included in the cohort, which was comprised of 9,575, 3,271, and 9,549 users of insulin glargine, detemir and NPH, respectively (figure 1). During the up to 12 years of follow-up (mean: 4.4 years), 321 incident breast cancers were diagnosed, corresponding to a crude incidence rate of 3.3 per 1,000 person-years. The baseline characteristics of long-acting insulin analogue users, compared with NPH insulin, were younger, at earlier stages of type 2 diabetes (shorter diabetes duration and used more first and second line antidiabetic medications), had lower prior duration of insulin use, lower BMI, less comorbidities, but had higher HbA1c and more alcohol-related diseases (table 1). Long-acting insulin analogue users had longer diabetes duration but used earlier antidiabetic medications such as metformin among new insulin users while prior insulin users had similar baseline characteristics as the overall cohort (appendix 3).

Effect estimates in the overall study population

We observed insulin glargine, but not insulin detemir, compared with NPH insulin users had a higher cumulative probability of developing breast cancer after five years of use as shown in the cumulative incidence curves (appendix 4). After adjustment in the Cox model (table 2), an association between insulin glargine, compared with NPH insulin, and breast cancer was observed (HR: 1.45, 95% CI: 1.12 to 1.87). In contrast, there was no increased risk of breast cancer associated with users of insulin detemir compared with NPH insulin (HR: 1.17, 95% CI: 0.78 to 1.79). In the overall cohort, the risk of breast cancer was elevated after five years of insulin glargine use compared with NPH insulin in the same category (HR: 2.25, 95% CI: 1.33 to 3.80), but not for insulin detemir (HR: 1.10, 95% CI: 0.43 to 2.80) (appendix 5). We presented restricted cubic splines of the HR and 95% confidence intervals as a function of time since the start of insulin glargine (figure 2) and detemir (appendix 6a) in the entire cohort, which corresponded with the duration-response observed. Similarly, a dose-response relationship was demonstrated, where a cumulative number of insulin glargine prescriptions of at least 30 prescriptions compared with NPH insulin in the same category was associated with a 2-fold increase in risk of breast cancer (HR: 2.28, 95% CI: 1.26 to 4.15). On the contrary, we observed no dose-response relationship with insulin detemir and breast cancer incidence (appendix 7).

Effect estimates among new and prior insulin users

The association between insulin glargine and breast cancer incidence remained among prior insulin users (HR: 1.54; 95% CI: 1.11 to 2.13), but not new insulin users (HR: 1.18; 95% CI: 0.77 to 1.81) (appendix 8). There was no observed increased risk for breast cancer for new (HR: 0.89; 95% CI: 0.42 to 1.88) or prior insulin users of detemir (HR: 1.41; 95% CI: 0.85 to 2.35). The risk of breast cancer was elevated after five or more years of insulin glargine use among prior insulin users (HR: 2.50, 95% CI: 1.29 to 4.84) (appendix 5). In appendix 6b-c, we illustrated restricted cubic splines of the HR and 95% CIs as a function of time since the start of long-acting insulin analogues among new and prior insulin users. Similarly, a dose-response relationship was demonstrated with prior insulin users only (appendix 9), where a cumulative number of insulin glargine prescriptions of at least 30 prescriptions was associated with a 3-fold increase in the risk of breast cancer (HR: 3.24, 95% CI: 1.47 to 7.13). In contrast, no dose-

response relationship was demonstrated with insulin detemir users in either new or prior insulin users.

Sensitivity analyses

We conducted several sensitivity analyses to evaluate the robustness of our estimates between long-acting insulin analogues and breast cancer incidence with the primary exposure. As summarised in figure 3, the estimates were robust for insulin glargine, but not insulin detemir, and breast cancer incidence. In all the sensitivity analyses, insulin glargine was associated with an increased risk of breast cancer with adjusted HRs ranging from 1.45 to 1.93. In contrast, insulin detemir was not associated with increased risk of breast cancer with adjusted HRs ranging from 0.79 to 1.78 (appendix 10). In appendix 11, we varied the prevalence of the unmeasured confounder in the insulin glargine group and strength of the unmeasured confounder-disease association that bias the HR of insulin glargine and breast cancer (1.45) away from the null. We observed a strong unmeasured confounder (RR_{CD}>2.0) and large imbalance in the unmeasured confounders between glargine and NPH insulin users (>10%) was required to bias the HR. For more detailed results of the sensitivity analyses, please see appendices 12-14.

5.7. Discussion

Statement of principal findings

In a cohort of women with type 2 diabetes treated with insulin, we observed the use of insulin glargine was associated with an increased risk for breast cancer. In our duration-response analysis, we found the risk of breast cancer increased after five or more years of insulin glargine use. In our cumulative dose-response analysis, we found the risk of breast cancer increased with more than 30 insulin glargine prescriptions. The insulin glargine users were associated with an increased risk of breast cancer among prior insulin users but not new insulin users. Our findings for insulin glargine and breast cancer incidence remained consistent in several sensitivity analyses. In contrast, the results between insulin detemir and breast cancer were inconclusive given the smaller sample size and short duration of use due to its recent introduction into the UK market.

Comparison with other studies

To our knowledge, 14 observational studies have been conducted to evaluate the association between insulin glargine and breast cancer incidence.[12-25] However, a majority of these studies had an insufficient duration of follow-up (<5 years).[31] Yet, three of the 14 studies observed an increased risk of breast cancer with insulin glargine use in the overall population,[14 19 21] while four studies observed an association among new or prior insulin users only (HRs ranged from 1.30 to 3.65).[12 18 23 25] Though these specific studies had additional methodological issues including prevalent user bias, used inappropriate comparators (other insulins) and lacked a lag period to account for cancer latency.[31] To date, there have only been four observational studies on insulin detemir and any cancer incidence which demonstrated null associations;[16 118 121 122] although, these studies had the same aforementioned methodological limitations.[31] Several RCTs have explored the relationship between long-acting insulin analogues and any and site-specific cancer incidence and observed null effects.[8-11] The most important RCT, the Outcomes Reduction Insulin Glargine Intervention (ORIGIN) trial, had important strengths including adjudication of cancer outcomes and had slightly longer follow-up in relation to other post hoc analyses of pooled RCTs, but it had insufficient power to detect an effect with site-specific cancers and it still had a relatively short follow-up to assess outcomes with long latency such as cancer (median: 6.2 years).[11]

Biological mechanisms

Long-acting insulin analogues have been demonstrated to be efficacious at controlling HbA1c levels, but it has also been shown to have mitogenic effects through the insulin or insulin-like growth factor (IGF) mediated cellular pathways. Though the evidence has not been consistent in experimental studies, these synthetic insulins have been reported to have altered receptor binding characteristics resulting in stronger binding affinity.[7 88 109 110 113 188-190] Furthermore, a majority of experimental studies have been conducted in breast cancer cell lines and suggested that insulin glargine binds with greater affinity to IGF receptors to activate mitogen-activated protein kinase (MAPK) pathways which then upregulates cellular proliferation and anti-apoptotic effects.[7 88] In contrast, insulin detemir and breast cancer cells have been shown to have less binding affinity.[7 88 113] Albeit, some studies have shown that insulin detemir can still initiate the same mitogenic cellular pathways through the insulin receptor family.[112 188]

Consequently, it is plausible the use of long-acting insulin analogues could increase the risk of breast cancer through the altered pharmacokinetics.

Strengths and weaknesses of this study

Our study has several important strengths. First, unlike previous studies, we had up to 12 years of follow-up. Importantly, we were able to study the effects among new and prior insulin users separately and accounted for prior insulin use duration to minimise the prevalent user bias. In addition, we evaluated several time windows for a lag period to account for cancer latency. Second, we were the first study to evaluate insulin detemir and breast cancer incidence. Third, given the use of CPRD, we were able to adjust for confounders that are not typically available in health administrative databases such as BMI and smoking status. Lastly, we conducted several sensitivity analyses that demonstrated our estimates were robust.

Despite the strengths of the study, we have several limitations. First, like many other observational studies, there is still potential for unmeasured confounding. However, when we used the "array approach" it showed that a strong unmeasured confounder and a large imbalance in the prevalence of the unmeasured confounder between insulin glargine and NPH were needed to bias the HR. Second, even though there is fairly complete documentation on prescriptions written in the CPRD it is unknown whether women adhered to the prescribed medications. However, since women with advanced type 2 diabetes depend on insulin therapy to manage their HbA1c levels, it is unlikely women would not adhere to their medications. Hence, exposure misclassification may be minimal. Third, there is potential for misclassification of the outcome but aforementioned a previous study has compared CPRD cancer diagnostic codes to the UK NCDR and observed high concordance (>90%).[155] Moreover, this outcome misclassification will likely be non-differential with respect to exposure status and potentially bias the results towards the null in the primary analysis. Fourth, although we made efforts to separate prior insulin users and adjusted for prior duration of insulin use, the types of prior insulin users included in our study may not reflect the real world population (women who switched from NPH and other insulins to either long-acting insulin analogue). In our study, we determined the study cohort eligibility of a woman using all insulin prescriptions; therefore, women could have started on long-acting insulin analogues prior to study cohort entry. Although we restricted prior insulin users to switchers from NPH or other insulins to long-acting insulin analogues only, we observed

that the estimates remained the same. Lastly, detection bias could have been present in our study as result of the earlier published studies, which suggested insulin glargine increased the risk of breast cancer and the inconclusive statements made by drug regulatory agencies in Canada, United States and Europe.[131 132 135] However, our lagged, age-stratified, and prior mammography screening stratified analyses yielded similar results, which suggested that detection bias alone, cannot explain the increased risk of breast cancer among insulin glargine users compared with NPH insulin users.

5.8. Conclusions

In summary, insulin glargine use was associated with an increased risk of breast cancer in a cohort of women with type 2 diabetes, particularly with long-term use. Despite these findings, the benefits and risks of insulin glargine must be considered by drug regulatory agencies before any changes to clinical practice be made. Given the fewer number of women and shorter duration of use due to its more recent introduction to the UK market, future studies are needed to further evaluate the relationship between long-term use of insulin detemir and breast cancer.

5.9. Tables

Table 5.1 Baseline characteristics of women with at least one prescription for insulin glargine, detemir or NPH according to insulin exposure at study cohort entry

Covariates	Glargine	Detemir	NPH
Number of women	9,575	3,271	9,549
Age (years) *	64.9 ± 12.9	63.5 ± 12.2	70.3 ± 11.6
Calendar year*			
2002-2005	2,901 (30.3)	149 (4.6)	4,123 (43.2)
2006-2008	3,778 (39.5)	1,646 (50.3)	3,066 (32.1)
2009-2012	2,896 (30.3)	1,476 (45.1)	2,360 (24.7)
Excessive alcohol use†	421 (4.4)	175 (5.4)	330 (3.5)
Smoking status†			
Ever	3,205 (33.5)	1,106 (33.8)	3,096 (32.4)
Never	6,245 (65.2)	2,157 (65.9)	6,124 (63.1)
Unknown	125 (1.3)	8 (0.2)	329 (3.5)
BMI (kg/m^2) ‡			
<25	2,066 (21.6)	621 (19.0)	1,493 (15.6)
25-30	2,699 (28.2)	850 (26.0)	2,440 (25.6)
≥ 30	4,296 (44.9)	1,711 (52.3)	4,575 (47.9)
Unknown	514 (5.4)	89 (2.7)	1,041 (10.9)
HbA1c (%)§			
<6.5	332 (3.5)	94 (2.9)	695 (7.3)
6.5-8.0	1,964 (20.5)	662 (20.2)	2,766 (29.0)
>8	6,750 (70.5)	2,414 (73.8)	5,057 (53.0)
Unknown	529 (5.5)	101 (3.1)	1,031 (10.8)
Diabetes duration (years) *	8.2 ± 5.4	8.5 ± 5.6	9.3 ± 5.6
Prior duration of insulin use*	2.8 ± 4.3	3.1 ± 4.5	4.3 ± 4.7
Non-insulin diabetes medication use§¶			
Metformin	5,830 (60.9)	2,046 (62.6)	4,837 (50.7)
Sulfonylurea	4,661 (48.7)	1,495 (45.7)	2,930 (30.7)
Thiazolidinedione	2,048 (21.4)	728 (22.3)	977 (10.2)
Others**	1,177 (12.3)	474 (14.5)	743 (7.8)
Charlson comorbidity score‖			
≤ 1 comorbidities	5,052 (52.8)	1,612 (49.3)	4,907 (51.4)
2-3	3,376 (35.3)	1,246 (38.1)	3,410 (35.7)
>3	1,147 (12.0)	413 (12.6)	1,232 (12.9)
Statin use§	6,795 (71.0)	2,547 (77.9)	6,392 (66.9)
Aspirin use§	4,722 (49.3)	1,638 (50.1)	5,269 (55.2)
NSAID use§	3,631 (37.9)	1,286 (39.3)	3,519 (36.9)

Data are presented as mean (SD) or n (%) unless otherwise specified.
*Measured at study cohort entry. †Measured at any time prior to study cohort entry. ‡Measured five years prior. §Measured two years prior to study cohort entry. ‖Measured one year prior to study cohort entry. ¶Not mutually exclusive. **Other non-insulin diabetes medications included meglitinides, DPP-4i, GLP-1, alpha-glucosidase inhibitors, guar gum, and SGLT2i. Abbreviations: BMI – body mass index; HbA1c – glycated haemoglobin; NSAID – nonsteroidal anti-inflammatory drugs

Table 5.2 Crude and adjusted hazard ratios of breast cancer associated with the use of long-acting insulin analogues compared to NPH insulin use, in the entire insulin cohort

Exposure	N	Breast cancer cases	Person-years	Incidence rate (95% CI)*	Hazard ratios (95% CI)	
					Crude	Adjusted†
Intent-to-treat						
NPH	9,549	108	35,077	3.1 (2.5, 3.7)	1.00	1.00 (reference)
Glargine	9,575	176	48,685	3.6 (3.1, 4.2)	1.19	1.45 (1.12, 1.87)
Detemir	3,271	37	14,834	2.5 (1.8, 3.4)	0.81	1.17 (0.78, 1.79)
Dose-response						
NPH						
<10	9,549	60	16,136	3.7 (2.9, 4.8)	1.00	1.00
10-30	5,133	33	11,306	2.9 (2.1, 4.1)	1.00	1.00
>30	2,293	15	7,635	2.0 (1.2, 3.3)	1.00	1.00
Glargine						
<10	9,575	73	19,871	3.7 (2.9, 4.6)	0.99	1.16 (0.81, 1.65)
10-30	6,382	61	17,124	3.6 (2.8, 4.6)	1.26	1.53 (0.99, 2.36)
>30	3,256	42	11,689	3.6 (2.7, 4.9)	1.85	2.28 (1.26, 4.15)

* Per 1,000 person-years. †The multivariable Cox proportional hazards model was adjusted for the following covariates prior to study cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, missing), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, missing), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, missing).

5.10. Figures

Figure 5.1 Flow chart describing the selection of 22,395 women 40 years or older with at least one prescription for any basal insulin between 1 September 2002 and 31 December 2012

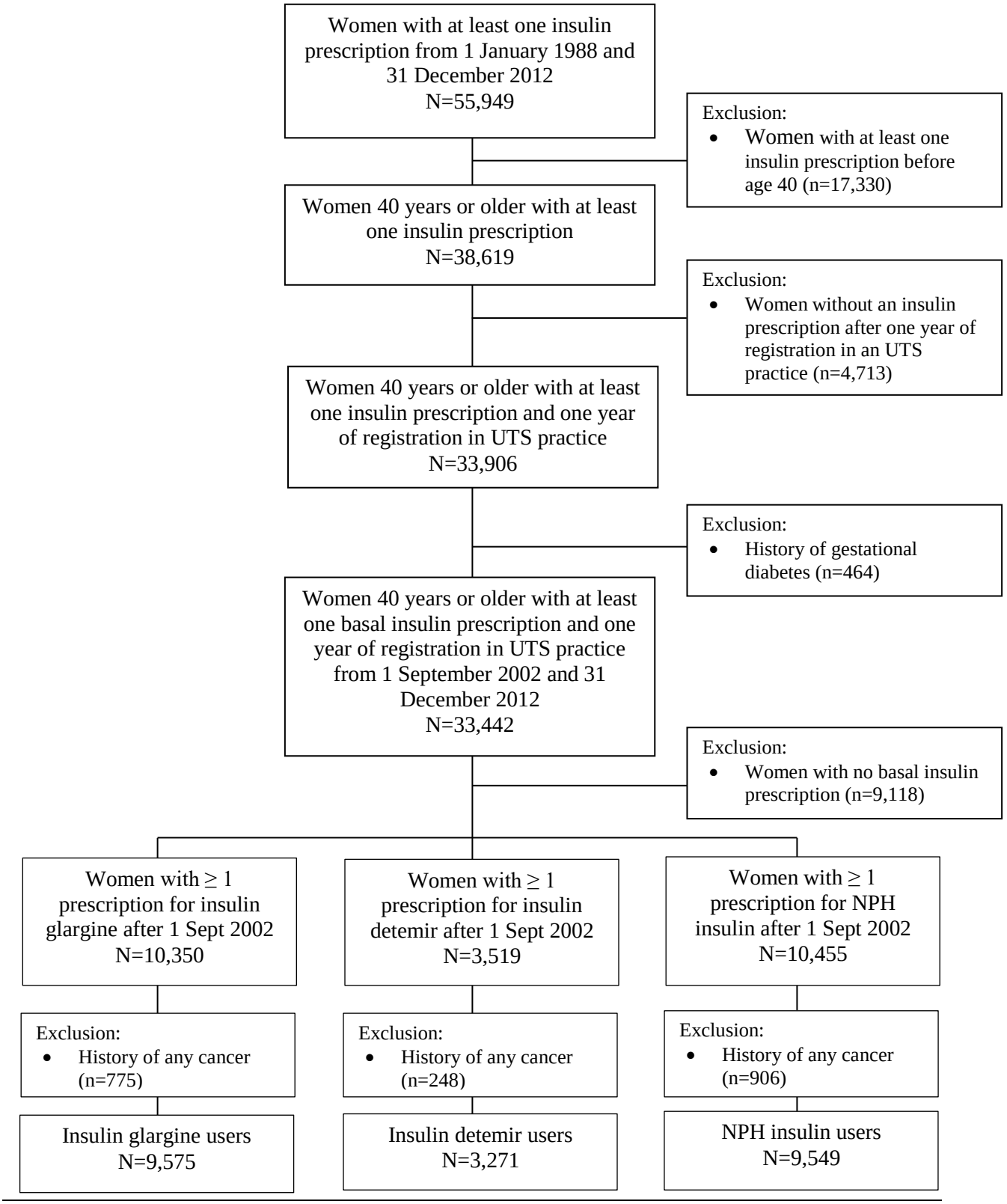


Figure 5.2 Restricted cubic splines of the adjusted hazard ratio (solid line) and 95% CIs (dotted lines) for breast cancer incidence as a function of duration since first insulin glargine prescription compared to NPH in the entire cohort with three knots placed at the 20th (0.7 years), 50th (2.4), and 80th (5.8) percentiles

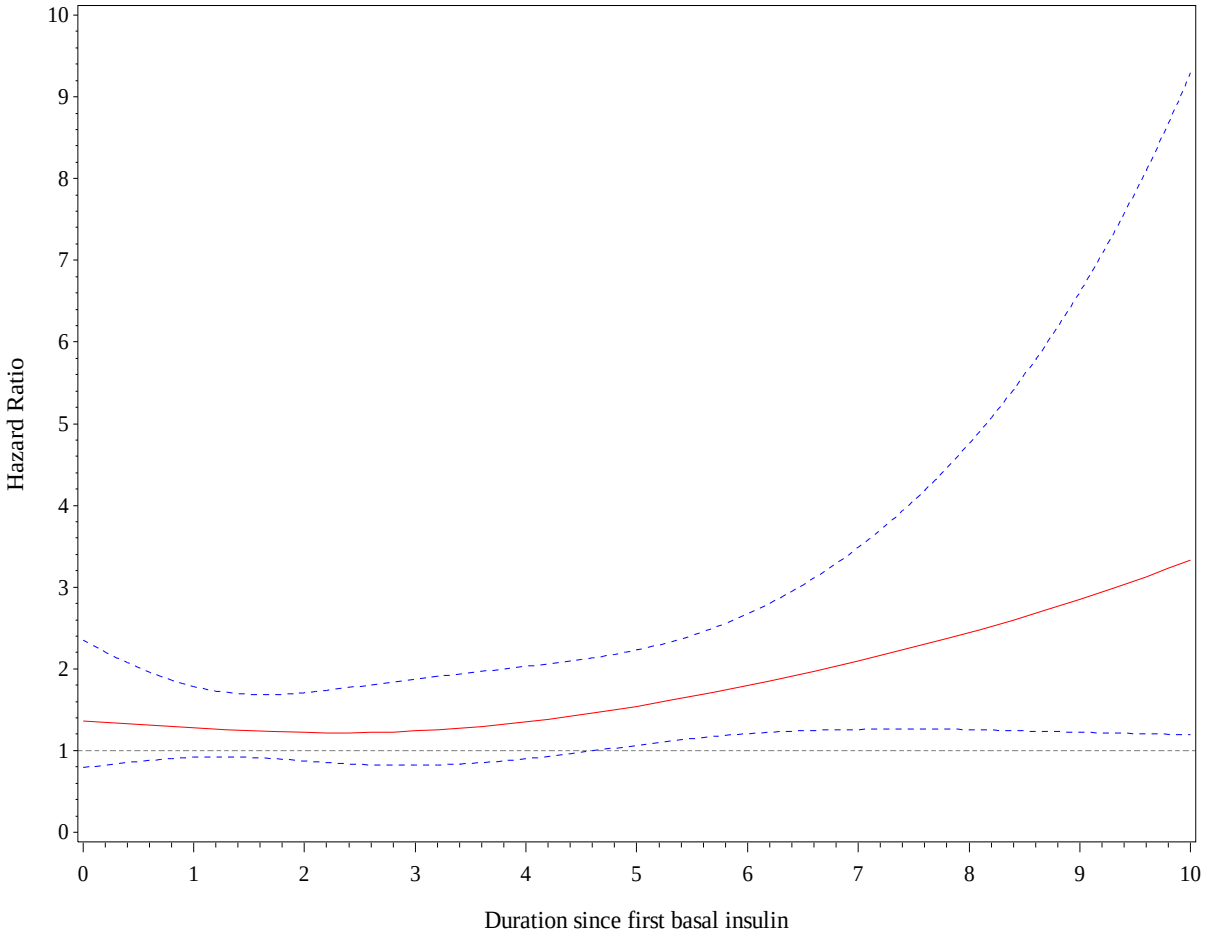
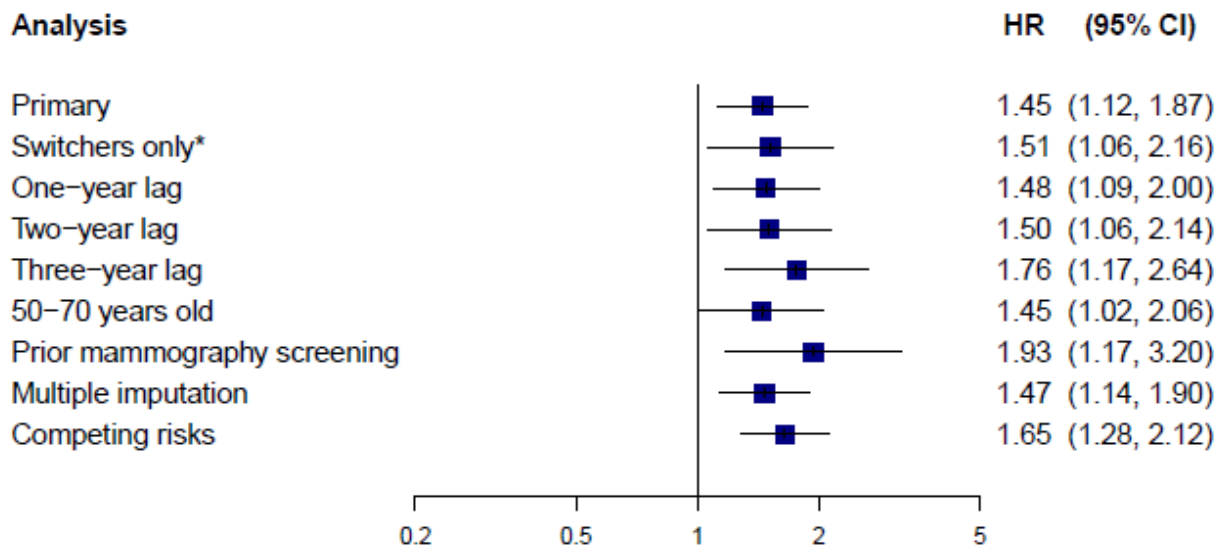


Figure 5.3 Forest plots of sensitivity analyses for insulin glargine and breast cancer



*Women who switched from NPH or other insulins to insulin glargine

5.11. Appendices

Appendix 1: Additional method details

Appendix 2: Additional sensitivity analyses details

Appendix 3: Baseline characteristics of long-acting insulin analogues and NPH insulin users among new and prior insulin users

Appendix 4: Cumulative incidence of breast cancer curves

Appendix 5: Duration-response of long-acting insulin analogues and breast cancer incidence

Appendix 6: Hazard ratio as a function of duration since first long-acting insulin analogue prescription

Appendix 7: Cumulative dose of insulin detemir and breast cancer incidence

Appendix 8: Intent-to-treat analyses for long-acting insulin analogues and breast cancer incidence among new and prior insulin users

Appendix 9: Cumulative dose of long-acting insulin analogues and breast cancer incidence among new and prior insulin users

Appendix 10: Forest plots of sensitivity analyses for insulin detemir

Appendix 11: Rule-out residual confounding

Appendix 12: Restricted prior insulin users to switchers only

Appendix 13: Detection bias analyses

Appendix 14: Multiple imputations

Appendix 1a Additional method details

Restricted cubic splines

For continuous variables, we tested the linearity assumption using restricted cubic splines as this was a method that offers the most flexibility and was commonly used to assess duration and dose-response relationships and test linearity assumptions in epidemiologic research.[191 192] We created the restricted cubic splines using a SAS macro developed by Heinzl and Kaider.[185] We used three knots placed at the 5th, 50th and 95th percentile as this has been previously recommended, particularly for testing linearity assumptions and to have a more parsimonious model (see results in appendix 1b).[191 193] To assess the HR and 95% CIs as a function of duration since first long-acting insulin analogue prescription, we also used restricted cubic splines with three knots placed at the 20th, 50th, and 80th percentiles (see results in figure 2 and appendix 6). We used different percentiles to place knots based on previous research and a priori assumptions of when the risk of cancer would most likely change over the follow-up.[25]

Appendix 1b Test of non-linearity using restricted cubic splines.* In the multivariable Cox models, we tested the continuous variables (age, calendar year, diabetes duration, and duration of prior insulin use) for non-linearity and found that the continuous variables were linear.

