ASSOCIATION BETWEEN OSTEOARTHRITIS AND INCREASED RISK OF CARDIOVASCULAR DISEASES: INVESTIGATION OF THE ROLE OF NSAIDS AS AN UNDERLYING MECHANISM

by

Mohammad Atiquzzaman

M. Pharm, University of Dhaka, 2001
B. Pharm, University of Dhaka, 2000

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(PHARMACEUTICAL SCIENCES)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

December 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Association between osteoarthritis and increased risk of cardiovascular diseases; investigation of the role of NSAIDs as an underlying mechanism

submitted by Mohammad Atiquzzaman in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Pharmaceutical Sciences

Examining Committee:

Professor Dr. Aslam Anis
Supervisor

Dr. Jacek Kopec
Supervisory Committee Member

Dr. Mohammad Ehsanul Karim
Supervisory Committee Member

Dr. Juan Antonio Avina Zubieta
University Examiner

Dr. Mark Harrison
University Examiner

Additional Supervisory Committee Members:

Dr. Hubert Wong
Supervisory Committee Member

Dr. Mary De Vera
Supervisory Committee Member

Abstract

Osteoarthritis (OA) has been reported as an independent risk factor for cardiovascular diseases (CVD). There is no cure for OA and non-steroidal anti-inflammatory drugs (NSAIDs) are the mainstay of OA treatment. NSAIDs are known to be associated with various cardiovascular adverse effects, but their direct impact on CVD risk among OA patients is not well studied. There is a need to understand better the underlying mechanism of the increased risk of CVD among OA patients and to what extent NSAIDs play a role. This thesis conducted three studies using health administrative data (HAD) from British Columbia (BC), Canada and the Canadian Community Health Survey (CCHS) focusing on the role of NSAIDs in the OA-CVD association. The objective of study 1 was to quantify the role of NSAIDs in the increased risk of CVD among OA patients. This longitudinal study performed a mediation analysis using a marginal structural model and showed that a substantial proportion of total CVD risk among OA patients was attributable to NSAID use. The objective of study 2 was to evaluate the overall cardiovascular safety of various NSAIDs that are used in treating OA patients. This retrospective cohort study used time-dependent Cox regression analysis to estimate CVD risk associated with NSAID use overall and four unique groups of NSAIDs, i.e., coxibs, naproxen, ibuprofen and other conventional NSAIDs. This study showed that exposure to NSAIDs substantially increased CVD risk compared to unexposed person-time. It also showed that, relative to unexposed person-time coxibs and naproxen may increase CVD risk more than conventional NSAIDs including ibuprofen. The objective of study 3 was to identify a valid approach in imputing body mass index (BMI), an important confounding variable in the OA-CVD relationship for which information is usually not available in HAD. Multiple imputation was compared with proportion-based imputation (PBI) approach using plasmode simulated dataset created from CCHS data.

This study showed that multiple imputation was superior to PBI in imputing BMI category using information from an external dataset. As a collective work, this thesis provides a better understanding of OA-CVD association that hopefully will improve clinical management of OA.

Lay Summary

The most common form of arthritis is OA (short for ‘osteoarthritis’). This disease is a major cause of pain and disability and there is no known cure. Research also suggests that patients with OA have a higher risk of cardiovascular diseases (CVD), although the reason for this is not known. Non-steroidal anti-inflammatory drugs (NSAIDs) are often used to control pain and inflammation among OA patients. However, NSAIDs are known to be associated with cardiovascular adverse effects. This PhD thesis looked at the role of NSAIDs in the increased risk of CVD among OA patients and evaluated the cardiovascular safety of various NSAIDs used to treat OA. This thesis created new knowledge to understand the link between OA and CVD better. Hopefully this will raise awareness among health care providers including doctors, researchers and OA patients and improve the safe treatment of this disease.

Preface

Sections of this thesis were written as multi-authored manuscripts for publication in peer-reviewed journals. All inferences, opinions, and conclusions drawn in this dissertation are those of the authors, and do not reflect the opinions or policies of the Data Steward(s). Chapter 2 has been published in Arthritis & Rheumatology (1). Two other chapters are submitted for publication (Chapters 3 and 4). Mohammad Atiquzzaman (MA) played the lead role in each of these chapters. The Behavioural Research Ethics Board at the University of British Columbia, Canada reviewed and approved the secondary data analysis in Chapters 2 and 3 (H15-00203). Ethics approval of Chapter 4 using publicly available CCHS data was covered by item 7.10.3 in University of British Columbia's Policy #89: Research and Other Studies Involving Human Subjects (2) and Article 2.2 in of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2) (3).

Chapter 1

MA conducted the literature review and developed the first draft of the chapter. Dr. Anis, Dr. Karim, Dr. Kopec, Dr. De Vera and Dr. Wong edited the manuscript and suggested revisions.

Chapter 2

This is the accepted version of the following article: Atiquzzaman M, Karim ME, Kopec J, Wong H, Anis AH. Role of Nonsteroidal Antiinflammatory Drugs in the Association Between Osteoarthritis and Cardiovascular Diseases: A Longitudinal Study. Arthritis Rheumatol. 2019;71(11):1835–43, available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/art.41027 (1). MA was responsible for the study concept and analysis plan under the guidance of Dr. Anis,

Dr. Karim and Dr. Wong and conducted the analyses. MA wrote the first draft of the manuscript. Dr. Karim, Dr. Kopec, Dr. Wong and Dr. Anis substantially contributed to interpreting the results and suggested revisions to the manuscript.

Chapter 3

A version of Chapter 3 is submitted for publication: Atiquzzaman M, Karim ME, Kopec J, Wong H, De Vera MA, Anis A. The cardiovascular safety of non-steroidal anti-inflammatory drugs (NSAIDs) among individuals with osteoarthritis: Findings from real-world claims data. MA developed the study concept under the guidance of Dr. Anis, Dr. Kopec and Dr. Karim. Dr. Wong and Dr. Karim guided MA in the statistical analyses. MA wrote the first draft of the manuscript. Dr. Karim, Dr. Wong, Dr. Anis, Dr. Kopec and Dr. De Vera guided in the interpretation of results, reviewed the manuscript and suggested revisions.

Chapter 4

A version of Chapter 4 is submitted for publication: Atiquzzaman M, Kopec J, Karim ME, De Vera MA, Wong H, Anis A. Using external data to estimate unmeasured confounders: a plasmode simulation study comparing alternative approaches to impute body mass index in a study of the relationship between osteoarthritis and cardiovascular disease. MA developed the hypothesis and analysis plan under the guidance of Dr. Kopec and Dr. Anis and conducted the data analysis under the guidance of Dr. Karim and Dr. Wong. MA wrote the first draft of the manuscript. Dr. Karim, Dr. Wong, Dr. Kopec, Dr. De Vera and Dr. Anis participated in the interpretation of results and reviewed the manuscript.

Chapter 5

MA wrote the first draft of this section and all subsequent revisions. Dr. Anis, Dr. Karim, Dr. Kopec, Dr. De Vera, and Dr. Wong reviewed the chapter and suggested revisions.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AIC: Akaike Information Criterion

AMI: acute myocardial infarction

BC: British Columbia

BMI: body mass index

CCHS: Canadian Community Health Survey

CI: confidence interval

CHF: Congestive heart failure

con-NSAIDs: conventional non-steroidal anti-inflammatory drugs

COPD: chronic obstructive pulmonary disease

CVD: cardiovascular diseases

Cox: Cyclooxygenase

HAD: health administrative data

HR: hazard ratio

ICD-9: International Classification of Diseases 9th revision

ICD-10: International Statistical Classification of Diseases and related health problems 10th revision

IHD: ischemic heart disease

ITT: intent-to-treat

MSP: medical services plan

NDE: natural direct effect

NIE: natural indirect effect

NSAIDs: non-steroidal anti-inflammatory drugs

ns-NSAIDs: non-selective NSAIDs

OA: osteoarthritis

OR: odds ratio

OTC: over-the-counter

PBI: proportion-based imputation

PPV: positive predictive value

PUD: peptic ulcer disease

RA: rheumatoid arthritis

RCTs: randomized controlled trials

RR: Rate Ratio

SES: Socio-economic status

TE: total effect

Acknowledgements

I must acknowledge a lot of contributions to this thesis. I express my heartfelt gratitude to Dr. Aslam Anis. This thesis would have not been possible without his supervision and support. My sincerest gratitude to the members of my supervisory committee; Dr. Hubert Wong, Dr. Jacek Kopec, Dr. Mary De Vera and Dr. Mohammad Ehsanul Karim. Thanks to Dr. Ujendra Kumar for chairing the supervisory committee. I have learnt a lot from your expertise over the years of my PhD training.