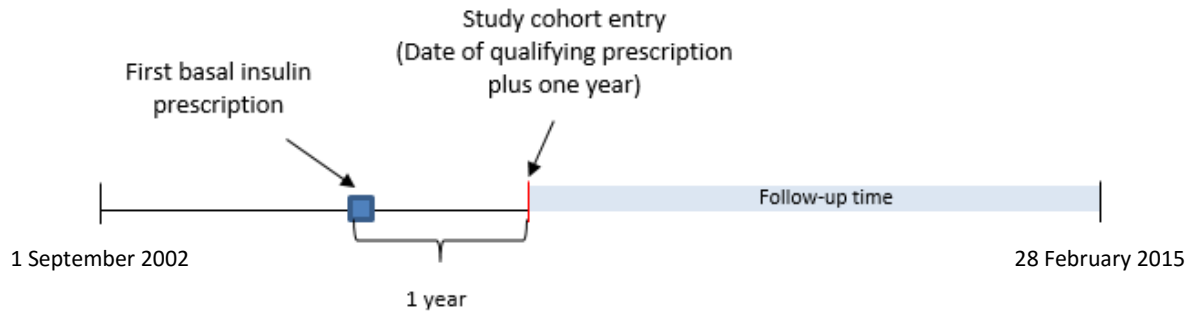
Continuous variables	Exposures in the outcome model	
	Glargine	Detemir
Age	0.227	0.296
Calendar year	0.919	0.709
Diabetes duration (year)	0.582	0.680
Duration of prior insulin use (year)	0.344	0.810

*Three knots placed at 5th, 50th, and 95th percentiles for the restricted cubic splines.

Appendix 2 Additional sensitivity analyses details

Lagged analyses

Lag periods are important to consider because they account for a minimum amount of time of onset between exposure and detection of the outcome.[194] Consequently, the lag periods can reduce the detection and protopathic bias that may be present.[31] Given the uncertainty of the appropriate time windows, we varied the lag period from one to three years. In our study, women were followed from study cohort date plus one, two or three years (see figure below). Cancer events occurring within the lag periods would be excluded from the analysis.



Age and prior mammography status stratified analyses

It is plausible that long-acting insulin analogues users may receive more mammography screenings compared to NPH insulin users resulting in a higher detection of breast cancer. To further investigate the potential for detection bias, we performed several stratified analyses. First, we evaluated the proportion of mammography screening by exposure status before and after study cohort entry and crude rate of mammography after study cohort entry. We first conducted an age-stratified analyses by <50, 50-70, and 70+ years and these cut-points were chosen based on the age range of the National Health Service Breast Screening Program.[195] To estimate the more direct effect of long-acting insulin analogues on breast cancer incidence (i.e. not mediated by mammography screening), we conducted a stratified analysis by mammography screening status in the three years and any time prior to study cohort entry.[196 197]

Multiple imputation

When it is minimal missing data (<5%), methods such as complete case analysis or using an indicator for missing could be used. However, using an indicator for missing can bias the

estimates even if the missingness is completely at random and the interpretation of this level of the categorical variable is not meaningful.[198] Therefore, we performed multiple imputations to account for the missing data in the analysis. In our study, HbA1c, smoking status and BMI had missing data. For each of the variables, missing data ranged from <1% to slightly over 5% in the entire cohort. The missing data patterns suggested it is missing at random because the mean of the covariates differed slightly but did not suggest an evident pattern (see results in appendix 14). We used multiple imputations by the chained equations or fully conditional specification to impute the three binary (smoking status) and categorical variables (HbA1c and BMI). The fully conditional specification method is more flexible in that it allows for imputing both continuous and categorical variables and it does not assume a joint multivariable normal distribution.[199 200] We used the multiple imputation procedure from SAS v9.4 with the fully conditional specific logistic regression method and ten imputations to impute values of the missing data for the three binary and categorical variables.[201] Based on the amount of missing information in the data, we used ten imputations as this has been shown to have high power and relative efficiency.[199 202] The logistic regression model included all the covariates in the multivariable Cox models and survival outcomes. The survival outcomes included a binary indicator for breast cancer and the Nelsen-Aalen estimator for cumulative hazards as this has been suggested to yield the least bias.[203 204]

Competing risks

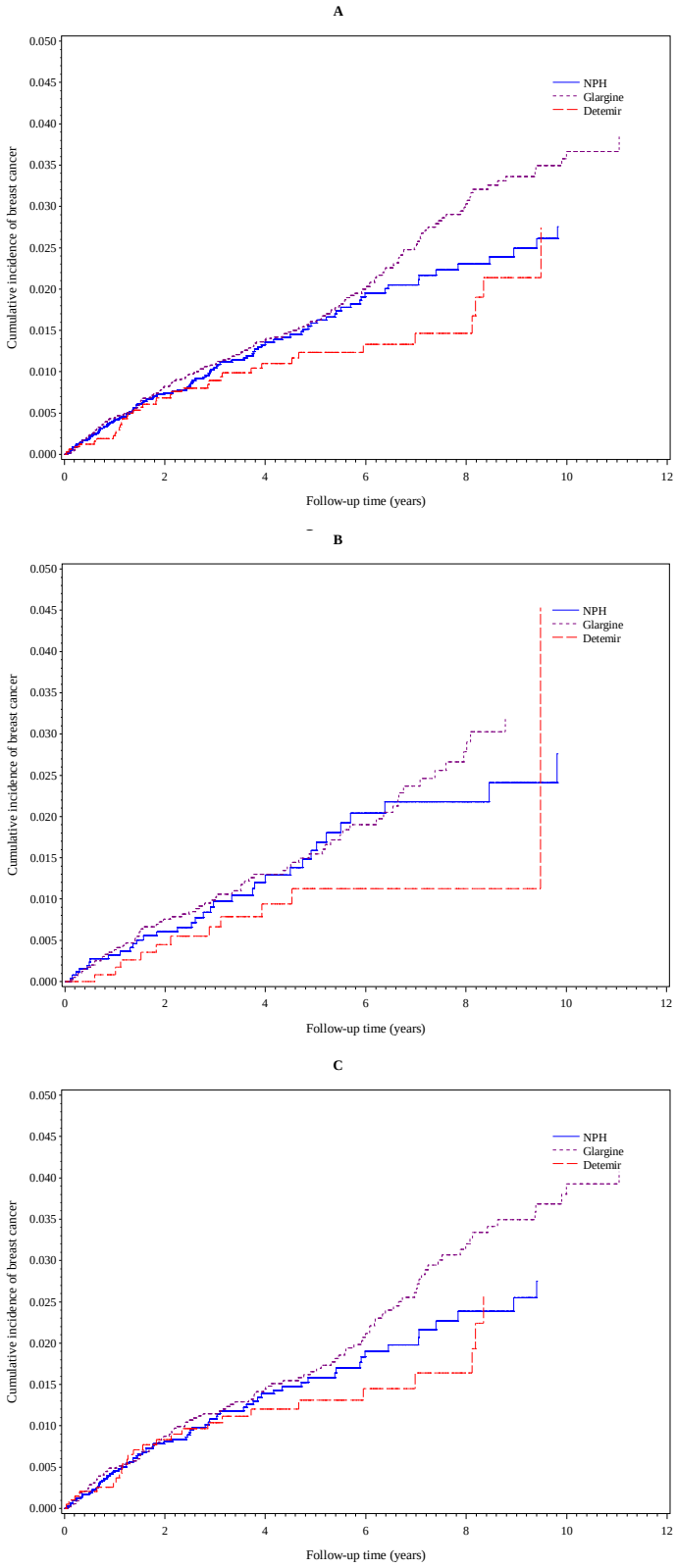
Women with type 2 diabetes have a higher risk of mortality from other diseases, such as cardiovascular, compared to those without diabetes.[177] Therefore, we performed a competing risk analysis to account for the possible competing risks due to death from any cause using the subdistribution hazards models as developed by Fine and Gray.[186] The subdistribution hazard models are commonly used in epidemiology and does not require the strong assumption that the event of interest and competing event need to be independent.[187]

Appendix 3 Baseline characteristics of women treated with basal insulin according to insulin exposure at cohort entry among new and prior insulin users.

Covariate	New insulin users			Prior insulin users[‡]		
	Glargine	Detemir	NPH	Glargine	Detemir	NPH
Number of women	4,148	1,262	2,652	5,427	2,009	6,897
Age (years) [*]	66.7 ± 12.8	64.6 ± 12.7	67.6 ± 12.3	63.6 ± 12.8	62.8 ± 11.9	71.3 ± 11.2
Calendar year[*]						
2002-2005	888 (21.4)	28 (2.2)	954 (36.0)	2,013 (37.1)	121 (6.0)	3,169 (46.0)
2006-2008	1,818 (43.8)	533 (42.2)	741 (27.9)	1,960 (36.1)	1,113 (55.4)	2,325 (33.7)
2009-2012	1,442 (34.8)	701 (55.6)	957 (36.1)	1,454 (26.8)	775 (38.6)	1,403 (20.3)
Excessive alcohol use[†]	147 (3.5)	50 (4.0)	124 (4.7)	274 (5.1)	125 (6.2)	206 (3.0)
Smoking status[†]						
Ever	1,245 (30.0)	334 (26.5)	792 (29.9)	1,960 (36.1)	772 (38.4)	2,304 (33.4)
Never	2,867 (69.1)	925 (73.3)	1,774 (66.9)	3,378 (62.2)	1,232 (61.3)	4,350 (63.1)
Unknown	36 (0.9)	3 (0.2)	86 (3.2)	89 (1.6)	5 (0.3)	243 (3.5)
BMI (kg/m^2) [‡]						
<25	714 (17.2)	197 (15.6)	443 (16.7)	1,352 (24.9)	424 (21.1)	1,050 (15.2)
25-30	1,177 (28.4)	347 (27.5)	679 (25.6)	1,522 (28.0)	503 (25.0)	1,761 (25.5)
≥ 30	2,104 (50.7)	685 (54.3)	1,276 (48.1)	2,192 (40.4)	1,026 (51.1)	3,299 (47.8)
Unknown	153 (3.7)	33 (2.6)	254 (9.6)	361 (6.7)	56 (2.8)	787 (11.4)
HbA1c (%) [mmol/mol] [§]						
<6.5 [47]	76 (1.8)	23 (1.8)	108 (4.1)	256 (4.7)	72 (3.6)	587 (8.5)
6.5-8.0 [48-64]	606 (14.6)	170 (13.5)	403 (15.2)	1,357 (25.0)	491 (24.4)	2,363 (34.3)
>8 [65]	3,242 (78.2)	1,013 (80.3)	1,759 (66.3)	3,509 (64.7)	1,401 (69.7)	3,298 (47.8)
Unknown	224 (5.4)	56 (4.4)	382 (14.4)	305 (5.6)	45 (2.2)	649 (9.4)
Diabetes duration (years) [*]	7.6 ± 5.0	7.6 ± 5.1	6.9 ± 5.3	8.6 ± 5.7	9.1 ± 5.9	10.2 ± 5.4
Prior duration of insulin use (years) [*]	--	--	--	4.9 ± 4.7	5.0 ± 4.8	5.9 ± 4.5
Non-insulin diabetes medication use[§¶]						
Metformin	3,464 (83.5)	1,033 (81.9)	1,872 (70.6)	2,366 (43.6)	1,013 (50.4)	2,965 (43.0)
Sulfonylurea	3530 (85.1)	1,063 (84.2)	1,926 (72.6)	1,131 (20.8)	432 (21.5)	1,004 (14.6)
Thiazolidinedione	1,616 (39.0)	524 (41.5)	745 (28.1)	432 (8.0)	204 (10.2)	232 (3.4)
Others[**]	894 (21.6)	352 (27.9)	521 (19.7)	283 (5.2)	122 (6.1)	222 (3.2)
Charlson comorbidity score[‖]						
≤ 1 comorbidities	2,144 (51.7)	631 (50.0)	1,278 (48.2)	2,908 (53.6)	981 (48.8)	3,629 (52.6)
2-3	1,489 (35.9)	484 (38.4)	997 (37.6)	1,887 (34.8)	762 (37.9)	2,413 (35.0)
>3	515 (12.4)	147 (11.7)	377 (14.2)	632 (11.7)	266 (13.2)	855 (12.4)
Statin use[§]	3,099 (74.7)	987 (78.2)	1,739 (65.6)	3,696 (68.1)	1,560 (77.7)	4,653 (67.5)
Aspirin use[§]	2,078 (50.1)	593 (47.0)	1,275 (48.1)	2,644 (48.7)	1,045 (52.0)	3,994 (57.9)
NSAID use[§]	1,566 (37.8)	486 (38.5)	975 (36.8)	2,065 (38.1)	800 (39.8)	2,544 (36.9)

Data are presented as mean (SD) or n (%) unless otherwise specified. [*]Measured at study cohort entry. [†]Measured at any time prior to study cohort entry. [‡]Measured five years prior to study cohort entry. [§]Measured two years prior to study cohort entry. [‖]Measured one year prior to study cohort entry. [¶]Not mutually exclusive. [**]Other non-insulin diabetes medications included meglitinides, DPP-4i, GLP-1, alpha-glucosidase inhibitors, guar gum, and SGLT2i. Abbreviations: BMI – body mass index; HbA1c – glycated haemoglobin; NSAID – nonsteroidal anti-inflammatory drugs

Appendix 4 Cumulative incidence of breast cancer curves for glargine (small dash line), detemir (large dash), or NPH (solid) insulin users among the entire cohort (A), and new (B) and prior (C) users of insulin

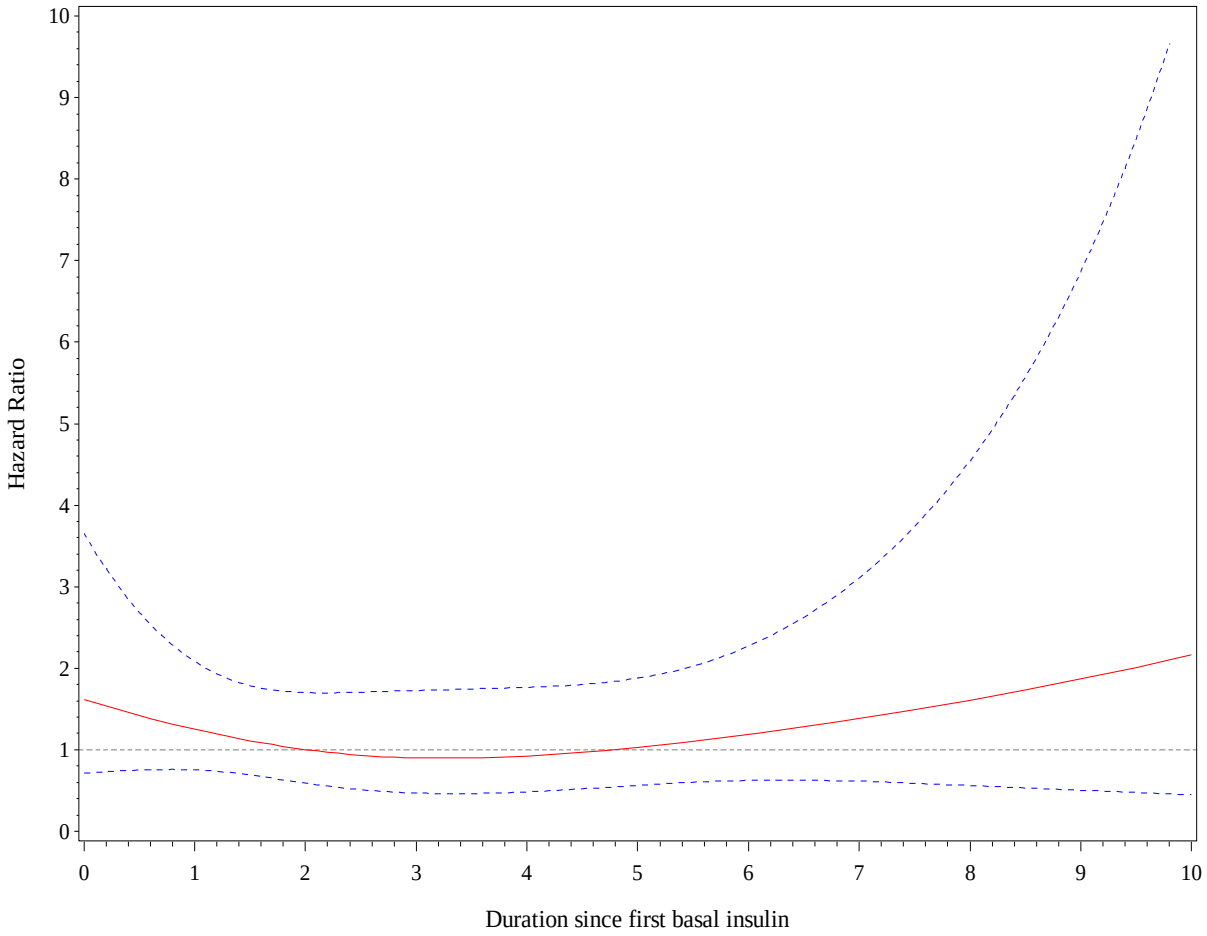


Appendix 5 Crude and adjusted hazard ratios of breast cancer for duration since first insulin prescription at study cohort entry in the entire cohort and among new and prior insulin users

Exposure (years)		N	Cancer cases	Person-years	Incidence rate (95% CI)*	Hazard ratios (95% CI) Crude	Adjusted†
All women							
NPH	<3	9,549	70	19,597	4.6 (2.8, 4.5)	1.00	1.00
	3-5	4,386	19	6,944	2.7 (1.7, 4.3)	1.00	1.00
	≥ 5	2,754	19	8,536	2.2 (1.4, 3.5)	1.00	1.00
Glargine	<3	9,575	88	23,842	3.7 (3.0, 4.5)	1.05	1.26 (0.91, 1.75)
	3-5	6,332	29	10,612	2.7 (1.9, 3.9)	1.00	1.22 (0.66, 2.19)
	≥ 5	4,391	59	14,231	4.1 (3.2, 5.4)	1.85	2.25 (1.33, 3.80)
Detemir	<3	3,271	25	8,202	3.0 (2.8, 4.5)	0.86	1.29 (0.78, 2.11)
	3-5	2,139	6	3,442	1.7 (0.8, 3.9)	0.64	0.92 (0.36, 2.34)
	≥ 5	1,330	6	3,190	1.9 (0.8, 4.2)	0.80	1.09 (0.43, 2.78)
New insulin users							
NPH	<3	2,652	20	6,257	3.2 (2.1, 5.0)	1.00	1.00
	3-5	1,477	7	2,339	3.0 (1.4, 6.3)	1.00	1.00
	≥ 5	941	7	3,148	2.2 (1.1, 4.7)	1.00	1.00
Glargine	<3	4,148	35	10,409	3.4 (2.4, 4.7)	1.05	1.09 (0.62, 1.90)
	3-5	2,741	13	4,505	2.9 (1.7, 5.0)	0.97	1.03 (0.41, 2.60)
	≥ 5	1,809	19	5,257	3.6 (2.3, 5.7)	1.52	1.60 (0.67, 3.87)
Detemir	<3	1,262	7	3,210	2.2 (1.0, 4.6)	0.68	0.90 (0.36, 2.23)
	3-5	839	S	1,282	2.3 (0.8, 7.3)	0.79	1.06 (0.26, 4.30)
	≥ 5	455	S	960	1.0 (0.1, 7.4)	0.44	0.58 (0.07, 4.95)
Prior insulin users‡							
NPH	<3	6,897	50	13,340	3.7 (2.8, 4.9)	1.00	1.00
	3-5	2,909	12	4,605	2.6 (1.5, 4.6)	1.00	1.00
	≥ 5	1,813	12	5,389	2.2 (1.3, 3.9)	1.00	1.00
Glargine	<3	5,427	53	13,434	3.9 (3.0, 5.2)	1.08	1.33 (0.89, 2.01)
	3-5	3,591	16	6,107	2.6 (1.6, 4.3)	1.01	1.25 (0.58, 2.67)
	≥ 5	2,582	40	8,974	4.5 (3.3, 6.1)	2.03	2.50 (1.29, 4.84)
Detemir	<3	2,009	18	4,992	3.6 (2.3, 5.7)	0.98	1.61 (0.89, 2.92)
	3-5	1,300	S	2,160	1.4 (0.4, 4.3)	0.54	0.83 (0.23, 2.99)
	≥ 5	875	S	2,230	2.2 (0.9, 5.4)	0.93	1.35 (0.47, 3.94)

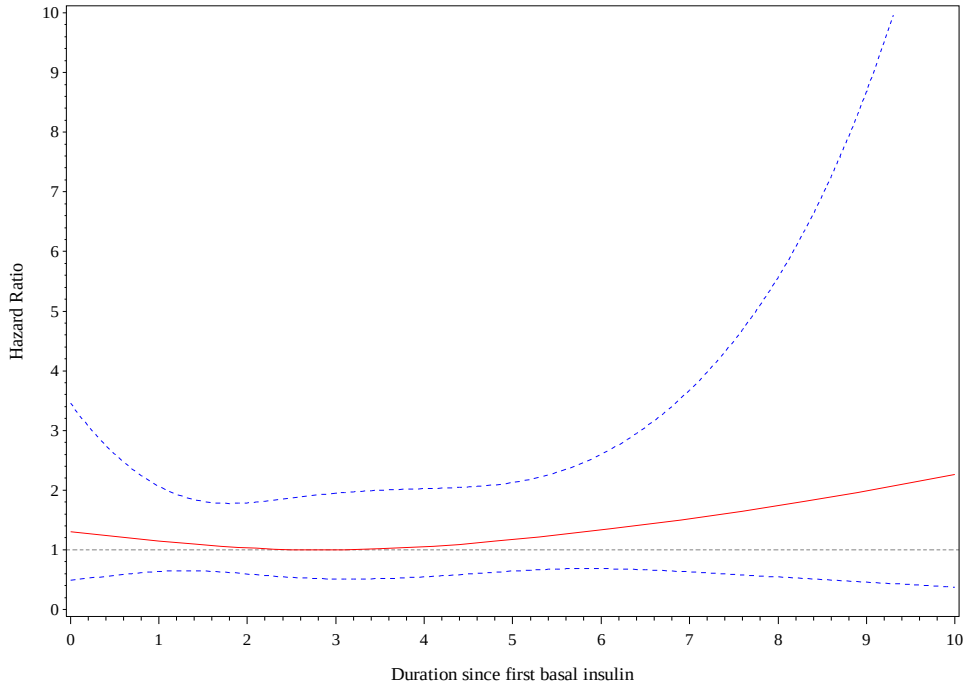
* Per 1,000 person-years. †The multivariable Cox proportional hazards model was adjusted for the following covariates prior to study cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, missing), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, missing), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, missing). ‡Prior insulin users are defined as women with at least any type of insulin before study cohort entry. §Cell numbers less than five were suppressed (S) in accordance with the confidentiality agreements of CPRD.