My special thanks go to Dr. Huiying Sun for her relentless help in developing syntaxes of my statistical analysis. Thank you Huiying! I also extend my special thanks to Dr. Kathy H. Li, Dr. Stephanie Harvard, Dr. Natalie McCormick, Dr. Eric Sayre and Mr. Timothy Schmidt for their selfless support over the years.

I am thankful to the Centre for Health Evaluation and Outcome Sciences (CHEOS) and all the members in this organization for the enormous support over the years. I also thank Arthritis Research Canada (ARC) and Collaboration for Outcomes Research and Evaluation (CORE) for their support. My gratitude to The BC Ministry of Health and the BC Vital Statistics Agency for approving access to and use of the administrative data. Thanks to Population Data BC for facilitating the data access.

I am grateful to the Andrew Nord Fellowship in Rheumatology at the University of British Columbia (UBC) for awarding the funding support during my training. I also acknowledge for

the financial support I received from CHEOS, the faculty of Pharmaceutical Sciences at UBC and the Arthritis Research Canada/Lilly Early Rheumatology Researcher Scholarships program.

Finally, I deeply remember my late parents who raised me with great values to be the person who I am today. May their souls rest in eternal peace! I cannot thank enough my wife, Shanjida Akter, for her love and support. I am also grateful to my family and friends for their encouragements.

Dedication

To my family members.

Chapter 1: Introduction

1.1 Thesis overview

1.1.1 Research statement

In this thesis the goal was to understand the role of non-steroidal anti-inflammatory drugs (NSAIDs) in the association between osteoarthritis (OA) and increased risk of cardiovascular disease (CVD) and to investigate whether CVD risk associated with different NSAIDs used in treating OA vary among OA patients. Previous research has found that people with OA experience higher risk of CVD compared to the general population, although the reason for this is not yet known. Several causal pathways have been hypothesized, including the possibility that people with OA tend to use more NSAIDs to control the symptoms of pain and inflammation than people without OA. This increased use of NSAIDs may contribute to the greater CVD risk among OA patients. The scientific knowledge about adverse cardiovascular effects associated with NSAIDs makes this hypothesis persuasive. However, no study has directly investigated the role of NSAIDs in the OA-CVD relationship. As there is no cure for OA and many different NSAIDs are frequently used in OA treatment, it is important to understand the cardiovascular safety of different NSAIDs used to treat OA patients.

1.1.2 Overview of research questions and thesis chapters

This thesis addresses two clinical research questions on NSAID use in OA. The first clinical question is “Do NSAIDs contribute to the increased risk of CVD among people with OA?”. The second clinical question is “Does cardiovascular risk among OA patients vary as a function of the type of NSAID used to treat them?".

This thesis also explores a methodological question that is relevant to answering the clinical research questions. In this context, being overweight or obese is a potentially important confounding variable in studies that aim to explore the association between OA and CVD. Information on body mass index (BMI), a variable that describes body weight in relation to height as a measure of potential overweight or obesity, is not recorded in the dataset that was used to investigate the research questions in this thesis. In recent years a proportion-based imputation method has been used in the literature to impute unavailable BMI data using information from external data sources (4,5). This approach involves assigning a BMI category to a study participant according to the proportion observed in survey data that are grouped based on a set of pre-defined variables (4). However, the extent of potential bias or confounding effect of using a BMI variable derived by proportion-based imputation has not been investigated. It has been hypothesized that imputing BMI at the individual level using the multiple imputation technique would introduce less bias than proportion-based imputation of BMI (4,5). To help successfully answer the major research questions in this thesis, the following methodological question was also explored: "Compared to proportion-based imputation does multiple imputation produce less biased estimates of a given relationship when an important study variable is imputed for everyone using data from external sources?".

This thesis contains two novel epidemiologic studies based on health administrative data from British Columbia, Canada and a plasmode simulation study based on Canadian Community Health Survey (CCHS) data.

The introductory chapter is a literature review focusing on the relationship between OA and CVD. First, this chapter provides general background on OA and the recent research on the association of OA with CVD. It then discusses the prevailing hypothesis regarding the causal pathway of the OA-CVD association and describes the relevance of the research conducted in the subsequent chapters. Finally, the specific research aims are presented for each chapter of the thesis. As all three studies presented in chapters 2 to 4 are based on secondary data analysis of administrative data, a detailed description of the relevant databases is included.

Chapter 2 is a population-based longitudinal study evaluating the mediating role of NSAIDs in the increased risk of CVD among people with OA compared to people without. This mediation analysis addresses the first research question and provides information on the proportion of the total increased risk of CVD associated with OA that was mediated through NSAID use.

Chapter 3 is a separate population-based retrospective cohort study assessing the risk of CVD associated with different NSAIDs used in the real-world treatment of OA. This study addresses the second research question.

Chapter 4 is a simulation-based study taking the OA-CVD relationship as an example in which BMI is an important confounding variable. This study addresses the methodological research question by comparing multiple imputation with proportion-based imputation in imputing unavailable BMI using data from external source.

Chapter 5 is the closing chapter that puts together the research findings from individual thesis chapters, discusses the implications of the research findings, and describes the strengths and limitations of the overall thesis.

1.2 Osteoarthritis

This thesis primarily focuses on osteoarthritis (OA), the most common form of arthritis. Worldwide, approximately 9.6% of men and 18% of women aged 60 years or more have some form of OA (6). It is a degenerative disease of the joint in which cartilage, the rubbery material that protects bone from friction, breaks down, resulting in pain and inflammation at the joints (7). This pain reduces mobility among OA patients. An estimated 80% of individuals with OA experience difficulty in movement and in performing daily life activities (6). OA is a non-curable chronic disorder and NSAIDs are the most commonly used medications to control pain and inflammation (8). In Canada more than a million NSAID prescriptions are written to OA patients each year (8).

1.2.1 Cardiovascular disease morbidity and mortality in OA

In recent years much research has been conducted investigating the risk of CVD associated with OA. Findings from epidemiologic studies indicate that people with OA experience a higher risk of CVD compared to the general population. The following sections summarize the findings on the OA-CVD relationship according to the hierarchy of evidence.

1.2.1.1 Findings from meta-analysis

Wang et al. conducted a systematic review and meta-analysis of observational studies on the OA-CVD relationship (9). The risk of CVD among OA patients was 24% higher compared to the general population (rate ratio 1.24 [95% CI 1.12 to 1.37]) (9).

1.2.1.2 Findings from cohort studies

Recently, a large prospective longitudinal study using population-based health administrative data from British Columbia, Canada suggested that OA is associated with an increased risk of CVD (4). The adjusted rate ratios (95% CI) were 1.13 (1.07–1.18), 1.30 (1.19, 1.42) and 1.15 (1.04, 1.28) for overall CVD, ischemic heart disease (IHD) and congestive heart failure (CHF) respectively (4). In a separate prospective cohort study, Veronese et al. reported a 22% increased risk of CVD among OA patients compared to those without OA (10). The adjusted hazard ratio (HR) and 95% CI was 1.22 (1.02, 1.49) (10). Haugen et al. found that the risk of developing incident coronary heart disease such as myocardial infarction (MI) was substantially higher among people with hand OA compared to people without OA (11). The adjusted HR and 95% CI was 2.26 (1.22, 4.18) (11).

Several studies have reported higher risk of CVD-specific mortality among OA patients. For example, in a prospective cohort study Tsuboi et al. found that, compared to a non-OA control population, people with knee OA experienced substantially higher risk of mortality from cerebrovascular or cardiovascular diseases (12). The adjusted OR and 95% CI was 2.77 (1.28, 5.99) (12). In a population-based cohort study Nuesch et al. reported significantly higher cardiovascular disease-specific mortality among people with knee or hip OA compared to

general population; the adjusted HR (95% CI) was 1.72 (1.22, 2.41) (13). In a prospective cohort study involving a representative sample of Finish population, Haara et al. reported a 42% greater risk of cardiovascular death among men with OA in any finger joint (14). In a separate prospective cohort study Kluzek et al. concluded that the risk of all-cause and CVD-specific mortality was significantly higher among women who experienced knee pain with or without radiographic OA compared to women who did not have knee pain or OA (15).