Appendix 6a Restricted cubic splines of the adjusted hazard ratio (solid line) and 95% CIs (dotted lines) for breast cancer incidence as a function of duration since first insulin detemir prescription compared to NPH in the entire cohort with three knots placed at the 20th (0.9 years), 50th (2.4), and 80th (5.8) percentiles

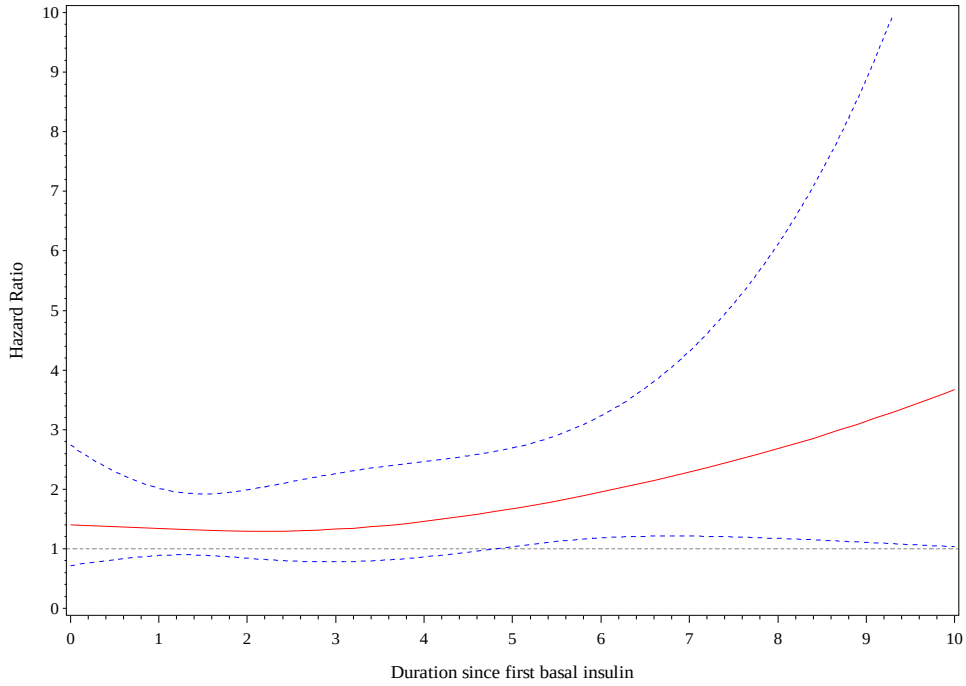


Appendix 6b Restricted cubic splines of the adjusted hazard ratio (solid line) and 95% CIs (dotted lines) for breast cancer incidence as a function of duration since first insulin glargine prescription compared to NPH among new (A) and prior insulin users (B) with three knots placed at the 20th (0.9 years), 50th (2.4), and 80th (5.8) percentiles

A

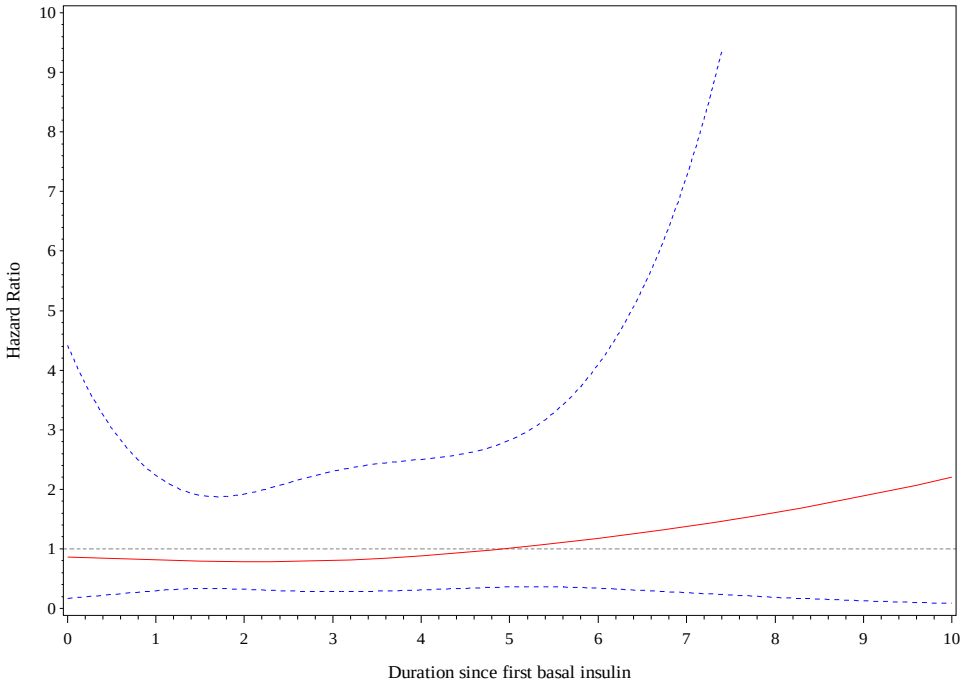


B

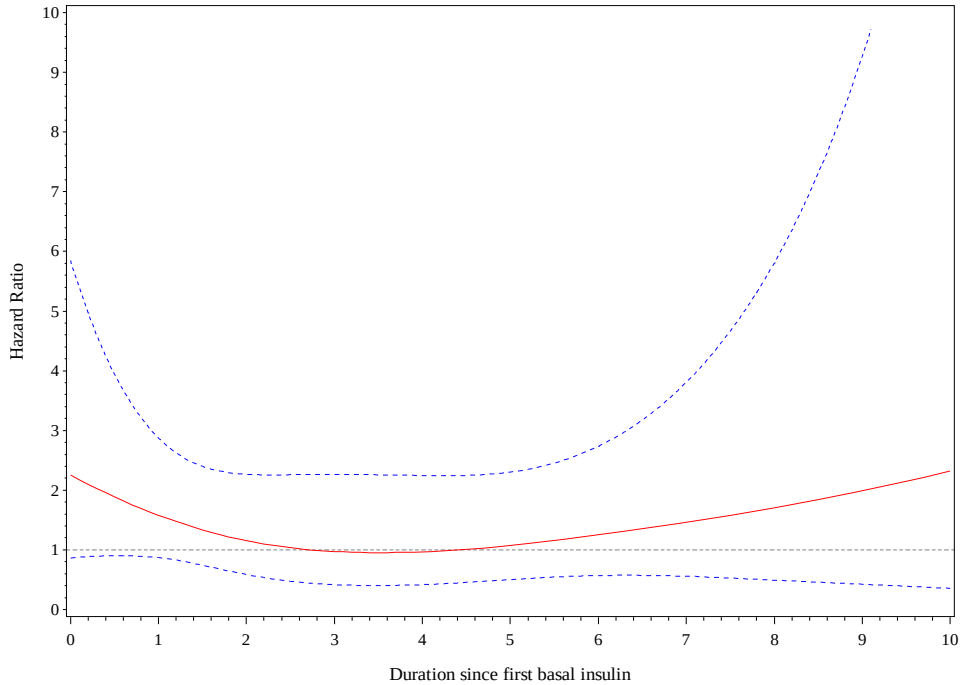


Appendix 6c Restricted cubic splines of the adjusted hazard ratio (solid line) and 95% CIs (dotted lines) for breast cancer incidence as a function of duration since first insulin detemir prescription compared to NPH among new (A) and prior insulin users (B) with three knots placed at the 20th (0.9 years), 50th (2.4), and 80th (5.8) percentiles

A



B



Appendix 7 Crude and adjusted hazard ratios of breast cancer associated with cumulative dose of insulin detemir compared with NPH insulin in the same category in the entire insulin cohort

Exposure	N	Breast cancer cases‡	Person -years	Incidence rate (95% CI)*	Hazard ratios (95% CI)	
					Crude	Adjusted†
Dose-response						
NPH						
<10	9,549	60	16,136	3.7 (2.9, 4.8)	1.00	1.00
10-30	5,133	33	11,306	2.9 (2.1, 4.1)	1.00	1.00
>30	2,293	15	7,635	2.0 (1.2, 3.3)	1.00	1.00
Detemir						
<10	3,271	18	6,872	2.6 (1.7, 4.2)	0.81	1.02 (0.58, 1.80)
10-30	2,082	14	5,158	2.7 (1.6, 4.6)	1.17	1.50 (0.78, 2.89)
>30	964	S	2,805	1.8 (0.7, 4.3)	0.99	1.24 (0.44, 3.48)

* Per 1,000 person-years. †The multivariable Cox proportional hazards model was adjusted for the following covariates prior to study cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, missing), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, missing), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, missing). ‡Cell numbers of less than five are suppressed (S) in accordance with the confidentiality agreements of CPRD.

Appendix 8 Crude and adjusted hazard ratios of breast cancer associated with use of long-acting insulin analogues compared to NPH insulin use among new and prior insulin users

Exposure	N	Breast cancer cases	Person-years	Incidence rate (95% CI)	Hazard ratios (95% CI)	
					Crude	Adjusted[†]
New insulin users						
NPH	2,652	34	11,743	2.9 (2.1, 4.1)	1.00	1.00 (reference)
Glargine	4,148	67	20,170	3.3 (2.6, 4.2)	1.13	1.18 (0.77, 1.81)
Detemir	1,262	11	5,452	2.0 (1.1, 3.6)	0.67	0.89 (0.42, 1.88)
Prior insulin users[‡]						
NPH	6,897	74	23,334	3.2 (2.5, 4.0)	1.00	1.00 (reference)
Glargine	5,427	109	28,515	3.8 (3.2, 4.6)	1.24	1.54 (1.11, 2.13)
Detemir	2,009	26	9,382	2.8 (1.9, 4.1)	0.89	1.41 (0.85, 2.35)

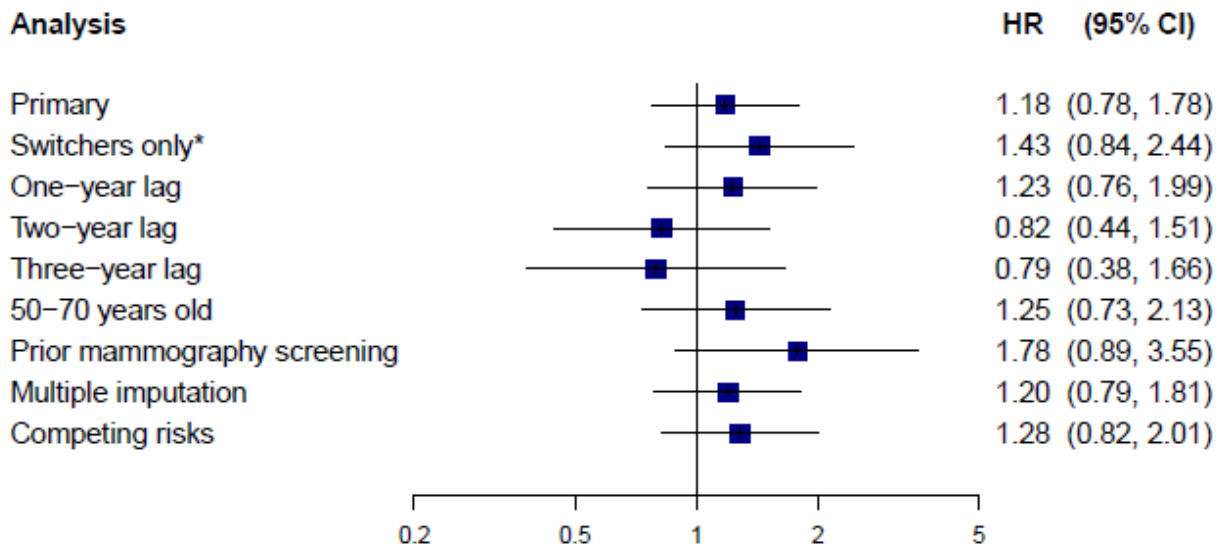
[*] Per 1,000 person-years. [†]The multivariable Cox proportional hazards model was adjusted for the following covariates prior to study cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, missing), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, missing), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, missing). [‡]Prior insulin users are defined as women with at least any type of insulin before study cohort entry.

Appendix 9 Crude and adjusted hazard ratios of breast cancer associated with cumulative dose of long-acting insulin glargine prescriptions compared with NPH insulin in the same category, among new and prior insulin users

	New insulin users						Prior insulin users‡					
Exposure	**N**	**Breast cancer cases**	**PT**	**IR (95% CI)***	**Hazard Ratios (95% CI)**		**N**	**Breast cancer cases**	**PT**	**IR (95% CI)***	**Hazard Ratios (95% CI)**	
					Crude	**Adjusted†**					**Crude**	**Adjusted†**
NPH												
<10	2,652	15	4,927	3.0 (1.8, 5.0)	1.00	1.00	6897	45	11,209	4.0 (3.0, 5.4)	1.00	1.00
10-30	1,685	12	3,671	3.3 (1.9, 5.8)	1.00	1.00	3448	21	7,634	2.8 (1.8, 4.2)	1.00	1.00
>30	890	7	3,144	2.2 (1.1, 4.7)	1.00	1.00	1403	8	4,491	1.8 (0.9, 3.6)	1.00	1.00
Glargine												
<10	4,148	37	8,938	4.1 (3.0, 5.7)	1.42	1.34 (0.72, 2.50)	5427	36	10,933	3.3 (2.4, 4.6)	0.81	0.98 (0.62, 1.55)
10-30	2,645	20	6,652	3.0 (1.9, 4.7)	0.95	0.99 (0.48, 2.50)	3737	41	10,473	3.9 (2.9, 5.3)	1.47	1.85 (1.07, 3.19)
>30	1,298	10	4,580	2.2 (1.2, 4.1)	1.00	1.12 (0.42, 2.98)	1958	32	7,109	4.5 (3.2, 6.4)	2.58	3.24 (1.47, 7.13)
Detemir												
<10	1,262	S	2,907	1.4 (0.5, 3.7)	0.60	0.67 (0.21, 2.16)	2,009	14	3,965	3.5 (2.1, 6.0)	0.92	1.31 (0.68, 2.51)
10-30	720	S	1,739	2.9 (1.2, 6.9)	1.13	1.26 (0.42, 3.81)	1,362	9	3,419	2.6 (1.4, 5.1)	1.17	1.72 (0.75, 3.91)
>30	308	S	807	2.5 (0.6, 9.9)	1.16	1.37 (0.27, 6.92)	656	S	1,998	1.5 (0.5, 4.7)	0.93	1.28 (0.33, 4.95)

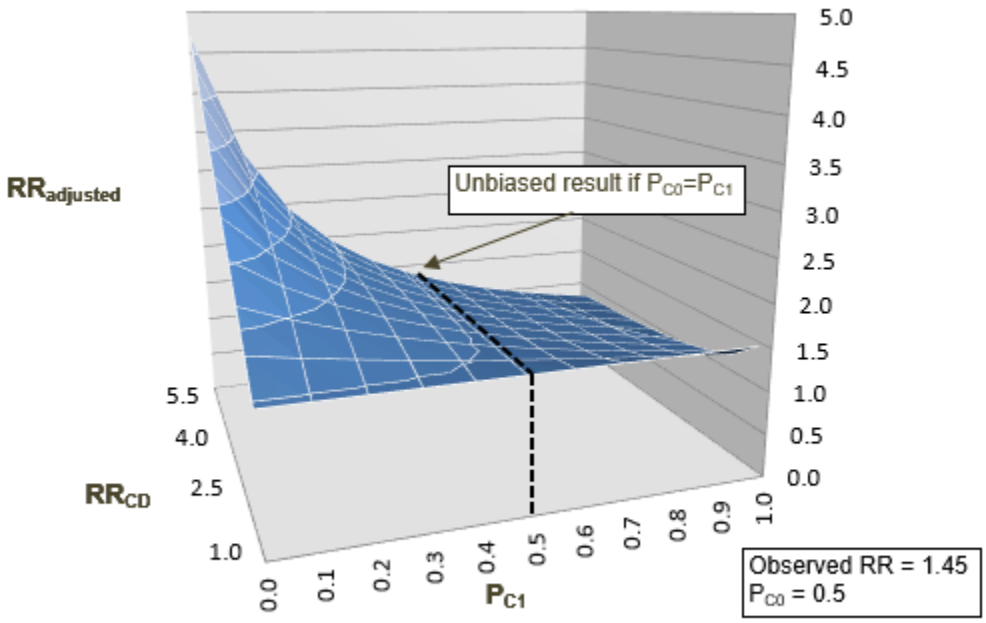
*Per 1,000 person-years. †The multivariable Cox proportional hazards model was adjusted for the following covariates prior to study cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, missing), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, missing), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, missing). ‡Prior insulin users are defined as women with at least any type of insulin before study cohort entry. §Cell numbers less than five were suppressed (S) in accordance with the confidentiality agreements of CPRD. Abbreviations: IR – Incidence rate (crude); PT – Person-time.

Appendix 10 Forest plots of sensitivity analyses for insulin detemir



*Women who switched from NPH or other insulins to insulin detemir

Appendix 11 Array approach for residual confounding based on an observed relative risk (RR) of 1.45 for breast cancer in glargine compared with NPH insulin users and a prevalence of an unmeasured confounder of 50% in the NPH insulin group (P_{C0}=0.50). The prevalence of the unmeasured confounder in the glargine insulin group was varied between 0 to 100% (P_{C1}), while the strength of confounder-disease association (RR_{CD}) was varied between 1.0 to 5.5. The dotted line represents the perfect balance of confounders between glargine and NPH insulin groups (P_{C1}=P_{C0}=0.5), leading to an unbiased observed RR.



Appendix 12a Types of prior insulin users

Types of prior insulin users	Exposure	N (%)
Prior NPH or other insulins*	NPH	6831 (99.0)
	Glargine	3,653 (67.3)
	Detemir	1,497 (74.5)
Prior glargine or detemir insulin	NPH	66 (1.0)
	Glargine	1,774 (32.7)
	Detemir	512 (25.5)

*Other insulins include rapid-acting insulin analogues, short-acting human insulin and animal insulins.

Appendix 12b Crude and adjusted hazard ratios of breast cancer associated with the use of long-acting insulin analogues compared to NPH insulin use among switchers*.

Exposure	N	Events	Person -years	Incidence rate (95% CI) †	HR (95% CI) Crude	HR (95% CI) Adjusted‡
NPH	6,831	74	23,212	3.2 (2.5, 4.0)	1.00	1.00
Glargine	3,653	76	21,077	3.6 (2.9, 4.5)	1.18	1.51 (1.06, 2.16)
Detemir	1,497	22	7,507	2.9 (1.9, 4.5)	0.95	1.43 (0.84, 2.44)

* Women who switched from NPH insulin or other insulins to a long-acting insulin analogue. †Per 1,000 person-years. ‡The multivariable Cox proportional hazards model was adjusted for the following covariates prior to study cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, unknown), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, unknown), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, unknown).

Appendix 13a Frequency and rate of mammography screening before and after study cohort entry

Exposure	Before study cohort entry		After study cohort entry	
	Frequency of mammography at any time	Frequency of mammography in the last three years	Frequency of mammography at any time	Crude rate (95% CI)*
NPH	41.8	21.3	19.9	8.8
Glargine	44.7	27.5	35.6	13.0
Detemir	51.3	33.6	38.6	14.0

* Per 100 person-years.
Before study cohort entry frequency was defined as having at least one mammogram at any time prior to study cohort entry within the specified time window.
After study cohort entry frequency was defined as having at least one mammogram at any time during follow-up.
Rate of mammography was defined as the total number of mammography screenings over the entire follow-up.

Appendix 13b Age-stratified analyses of long-acting insulin analogues and breast cancer incidence in the entire cohort

Age group (years)	Exposure	N	Events‡	Person-years	Incidence rate (95% CI) *	HR (95% CI)	
						Un-adjusted	Adjusted†
<50	NPH	489	S	2,172	0.9 (0.2, 3.7)	1.00	1.00
	Glargine	1,261	16	7,227	2.2 (1.4, 3.6)	2.38	2.51 (0.52, 12.01)
	Detemir	476	S	2,362	0.8 (0.2, 3.4)	0.99	0.34 (0.02, 7.71)
50-70	NPH	4,038	51	17,704	2.9 (2.2, 3.8)	1.00	1.00
	Glargine	4,956	102	28,921	3.5 (2.9, 4.3)	1.24	1.45 (1.02, 2.06)
	Detemir	1,802	25	8,723	2.9 (1.9, 4.2)	0.96	1.25 (0.73, 2.14)
>70	NPH	5,022	55	15,200	3.6 (2.8, 4.7)	1.00	1.00
	Glargine	3,358	58	12,536	4.6 (3.6, 6.0)	1.30	1.38 (0.93, 2.05)
	Detemir	993	10	3,748	2.7 (1.4, 5.0)	0.76	1.10 (0.54, 2.26)

* Per 1,000 person-years. †The multivariable Cox proportional hazards model was adjusted for the following covariates prior to study cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, unknown), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, unknown), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, unknown). ‡Cell numbers less than five were suppressed (S) in accordance with the confidentiality agreements of CPRD.

Appendix 13c Prior mammography screening stratified analyses of long-acting insulin analogues and breast cancer incidence in the entire cohort.

Prior mammography screening	Exposure	N	Events	Person-years	Incidence rate (95% CI)*	HR (95% CI)	
						Unadjusted	Adjusted†
At any time prior to study cohort entry							
	NPH	5,569	65	18,142	3.6 (2.8, 4.6)	1.00	1.00
No	Glargine	5,294	81	25,028	3.2 (2.6, 4.0)	0.91	1.10 (0.77, 1.56)
	Detemir	1,592	14	6,883	2.0 (1.8, 3.4)	0.57	0.89 (0.47, 1.68)
	NPH	3,980	43	16,935	2.5 (1.9, 3.4)	1.00	1.00
Yes	Glargine	4,281	95	23,657	4.0 (3.3, 4.9)	1.62	1.98 (1.35, 2.89)
	Detemir	1,679	23	7,952	2.9 (1.9, 4.4)	1.13	1.48 (0.85, 2.59)
Three years prior to study cohort entry							
	NPH	7,516	85	25,985	3.3 (2.6, 4.0)	1.00	1.00
No	Glargine	6,943	115	33,318	3.5 (2.9, 4.1)	1.07	1.29 (0.95, 1.75)
	Detemir	2,173	18	9,429	1.9 (1.2, 3.0)	0.58	0.89 (0.51, 1.55)
	NPH	2,033	23	9,091	2.5 (1.7, 3.8)	1.00	1.00
Yes	Glargine	2,632	61	15,367	4.0 (3.1, 5.1)	1.59	1.93 (1.17, 3.20)
	Detemir	1,098	19	5,405	3.5 (2.2, 5.5)	1.36	1.78 (0.89, 3.55)

* Per 1,000 person-years. †The multivariable Cox proportional hazards model was adjusted for the following covariates prior to study cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, unknown), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, unknown), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, unknown).

Appendix 14 Missing data patterns for glargine and detemir*

Missing data pattern	N (%)	Age	Calendar year	Diabetes duration (years)	Prior insulin duration (years)
Glargine					
No missing	16,447 (86.0)	67.1	2006.6	9.1	3.4
HbA1c	974 (5.1)	64.7	2005.1	5.8	3.0
BMI	834 (4.4)	75.0	2005.7	7.6	3.6
Smoking status	122 (0.6)	68.3	2003.6	7.3	3.0
HbA1c, BMI	415 (2.2)	70.5	2004.8	5.1	2.9
HbA1c, Smoking	26 (0.1)	68.0	2002.8	4.7	1.8
BMI, Smoking	161(0.8)	78.2	2004.0	6.0	3.2
HbA1c, BMI, smoking	145 (0.8)	72.2	2003.4	4.0	2.5
Detemir					
No missing	10,922 (85.2)	68.1	2006.8	9.6	4.1
HbA1c	676 (5.3)	65.7	2004.9	5.9	2.9
BMI	574 (4.5)	75.7	2005.5	7.7	3.8
Smoking status	73 (0.6)	71.6	2003.4	7.5	2.6
HbA1c, BMI	311 (2.4)	70.6	2004.6	5.0	2.8
HbA1c, Smoking	19 (0.2)	69.2	2002.7	4.3	1.4
BMI, Smoking	119 (0.9)	79.6	2003.6	5.8	3.1
HbA1c, BMI, smoking	123 (1.0)	73.3	2003.3	3.8	2.3

Abbreviations: BMI – body mass index; HbA1c – glycated haemoglobin.
*Variables are represented as mean values of the study population.

Chapter 6. Manuscript 3 – Alternative methods to assess risk in a comparative effectiveness study: an illustration with insulin glargine and breast cancer

6.1. Preamble

In the background chapter, we discussed the goal of comparative effectiveness research is to assess the risks and benefits of a newly available treatment in the real world setting using observational studies. In chapter 4, we observed a few comparative effectiveness observational studies on long-acting insulin analogues and cancer incidence using new user cohort study designs or matched on prior history of treatment.[12 13 15 16 23-25] These study designs, particularly the new user cohort study, have been recommended for comparative effectiveness research.[139 141 142] However, new user study designs reduce the sample size and may prevent a detection of an effect on outcomes with a long latency. Importantly, we cannot fully assess the safety profile of treatments in the entire patient population, which is contradictory to the purpose of comparative effectiveness research. Matching techniques that are currently used in practice to account for patients with a prior history of treatment may be infeasible with increasing number of patients using the newer treatments compared with older treatments over the calendar period. Hence, the primary motivation for this last manuscript was to explore and compare two different study designs and data analytical techniques – time-dependent and pseudo-matching approach – to account for all types of patients (new and prior treatment users) in a comparative effectiveness study using insulin glargine and breast cancer as an illustration. The findings from this manuscript will provide alternative methods to account for all patients in the real world setting in comparative effectiveness observational studies. In addition, we discuss the strengths and limitations of each study design and data analytical approach.