1.2.1.3 Findings from case-control and cross-sectional studies

In a cross-sectional study using CCHS data, prevalent OA was found to be associated with self-reported heart disease, particularly angina, and CHF (16). The overall adjusted OR (95% CI) was 1.45 (1.36, 1.54), 1.83 (1.62, 2.06) and 1.72 (1.46, 2.01) for heart diseases, angina and CHF, respectively (16). In a case-control study Kadam et al. found that the odds of having IHD was 1.73 times higher (OR 1.73 (1.13, 2.66)) and the odds of having heart failure (HF) was 1.28 times higher (OR 1.28 (1.07, 1.52)) among people with OA compared to non-OA controls (17). Jonsson et al. reported that hand OA was significantly associated with atherosclerosis. The odds of having coronary calcifications was 1.42 times higher (OR 1.42 (1.14, 1.76)) among women with hand OA compared to women without OA (18). Also, the OR (95% CI) of having moderate or severe carotid plaques was 1.25 (1.04, 1.49) (18).

1.3 Potential causes for the increased CVD risk in OA

Potential causes for the increased risk of CVD among OA individuals include the cardiovascular adverse effects of NSAIDs, decreased mobility and chronic inflammation (4,19–21). NSAIDs are frequently used in the treatment of OA (8,22,23). NSAIDs are known to be associated with

cardiovascular adverse effects (24–26). The use of NSAIDs in OA treatment is hypothesized as a contributing factor in the development of CVD (4,19,20). Another hypothesis is that the pain associated with OA may lead to reduced physical activity that may contribute to developing CVD (4,21). Sellam et al. reported that synovitis, inflammation of the synovial membrane, occurs in the early stage of OA and continues throughout the different phases of OA (27). This chronic inflammation is also hypothesised as a cause for increased CVD risk (4).

1.4 Nonsteroidal anti-inflammatory drugs (NSAIDs)

The NSAIDs are a group of medications used as painkillers. At present, more than 20 NSAIDs are available in Canada and worldwide (28). The history of treating pain and inflammation dates back to as early as 400 B.C. when Hippocrates, a Greek physician and the father of modern medicine, used an extract from the bark and leaves of the Willow tree in treating headaches, other forms of pain and fevers (29,30). In 1828, Dr. Johann Buchner, a German pharmacologist, first isolated salicin from willow bark (29). This is the compound that relieved pain. In 1838, Raffaele Piria first extracted salicylic acid in its pure form (29). Salicylic acid was associated with serious adverse effects in the gastrointestinal tract (29).

Although Charles Frederic Gerhardt, a French chemist, first discovered the buffering of salicylic acid to sodium salicylate and acetylsalicylic acid in 1853, he abandoned the discovery and did not market the compounds (29). It was Felix Hoffmann, a chemist at Bayer Research Laboratories, who rediscovered Gerhardt's formula and synthesised aspirin (acetylsalicylic acid) in 1899 (29). The new era of the development of NSAIDs started in 1971 when Sir John Robert Vane, an English pharmacologist, discovered the mechanism of action of aspirin-like drugs (31).

John Vane concluded that these drugs decrease prostaglandin (an inflammatory mediator) production by inhibiting cyclooxygenase (COX) enzyme (31). This discovery resulted in the development of several NSAIDs over two decades. Another breakthrough in NSAID development occurred in 1991 when Dr. Simmons discovered the COX-2 enzyme (30). The first COX-2 inhibitor, celecoxib, was introduced in 1999 followed by rofecoxib (30).

Table 1-1 List of NSAIDs available in Canada (32)

Name of NSAID	Schedule
Acetic Acid derivative	
Diclofenac	Rx*
Etodolac	Rx
Indomethacin	Rx
Ketorolac	Rx
Sulindac	Rx
Tolmetin	Rx (Discontinued)
Coxibs	
Celecoxib	Rx
Lumiracoxib	Rx (Discontinued)
Rofecoxib	Rx (Discontinued)
Valdecoxib	Rx (Discontinued)
Saicylates derivatives	
Aspirin (ASA)	OTC**
Diflunisal	Rx

Name of NSAID	Schedule
Naphthylalkanones	
Nabumetone	
Propionic Acid derivatives	
Flurbiprofen	Rx
Fenoprofen	Rx (Discontinued)
Ibuprofen	OTC: 200, 300, 400 mg and Rx: 600 mg
Ketoprofen	Rx
Naproxen	OTC: 220mg and Rx: 250, 275, 375, 500, 550, 750mg
Oxaprozin	Rx
Tiaprofenic Acid	Rx
Oxicam derivatives	
Meloxicam	Rx
Piroxicam	Rx
Tenoxicam	Rx
Pyrazolone derivatives	
Phenylbutazone	Rx (Discontinued)
Oxyphenbutazone	Rx (Discontinued)
* Rx: Prescription ** OTC: Over-the-counter	

1.5 Clinical research question 1: NSAID as a contributing risk factor for CVD in OA

The cardiovascular adverse effects of NSAIDs are hypothesized as the primary causal pathway explaining the increased CVD risk among OA patients (4,19,20). Direct adverse effects of decreased physical activity among OA patients are not evident (33). OA is a chronic disease condition and mobility and physical activity are usually reduced at only advanced stages (33–35). Thus, it is unlikely that reduced physical activity among OA patients would have a substantial role in the increased risk of CVD. Although chronic inflammation in OA patients has been reported in the literature (27), this inflammation was linked to cartilage breakdown and clinical symptoms, including swelling and pain in the joint. Unlike rheumatoid arthritis (RA) in which several similarities exist between the mechanisms of inflammation in the pathogenesis of atherosclerosis and RA (36–38), the link between inflammation in OA and CVD is not established. Rather, it provides a rationale for the frequent use of NSAID among OA patients. Chronic inflammation in OA patients may cause more cartilage breakdown resulting in more pain leading to increased need for NSAIDs. In contrast, findings from randomized controlled trials (RCTs) and observational studies suggest that NSAIDs, the most commonly used medications among OA patients, are an independent risk factor for CVD (24–26). Thus, it is logical to hypothesize that the use of NSAIDs may substantially contribute to CVD risk among OA patients.

1.5.1 Risk of CVD associated with NSAIDs

The risk of CVD, ischemic heart diseases and stroke associated with NSAIDs has been well documented. Below is a summary of the study findings according to the hierarchy of evidence.

1.5.1.1 Findings from systematic review and meta-analysis

1.5.1.1.1 Findings from systematic review and meta-analysis of RCTs on the association between NSAIDs and CVD

In a network meta-analysis involving 31 trials in 116,429 patients, Trelle et al. reported that ibuprofen, diclofenac, etoricoxib and lumiracoxib were associated with a greater than 30% increase in relative risk for various CVDs (25). In another meta-analysis of RCTs by Kearney et al., selective COX-2 inhibitors were found to be associated with a 42% increased risk for incidence of serious vascular events compared to placebo (26).

1.5.1.1.2 Findings from systematic review of observational studies on the association between NSAIDs and CVD

A systematic review of population-based observational studies on the cardiovascular risk with NSAIDs suggested that the use of NSAIDs is associated with increased risk of CVD (24). McGettigan et al. reported that among the NSAIDs that were investigated in at least 10 observational studies, the risk of CVD was highest for rofecoxib [RR (95% CI): 1.45 (1.33, 1.59)] followed by diclofenac [1.40 (1.27, 1.55)], ibuprofen [1.18 (1.11, 1.25)] and naproxen [1.09 (1.02, 1.16)] (24).

1.5.1.1.3 Findings from systematic review and meta-analysis of observational studies on the association between NSAID and ischemic heart diseases

In a systematic review and meta-analysis of observational studies investigating the risk of acute myocardial infarction (AMI) associated with NSAID use, Varas-Lorenzo et al. reported that NSAIDs increased the risk of AMI compared to non-use (39). The RR (95% CI) for various

NSAIDs were as follows: etoricoxib 1.97 (1.35, 2.89), etodolac 1.55 (1.16, 2.06), indometacin 1.40 (1.21, 1.62), diclofenac 1.38 (1.26, 1.52), rofecoxib 1.34 (1.22, 1.48), meloxicam 1.25 (1.04, 1.49), ibuprofen 1.14 (0.98, 1.31), celecoxib 1.12 (1.00, 1.24) and naproxen 1.06 (0.94, 1.20) (39). In a separate meta-analysis, non-selective NSAIDs (ns-NSAIDs) were found to be associated with a 19% increased risk of AMI (40). In a meta-analysis of observational studies investigating the risk of AMI with NSAIDs, Bally et al. reported that the odds of MI was higher among current NSAID users compared to non-users (41). The OR (95% CI) was 1.24 (0.91, 1.82), 1.58 (1.07, 2.17), 1.53 (1.07, 2.33), 1.48 (1.00, 2.26) and 1.50 (1.06, 2.04) for celecoxib, refecoxib, naproxen, ibuprofen and diclofenac, respectively (41).