This manuscript is being submitted to the *Pharmacoepidemiology and Drug Safety* journal.

6.2. Title page

Title: Alternative methods to assess risks in a comparative effectiveness study: an illustration with insulin glargine and breast cancer

Authors: Jennifer W. Wu[1,2], Laurent Azoulay[1,2,3], and Samy Suissa[1,2,4]

Affiliations:

[1] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, Montreal, Canada

[2] Centre for Clinical Epidemiology, Lady Davis Institute, Jewish General Hospital, Montreal, Canada

[3] Gerald Bronfman Department of Oncology, McGill University, Montreal, Canada

[4] Division of Clinical Epidemiology, Department of Medicine, McGill University, Montreal, Canada

Corresponding author:

Dr. Samy Suissa
Centre for Clinical Epidemiology
Lady Davis Institute, Jewish General Hospital
3755 Cote Ste-Catherine, H-461
Montreal, Quebec, Canada, H3T 1E2
Tel: (514) 340-7593
Fax: (514) 340-7564
Email: samy.suissa@mcgill.ca

6.3. Abstract

Purpose: Comparative effectiveness observational studies are conducted to evaluate the safety profile of a new treatment in all patients in the real world. Therefore, we explore two study designs and data analytical techniques to account for all patients (new and prior treatment users) in a comparative effectiveness study of insulin glargine compared with Neutral Protamine Hagedorn (NPH) and breast cancer risk.

Methods: Two approaches were presented and illustrated with data from the United Kingdom's Clinical Practice Research Datalink. The time-dependent approach used a study cohort of women with type 2 diabetes using insulin glargine or NPH and classified women as unexposed until the first insulin glargine prescription and exposed subsequently. The pseudo-matching approach defined a study cohort from first insulin glargine prescription and randomly selected an NPH insulin prescription (comparator) to be similar to insulin glargine users regarding the prior history of insulin treatment.

Results: In the time-dependent approach, 7,042 and 13,261 women with first insulin glargine and NPH prescriptions, respectively, were included. In the pseudo-matching approach, 9,575 insulin glargine and 9,550 NPH users were included. The hazard ratio for breast cancer with insulin glargine use compared with NPH users was 1.46 (95% confidence intervals [CI]: 1.16, 1.84) and 1.44 (95% CI: 1.11, 1.85) using the time-dependent and pseudo-matching approaches, respectively.

Conclusion: The time-dependent and pseudo-matching approaches provide alternatives to including new and prior treatment users in comparative effectiveness studies. However, there are complexities in each study design and data analytical technique that needs to be considered.

6.4. Introduction

Comparative effectiveness research involves comparing newer with older interventions or treatments to assess the benefits and risks in the real world setting.[137] To study the risk or safety profile of a new treatment, there are many approaches including spontaneous reports, meta-analyses of randomised controlled trials (RCTs), and observational studies.[138] Observational studies have been recommended because these types of studies are better at evaluating the safety profile of a new treatment in all types of patients (new and prior treatment users) in the real world.[138 205 206] Comparative effectiveness observational studies have typically employed two study designs: new user or matched cohort study design.[139 141 142] Both these types of study designs have been used to assess the effect of insulin glargine on the risk of breast cancer.[12 13 15 16 23-25]

A new user cohort study design restricts the study population to new users of a treatment defined by a minimum time period of non-use (of the treatment or treatment class of interest) prior to cohort entry.[141] This study design provides the opportunity to study the risk profile from treatment initiation.[141 180] However, this approach reduces the sample size and shorten the duration of follow-up preventing the detection of an effect on an outcome with a long latency.[139 141 180] More importantly, the new user study design cannot fully assess the safety profile of the treatment in all patients in the real world, which is the goal of comparative effectiveness research. An alternative approach which allows for the inclusion of both new and prior treatment users is to match on the duration of prior treatment or number of prior prescriptions.[142] Albeit, this method becomes infeasible when the newer treatment is prescribed more frequently over the calendar period such is the case for insulin glargine given its reduced adverse effects of nocturnal hypoglycaemia compared with the intermediate-acting Neutral Protamine Hagedorn (NPH) insulin. Thus, this could create an imbalance of users of newer compared with older treatment making it difficult to match on prior history of treatment as it could exclude a large number of treated individuals.[207]

Given these issues, we explored two different study designs and data analytical to assess the safety profile of the entire patient population in a comparative effectiveness study of insulin glargine on the risk of breast cancer.

6.5. Methods

Data source

To illustrate the two methods, we evaluated the effects of insulin glargine on the risk of breast cancer using data from the United Kingdom (UK) Clinical Practice Research Datalink (CPRD). In brief, the CPRD is a large database of electronic medical records from general physician practices across the UK, Scotland, and Northern Ireland.[143] The CPRD currently has over 14 million patients from nearly 700 practices and includes medical (i.e. diagnostic and procedural) codes and prescriptions based on the Read codes and British National Formulary classification system, respectively.[143] This study was approved by the International Scientific Advisory Committee of CPRD (protocol number 16_213R2), Institutional Review Board of McGill University, and Research Ethics Board of Jewish General Hospital in Montreal, Quebec.

Base cohort

We constructed a base cohort of women 40 years or older with at least one prescription for any insulin, which included rapid-acting insulin analogues, short-acting insulins, basal insulins (intermediate-acting insulins or long-acting insulin analogues), and premixed insulins, from 1 January 1988 to 31 December 2012.

Time-dependent approach

In the time-dependent approach, we defined a study cohort of women based on first insulin glargine or NPH prescription (study cohort entry), whichever prescription appeared first in the base cohort, between 1 September 2002 (the year insulin glargine entered the UK market) and 31 December 2012. New insulin users (first-time users or initiators of insulin) were defined as women with no prescription for any insulin prior to study cohort entry whereas prior insulin users (prevalent insulin users) were defined as women with at least one prescription for any insulin prior to study cohort entry. We excluded women with a previous history of any cancers prior to study cohort entry.

We used a time-dependent exposure definition where women were considered unexposed to insulin glargine until the time of the first prescription for insulin glargine and considered exposed until the end of the follow-up (figure 1a). If women switched to insulin detemir (another

type of long-acting insulin analogue), the time-dependent exposure variable included a category indicating a switch to insulin detemir. This time-dependent exposure definition allows the patient to contribute both to the unexposed (NPH) and exposed (glargine) person-time. Women were followed from study cohort entry to breast cancer event, death due to any causes, end of CPRD registration, or end of study date (February 28 2015), whichever came first.

Pseudo-matching approach

In the pseudo-matching approach, we created a study cohort of women with at least one prescription for insulin glargine or NPH from 1 September 2002 to 31 December 2012 from the base cohort. For the exposure definition, women with at least one prescription for insulin glargine would be considered as exposed to insulin glargine and the remaining women with at least one NPH insulin prescription would be considered exposed to NPH insulin (reference category). Similar to the time-dependent approach, new insulin users (first-time users or initiators of insulin) were defined as women with no prescription for any insulin prior to the first insulin glargine or NPH insulin prescription. Otherwise, women were considered prior insulin users (prevalent insulin users).

For the new insulin users, study cohort entry was defined based on the date of first insulin glargine or NPH prescription. For the prior insulin users, we used a hierarchical exposure definition and consequently insulin glargine users might enter the study cohort later whereas NPH insulin users enter earlier in the study cohort. To ensure that the NPH insulin users has a similar study cohort entry as the insulin glargine users, we randomly selected a NPH insulin prescription among all the prescriptions for NPH insulin between 2002 and 2012 for each patient and use this randomly selected prescription as the study cohort entry date for the NPH insulin user (figure 1b). We replicated the random selection of NPH insulin prescriptions ten times to determine how the selection might alter the estimates using PROC SURVEYSELECT with the simple random sampling method. Study cohort entry was based on first insulin glargine or randomly selected NPH insulin prescription among the prior insulin users.

We excluded women with a previous history of any cancers prior to study cohort entry date. We employed an approach analogous to intent-to-treat (ITT) analyses where we compared women exposed to insulin glargine with women exposed to NPH insulin. Women were followed

from study cohort entry to breast cancer event, death due to any causes, end of CPRD registration, or end of study date (February 28, 2015), whichever came first.

Outcome definition

For both the time-dependent and pseudo-matching approach, we defined breast cancer event as first, primary malignant breast cancer during follow-up using diagnostic codes only (available upon request).

Covariates

For both the pseudo-matching and time-dependent approach, we adjusted for potential confounders, defined prior to study cohort entry, based on a priori knowledge of the relationship between insulin and cancer.[4] We included age, year of study cohort entry, excessive alcohol use, smoking status (never (reference), ever, missing), body mass index (BMI; <25 (reference), 25-30, ≥ 30 kg/m^2, missing), glycated haemoglobin (HbA1c; (<6.5 (reference), 6.5-8.0, >8.0%, missing), diabetes duration (time between study cohort entry date and date of first non-insulin antidiabetic medication or insulin prescription, diagnosis of type 2 diabetes, or HbA1c ≥6.5% value), prior insulin use duration (time between study cohort entry date and date of first-ever of any type of insulin prescription), non-insulin antidiabetic medication use vs. no use (metformin, sulfonylureas, thiazolidinedione [TZD], and others), Deyo's Charlson comorbidity score (≤ 1 (reference), 2-3, >3; excludes previous cancers),[165] and other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications [NSAID]). Other antidiabetic medications included meglitinides, dipeptidyl-peptidase-4 (DPP-4) inhibitors, glucagon-like-peptide-1 analogues (GLP-1), alpha-glucosidase inhibitors, guar gum, and sodium/glucose cotransporter-2 inhibitors (SGLT2i). In the primary analysis, we included an indicator for missingness for smoking, BMI and HbA1c categorical variables because missingness was minimal (~5%).

Data analysis

For both the time-dependent and pseudo-matching approach, we used the time-dependent and independent Cox proportional hazards model, respectively, to calculate the unadjusted and adjusted hazard ratios (HR) and 95% confidence intervals (CI) of breast cancer associated with

use of insulin glargine compared with NPH insulin. Duration of follow-up was the timescale for both Cox models. We also performed a stratified analysis by new and prior insulin users. For the pseudo-matching approach and among prior insulin users only, we presented the mean number of NPH insulin prescriptions that were available for the random selection process, the mean number of NPH insulin prescriptions between the first NPH insulin prescription and study cohort entry date (i.e. first insulin glargine or randomly selected NPH insulin prescription), and the mean time from the first NPH insulin prescription to study cohort entry date. All data analyses were conducted with SAS version 9.4 (SAS Institute, Cary, NC). Plots were constructed with the 'ggplot2' package from R version 3.3.1 (R Development Core Team, Vienna, Austria).

6.6. Results

Time-dependent approach

In the time-dependent approach, a total of 7,042 and 13,261 women with first insulin glargine and NPH prescription, respectively, were included in the study (appendix 1). In general, insulin glargine users compared with NPH were younger, started after 2006, lower BMI but had higher HbA1c and more alcohol-related diseases (table 1). Diabetes duration was more similar in the two insulin groups with this particular approach.

In the time-dependent approach, we observed insulin glargine increased the risk of breast cancer in the entire cohort (HR: 1.46; 95% CI: 1.16-1.84) (table 2). The risk of breast cancer was also observed among prior insulin users (HR: 1.55; 95% CI: 1.16, 2.06) but not new insulin users (HR: 1.16; 95% CI: 0.79-1.72).

Pseudo-matching approach

From 2002-2012, the number of prescriptions for long-acting insulin analogues, particularly glargine, increased while the number of prescriptions for NPH insulin decreased (figure 2). In the pseudo-matching approach, 9,575 insulin glargine users (included women who switched from NPH or other insulins) and range from 9,539 to 9,567 NPH insulin users (included women using NPH insulin only) were identified in each of the ten replicates (appendix 2). Similar to the time-dependent approach, insulin glargine users compared with NPH (in one replicate) were younger, started after 2006, earlier stages of type 2 diabetes (short diabetes

duration and used earlier lines of antidiabetic medication), lower BMI, fewer comorbidities, but had higher HbA1c and more alcohol-related diseases (table 1). In the other replicates, the NPH insulin users had similar baseline characteristics (appendix 3).

Among NPH insulin users with a prior history of insulin use, the mean number of NPH insulin prescriptions between the first and randomly selected NPH insulin prescription was 16.6 ± 20.1 (out of a mean number of 33 ± 32 prescriptions available for random selection) and the mean time was 1.9 ± 2.2 years (appendix 4). Similarly, among switchers from NPH insulin to glargine, the mean average number of NPH insulin prescriptions between the first NPH insulin and first insulin glargine prescription was 15.3 ± 18.7 and mean time was 2.3 ± 2.2 years.

Similar to the time-dependent approach, we observed, after adjustment for potential confounders, insulin glargine was associated with an increased risk of breast cancer in the entire cohort (HR: 1.44; 95% CI: 1.11-1.85) (table 2). The risk of breast cancer was also observed among prior insulin users (HR: 1.52; 95% CI: 1.10, 2.10) but not new insulin users (HR: 1.18; 95% CI: 0.77-1.81). When we replicated the random selection of NPH insulin prescription, we observed similar HRs of 1.32-1.47 (figure 3).

6.7. Discussion

Summary

In this study, we explored two alternative approaches, time-dependent and pseudo-matching, which could be used to conduct comparative effectiveness research to assess the safety profile of a new treatment in all types of patients – new and prior treatment users – in the real world setting. We demonstrated using insulin glargine and breast cancer as an example that both the pseudo-matching and time-dependent approaches yielded similar results and provided an alternative to evaluating the safety profile of a new treatment in all patients. However, there were complexities to both approaches that need to be considered when choosing one approach over the other.

Current approaches in comparative effectiveness observational studies

New user and matched cohort study designs in comparative effectiveness observational studies have been used. In the 14 observational studies of insulin glargine and breast cancer

risk,[12-25] nine studies have employed either a new user or matched (on prior history of insulin use) cohort study designs.[12-16 18 23-25] Although the new user cohort study design eliminates prevalent user bias,[141] it does reduce the sample size and shorten the duration of follow-up preventing the ability to detect an effect on outcomes with a long latency.[141 180] More importantly, the goal of comparative effectiveness research is to evaluate the safety profile of a treatment in all patients in the real world setting. For example, five out of the seven studies on insulin glargine and breast cancer risk produced null associations as a result of the new user cohort study design and reducing the duration of follow-up time.[13-16 24] In contrast, a previous observational study using a matched cohort study design observed a difference in risk between new and prior insulin users, particularly after five years of use among women with a prior history of insulin use, with a HR 2.7 (95% CI: 1.1-6.5).[25] Given the difference in risk among new and prior treatment users, it is important to consider all patients in the real world setting to fully evaluate the safety profile of a new treatment.

An alternative approach which allows for the inclusion of both new and prior treatment users is to match on prior history of treatment. More specifically, this study design technique matches patients using newer to older treatments on either duration or number of prescriptions of prior treatment.[142] However, a challenge arises when a new treatment enters the market, and patients are prescribed the new treatment more frequently over time because it is more effective. For example, insulin glargine has been more effective at reducing nocturnal hypoglycaemia resulting in physicians prescribing this medication more frequently over time.[104 105] As we observed in our study, from 2002-2012, there has been increasing number of prescriptions for long-acting insulin analogues compared with NPH insulin. Thus, this increasing trend of long-acting insulin analogues created an imbalance of users making it infeasible to match on prior history of treatment as it could exclude a large number of treated individuals reducing the sample size.[207]

Simplicities and complexities of the time-dependent approach

In the time-dependent approach, the design is simple, but the data analysis is complex. With this approach, we can simply start with a cohort of old (NPH insulin) and new (insulin glargine) treatments and determine cohort entry based on the date of the first old or new prescription, whichever occurred first. However, in the data analysis, we would need to use a

time-dependent exposure and depend on the unit of time and the database structure it could increase the computational time. A simple method to mitigate this issue is to calculate and cumulate the time prior to the insulin glargine prescription as unexposed (one line of observation) and the time after the switch to insulin glargine (second line of observation) as exposed until the end of follow-up. Consequently, this could reduce the computational time. Another issue to consider is if there is another newer treatment that patient can switch to (e.g. insulin detemir) the author needs to consider censoring the switch or adding another indicator for switching. The latter may be preferred because if the patient is censored for switching and censoring is informative (i.e. switching to this alternative treatment is affected by both exposure and outcome), then we may introduce selection bias.[208] Another important issue with the time-dependent approach is the potential for residual confounding due to the lack of adjustment for time-dependent confounders. Indeed, we cannot adjust for covariates at the exact moment when patients are switching from NPH insulin to glargine in a conventional Cox model since the model calculates the hazard ratios at the time an event is occurring and not when treatment switching is occurring. Although one can simply add time-dependent confounders into the Cox proportional hazard models; however, if the time-dependent confounder is also a mediator in the causal pathway then other methods such as marginal structural Cox models may be needed.[209] Albeit the study design is more straightforward; there are a lot more consideration and complexity in the data analysis stage.

Simplicities and complexities of the pseudo-matching approach

In the pseudo-matching approach, the design of this approach is complex while the data analysis is simple. First, in the pseudo-matching approach, one has to identify patients who are using or switching to the treatment of interest (insulin glargine) within the cohort. The remaining patients in the cohort using the comparator (NPH insulin) would be the referent. This is a hierarchical exposure definition, and we are "looking into the future" and by doing so, we could be introducing immortal time bias.[210] However, to minimise the immortal time bias that may be present, in our second step, we randomly selected a NPH insulin prescription as the new cohort entry among the referent with a prior history of insulin use. The goal of this step is to "match" or pseudo-match on the time between first starting NPH insulin and switching to insulin glargine. Aforementioned, matching on prior history of insulin use either on time since initiation of insulin

or number of insulin prescriptions is another proposed technique to mitigate immortal time bias further.[142] Unlike the matching technique, one caveat of the pseudo-matching approach is that it would not work for fatal outcomes such as death. For example, if patients using the active comparator are more likely to die early, before they had a chance to switch to the study drug, for instance, the randomly selected prescriptions may be skewed to the early follow-up. Consequently, immortal time bias may not be minimised using the pseudo-matching approach and immortal time bias may persist.

Another added level of complexity is to ensure that during this random selection of the new cohort entry dates among the referent group selection bias is not introduced. In cohort studies such as insulin glargine and the risk of breast cancer, it is common to exclude patients with a prior history of any or breast cancer. Without a step-wise process, it is possible to misclassify a prior (to cohort entry) cancer event as an outcome and vice versa. Consequently, it is important in the pseudo-matching approach that we exclude patients with a prior history of cancer event after the random selection of the NPH insulin prescription as the cohort entry date and not during the process of random selection. This selection bias is not specific to the pseudo-matching approach; it has been discussed with the matched cohort study design approach.[142] Furthermore, an illustration of this selection bias occurred in a cohort study of metformin and cancer risk where metformin users were matched to nonusers.[211] During the matching process, comparators with a prior history of cancer were excluded for one metformin user but may be available to another, and the event would be reclassified as an outcome. However, all metformin users with a prior history of cancer were excluded, and metformin users were not "reused". This differential exclusion, selection bias, yielded a reduction of cancer risk by 40% for metformin users compared with nonusers.[211] This reshuffling of events could either increase or decrease the number of events in the comparator group potentially biasing the results away from the null in either direction.

Therefore, with the pseudo-matching approach, one needs to consider the hierarchical exposure definition used and exclusion of prior history for the outcome of interest and how these two aspects could bias the estimates if the study design was not done systematically. Albeit, once the random selection and exclusions are completed and conducted appropriately, the data analysis is simple in that one can employ the analogous approach of an ITT analysis.

Strengths and limitations of the study

Our study has several strengths. First, this study provided alternative study designs and data analytical techniques to allow for the assessment of the safety profile of a treatment of all patients in the real world setting. Second, since we included all types of patients (new and prior treatment users) it allows the results to be more generalizable. This is particularly important given the goal of comparative effectiveness research.[137]

Despite the strengths of this study, there are some limitations of this methodological study. First, left truncation is an issue for our study such that we are uncertain whether new users of insulin glargine or NPH are truly new users as patients can receive medications from other healthcare professionals (e.g. endocrinologists). However, this is an inherent issue with all observational studies using health administrative databases and electronic medical records.[212 213] Moreover, if new insulin users could be prior insulin users, we would have observed an effect of insulin glargine on the risk of breast cancer among new insulin users as well. Second, we cannot provide a direct comparison of whether the previous methods such as matching on prior treatment history will yield similar results. However, we stratified the analyses by new and prior insulin users in both approaches. Our estimates reported in this study conferred with previous null finding in the literature of insulin glargine and breast cancer using new user study designs.[13-16 24] This secondary analysis provided a direct comparison to existing studies using the established new user study design. The results from our stratified analyses among new insulin users further justified why we need to include prior insulin users given the difference in the risk of breast cancer. Third, in the pseudo-matching approach, despite our attempts to randomly select an NPH insulin prescription to pseudo-match on the prior history of insulin use, we were not able to match exactly on the prior duration of insulin use or number of prior insulin prescriptions. However, the difference in the prior duration of insulin use between insulin glargine and NPH users was smaller compared with the time-dependent approach. Fourth, again with the pseudo-matching approach, it is possible that we did not fully account for all the immortal time that could be introduced due to the hierarchical exposure definition. If patients had only one prescription, for example, the random selection process would select the same prescription, and as a result, we may not have fully accounted for the time between switching from NPH or other insulins to insulin glargine. However, our study identified women had a mean of 33 NPH insulin prescriptions available for the random selection process. Future research

studies need to test the limits of the pseudo-matching design with a variable proportion of patients with only one or a small number of prescriptions for the random selection process and its impact on the effect estimates.

6.8. Conclusion

In summary, the pseudo-matching and time-dependent approaches are alternative solutions to including all types of patients in the real world setting to assess the safety profile of a new treatment. This is particularly important given that the goal of comparative effectiveness research is to evaluate the safety profile of a new treatment in all patients. However, important considerations should be given on the complexity of the study designs and data analytical techniques.

6.9. Tables

Table 6.1 Baseline characteristics of women with at least one prescription for insulin glargine or NPH according to insulin exposure at study cohort entry between 2002-2012

Covariates	Study cohort			
	Time-dependent approach		Pseudo-matching approach	
	Glargine	NPH*	Glargine	NPH
Number of women	7,042	13,261	9,575	9,550
Age †	64.7 ± 13.1	67.1 ± 11.9	64.9 ± 12.9	70.2 ± 11.6
Calendar year†				
2002-2005	1,642 (23.3)	9,432 (71.1)	2,901 (30.3)	4,203 (44.0)
2006-2008	2,908 (41.3)	2,162 (16.3)	3,778 (39.5)	2,997 (31.4)
2009-2012	2,492 (35.4)	1,667 (12.6)	2,896 (30.3)	2,350 (24.6)
Excessive alcohol use‡	363 (5.2)	371 (2.6)	421 (4.4)	331 (3.5)
Smoking status‡				
Ever	4,530 (64.3)	8,096 (61.1)	6,245 (65.2)	6,108 (64.0)
Never	2,428 (34.5)	4,463 (33.7)	3,205 (33.5)	3,109 (32.6)
Unknown	84 (1.2)	702 (5.3)	125 (1.3)	333 (3.5)
BMI (kg/m^2) §				
<25	1,444 (20.5)	2,287 (17,3)	2,066 (21.6)	1,495 (15.7)
25-30	1,966 (27.9)	3,435 (25.9)	2,699 (28.2)	2,432 (25.5)
≥ 30	3,259 (46.3)	5,906 (44.5)	4,296 (44.9)	4,579 (48.0)
Unknown	373 (5.3)	1,633 (12.3)	514 (5.4)	1,044 (10.9)
HbA1c (%)‖				
<6.5	222 (3.2)	687 (5.2)	332 (3.5)	735 (7.7)
6.5-8.0	1,374 (19.5)	3,085 (23.3)	1,964 (20.5)	2,836 (29.7)
>8	5,065 (71.9)	6,768 (51.0)	6,750 (70.5)	4,946 (51.8)
Unknown	381 (5.4)	2,721 (20.5)	529 (5.5)	1,033 (10.8)
Diabetes duration †	7.5 ± 5.3	7.7 ± 4.9	8.2 ± 5.4	9.2 ± 5.5
Prior duration of insulin use†	1.3 ± 3.0	3.1 ± 1.4	2.8 ± 4.3	4.3 ± 4.6
Non-insulin diabetes medication use‖**				
Metformin	4,807 (68.3)	6,552 (49.4)	5,830 (60.9)	4,830 (50.6)
Sulfonylurea	4,231 (60.1)	4,702 (35.5)	4,661 (48.7)	2,916 (30.5)
Thiazolidinedione	1,884 (26.8)	1,438 (10.8)	2,048 (21.4)	986 (10.3)
Others††	1,077 (15.3)	1,112 (8.4)	1,177 (12.3)	744 (7.8)
Charlson comorbidity score¶				
≤ 1 comorbidities	3,597 (51.1)	8,056 (60.8)	5,052 (52.8)	4,988 (52.2)
2-3	2,557 (36.3)	4,067 (30.7)	3,376 (35.3)	3,359 (35.2)
>3	888 (12.6)	1,138 (8.6)	1,147 (12.0)	1,203 (12.6)
Statin use‖	5,079 (72.1)	7,453 (56.2)	6,795 (71.0)	6,367 (66.7)
Aspirin use‖	3,431 (48.7)	6,338 (47,8)	4,722 (49.3)	5,257 (55.1)
NSAID use‖	2,592 (36.8)	5,172 (39.0)	3,631 (37.9)	3,523 (36.9)

Data are presented as mean (SD) or n (%) unless otherwise specified.
*NPH insulin users included women who switch to either long-acting insulin analogue glargine or detemir. †Measured at study cohort entry. ‡Measured at any time prior to study cohort entry. §Measured five years prior to study cohort entry. ‖Measured two years prior to study cohort entry. ¶Measured one year prior to study cohort entry. **Not mutually exclusive. ††Other non-insulin diabetes medications included meglitinides, DPP-4i, GLP-1, alpha-glucosidase inhibitors, guar gum, and SGLT2i. Abbreviations: BMI – body mass index; HbA1c – glycated haemoglobin; NSAID – nonsteroidal anti-inflammatory drugs

Table 6.2 Crude and adjusted hazard ratios of breast cancer associated with use of insulin glargine compared with NPH, in the entire cohort and among new and prior insulin users

Types of users	Exposure	N	Cases	Person-years	Incidence rate (95% CI)*	Hazard ratio (95% CI)	
						Crude	Adjusted†
Time-dependent approach‡§							
Overall	NPH	13,261	179	58,613	3.1 (2.6,3.5)	1.00	1.00
	Glargine	9,717	177	55,704	3.2 (2.7, 3.7)	1.19	1.46 (1.16, 1.84)
New insulin user	NPH	3,191	42	12,871	3.3 (2.4, 4.4)	1.00	1.00
	Glargine	4,528	77	22,686	3.4 (2.7, 4.2)	1.08	1.16 (0.79, 1.72)
Prior insulin user	NPH	10,070	137	45,742	3.0 (2.5, 3.5)	1.00	1.00
	Glargine	5,189	100	33,019	3.0 (2.5, 3.7)	1.24	1.55 (1.16, 2.06)
Pseudo-matching approach							
Overall	NPH	9,550	109	35,374	2.9 (2.5, 3.4)	1.00	1.00
	Glargine	9,575	176	48,685	3.6 (3.1, 4.2)	1.19	1.44 (1.11, 1.85)
New insulin user	NPH	2,652	34	11,743	2.6 (2.0, 3.5)	1.00	1.00
	Glargine	4,148	67	20,170	3.3 (2.6, 4.2)	1.13	1.18 (0.77, 1.81)
Prior insulin user	NPH	6,898	75	23,631	3.1 (2.5, 3.7)	1.00	1.00
	Glargine	5,427	109	28,515	3.8 (3.2, 4.6)	1.23	1.52 (1.10, 2.10)

* Per 1,000 person-years. †The multivariable Cox proportional hazards model was adjusted for the following covariates prior to cohort entry: age (years), calendar year, diabetes duration (years), prior duration of insulin use (years), HbA1c (<6.5 (reference), 6.5-8.0, >8.0%, unknown), use of antidiabetic medications vs. no use (metformin, sulfonylureas, TZD, and others), other medication use vs. no use (statins, aspirin and non-steroidal anti-inflammatory medications), BMI (<25 (reference), 25-30, ≥ 30 kg/m^2, unknown), Charlson comorbidity score (≤ 1 (reference), 2-3, >3), excessive alcohol use vs. no use, smoking status (never (reference), ever, unknown). ‡The time-dependent exposure in the Cox models included an indicator for women who switched from insulin NPH to detemir. §Women can contribute person-time to both insulin NPH and glargine.

6.10. Figures

Figure 6.1 Illustration of the time-dependent (A) and pseudo-matching approach (B) using insulin glargine compared with NPH insulin as an example

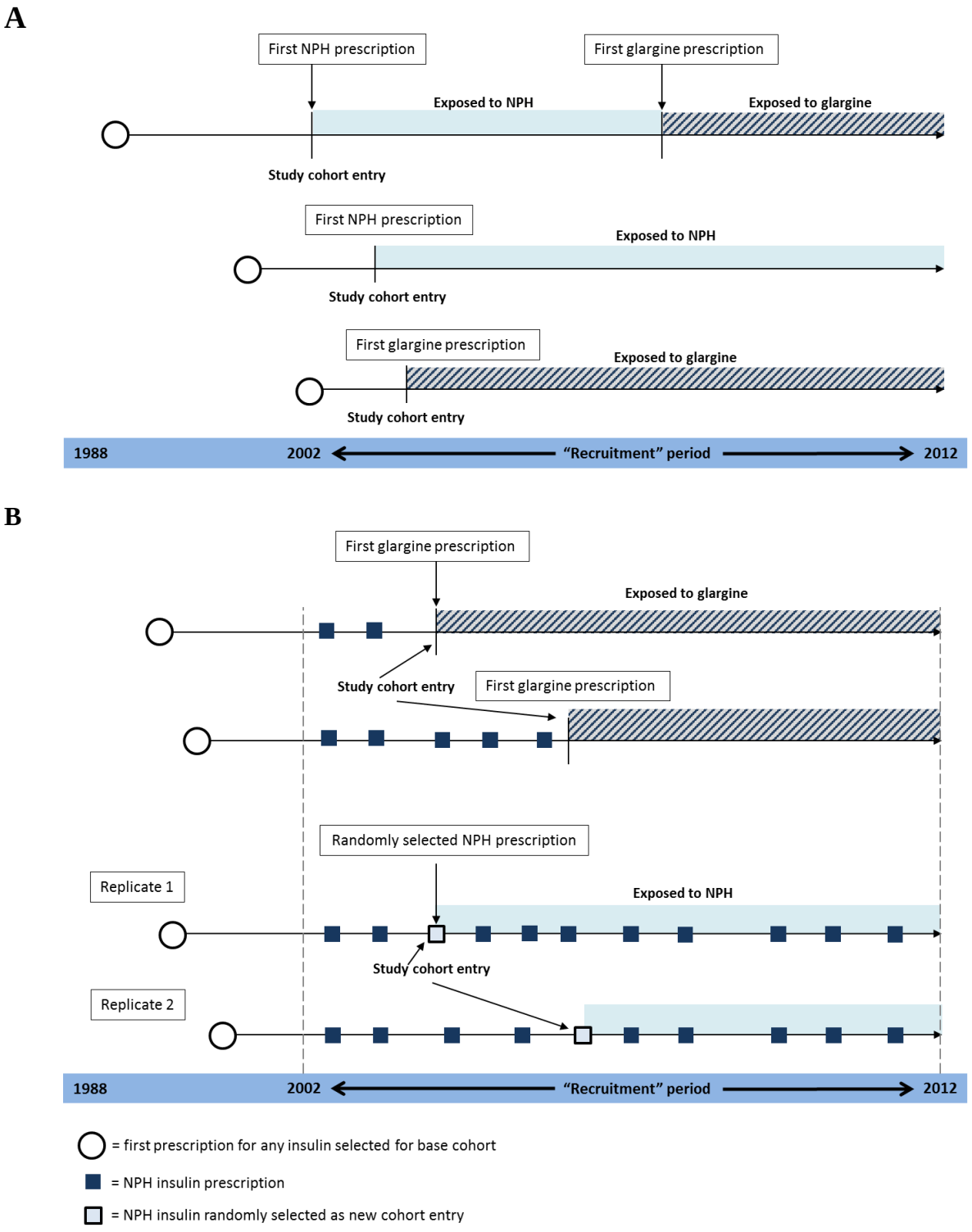


Figure 6.2 Basal insulin prescription from 2002-2012 among women with type 2 diabetes using insulin

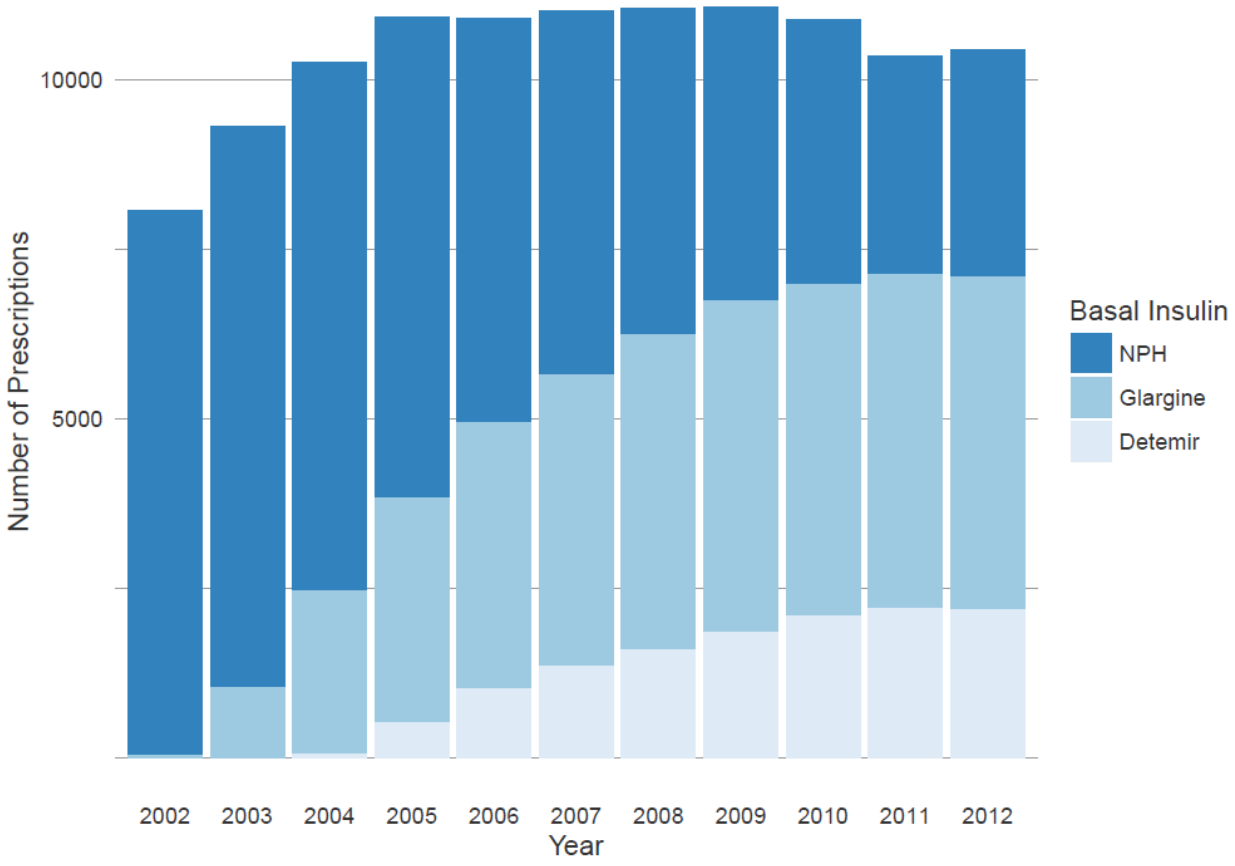
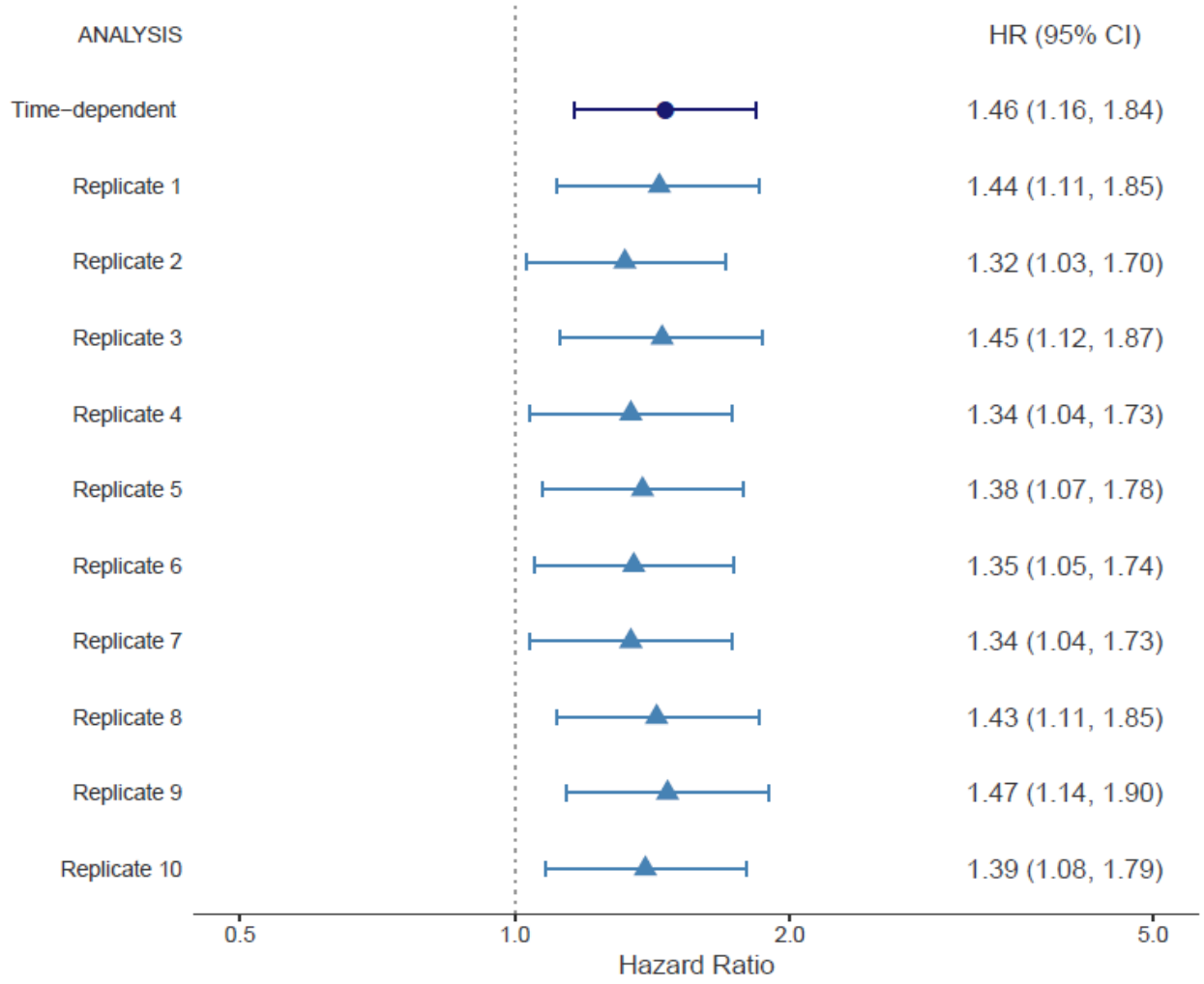
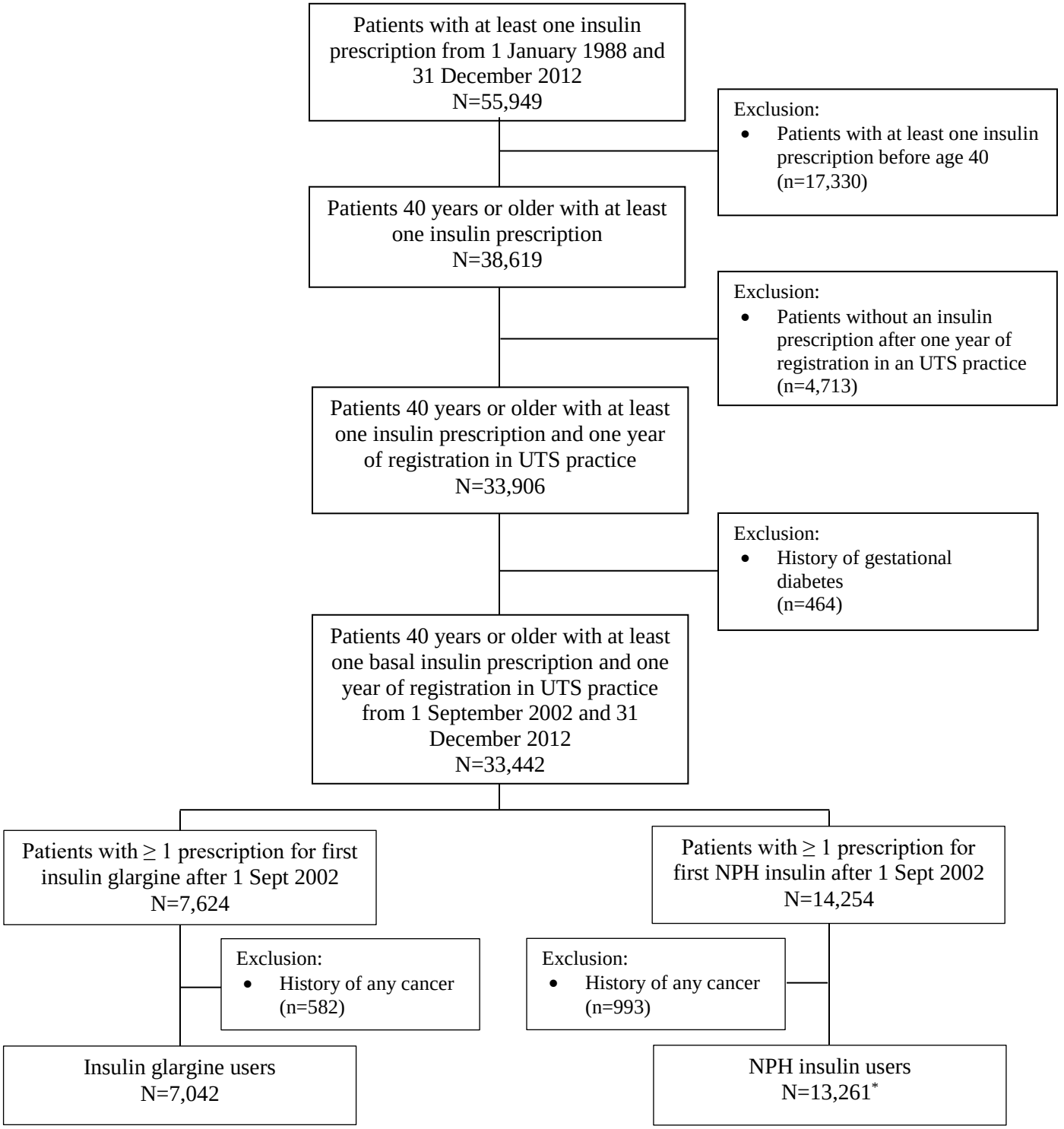


Figure 6.3 Forest plots of hazard ratios and 95% CIs of the time-dependent and pseudo-matching approaches (replicate 1-10)



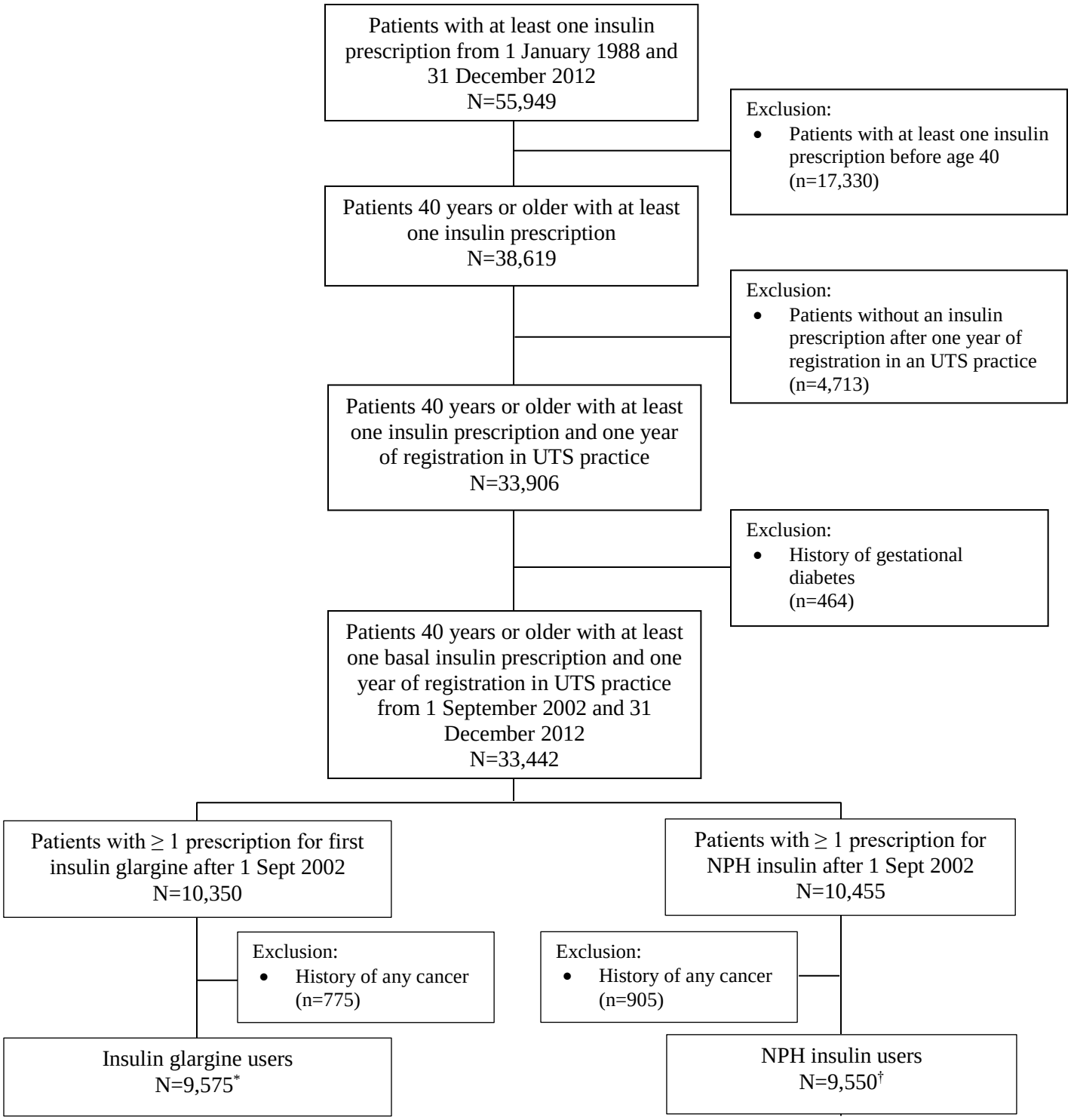
6.11. Appendices

Appendix 1 Flow chart describing the selection of 20,403 women 40 years or older with at least one prescription for any basal insulin between 1 September 2002 and 31 December 2012 for the time-dependent approach



*NPH insulin users contain switchers (i.e. women who switched to insulin glargine or detemir)

Appendix 2a Flow chart describing the selection of 19,125 women 40 years or older with at least one prescription for any basal insulin between 1 September 2002 and 31 December 2012 for the pseudo-matching approach based on one random sample (replicate 1)



*Includes women who switched from NPH or other insulins; †Includes NPH insulin users only

Appendix 2b Final number of NPH insulin users included in the study after excluding women with a history of any cancers prior to cohort entry in nine replications

	Replicates								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Exclusion: History of any cancer	888	912	908	907	906	916	914	912	903
Final NPH insulin user	9,567	9,543	9,547	9,548	9,550	9,539	9,541	9,543	9,552