1.5.1.1.4 Findings from systematic review of observational studies on the association between NSAIDs and stroke

In a systematic review of observational studies, Varas-Lorenzo et al. reported that compared to non-use of NSAIDs, current use of rofecoxib and diclofenac was associated with an increased risk of incident stroke; the RR (95% CI) was 1.64 (1.15, 2.33) and 1.27 (1.08, 1.48), respectively (42).

1.5.1.2 Findings from cohort studies

In a cohort study involving older adults with arthritis who were treated with NSAIDs, Solomon et al. found that coxibs were associated with a 28% higher risk of CVD events when compared to non-selective NSAIDs (43). In a separate cohort study, Cunnington et al. found that chronic exposure to rofecoxib increased the risk of ischemic events by 25% compared to non-users (44). In a retrospective cohort study, Rahme et al. reported that the risk of hospitalization for AMI was

substantially higher among NSAID users compared with acetaminophen users (45). In a retrospective cohort study using health administrative data from Ontario Canada, Mamdani et al. reported that compared to non-NSAID users, rofecoxib and non-selective NSAIDs were associated with a higher risk of congestive heart failure. The RR (95% CI) was 1.8 (1·5, 2·2) and 1.4 (1.0, 1.9), respectively (46).

1.5.1.3 Findings from case-control studies

In a nested case-control study, Andersohn et al. observed that current use of coxibs and diclofenac significantly increased the risk of AMI compared to non-NSAID use within one year from the outcome event date (47). The adjusted RRs (95% CI) were 2.09 (1.10, 3.97), 1.29 (1.02, 1.63), 1.56 (1.22, 2.00) and 1.37 (1.17, 1.59) for etoricoxib, rofecoxib, celecoxib and diclofenac, respectively (47). In a case-control study investigating the risk of AMI among elderly adults, Levesque et al. found that rofecoxib was associated with an increased risk of AMI (48). In a nationwide case-control study in Finland, Helin-Salmivaara et al. found that the odds of MI was 1.40 times higher among current NSAID users compared to non-users (49). In a population-based case-control study, Poza et al. reported that the odds of ischemic stroke was 1.53 times higher among current users of diclofenac compared to non-users (50).

1.5.2 NSAID use in OA-CVD association: confounder versus mediator?

The following table contains detailed description of the observational studies that investigated the relationship between OA and CVD. Out of 15 studies 8 did not consider the CVD related adverse effect of NSAIDs. Six studies accounted for NSAID use by merely adjusting for it as a

confounding variable in the analysis (9). One study evaluated the mediating role of NSAIDs in the relationship between hip OA and mortality (51).

Table 1-2 Summary of observational studies investigating the association between osteoarthritis and cardiovascular diseases

Author	*Year published*	*Study design*	*Exposure*	*Outcome*	*How is the effect of NSAID on outcome accounted for?*
Kluzek et al. (15)	2016	Cohort	Knee and hand OA	All-cause and disease-specific mortality including CVD-mortality	Adjusted for self-reported use of NSAID at baseline as a confounding variable
Veronese et al. (10)	2016	Cohort	OA	CVD	Adjusted for NSAID use at baseline as a confounding variable
Barbour et al. (51)	2015	Cohort	Hip OA	All-cause and disease-specific mortality including CVD-mortality	Examined the mediating role of baseline NSAID use in the hip OA and mortality among older white women
Haugen et al. (11)	2015	Cohort	Hand OA	Mortality and incident cardiovascular events	Adjusted for current use of NSAIDs as a confounding variable
Hoeven et al. (52)	2015	Cohort	OA	CVD	Did not account for NSAID use
Rahman et al. (4)	2013	Cohort	OA	CVD	Did not account for NSAID use
Rahman et al. (16)	2013	Cross-sectional	Self-reported OA	Self-reported heart diseases	Adjusted for self-reported pain medication including NSAID use within the last 30 days from the survey date as a confounding variable
Ong et al. (53)	2013	Cross-sectional	Arthritis (OA and rheumatoid arthritis)	CVD	Did not account for NSAID use
Nielen et al. (54)	2012	Cross-sectional	OA, diabetes mellitus and inflammatory arthritis	CVD	Did not account for NSAID use