Appendix 3 Baseline characteristics of NPH insulin users for replicates 2-10

Covariates	Replicates								
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Measured at cohort entry									
Number of women	9,567	9,543	9,547	9,548	9,550	9,539	9,541	9,543	9,552
Age (years)	70.2 ± 11.6	70.2 ± 11.6	70.3 ± 11.6	70.2 ± 11.6	70.2 ± 11.6	70.2 ± 11.6	70.2 ± 11.6	70.2 ± 11.6	70.2 ± 11.6
Calendar year									
2002-2005	4,208 (44.0)	4,176 (43.8)	4,115 (43.1)	4,171 (43.7)	4,152 (43.5)	4,155 (43.6)	4,171 (43.7)	4,188 (43.9)	4,158 (43.5)
2006-2008	2,939 (30.7)	3,015 (31.6)	3,023 (31.7)	3,051 (32.0)	3,012 (31.5)	3,018 (31.6)	3,013 (31.6)	3,028 (31.7)	3,062 (32.1)
2009-2012	2,420 (25.3)	2,352 (24.7)	2,409 (25.2)	2,326 (24.4)	2,386 (25.0)	2,366 (24.8)	2,357 (24.7)	2,327 (24.4)	2,332 (24.4)
Diabetes duration (years)	9.2 ± 5.6	9.3 ± 5.5	9.3 ± 5.6	9.2 ± 5.5	9.3 ± 5.6	9.3 ± 5.6	9.3 ± 5.6	9.3 ± 5.5	9.3 ± 5.5
Prior duration of insulin use (years)	4.3 ± 4.6	4.3 ± 4.6	4.3 ± 4.6	4.3 ± 4.6	4.3 ± 4.7	4.3 ± 4.7	4.3 ± 4.6	4.2 ± 4.6	4.2 ± 4.6
Measured prior to cohort entry									
BMI (kg/m^2) *									
<25	1,491 (15.7)	1,504 (15.7)	1,482 (15.5)	1,513 (15.9)	1,501 (15.7)	1,495 (15.7)	1,498 (15.7)	1,491 (15.6)	1,501 (15.7)
25-30	2,445 (25.6)	2,424 (25.4)	2,414 (25.3)	2,417 (25.3)	2,435 (25.5)	2,422 (25.4)	2,421 (25.4)	2,418 (25.3)	2,401 (25.1)
≥ 30	4,578 (47.9)	4,569 (47.9)	4,612 (48.3)	4,587 (48.0)	4,585 (48.0)	4,571 (47.9)	4,582 (48.0)	4,589 (48.1)	4,605 (48.2)
Unknown	1,047 (10.9)	1,046 (11.0)	1,039 (10.9)	1,031 (10.8)	1,029 (10.8)	1,051 (11.0)	1,040 (10.9)	1,045 (11.0)	1,045 (10.9)
Haemoglobin A1c (%)†									
<6.5	720 (7.5)	706 (7.4)	709 (7.4)	687 (7.2)	730 (7.6)	717 (7.5)	714 (7.5)	705 (7.4)	712 (7.5)
6.5-8.0	2,849 (29.8)	2,822 (29.6)	2,854 (29.9)	2,791 (29.2)	2,761 (28.9)	2,778 (29.1)	2,801 (29.4)	2,828 (29.6)	2,785 (29.2)
>8	4,970 (52.0)	4,957 (51.9)	4,927 (51.6)	5,040 (52.8)	5,006 (52.4)	4,986 (52.3)	4,988 (52.3)	4,978 (52.2)	5,000 (52.4)
Unknown	1,028 (10.8)	1,058 (11.1)	1,057 (11.1)	1,030 (10.8)	1,053 (11.0)	1,058 (11.1)	1,038 (10.9)	1,032 (10.8)	1,055 (11.0)
Alcohol-related diseases‡	340 (3.6)	329 (3.5)	335 (3.5)	332 (3.5)	333 (3.5)	333 (3.5)	332 (3.5)	327 (3.4)	335 (3.5)
Smoking status‡									
Ever	6,152 (64.3)	6,106 (64.0)	6,118 (64.1)	6,122 (64.1)	6,141 (64.3)	6,112 (64.1)	6,128 (64.2)	6,103 (64.0)	6,121 (64.1)
Never	3,091 (32.3)	3,110 (32.6)	3,109 (32.6)	3,097 (32.4)	3,095 (32.4)	3,104 (32.5)	3,082 (32.3)	3,116 (32.7)	3,116 (32.6)
Unknown	324 (3.4)	327 (3.4)	320 (3.4)	329 (3.5)	314 (3.3)	323 (3.4)	331 (3.5)	324 (3.4)	315 (3.3)
Non-insulin diabetes medication use†‖									
Metformin	4,841 (50.6)	4,835 (50.7)	4,841 (50.7)	4,835 (50.6)	4,851 (50.8)	4,826 (50.6)	4,822 (50.5)	4,843 (50.8)	4,846 (50.7)
Sulfonylurea	2,917 (30.5)	2,903 (30.4)	2,904 (30.4)	2,930 (30.7)	2,908 (30.5)	2,922 (30.6)	2,907 (30.5)	2,910 (30.5)	2,910 (30.5)
Thiazolidinedione	977 (10.2)	966 (10.1)	977 (10.2)	984 (10.3)	984 (10.3)	1,001 (10.5)	984 (10.3)	990 (10.4)	988 (10.3)
Others¶	740 (7.7)	730 (7.7)	739 (7.7)	747 (7.8)	736 (7.7)	741 (7.8)	740 (7.8)	739 (7.7)	745 (7.8)

Charlson comorbidity score[§]									
≤ 1 comorbidities	5,000 (52.3)	4,966 (52.0)	4,996 (52.3)	5,002(52.4)	4,943 (51.8)	4,944 (51.8)	4,947 (51.9)	4,971 (52.1)	4,987 (52.2)
2-3	3,345 (35.0)	3,373 (35.4)	3,334 (34.9)	3,356 (35.2)	3,389 (35.5)	3,360 (35.2)	3,403 (35.7)	3,349 (35.1)	3,334 (34.9)
>3	1,222 (12.8)	1,204 (12.6)	1,217 (12.8)	1,190 (12.5)	1,218 (12.8)	1,235 (13.0)	1,191 (12.5)	1,223 (12.8)	1,231 (12.9)
Statin use[‡]	6,389 (66.8)	6,347 (66.5)	6,381 (66.8)	6,379 (66.8)	6,377 (66.8)	6,350 (66.6)	6,354 (66.6)	6,362 (66.7)	6,384 (66.8)
Aspirin use[‡]	5,224 (54.6)	5,239 (54.9)	5,228 (54.8)	5,230 (54.8)	5,235 (54.8)	5,241 (54.9)	5,228 (54.8)	5,224 (54.7)	5,222 (54.7)
NSAID use[‡]	3,504 (36.3)	3,502 (36.7)	3,501 (36.7)	3,525 (36.9)	3,491 (36.6)	3,504 (36.7)	3,501 (36.7)	3,512 (36.8)	3,479 (36.4)

Data are presented as mean (SD) or n (%) unless otherwise specified. [*] Measured five years prior. [†]Measured two years prior. [‡]Measured at any time prior to cohort entry. [§]Measured one year prior. [‖] Not mutually exclusive. [¶] Other non-insulin diabetes medications included meglitinides, DPP-4i, GLP-1, alpha-glucosidase inhibitors, guar gum, and SGLT2i. Abbreviations: BMI – body mass index; HbA1c – glycated haemoglobin; NSAID – nonsteroidal anti-inflammatory drugs

Appendix 4 Mean number of prescriptions and time between first NPH insulin prescription and study cohort entry (date of first insulin glargine or randomly selected NPH insulin prescription) in the pseudo-matching approach between 2002-2012 among prior insulin users

Glargine			NPH*		
N†	**Time**	**Number of prescriptions**	**N**	**Time**	**Number of prescriptions**
2,533	2.3 ± 2.2	15.3 ± 18.7	6,898	1.9 ± 2.2	16.6 ± 20.1

*Based on replicate 1.

†Restricted to women who switched from NPH insulin after 2002 to insulin glargine.

Chapter 7. Discussion

7.1. Summary of findings

The overall goal of my thesis was to determine the relationship between long-acting insulin analogues and the risk of cancer, specifically breast while addressing some of the methodological issues that have been identified in the literature. The first step was to systematically review the literature on long-acting insulin analogues and the risk of cancer. In addition, we developed a quality assessment criterion to evaluate the quality of these studies for pharmacoepidemiology biases. Once the literature was reviewed, and the research gaps were identified, we conducted a study to assess the relationship between long-acting insulin analogues on the risk of breast cancer in women with type 2 diabetes using insulin therapy. Finally, we shifted our focus to exploring alternative study designs and data analytical techniques to account for all types of patients – new and prior treatment users – in comparative effectiveness observational studies. In this last objective, we also discussed the strengths and limitations of the proposed alternative approaches.

In the first manuscript, entitled "The effect of long-acting insulin analogues on the risk of cancer: a systematic review of observational studies", we examined observational studies evaluating the association of long-acting insulin analogues on the risk of any and site-specific cancers, with a particular focus on methodological strengths and weaknesses of these studies. In this systematic review, we identified 16 cohort and 3 case-control studies on long-acting insulin analogues and any and common site-specific cancer risks including breast, colorectal and prostate. We did not find any observational studies on insulin detemir and site-specific cancer risk. The RRs reported in the existing literature suggested there was no increased risk for any cancers and colorectal or prostate cancers but four studies observed an increased risk with breast cancer when comparing insulin glargine with other insulins. In the quality assessment, we demonstrated that seven studies included prevalent users, 11 did not incorporate a lag period, six were subject to time-related biases (four of which had time-lag bias) and 16 had short duration of follow-up between insulin initiation and cancer incidence (< 5 years). Due to the methodological issues, the conclusions that can be drawn from observational studies on long-acting insulin analogues and cancer risk, particularly breast, were limited. Moreover, there were no studies that evaluated insulin detemir on the risk of breast cancer.

In the second manuscript, entitled “Long-term effect of long-acting insulin analogues on breast cancer incidence in women with type 2 diabetes”, we assessed the relationship between long-acting insulin analogues and breast cancer risk among women with type 2 diabetes using insulin therapy. In our cohort study, we observed the use of insulin glargine was associated with an increased risk of breast cancer (HR: 1.45, 95% CI: 1.12-1.87). In our duration-response analysis, we found the risk of breast cancer increased after five or more years of insulin glargine use (HR: 2.25, 95% CI: 1.33-3.80). We also demonstrated a dose-response relationship, where women using more than 30 insulin glargine prescriptions was associated with an increased risk of breast cancer (HR: 2.28, 95% CI: 1.26-4.15). A similar association between insulin glargine users and breast cancer incidence was observed among prior insulin users only. On the other hand, we did not observe an association between insulin detemir and breast cancer risk in neither the overall nor duration- or dose-response analysis. However, the relationship between insulin detemir and breast cancer remains uncertain due to its more recent introduction in the UK market yielding a smaller number of women and shorter duration of use.

In the third manuscript, entitled “Alternative methods to assess risk in a comparative effectiveness study: an illustration with insulin glargine and breast cancer”, we explored study designs and data analytical techniques to account for all patients – new and prior treatment users – in a comparative effectiveness observational study. More specifically, in this study, we explored two alternative approaches, time-dependent and pseudo-matching, to assess the safety profile of insulin glargine in all patients in the real world setting. The time-dependent approach used a study cohort of women 40 years or older using insulin glargine or NPH and was classified as unexposed until the first glargine prescription and exposed until the end of the follow-up. The pseudo-matching approach defined a study cohort from first insulin glargine prescription, and an NPH insulin prescription (comparator) was randomly selected to be similar to insulin glargine users in terms of the prior history of insulin treatment. The HR for breast cancer with insulin glargine use compared with NPH users was 1.46 (95% CI: 1.16, 1.84) and 1.44 (95% CI: 1.11, 1.85) using the time-dependent and pseudo-matching approaches, respectively. The time-dependent or pseudo-matching approach provides an alternative to assessing the safety profile of the entire patient population in the real world setting. However, there are complexities, such as constructing the cohort and increased computational time for the data analysis, to both approaches that need to be considered when choosing one approach over the other.

7.2. Limitations and research challenges

There were strengths and limitations in this thesis which were discussed in manuscripts 1-3. However, there were certain limitations and research challenges that we will elaborate further given their importance and implications for our study results and conclusions. We specifically focused on following limitations: exposure misclassification, outcome misclassification, left truncation in the CPRD, types of prior insulin users, detection bias, and the complex relationship between diabetes duration, insulin therapy and cancer.

Exposure misclassification

In manuscripts 2-3, we defined women as being exposed to long-acting insulin analogues or NPH insulin based on one prescription. We could have potentially misclassified exposure status if women switched or discontinued from, or did not adhere to the prescribed insulin prescription during the follow-up. Moreover, we used an approach which was analogous to an ITT analysis in RCTs, which does not account for this switching, discontinuation, and adherence, for our primary analysis.[214] Given the use an active comparator (NPH insulin), it was possible we could have overestimated the effect estimates if women switched, discontinued, or did not adhere to the prescribed basal insulin therapy. As we discussed in manuscript 2, the lack of adherence may not be an issue because women with advanced stages of type 2 diabetes depend on insulin therapy. Although, it is still possible we could have higher adherence with long-acting insulin analogue users compared with NPH because long-acting insulin analogues require less frequent injections given the longer duration of effect. Hence, the risk of breast cancer may appear to be greater among insulin glargine users as a result of the differential proportion of adherers. Despite this, since we are studying incident breast cancer as an outcome, it was more appropriate to use an ITT analysis due to the long latency of cancer. Furthermore, in manuscript 2, when we accumulated the number of insulin glargine or NPH prescriptions over the follow-up, we observed a dose-response relationship. This suggested the potential exposure misclassification and differential adherence alone cannot explain the risk of breast cancer associated with insulin glargine use in the primary analysis.

Outcome misclassification

One of the limitations of the CPRD is that there is no consensus of the types of medical codes that should be included in an outcome definition.[143] In manuscripts 2-3, we defined an incident breast cancer case using the first medical code indicating malignant neoplasm of the breast. A previous study compared using diagnostic codes versus an algorithm (combination of medical and product codes) to identify breast cancer cases and observed 95% of cases had a diagnostic code for cancer.[158] Furthermore, as discussed in chapter 3, cancer cases using medical codes identified in CPRD have been compared with the UK's cancer registry, NCDR, and it was observed that over 90% of the breast cancer cases identified in the CPRD were also confirmed in NCDR.[155] Importantly, even if there was outcome misclassification, we do not anticipate this to be differential with respect to our exposure definitions in the study in the primary analysis using the approach analogous to an ITT analysis.

Breast cancer is a heterogeneous disease with five major subtypes: luminal A, luminal B, human epidermal growth factor receptor 2 (HER2), basal, and unclassified.[215] Epidemiological studies have demonstrated different associations between risk factors and breast cancer subtypes and have suggested studying breast cancer subtypes separately.[216-218] Moreover, experimental studies have shown different expression and roles IGF receptors play in the prognosis of the different breast cancer subtypes.[219] Unfortunately, the information on hormone receptor status (which define luminal A and B breast cancers) and HER2 in the CPRD was incomplete. In addition, we could not link CPRD to NCDR due to the additional costs associated with the linkage. Although the NCDR would have provided additional information on stage and grade of breast cancer tumour, it has been reported stage and grade were missing in >50% and >15%, respectively, among the study population.[220] Moreover, information hormone receptor status and HER2 is also not available in the NCDR.[221] Furthermore, experimental and observational studies of long-acting insulin analogues and breast cancer risk have not suggested differential effects on breast cancer subtypes.

Left truncation in the CPRD

In the context of electronic medical records or health administrative databases, left truncation, a form of missing data, occurs when the patient's medical information prior to the or gaps in their membership of the databases are not observed.[213 222] In manuscripts 2-3, left

truncation could have occurred when women were not yet a member of the CPRD (i.e. seen in GP practices outside of the CPRD) or women could have been treated in secondary care settings such as an endocrinologist prior to or during CPRD membership. These various forms of left truncation can be problematic particularly when we were trying to define women with or without a prior history of insulin use (new and prior insulin users). It is possible that new insulin users may not truly be new insulin users if we have not captured all their prescriptions in the CPRD. Similarly, when we were defining prior duration of insulin use, the first-ever insulin prescription may not truly be the first-ever, and so we may be underestimating the prior duration of insulin use. However, if there were a large proportion of women with left truncation, we would have expected to see an association among new insulin users. In fact, our results among new insulin users conferred with studies using a new user cohort study design.[12 13 15 16 23 24] Moreover, as discussed in chapter 3, although women could have visited their endocrinologists, a consultation note is sent back to the GP. Importantly, GPs are primary gatekeepers to the health care system in the UK. Therefore, the potential for left truncation in the CPRD may not be too substantial.

Types of prior insulin users

The goal of comparative effectiveness research is to evaluate the benefits and risks of all types of patients in the real world setting. Although we have attempted to include both new and prior insulin users, we have included a few different types of prior insulin users in manuscripts 2-3 by design of the cohort study. The typical prior insulin users that may exist in the real world setting are the switchers. Switchers are essentially patients who started on NPH (or other types) insulin and switched to one of the long-acting insulin analogue glargine or detemir. It is also possible that long-acting insulin analogue users could switch to NPH insulin; however, given the benefits of long-acting insulin analogues the percentage of this type of switch was small (appendix 12a of manuscript 2). In our cohort study, we may have included women who are not switchers. In manuscripts 2-3, we created a cohort of women with type 2 diabetes using insulin therapy. Women had to be registered in a UTS practice within the CPRD for at least one year to qualify for study cohort entry. We determined the study cohort eligibility of a patient using all insulin prescriptions, and as a result, we may have artificially created various types of prior insulin users (e.g. prior insulin users could be using long-acting insulin analogues, NPH insulin,

and other insulins). However, we conducted a sensitivity analysis in manuscript 2 where we restricted the study population to switchers (i.e. women who switched from NPH or types of insulin to either long-acting insulin analogues) and found similar results except estimates were imprecise due to the reduced sample size.

Detection bias

It has been several years since the four cohort studies have been published suggesting insulin glargine increases the risk of cancer, particularly breast.[14 15 19 117] In the wake of these studies, as discussed in the background, U.S. FDA and Health Canada has requested for more epidemiological evidence but have not suggested a change in clinical practice.[131 132] In contrast, the EMA has concluded that there is no relationship between insulin glargine and any cancer incidence.[135] Given the response of the drug regulatory agencies, we cannot exclude the possibility that GPs may be more likely to send their women for a mammography screening after prescribing their women with a long-acting insulin analogue. Consequently, perhaps during this increased mammography screening process the GPs inadvertently detect more cancers among long-acting insulin analogue users compared with NPH. However, again in manuscript 2, we reported that the association between insulin glargine and breast cancer incidence remained consistent in the one to three years lagged analyses, women within the age range of mammography screening (between 50-70 years), and those with a prior mammography screening. Based on these sensitivity analyses, it suggested that risk of breast cancer associated with insulin glargine was not necessarily a by-product of detection bias.

The complexities of diabetes duration, insulin therapy, and cancer

In manuscript 2, we observed both a duration-response with insulin glargine use on breast cancer risk, particularly after five years of use. Although it is biologically plausible that long-term use of insulin glargine may increase the risk of breast cancer, we cannot discount the possibility that diabetes duration could still confound the relationship even though we adjusted for diabetes duration in all multivariable models. Aforementioned, we defined diabetes duration as the time between the first occurrence of type 2 diabetes diagnosis, a prescription for any antidiabetic medication (including insulin) and HbA1c value (≥ 6.5%) and study cohort entry. However, we could have underestimated diabetes duration based on this definition. For example,

if GPs sent women to the endocrinologists for further investigations, diagnoses made at the secondary care may not be documented in the CPRD even though GPs receive consultation notes from the endocrinologists. Consequently, for women without a diagnosis of type 2 diabetes, we may not truly know the onset of type 2 diabetes. Although it is also possible the high HbA1c values may lead the GP to refer the patient to an endocrinologist, which could be more indicative of the onset of type 2 diabetes. Moreover, even if a diagnostic code of type 2 diabetes may be recorded in the CPRD, the disease could have gone undiagnosed for months or years and so the true onset of type 2 diabetes may still not be captured in the CPRD. Despite this, a previous observational study has evaluated the relationship between diabetes duration, insulin therapy, and several site-specific cancer risks and found patients with diabetes using insulin therapies or not had a higher incidence of cancer compared with non-diabetics and the rate ratio remained constant over the diabetes duration.[223] This evidence suggested insulin therapies such as long-acting insulin analogues could still augment the risk of cancer independent of diabetes duration. Additional research is needed to further elucidate the relationship between diabetes duration, long-acting insulin analogues, and breast cancer specifically.

Despite the limitations we have highlighted, the results from these studies confer with the biological plausibility of long-acting insulin analogues on breast cancer cells.[7 55-61] This suggests the potential limitations and biases may have limited impact on conclusions of our studies. Therefore, the findings of this thesis may have important research, public health and clinical implications.

7.3. Implications for public health and future direction

Quality assessments of observational studies in pharmacoepidemiology

In the first manuscript, we identified the need for thorough quality assessment tool for pharmacoepidemiology biases in observational studies. More specifically, we observed that current quality assessment tools would not have captured some of the methodological limitations that were present in the long-acting insulin analogues and cancer incidence literature. Although pharmacoepidemiology biases are variations of selection bias, information bias and confounding issues as elaborated in the background chapter, the pharmacoepidemiology biases are framed

differently and may require a different set of guidelines to help users (i.e. assessor of study quality) thoroughly assess the quality of studies in pharmacoepidemiology. Aforementioned, previous systematic reviews have erroneously found moderate to high-quality studies, including those four highly criticised cohort studies on insulin glargine and cancer risk, using quality assessment tools such as NOS.[92 93 95] Currently, Cochrane has published a new quality assessment tool specifically for observational studies known as ROBINS-I and it is more comprehensive compared with NOS.[127] However, it still does not evaluate some of the methodological limitations we have included and identified in our systematic review such as time-related biases. For example, immortal time bias is considered as a selection bias in ROBINS-I, but it could be either selection or information bias depending on the study design.[128] Consequently, there is still a need for the research community to consider these subtle nuances in pharmacoepidemiology biases in a new or improved quality assessment tool such that observational studies in pharmacoepidemiology are adequately evaluated for quality. At the moment, our systematic review could serve as a guideline for future systematic reviews of observational studies in pharmacoepidemiology or template for the development of a quality assessment tool for pharmacoepidemiology studies.

Benefit risk assessment at the population level

The findings from our thesis suggested insulin glargine was associated with an increased risk of breast cancer. As we discussed previously, the current observational studies on insulin glargine and its effects on the risk of breast cancer have methodological limitations. Even though we have designed a study to minimise the biases that were present in the previous studies, our study alone does not warrant policymakers, health care professionals, and other relevant stakeholders to consider withdrawing or recommend not prescribing insulin glargine immediately. This is because the benefit of insulin glargine must also have to be considered. Previous meta-analyses of RCTs have reported that both long-acting insulin analogue glargine and detemir compared with NPH are at a reduced risk of nocturnal hypoglycaemia by at least 50%, preventing fatal outcomes such as death.[104 105] Moreover, due to the longer duration of effect of insulin glargine it reduces the frequency of injections, which can potentially improve adherence.[224] Presently, our findings suggest drug regulatory agencies should consider re-evaluating all the evidence on the effectiveness and safety of long-acting insulin analogues,

particularly insulin glargine. In addition, we recommend additional studies be conducted on the effects of long-acting insulin analogues on breast cancer risk with longer follow-up and to include all types of patients (i.e. include new and prior insulin users). Future studies should also consider using our proposed alternative study designs and data analytical techniques to better assess the safety profile of long-acting insulin analogues in all types of patients in the real world setting. Finally, there is also a need for continued surveillance on the relationship between long-acting insulin analogues and breast cancer incidence.

Post-approval surveillance of drug safety

While the results from the second manuscript are important and contribute to the understanding of long-acting insulin analogues and risk of breast cancer, we still need to survey the data continually. Post-approval surveillance of drug safety relies on various sources of information including spontaneous reports, meta-analyses of RCTs, and observational studies using health administrative databases.[138] Each of these sources of information used for identifying safety signals has advantages and disadvantages. More specifically, spontaneous reports and meta-analyses of RCTs have limitations, and it has been recommended that we move towards conducting observational studies using health administrative databases.[138] This is because spontaneous report systems can be biased due to confounding by indication, systematic over- or under-reporting, generalizability, duplication of reports, and attribution an adverse event to a single drug in the context of polypharmacy. Similarly, meta-analyses of RCTs are not designed for safety outcomes resulting in smaller sample size and short duration of follow-up.[138] [205] Currently, the FDA has started the Sentinel Initiative whereas Health Canada has established the Canadian Network for Observational Drug Effect Studies (CNODES) to address the knowledge gaps in regards to safety and effectiveness of drugs.[225] [226] The Sentinel Initiative and CNODES are a large network of health administrative databases designed to surveying adverse events associated with certain drugs. This safety signal identified in our observational study perhaps warrants the FDA and Health Canada to further investigate the relationship between long-acting insulin analogues and the risk of breast cancer in multiple health administrative databases.

7.4. Final conclusions

My thesis furthers our understanding of the relationship between long-acting insulin analogues and the risk of cancer, particularly breast. Importantly, the results from the first two manuscripts can be used to inform drug regulatory agencies of the safety profile of long-acting insulin analogues. More specifically, we have a better understanding of insulin glargine and its effects on the risk of breast cancer. On the contrary, we recommend future research should be conducted on insulin detemir given its more recent introduction in the market. Additionally, my thesis provides alternative study designs and data analytical approaches that could be used in comparative effectiveness observational studies to assess the safety in all patient populations in the real world setting. In summary, the results from this thesis warrant drug regulatory agencies to re-evaluate the totality of the evidence on long-acting insulin analogues with respect to effectiveness and safety. We urge for continued post-approval surveillance on long-acting insulin analogues. Further, additional research should be conducted to evaluate the safety of long-acting insulin analogues.

Reference List

1. World Health Organization (WHO). Global Report on Diabetes, 2016.

2. World Health Organization (WHO). Cancer [Internet]. 2015 [Available from: http://www.who.int/mediacentre/factsheets/fs297/en/ accessed October 13 2016].

3. Ferlay J., Soerjomataram I., Ervik M., et al. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. 2012 [Available from: http://globocan.iarc.fr/ accessed August 22 2014].

4. Giovannucci E, Harlan DM, Archer MC, et al. Diabetes and cancer: a consensus report. *Diabetes care* 2010;33(7):1674-85. doi: 10.2337/dc10-0666 [published Online First: 2010/07/01]

5. Onitilo AA, Engel JM, Glurich I, et al. Diabetes and cancer II: role of diabetes medications and influence of shared risk factors. *Cancer causes & control : CCC* 2012;23(7):991-1008. doi: 10.1007/s10552-012-9971-4 [published Online First: 2012/04/25]

6. National Institute for Health and Care Excellence (NICE). Type 2 diabetes in adults: management, 2016.

7. Bronsveld HK, ter Braak B, Karlstad O, et al. Treatment with insulin (analogues) and breast cancer risk in diabetics; a systematic review and meta-analysis of in vitro, animal and human evidence. *Breast Cancer Res* 2015;17(1):100. doi: 10.1186/s13058-015-0611-2

8. Dejgaard A, Lynggaard H, Råstam J, et al. No evidence of increased risk of malignancies in patients with diabetes treated with insulin detemir: a meta-analysis. *Diabetologia* 2009;52(12):2507-12.

9. Home PD, Lagarenne P. Combined randomised controlled trial experience of malignancies in studies using insulin glargine. *Diabetologia* 2009;52(12):2499-506. doi: 10.1007/s00125-009-1530-5 [published Online First: 2009/09/17]

10. Rosenstock J, Fonseca V, McGill JB, et al. Similar risk of malignancy with insulin glargine and neutral protamine Hagedorn (NPH) insulin in patients with type 2 diabetes: findings from a 5 year randomised, open-label study. *Diabetologia* 2009;52(9):1971-3. doi: 10.1007/s00125-009-1452-2 [published Online First: 2009/07/18]

11. Bordeleau L, Yakubovich N, Dagenais GR, et al. The association of basal insulin glargine and/or n-3 fatty acids with incident cancers in patients with dysglycemia. *Diabetes care* 2014;37(5):1360-6. doi: 10.2337/dc13-1468

12. Peeters PJ, Bazelier MT, Leufkens HG, et al. Insulin glargine use and breast cancer risk: Associations with cumulative exposure. *Acta oncologica* 2016;55(7):851-8. doi: 10.3109/0284186X.2016.1155736

13. Chang CH, Toh S, Lin JW, et al. Cancer risk associated with insulin glargine among adult type 2 diabetes patients--a nationwide cohort study. *PloS one* 2011;6(6):e21368. doi: 10.1371/journal.pone.0021368 [published Online First: 2011/07/09]

14. Colhoun HM, Group SE. Use of insulin glargine and cancer incidence in Scotland: a study from the Scottish Diabetes Research Network Epidemiology Group. *Diabetologia* 2009;52(9):1755-65. doi: 10.1007/s00125-009-1453-1 [published Online First: 2009/07/16]

15. Currie CJ, Poole CD, Gale EA. The influence of glucose-lowering therapies on cancer risk in type 2 diabetes. *Diabetologia* 2009;52(9):1766-77. doi: 10.1007/s00125-009-1440-6 [published Online First: 2009/07/03]

16. Fagot JP, Blotiere PO, Ricordeau P, et al. Does insulin glargine increase the risk of cancer compared with other basal insulins?: A French nationwide cohort study based on national administrative databases. *Diabetes care* 2013;36(2):294-301. doi: 10.2337/dc12-0506

17. Grimaldi-Bensouda L, Cameron D, Marty M, et al. Risk of breast cancer by individual insulin use: an international multicenter study. *Diabetes care* 2014;37(1):134-43.

18. Habel LA, Danforth KN, Quesenberry CP, et al. Cohort study of insulin glargine and risk of breast, prostate, and colorectal cancer among patients with diabetes. *Diabetes care* 2013;36(12):3953-60. doi: 10.2337/dc13-0140 [published Online First: 2013/10/31]

19. Jonasson JM, Ljung R, Talback M, et al. Insulin glargine use and short-term incidence of malignancies-a population-based follow-up study in Sweden. *Diabetologia* 2009;52(9):1745-54. doi: 10.1007/s00125-009-1444-2 [published Online First: 2009/07/10]

20. Lind M, Fahlen M, Eliasson B, et al. The relationship between the exposure time of insulin glargine and risk of breast and prostate cancer: an observational study of the time-dependent effects of antidiabetic treatments in patients with diabetes. *Primary care diabetes* 2012;6(1):53-9. doi: 10.1016/j.pcd.2011.10.004

21. Ljung R, Talback M, Haglund B, et al. Insulin glargine use and short-term incidence of malignancies - a three-year population-based observation. *Acta oncologica* 2011;50(5):685-93. doi: 10.3109/0284186X.2011.558913 [published Online First: 2011/04/22]

22. Morden NE, Liu SK, Smith J, et al. Further exploration of the relationship between insulin glargine and incident cancer: a retrospective cohort study of older Medicare patients. *Diabetes care* 2011;34(9):1965-71. doi: 10.2337/dc11-0699 [published Online First: 2011/07/22]

23. Ruiter R, Visser LE, van Herk-Sukel MP, et al. Risk of cancer in patients on insulin glargine and other insulin analogues in comparison with those on human insulin: results from a large population-based follow-up study. *Diabetologia* 2012;55(1):51-62. doi: 10.1007/s00125-011-2312-4 [published Online First: 2011/10/01]

24. Sturmer T, Marquis MA, Zhou H, et al. Cancer incidence among those initiating insulin therapy with glargine versus human NPH insulin. *Diabetes care* 2013;36(11):3517-25. doi: 10.2337/dc13-0263 [published Online First: 2013/07/24]

25. Suissa S, Azoulay L, Dell'Aniello S, et al. Long-term effects of insulin glargine on the risk of breast cancer. *Diabetologia* 2011;54(9):2254-62. doi: 10.1007/s00125-011-2190-9 [published Online First: 2011/05/27]

26. Gale EA. Insulin glargine and cancer: another side to the story? *Lancet* 2009;374(9689):521. doi: 10.1016/S0140-6736(09)61477-X

27. Garg SK, Hirsch IB, Skyler JS. Insulin glargine and cancer--an unsubstantiated allegation. *Diabetes technology & therapeutics* 2009;11(8):473-6. doi: 10.1089/dia.2009.1705

28. Hernandez-Diaz S, Adami HO. Diabetes therapy and cancer risk: causal effects and other plausible explanations. *Diabetologia* 2010;53(5):802-8. doi: 10.1007/s00125-010-1675-2 [published Online First: 2010/02/24]

29. Johnson JA, Carstensen B, Witte D, et al. Diabetes and cancer (1): evaluating the temporal relationship between type 2 diabetes and cancer incidence. *Diabetologia* 2012;55(6):1607-18. doi: 10.1007/s00125-012-2525-1

30. Pocock SJ, Smeeth L. Insulin glargine and malignancy: an unwarranted alarm. *Lancet* 2009;374(9689):511-3. doi: 10.1016/S0140-6736(09)61307-6

31. Wu JW, Filion KB, Azoulay L, et al. Effect of Long-Acting Insulin Analogs on the Risk of Cancer: A Systematic Review of Observational Studies. *Diabetes care* 2016;39(3):486-94. doi: 10.2337/dc15-1816

32. Guariguata L, Whiting DR, Hambleton I, et al. Global estimates of diabetes prevalence for 2013 and projections for 2035. *Diabetes research and clinical practice* 2014;103(2):137-49. doi: 10.1016/j.diabres.2013.11.002

33. Vigneri P, Frasca F, Sciacca L, et al. Diabetes and cancer. *Endocrine-related cancer* 2009;16(4):1103-23. doi: 10.1677/ERC-09-0087

34. Alberti KGMM, Zimmet Pf. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation. *Diabetic medicine* 1998;15(7):539-53.

35. Batabyal P, Vander Hoorn S, Christophi C, et al. Association of diabetes mellitus and pancreatic adenocarcinoma: a meta-analysis of 88 studies. *Ann Surg Oncol* 2014;21(7):2453-62. doi: 10.1245/s10434-014-3625-6

36. De Bruijn K, Arends L, Hansen B, et al. Systematic review and meta‐analysis of the association between diabetes mellitus and incidence and mortality in breast and colorectal cancer. *British Journal of Surgery* 2013;100(11):1421-29.

37. Fang H, Yao B, Yan Y, et al. Diabetes mellitus increases the risk of bladder cancer: an updated meta-analysis of observational studies. *Diabetes technology & therapeutics* 2013;15(11):914-22. doi: 10.1089/dia.2013.0131

38. Gong Y, Wei B, Yu L, et al. Type 2 diabetes mellitus and risk of oral cancer and precancerous lesions: a meta-analysis of observational studies. *Oral Oncol* 2015;51(4):332-40. doi: 10.1016/j.oraloncology.2015.01.003

39. Gu J, Yan S, Wang B, et al. Type 2 diabetes mellitus and risk of gallbladder cancer: a systematic review and meta‐analysis of observational studies. *Diabetes/metabolism research and reviews* 2016;32(1):63-72.

40. Jian Gang P, Mo L, Lu Y, et al. Diabetes mellitus and the risk of prostate cancer: an update and cumulative meta-analysis. *Endocrine research* 2015;40(1):54-61.

41. Tsilidis KK, Kasimis JC, Lopez DS, et al. Type 2 diabetes and cancer: umbrella review of meta-analyses of observational studies. *Bmj* 2015;350:g7607. doi: 10.1136/bmj.g7607

42. Xu X, Wu J, Mao Y, et al. Diabetes mellitus and risk of bladder cancer: a meta-analysis of cohort studies. *PloS one* 2013;8(3):e58079. doi: 10.1371/journal.pone.0058079

43. Yeo Y, Ma SH, Hwang Y, et al. Diabetes mellitus and risk of thyroid cancer: a meta-analysis. *PloS one* 2014;9(6):e98135. doi: 10.1371/journal.pone.0098135

44. Zhang ZH, Su PY, Hao JH, et al. The role of preexisting diabetes mellitus on incidence and mortality of endometrial cancer: a meta-analysis of prospective cohort studies. *Int J Gynecol Cancer* 2013;23(2):294-303. doi: 10.1097/IGC.0b013e31827b8430

45. Gordon-Dseagu VL, Shelton N, Mindell JS. Epidemiological evidence of a relationship between type-1 diabetes mellitus and cancer: a review of the existing literature. *International journal of cancer Journal international du cancer* 2013;132(3):501-8. doi: 10.1002/ijc.27703

46. Cohen DH, LeRoith D. Obesity, type 2 diabetes, and cancer: the insulin and IGF connection. *Endocrine-related cancer* 2012;19(5):F27-45. doi: 10.1530/ERC-11-0374

47. Klil-Drori AJ, Azoulay L, Pollak MN. Cancer, obesity, diabetes, and antidiabetic drugs: is the fog clearing? *Nat Rev Clin Oncol* 2016 doi: 10.1038/nrclinonc.2016.120

48. Garg SK, Maurer H, Reed K, et al. Diabetes and cancer: two diseases with obesity as a common risk factor. *Diabetes, obesity & metabolism* 2014;16(2):97-110. doi: 10.1111/dom.12124

49. Xue F, Michels KB. Diabetes, metabolic syndrome, and breast cancer: a review of the current evidence. *The American journal of clinical nutrition* 2007;86(3):s823-35.

50. Calle EE, Kaaks R. Overweight, obesity and cancer: epidemiological evidence and proposed mechanisms. *Nature reviews Cancer* 2004;4(8):579-91. doi: 10.1038/nrc1408

51. Vigneri R, Goldfine ID, Frittitta L. Insulin, insulin receptors, and cancer. *Journal of endocrinological investigation* 2016 doi: 10.1007/s40618-016-0508-7

52. Pollak MN, Schernhammer ES, Hankinson SE. Insulin-like growth factors and neoplasia. *Nature reviews Cancer* 2004;4(7):505-18. doi: 10.1038/nrc1387

53. Gallagher EJ, LeRoith D. Minireview: IGF, Insulin, and Cancer. *Endocrinology* 2011;152(7):2546-51. doi: 10.1210/en.2011-0231

54. Clayton PE, Banerjee I, Murray PG, et al. Growth hormone, the insulin-like growth factor axis, insulin and cancer risk. *Nat Rev Endocrinol* 2011;7(1):11-24. doi: 10.1038/nrendo.2010.171

55. Jones RA, Campbell CI, Wood GA, et al. Reversibility and recurrence of IGF-IR-induced mammary tumors. *Oncogene* 2009;28(21):2152-62. doi: 10.1038/onc.2009.79

56. Surmacz E. Function of the IGF-I receptor in breast cancer. *J Mammary Gland Biol Neoplasia* 2000;5(1):95-105.

57. Papa V, Pezzino V, Costantino A, et al. Elevated insulin receptor content in human breast cancer. *J Clin Invest* 1990;86(5):1503-10. doi: 10.1172/JCI114868

58. Osborne CK, Monaco ME, Lippman ME, et al. Correlation among insulin binding, degradation, and biological activity in human breast cancer cells in long-term tissue culture. *Cancer research* 1978;38(1):94-102.

59. Frittitta L, Vigneri R, Papa V, et al. Structural and functional studies of insulin receptors in human breast cancer. *Breast cancer research and treatment* 1993;25(1):73-82.

60. Chappell J, Leitner JW, Solomon S, et al. Effect of insulin on cell cycle progression in MCF-7 breast cancer cells. Direct and potentiating influence. *The Journal of biological chemistry* 2001;276(41):38023-8. doi: 10.1074/jbc.M104416200

61. Pollak MN, Polychronakos C, Yousefi S, et al. Characterization of insulin-like growth factor I (IGF-I) receptors of human breast cancer cells. *Biochem Biophys Res Commun* 1988;154(1):326-31.

62. Pisani P. Hyper-insulinaemia and cancer, meta-analyses of epidemiological studies. *Archives of physiology and biochemistry* 2008;114(1):63-70. doi: 10.1080/13813450801954451

63. Xu J, Ye Y, Wu H, et al. Association between markers of glucose metabolism and risk of colorectal cancer. *BMJ Open* 2016;6(6):e011430. doi: 10.1136/bmjopen-2016-011430

64. Chen L, Li L, Wang Y, et al. Circulating C-peptide level is a predictive factor for colorectal neoplasia: evidence from the meta-analysis of prospective studies. *Cancer causes & control : CCC* 2013;24(10):1837-47. doi: 10.1007/s10552-013-0261-6

65. Hernandez AV, Guarnizo M, Miranda Y, et al. Association between insulin resistance and breast carcinoma: a systematic review and meta-analysis. *PloS one* 2014;9(6):e99317. doi: 10.1371/journal.pone.0099317

66. Autier P, Koechlin A, Boniol M, et al. Serum insulin and C-peptide concentration and breast cancer: a meta-analysis. *Cancer causes & control : CCC* 2013;24(5):873-83. doi: 10.1007/s10552-013-0164-6

67. Lai GY, Helzlsouer KJ, Clipp SL, et al. Association between C-peptide concentration and prostate cancer incidence in the CLUE II cohort study. *Cancer Prev Res (Phila)* 2010;3(10):1334-41. doi: 10.1158/1940-6207.CAPR-10-0053

68. Lai GY, Giovannucci EL, Pollak MN, et al. Association of C-peptide and leptin with prostate cancer incidence in the Health Professionals Follow-up Study. *Cancer causes & control : CCC* 2014;25(5):625-32. doi: 10.1007/s10552-014-0369-3

69. Chi F, Wu R, Zeng YC, et al. Circulation insulin-like growth factor peptides and colorectal cancer risk: an updated systematic review and meta-analysis. *Mol Biol Rep* 2013;40(5):3583-90. doi: 10.1007/s11033-012-2432-z

70. Rowlands MA, Gunnell D, Harris R, et al. Circulating insulin-like growth factor peptides and prostate cancer risk: a systematic review and meta-analysis. *International journal of cancer Journal international du cancer* 2009;124(10):2416-29. doi: 10.1002/ijc.24202

71. Roddam AW, Allen NE, Appleby P, et al. Insulin-like growth factors, their binding proteins, and prostate cancer risk: analysis of individual patient data from 12 prospective studies. *Annals of internal medicine* 2008;149(7):461-71, W83-8.

72. Shi R, Berkel HJ, Yu H. Insulin-like growth factor-I and prostate cancer: a meta-analysis. *British journal of cancer* 2001;85(7):991-6. doi: 10.1038/sj.bjc.6691961

73. Renehan AG, Zwahlen M, Minder C, et al. Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis. *Lancet* 2004;363(9418):1346-53. doi: 10.1016/S0140-6736(04)16044-3

74. Rinaldi S, Rohrmann S, Jenab M, et al. Glycosylated hemoglobin and risk of colorectal cancer in men and women, the European prospective investigation into cancer and nutrition. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2008;17(11):3108-15. doi: 10.1158/1055-9965.EPI-08-0495

75. Hormones TE, Group BCC. Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. *The lancet oncology* 2010;11(6):530-42.

76. Chen B, Liu S, Xu W, et al. IGF-I and IGFBP-3 and the risk of lung cancer: a meta-analysis based on nested case-control studies. *J Exp Clin Cancer Res* 2009;28:89. doi: 10.1186/1756-9966-28-89

77. Li Y, Li Y, Zhang J, et al. Circulating Insulin-Like Growth Factor-1 Level and Ovarian Cancer Risk. *Cell Physiol Biochem* 2016;38(2):589-97. doi: 10.1159/000438652

78. Wang Q, Bian CE, Peng H, et al. Association of circulating insulin-like growth factor 1 and insulin-like growth factor binding protein 3 with the risk of ovarian cancer: A systematic review and meta-analysis. *Mol Clin Oncol* 2015;3(3):623-28. doi: 10.3892/mco.2015.516

79. Cao H, Wang G, Meng L, et al. Association between circulating levels of IGF-1 and IGFBP-3 and lung cancer risk: a meta-analysis. *PloS one* 2012;7(11):e49884. doi: 10.1371/journal.pone.0049884

80. Baglietto L, English DR, Hopper JL, et al. Circulating insulin-like growth factor-I and binding protein-3 and the risk of breast cancer. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2007;16(4):763-8. doi: 10.1158/1055-9965.EPI-06-0960

81. Sugumar A, Liu YC, Xia Q, et al. Insulin-like growth factor (IGF)-I and IGF-binding protein 3 and the risk of premenopausal breast cancer: a meta-analysis of literature. *International journal of cancer Journal international du cancer* 2004;111(2):293-7. doi: 10.1002/ijc.20253

82. Canadian Diabetes Association (CDA). Canadian Diabetes Association 2013 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. *Canadian Journal of Diabetes* 2013;37(1):1-227.

83. Inzucchi SE, Bergenstal RM, Buse JB, et al. Management of hyperglycemia in type 2 diabetes, 2015: a patient-centered approach: update to a position statement of the American Diabetes Association and the European Association for the Study of Diabetes. *Diabetes care* 2015;38(1):140-9. doi: 10.2337/dc14-2441

84. International Diabetes Federation (IDF). Global Guideline for Type 2 Diabetes [Internet]. 2012 [Available from: www.idf.org accessed November 28 2016].

85. University of California San Francisco (UCSF). Types of Insulin [Internet]. 2016 [Available from: https://dtc.ucsf.edu/types-of-diabetes/type2/treatment-of-type-2-diabetes/medications-and-therapies/type-2-insulin-rx/types-of-insulin/ accessed October 31 2016].

86. University of California San Francisco (UCSF). Pre-Mixed Insulin [Internet]. 2016 [Available from: https://dtc.ucsf.edu/types-of-diabetes/type2/treatment-of-type-2-diabetes/medications-and-therapies/type-2-insulin-rx/types-of-insulin/pre-mixed-insulin/ accessed October 31 2016].

87. Canadian Diabates Association (CDA). Canadian Diabetes Association 2008 Clinical Practice Guidelines for the Prevention and Management of Diabetes in Canada. *Canadian Journal of Diabetes* 2008;32(1):1-215.

88. Sciacca L, Le Moli R, Vigneri R. Insulin analogs and cancer. *Front Endocrinol (Lausanne)* 2012;3:21. doi: 10.3389/fendo.2012.00021

89. National Institute for Health and Care Excellence (NICE). Type 2 diabetes: insulin degludec [Internet]. 2013 [Available from: https://www.nice.org.uk/advice/esnm25/chapter/key-points-from-the-evidence accessed October 19 2016].

90. Novo Nordisk. Novo Nordisk Launches Tresiba® (insulin degludec injection 200 Units/mL) in the United States [Internet]. 2016 [Available from: http://press.novonordisk-us.com/2016-01-

26-Novo-Nordisk-Launches-Tresiba-insulin-degludec-injection-200-Units-mL-in-the-United-States accessed October 19 2016].

91. Berard L, MacNeill G. Insulin degludec, a long-acting once-daily basal analogue for type 1 and type 2 diabetes mellitus. *Can J Diabetes* 2015;39(1):4-9. doi: 10.1016/j.jcjd.2014.02.019

92. Chen YB, Chen Q, Wang Z, et al. Insulin therapy and risk of prostate cancer: a systematic review and meta-analysis of observational studies. *PloS one* 2013;8(11):e81594. doi: 10.1371/journal.pone.0081594

93. Colmers IN, Bowker SL, Tjosvold LA, et al. Insulin use and cancer risk in patients with type 2 diabetes: a systematic review and meta-analysis of observational studies. *Diabetes & metabolism* 2012;38(6):485-506. doi: 10.1016/j.diabet.2012.08.011 [published Online First: 2012/11/20]

94. Janghorbani M, Dehghani M, Salehi-Marzijarani M. Systematic review and meta-analysis of insulin therapy and risk of cancer. *Hormones & Cancer* 2012;3(4):137-46.

95. Karlstad O, Starup-Linde J, Vestergaard P, et al. Use of insulin and insulin analogs and risk of cancer - systematic review and meta-analysis of observational studies. *Curr Drug Saf* 2013;8(5):333-48.

96. Singh S, Singh H, Singh PP, et al. Anti-diabetic medications and colorectal cancer risk: A systematic review and meta-analysis. *Gastroenterology* 2013;1):S18.

97. Singh S, Singh PP, Singh AG, et al. Anti-diabetic medications and risk of pancreatic cancer in patients with diabetes mellitus: a systematic review and meta-analysis. *American Journal of Gastroenterology* 2013;108(4):510-9; quiz 20.

98. Singh S, Singh PP, Singh AG, et al. Anti-diabetic medications and the risk of hepatocellular cancer: a systematic review and meta-analysis. *American Journal of Gastroenterology* 2013;108(6):881-91; quiz 92.

99. Sun A, Liu R, Sun G. Insulin therapy and risk of colorectal cancer: an updated meta-analysis of epidemiological studies. *Current medical research and opinion* 2014;30(3):423-30. doi: 10.1185/03007995.2013.858622

100. Wang L, Cai S, Teng Z, et al. Insulin therapy contributes to the increased risk of colorectal cancer in diabetes patients: a meta-analysis. *Diagn Pathol* 2013;8:180. doi: 10.1186/1746-1596-8-180

101. Wang Z, Bao C, Su C, et al. Association between diabetes or antidiabetic therapy and lung cancer: A meta‐analysis. *Journal of diabetes investigation* 2013;4(6):659-66.

102. Yang YX, Hennessy S, Lewis JD. Insulin therapy and colorectal cancer risk among type 2 diabetes mellitus patients. *Gastroenterology* 2004;127(4):1044-50.

103. Canadian Diabetes Association (CDA). Table 1. Types of insulin [Internet]. 2015 [Available from: http://guidelines.diabetes.ca/cdacpg_resources/Ch12_Table1_Types_of_Insulin_updated_Aug_5.pdf accessed November 22 2016].

104. Rosenstock J, Dailey G, Massi-Benedetti M, et al. Reduced hypoglycemia risk with insulin glargine: a meta-analysis comparing insulin glargine with human NPH insulin in type 2 diabetes. *Diabetes care* 2005;28(4):950-5.

105. Brunton SA. Nocturnal hypoglycemia: answering the challenge with long-acting insulin analogs. *MedGenMed* 2007;9(2):38.

106. Freeman JS. Insulin analog therapy: improving the match with physiologic insulin secretion. *J Am Osteopath Assoc* 2009;109(1):26-36.

107. Sciacca L, Cassarino MF, Genua M, et al. Biological effects of insulin and its analogs on cancer cells with different insulin family receptor expression. *J Cell Physiol* 2014;229(11):1817-21. doi: 10.1002/jcp.24635

108. Pierre-Eugene C, Pagesy P, Nguyen TT, et al. Effect of insulin analogues on insulin/IGF1 hybrid receptors: increased activation by glargine but not by its metabolites M1 and M2. *PloS one* 2012;7(7):e41992. doi: 10.1371/journal.pone.0041992 [published Online First: 2012/08/01]

109. Yehezkel E, Weinstein D, Simon M, et al. Long-acting insulin analogues elicit atypical signalling events mediated by the insulin receptor and insulin-like growth factor-I receptor. *Diabetologia* 2010;53(12):2667-75. doi: 10.1007/s00125-010-1899-1

110. Sommerfeld MR, Muller G, Tschank G, et al. In vitro metabolic and mitogenic signaling of insulin glargine and its metabolites. *PloS one* 2010;5(3):e9540. doi: 10.1371/journal.pone.0009540 [published Online First: 2010/03/09]

111. Mayer D, Chantelau E. Treatment with insulin glargine (Lantus) increases the proliferative potency of the serum of patients with type-1 diabetes: a pilot study on MCF-7 breast cancer cells. *Archives of physiology and biochemistry* 2010;116(2):73-8. doi: 10.3109/13813451003631439

112. Mayer D, Shukla A, Enzmann H. Proliferative effects of insulin analogues on mammary epithelial cells. *Archives of physiology and biochemistry* 2008;114(1):38-44. doi: 10.1080/13813450801900645

113. Kurtzhals P, Schaffer L, Sorensen A, et al. Correlations of receptor binding and metabolic and mitogenic potencies of insulin analogs designed for clinical use. *Diabetes* 2000;49(6):999-1005. [published Online First: 2000/06/24]

114. Agin A, Jeandidier N, Gasser F, et al. Glargine blood biotransformation: in vitro appraisal with human insulin immunoassay. *Diabetes & metabolism* 2007;33(3):205-12. doi: 10.1016/j.diabet.2006.12.002 [published Online First: 2007/03/16]

115. Kuerzel GU, Shukla U, Scholtz HE, et al. Biotransformation of insulin glargine after subcutaneous injection in healthy subjects. *Current medical research and opinion* 2003;19(1):34-40. [published Online First: 2003/03/29]

116. De Meyts P, Christoffersen CT, Ursø B, et al. Role of the time factor in signaling specificity: application to mitogenic and metabolic signaling by the insulin and insulin-like growth factor-I receptor tyrosine kinases. *Metabolism: clinical and experimental* 1995;44:2-11.

117. Hemkens LG, Grouven U, Bender R, et al. Risk of malignancies in patients with diabetes treated with human insulin or insulin analogues: a cohort study. *Diabetologia* 2009;52(9):1732-44. doi: 10.1007/s00125-009-1418-4 [published Online First: 2009/07/01]

118. van Staa TP, Patel D, Gallagher AM, et al. Glucose-lowering agents and the patterns of risk for cancer: a study with the General Practice Research Database and secondary care data. *Diabetologia* 2012;55(3):654-65. doi: 10.1007/s00125-011-2390-3 [published Online First: 2011/12/01]

119. Blin P, Lassalle R, Dureau-Pournin C, et al. Insulin glargine and risk of cancer: a cohort study in the French National Healthcare Insurance Database. *Diabetologia* 2012;55(3):644-53. doi: 10.1007/s00125-011-2429-5 [published Online First: 2012/01/10]

120. Mannucci E, Monami M, Balzi D, et al. Doses of insulin and its analogues and cancer occurrence in insulin-treated type 2 diabetic patients. *Diabetes care* 2010;33(9):1997-2003. doi: 10.2337/dc10-0476 [published Online First: 2010/06/17]

121. Simo R, Plana-Ripoll O, Puente D, et al. Impact of glucose-lowering agents on the risk of cancer in type 2 diabetic patients. The barcelona case-control study. *PloS one* 2013;8(11):e79968. doi: 10.1371/journal.pone.0079968

122. Buchs AE, Silverman BG. Incidence of malignancies in patients with diabetes mellitus and correlation with treatment modalities in a large Israeli health maintenance organization: a

historical cohort study. *Metabolism: clinical and experimental* 2011;60(10):1379-85. doi: 10.1016/j.metabol.2011.05.002

123. Lim S, Stember KG, He W, et al. Electronic medical record cancer incidence over six years comparing new users of glargine with new users of NPH insulin. *PloS one* 2014;9(10):e109433. doi: 10.1371/journal.pone.0109433

124. Du X, Zhang R, Xue Y, et al. Insulin glargine and risk of cancer: a meta-analysis. *The International journal of biological markers* 2012;27(3):e241-6. doi: 10.5301/JBM.2012.9349 [published Online First: 2012/08/07]

125. Tang X, Yang L, He Z, et al. Insulin glargine and cancer risk in patients with diabetes: a meta-analysis. *PloS one* 2012;7(12):e51814. doi: 10.1371/journal.pone.0051814 [published Online First: 2013/01/04]

126. Suissa S, Azoulay L. Metformin and the risk of cancer: time-related biases in observational studies. *Diabetes care* 2012;35(12):2665-73. doi: 10.2337/dc12-0788 [published Online First: 2012/11/23]

127. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *Bmj* 2016;355:i4919. doi: 10.1136/bmj.i4919

128. Levesque LE, Hanley JA, Kezouh A, et al. Problem of immortal time bias in cohort studies: example using statins for preventing progression of diabetes. *Bmj* 2010;340:b5087. doi: 10.1136/bmj.b5087 [published Online First: 2010/03/17]

129. Gerstein HC, Bosch J, Dagenais GR, et al. Basal insulin and cardiovascular and other outcomes in dysglycemia. *The New England journal of medicine* 2012;367(4):319-28. doi: 10.1056/NEJMoa1203858 [published Online First: 2012/06/13]

130. Dejgaard A, Lynggaard H, Rastam J, et al. No evidence of increased risk of malignancies in patients with diabetes treated with insulin detemir: a meta-analysis. *Diabetologia* 2009;52(12):2507-12. doi: 10.1007/s00125-009-1568-4

131. Health Canada. Health Canada Update About Safety of Lantus (Insulin Glargine) [Internet]. 2009 [Available from: http://www.healthycanadians.gc.ca/recall-alert-rappel-avis/hc-sc/2009/13315a-eng.php accessed July 18 2016].

132. Food and Drug Administration (FDA). FDA Drug Safety Podcast for Healthcare Professionals: Update to ongoing safety review of Lantus (insulin glargine) and possible risk of cancer [Internet]. 2013 [Available from: http://www.fda.gov/Drugs/DrugSafety/DrugSafetyPodcasts/ucm240508.htm accessed December 27 2016].

133. Food and Drug Administration (FDA). Early communication about safety of Lantus (insulin glarigne) [Internet]. 2009 [Available from: http://www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHeathcareProfessionals/ucm169722.htm accessed December 27 2016].

134. Food and Drug Administration (FDA). FDA Drug Safety Communication: Update to ongoing safety review of Lantus (insulin glargine) and possible risk of cancer [Internet]. 2011 [Available from: http://www.fda.gov/Drugs/DrugSafety/ucm239376.htm accessed December 27 2016].

135. European Medicines Agency (EMA). Outcome of review of new safety data on insulin glargine [Internet]. 2013 [Available from: http://www.ema.europa.eu/docs/en_GB/document_library/Medicine_QA/2013/05/WC500143823.pdf accessed December 13, 2013].

136. European Medicines Agency (EMA). Lantus: Procedural steps taken and scientific information after the authorisation, 2015.

137. Strom BL. Pharmacoepidemiology: Wiley. com 2012.

138. Gibbons RD, Amatya AK, Brown CH, et al. Post-approval drug safety surveillance. *Annu Rev Public Health* 2010;31:419-37. doi: 10.1146/annurev.publhealth.012809.103649

139. Johnson ES, Bartman BA, Briesacher BA, et al. The incident user design in comparative effectiveness research. *Pharmacoepidemiology and drug safety* 2013;22(1):1-6. doi: 10.1002/pds.3334

140. Dreyer NA, Schneeweiss S, McNeil BJ, et al. GRACE principles: recognizing high-quality observational studies of comparative effectiveness. *The American journal of managed care* 2010;16(6):467-71.

141. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *American journal of epidemiology* 2003;158(9):915-20.

142. Suissa S, Moodie E, Dell'Aniello S. Prevalent new-user cohort designs for comparative drug effect studies by time-conditional propensity scores. *Pharmacoepidemiology and drug safety* 2016 doi: 10.1002/pds.4107

143. Herrett E, Gallagher AM, Bhaskaran K, et al. Data Resource Profile: Clinical Practice Research Datalink (CPRD). *International journal of epidemiology* 2015;44(3):827-36. doi: 10.1093/ije/dyv098

144. Walley T, Mantgani A. The UK General Practice Research Database. *Lancet* 1997;350(9084):1097-9. doi: 10.1016/S0140-6736(97)04248-7 [published Online First: 1999/04/23]

145. Garcia Rodriguez LA, Perez Gutthann S. Use of the UK General Practice Research Database for pharmacoepidemiology. *British journal of clinical pharmacology* 1998;45(5):419-25. [published Online First: 1998/06/27]

146. Mathur R, Bhaskaran K, Chaturvedi N, et al. Completeness and usability of ethnicity data in UK-based primary care and hospital databases. *Journal of public health* 2013 doi: 10.1093/pubmed/fdt116

147. Medicines and Healthcare Products Regulatory Agency. General Practice Research Database (GPRD) [Internet]. [Available from: http://www.lshtm.ac.uk/study/applicants/shortcourses/practical_pharmacoepidemiology__gprd_background.pdf accessed October 25 2016].

148. Chisholm J. The Read clinical classification. *Bmj* 1990;300(6732):1092.

149. Scottish Clinical Information Management in Practice. Guide to Read Codes [Internet]. 2014 [Available from: http://www.scimp.scot.nhs.uk/better-information/clinical-coding/scimp-guide-to-read-codes/ accessed October 26 2016].

150. National Health Service (NHS). Quality and Outcomes Framework [Internet]. [Available from: http://content.digital.nhs.uk/qof accessed October 27 2016].

151. Bhaskaran K, Forbes HJ, Douglas I, et al. Representativeness and optimal use of body mass index (BMI) in the UK Clinical Practice Research Datalink (CPRD). *BMJ Open* 2013;3(9):e003389. doi: 10.1136/bmjopen-2013-003389

152. Khan NF, Harrison SE, Rose PW. Validity of diagnostic coding within the General Practice Research Database: a systematic review. *Br J Gen Pract* 2010;60(572):e128-36. doi: 10.3399/bjgp10X483562

153. Herrett E, Thomas SL, Schoonen WM, et al. Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *British journal of clinical pharmacology* 2010;69(1):4-14. doi: 10.1111/j.1365-2125.2009.03537.x [published Online First: 2010/01/19]

154. Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *Bmj* 1991;302(6779):766-8.

155. Boggon R, van Staa TP, Chapman M, et al. Cancer recording and mortality in the General Practice Research Database and linked cancer registries. *Pharmacoepidemiology and drug safety* 2013;22(2):168-75. doi: 10.1002/pds.3374

156. electronic Medicines Compendium (emc). ABASAGLAR 100 units/mL solution for injection in cartridge & pre-filled pen [Internet]. 2015 [Available from: https://www.medicines.org.uk/emc/medicine/30494 accessed November 24 2016].

157. electronic Medicines Compendium (emc). Toujeo 300 units/ml solution for injection in a pre-filled pen [Internet]. 2016 [Available from: https://www.medicines.org.uk/emc/medicine/30586 accessed November 24 2016].

158. Opatrny L, Dell'Aniello S, Assouline S, et al. Hormone replacement therapy use and variations in the risk of breast cancer. *BJOG* 2008;115(2):169-75; discussion 75. doi: 10.1111/j.1471-0528.2007.01520.x

159. Savitz DA, Baron AE. Estimating and correcting for confounder misclassification. *American journal of epidemiology* 1989;129(5):1062-71.

160. Greenland S. The effect of misclassification in the presence of covariates. *American journal of epidemiology* 1980;112(4):564-9.

161. Brunelli SM, Gagne JJ, Huybrechts KF, et al. Estimation using all available covariate information versus a fixed look-back window for dichotomous covariates. *Pharmacoepidemiology and drug safety* 2013;22(5):542-50. doi: 10.1002/pds.3434

162. Tagalakis V, Kahn SR. Determining the test characteristics of claims-based diagnostic codes for the diagnosis of venous thromboembolism in a medical service claims database. *Pharmacoepidemiology and drug safety* 2011;20(3):304-7. doi: 10.1002/pds.2061

163. National Health Service (NHS). HbA1c Standardisation for Laboratory Professionals [Internet]. [Available from: https://www.diabetes.org.uk/upload/Professionals/Key%20leaflets/53130HbA1cLableaflet.pdf accessed November 25 2016].

164. InDependent Diabetes Trust (IDDT). New HbA1c units [Internet]. 2016 [Available from: http://www.iddt.org/news/new-hba1c-units accessed November 25 2016].

165. Deyo RA, Cherkin DC, Ciol MA. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of clinical epidemiology* 1992;45(6):613-9.

166. National Health Service (NHS). Height/weight chart [Internet]. 2015 [Available from: http://www.nhs.uk/livewell/loseweight/pages/height-weight-chart.aspx accessed October 28 2016].

167. Smith U, Gale EA. Does diabetes therapy influence the risk of cancer? *Diabetologia* 2009;52(9):1699-708. doi: 10.1007/s00125-009-1441-5

168. Liberati A, Altman D, Tetzlaff J, et al. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS medicine* 2009;6(7):e1000100.

169. Scottish Intercollegiate Guidelines Network (SIGN). Search Filters [Internet]. 2014 [Available from: http://www.sign.ac.uk/methodology/filters.html accessed January 16 2015].

170. Suissa S. Immortal time bias in pharmaco-epidemiology. *American journal of epidemiology* 2008;167(4):492-9. doi: 10.1093/aje/kwm324 [published Online First: 2007/12/07]

171. Suissa S, Dell'aniello S, Vahey S, et al. Time-window bias in case-control studies: statins and lung cancer. *Epidemiology* 2011;22(2):228-31. doi: 10.1097/EDE.0b013e3182093a0f

172. Tamim H, Monfared AA, LeLorier J. Application of lag-time into exposure definitions to control for protopathic bias. *Pharmacoepidemiology and drug safety* 2007;16(3):250-8. doi: 10.1002/pds.1360

173. Johnson JA, Bowker SL, Richardson K, et al. Time-varying incidence of cancer after the onset of type 2 diabetes: evidence of potential detection bias. *Diabetologia* 2011;54(9):2263-71. doi: 10.1007/s00125-011-2242-1

174. Stroup DF, Berlin JA, Morton SC, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis Of Observational Studies in Epidemiology (MOOSE) group. *JAMA : the journal of the American Medical Association* 2000;283(15):2008-12.

175. Berry SD, Ngo L, Samelson EJ, et al. Competing risk of death: an important consideration in studies of older adults. *J Am Geriatr Soc* 2010;58(4):783-7. doi: 10.1111/j.1532-5415.2010.02767.x

176. Satagopan JM, Ben-Porat L, Berwick M, et al. A note on competing risks in survival data analysis. *British journal of cancer* 2004;91(7):1229-35. doi: 10.1038/sj.bjc.6602102

177. Haffner SM, Lehto S, Rönnemaa T, et al. Mortality from coronary heart disease in subjects with type 2 diabetes and in nondiabetic subjects with and without prior myocardial infarction. *New England journal of medicine* 1998;339(4):229-34.

178. Centers for Disease Control and Prevention (CDC). National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2014. In: Services USDoHaH, ed. Atlanta, GA, 2014.

179. Schneeweiss S. Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and drug safety* 2006;15(5):291-303. doi: 10.1002/pds.1200

180. Vandenbroucke J, Pearce N. Point: incident exposures, prevalent exposures, and causal inference: does limiting studies to persons who are followed from first exposure onward damage epidemiology? *American journal of epidemiology* 2015;182(10):826-33. doi: 10.1093/aje/kwv225

181. Schneeweiss S, Patrick AR, Sturmer T, et al. Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Medical care* 2007;45(10 Supl 2):S131-42. doi: 10.1097/MLR.0b013e318070c08e [published Online First: 2007/10/25]

182. Pollak M. The insulin and insulin-like growth factor receptor family in neoplasia: an update. *Nature reviews Cancer* 2012;12(3):159-69. doi: 10.1038/nrc3215

183. National Health Service (NHS). Type 2 Diabetes - Causes [Internet]. 2016 [Available from: http://www.nhs.uk/Conditions/Diabetes-type2/Pages/Causes.aspx accessed July 26 2016].

184. Kleinbaum DG, Klein M. Survival analysis. 3rd Edition ed: Springer 2012.

185. Heinzl H, Kaider A. Gaining more flexibility in Cox proportional hazards regression models with cubic spline functions. *Comput Methods Programs Biomed* 1997;54(3):201-8.

186. Fine JP, Gray RJ. A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association* 1999;94(446):496-509.

187. Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *American journal of epidemiology* 2009;170(2):244-56. doi: 10.1093/aje/kwp107

188. Weinstein D, Simon M, Yehezkel E, et al. Insulin analogues display IGF-I-like mitogenic and anti-apoptotic activities in cultured cancer cells. *Diabetes/metabolism research and reviews* 2009;25(1):41-9. doi: 10.1002/dmrr.912 [published Online First: 2009/01/16]

189. Sciacca L, Cassarino MF, Genua M, et al. Insulin analogues differently activate insulin receptor isoforms and post-receptor signalling. *Diabetologia* 2010;53(8):1743-53. doi: 10.1007/s00125-010-1760-6

190. Gallagher EJ, Zelenko Z, Tobin-Hess A, et al. Non-metabolisable insulin glargine does not promote breast cancer growth in a mouse model of type 2 diabetes. *Diabetologia* 2016:1-8.

191. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. *Statistics in medicine* 2010;29(9):1037-57. doi: 10.1002/sim.3841

192. Greenland S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. *Epidemiology* 1995;6(4):356-65.

193. Harrell F. Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis: Springer 2015.

194. Rothman KJ. Induction and latent periods. *American journal of epidemiology* 1981;114(2):253-9.

195. Cancer Research UK. Breast cancer screening in the UK [Internet]. 2014 [Available from: http://www.cancerresearchuk.org/about-cancer/type/breast-cancer/about/screening/who-is-screened-for-breast-cancer accessed July 27 2016].

196. Joffe MM, Byrne C, Colditz GA. Postmenopausal hormone use, screening, and breast cancer: characterization and control of a bias. *Epidemiology* 2001;12(4):429-38.

197. Joffe MM, Colditz GA. Restriction as a method for reducing bias in the estimation of direct effects. *Statistics in medicine* 1998;17(19):2233-49.

198. Vach W, Blettner M. Biased estimation of the odds ratio in case-control studies due to the use of ad hoc methods of correcting for missing values for confounding variables. *American journal of epidemiology* 1991;134(8):895-907. [published Online First: 1991/10/15]

199. Azur MJ, Stuart EA, Frangakis C, et al. Multiple imputation by chained equations: what is it and how does it work? *International journal of methods in psychiatric research* 2011;20(1):40-9. doi: 10.1002/mpr.329

200. Lee KJ, Carlin JB. Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. *American journal of epidemiology* 2010;171(5):624-32. doi: 10.1093/aje/kwp425

201. University of California Los Angeles (UCLA). Missing Data in SAS Part 1 [Internet]. 2016 [Available from: http://www.ats.ucla.edu/stat/sas/seminars/missing_data/mi_new_1.htm accessed July 8 2016].

202. Graham JW, Olchowski AE, Gilreath TD. How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prev Sci* 2007;8(3):206-13. doi: 10.1007/s11121-007-0070-9

203. White IR, Royston P. Imputing missing covariate values for the Cox model. *Statistics in medicine* 2009;28(15):1982-98. doi: 10.1002/sim.3618

204. SAS and R. Example 7.39: Nelson-Aalen estimate of cumulative hazard [Internet]. 2010 [Available from: http://sas-and-r.blogspot.ca/2010/05/example-739-nelson-aalen-estimate-of.html accessed July 8 2016].

205. Curtin F, Schulz P. Assessing the benefit: risk ratio of a drug--randomized and naturalistic evidence. *Dialogues Clin Neurosci* 2011;13(2):183-90.

206. Concato J, Lawler EV, Lew RA, et al. Observational methods in comparative effectiveness research. *Am J Med* 2010;123(12 Suppl 1):e16-23. doi: 10.1016/j.amjmed.2010.10.004

207. Rassen JA, Shelat AA, Myers J, et al. One-to-many propensity score matching in cohort studies. *Pharmacoepidemiology and drug safety* 2012;21 Suppl 2:69-80. doi: 10.1002/pds.3263

208. Hernan MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15(5):615-25.

209. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* 2000;11(5):550-60.

210. Suissa S. Inhaled steroids and mortality in COPD: bias from unaccounted immortal time. *The European respiratory journal* 2004;23(3):391-5.

211. Libby G, Donnelly LA, Donnan PT, et al. New users of metformin are at low risk of incident cancer: a cohort study among people with type 2 diabetes. *Diabetes care* 2009;32(9):1620-5. doi: 10.2337/dc08-2175

212. Hernan MA. Counterpoint: epidemiology to guide decision-making: moving away from practice-free research. *American journal of epidemiology* 2015;182(10):834-9. doi: 10.1093/aje/kwv215

213. . In: Gliklich RE, Dreyer NA, Leavy MB, eds. Registries for Evaluating Patient Outcomes: A User's Guide. 3rd ed. Rockville (MD)2014.

214. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012;9(1):48-55. doi: 10.1177/1740774511420743

215. Breast Cancer Fund. Breast cancer subtypes [Internet]. [Available from: http://www.breastcancerfund.org/clear-science/biology-of-breast-cancer/breast-cancer-subtypes/?referrer=https://www.google.ca/ accessed November 6 2016].

216. Crispo A, Augustin L, Grimaldi M, et al. Risk Differences Between Prediabetes And Diabetes According To Breast Cancer Molecular Subtypes. *Journal of Cellular Physiology* 2016

217. Tamimi RM, Colditz GA, Hazra A, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. *Breast cancer research and treatment* 2012;131(1):159-67. doi: 10.1007/s10549-011-1702-0

218. Yang XR, Sherman ME, Rimm DL, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* 2007;16(3):439-43. doi: 10.1158/1055-9965.EPI-06-0806

219. Farabaugh SM, Boone DN, Lee AV. Role of IGF1R in Breast Cancer Subtypes, Stemness, and Lineage Differentiation. *Front Endocrinol (Lausanne)* 2015;6:59. doi: 10.3389/fendo.2015.00059

220. Hicks BM, Murray LJ, Hughes C, et al. Clopidogrel use and cancer-specific mortality: a population-based cohort study of colorectal, breast and prostate cancer patients. *Pharmacoepidemiology and drug safety* 2015;24(8):830-40. doi: 10.1002/pds.3807

221. Murray LJ, Cooper JA, Hughes CM, et al. Post-diagnostic prescriptions for low-dose aspirin and breast cancer-specific survival: a nested case-control study in a breast cancer cohort from the UK Clinical Practice Research Datalink. *Breast Cancer Res* 2014;16(2):R34. doi: 10.1186/bcr3638

222. Klein JP, Moeschberger ML. Survival analysis: techniques for censored and truncated data: Springer Science & Business Media 2005.

223. Carstensen B, Witte DR, Friis S. Cancer occurrence in Danish diabetic patients: duration and insulin effects. *Diabetologia* 2012;55(4):948-58. doi: 10.1007/s00125-011-2381-4

224. Davies MJ, Gagliardino JJ, Gray LJ, et al. Real-world factors affecting adherence to insulin therapy in patients with Type 1 or Type 2 diabetes mellitus: a systematic review. *Diabetic medicine : a journal of the British Diabetic Association* 2013;30(5):512-24. doi: 10.1111/dme.12128 [published Online First: 2013/01/18]

225. Food and Drug Administration (FDA). Senitinel Initiative [Internet]. 2016 [Available from: http://www.fda.gov/Safety/FDAsSentinelInitiative/default.htm accessed November 9 2016].

226. Canadian Network for Observational Drug Effect Studies (CNODES). About CNODES [Internet]. 2016 [Available from: https://www.cnodes.ca/about-cnodes/ accessed November 11 2016].

Certificates of Ethical Approval

Manuscript 1: The effect of long-acting insulin analogues on the risk of cancer: a systematic review of observational studies

- McGill University Institutional Review Board (Study No. A11-M114-14B)
- Jewish General Hospital Research Ethics Board approved

Manuscript 2: Long-term effect of long-acting insulin analogues on breast cancer incidence in women with type 2 diabetes

- McGill University Institutional Review Board (Study No. A11-M114-14B)
- Jewish General Hospital Research Ethics Board approved
- CPRD International Scientific Advisory Committee (Protocol No. 15_005R)

Manuscript 3: Alternative methods to account for patients with a prior history of treatment: an illustration with insulin glargine and breast cancer

- McGill University Institutional Review Board (Study No. A11-M114-14B)
- Jewish General Hospital Research Ethics Board approved
- CPRD International Scientific Advisory Committee (Protocol No. 16_213R2)