Author	*Year published*	*Study design*	*Exposure*	*Outcome*	*How is the effect of NSAID on outcome accounted for?*
Nuesch et al. (13)	2011	Cohort	Knee or hip OA	All-cause and disease-specific mortality	Adjusted for NSAID use at baseline as a confounding variable
Tsuboi et al. (12)	2011	Cohort	Musculoskeletal degenerative disease (knee OA, lumbar spondylosis, and osteoporosis)	Disease-related mortality including cerebrovascular and cardiovascular-related death	Did not account for NSAID use
Kishimoto et al. (55)	2009	Cohort	Arthritis	Coronary heart disease	Adjusted for aspirin or NSAID use within the last two weeks from the medical interview date as a confounding variable
Jonsson et al. (18)	2009	Case-control	Hand OA	Atherosclerosis	Did not account for NSAID use
Kadam et al. (17)	2004	Case-control	OA	Clinical co-morbidities including ischaemic heart disease, angina pectoris and heart failure	Did not account for NSAID use
Haara et al. (14)	2003	Cohort	Finger joint OA	Total and cardiovascular mortality	Did not account for NSAID use

The use of NSAID lies in the causal pathway of the OA-CVD relationship. Instead of adjusting for NSAID use in the outcome model, a mediation analysis approach is more appropriate in investigating the role of NSAID in the OA-CVD association. The first research question in the thesis is developed based on the NSAID use in treating OA patients and the demonstrated CVD risk associated with NSAIDs. The objective is to investigate the role of NSAIDs in the increased risk of CVD among OA patient compared to non-OA individuals using a population-based cohort and sophisticated methods of mediation analysis.

1.5.3 Mediation analysis using Marginal Structural Model (MSM)

Baron and Kenny proposed the difference of coefficients approach for mediation analysis in which two separate models, with and without the mediator, are fitted and the difference in the coefficients of the exposure variable is considered as the mediated effect (56). This approach is usually applicable for normally distributed outcomes and mediators (56) and has several limitations in conducting mediation analysis in a survival analysis context (57,58). Survival times are generally right-censored, may not be normally distributed and proportional hazards assumption may not be satisfied for the associated models. Moreover, a causal interpretation cannot be made from the changes in coefficients (56,58–62). As a result, assessment of mediation in the survival analysis context would require a more sophisticated analysis technique. Mediation analysis using Marginal Structural Model (MSM) is becoming popular in epidemiological studies, including pharmacoepidemiologic studies involving OA patients. For example, in a population-based longitudinal study Liu et al. used the MSM approach to investigate the mediating role of NSAID use and walking disability in the association between knee OA and all-cause mortality (63). The authors reported that knee OA was associated with a

51% increase in the risk of all-cause mortality (63). Approximately 76% of this total risk was mediated through NSAIDs use (63). In a prospective cohort study Barbour et al. also used MSM to reveal the mediating role of NSAIDs in the hip OA and all-cause mortality relationship (51). The study concluded that there was no evidence to support the mediating role of NSAID in this complex relationship (51). The following table presents examples of epidemiological studies that have used the MSM approach in mediation analysis.

Table 1-3 Examples of epidemiological studies that used the marginal structural model as a tool for mediation analysis

Author	Title	Publication year	Journal published
Liu et al. (63)	Knee Symptomatic Osteoarthritis, Walking Disability, NSAIDs Use and All-cause Mortality: Population based Wuchuan Osteoarthritis Study	2017	Nature Scientific Reports
Barbour et al. (51)	Hip Osteoarthritis and the Risk of All-Cause and Disease-Specific Mortality in Older Women	2015	Arthritis & Rheumatology
Nordahl et al. (64)	Education and Cause-specific Mortality The Mediating Role of Differential Exposure and Vulnerability to Behavioral Risk Factors	2014	Epidemiology
Rochon et al. (58)	Mediation analysis of the relationship between institutional research activity and patient survival	2012	BMC (BioMed Central)
Lange et al. (57)	A Simple Unified Approach for Estimating Natural Direct and Indirect Effects	2011	American Journal of Epidemiology
Lange et al. (62)	Direct and Indirect Effects in a Survival Context	2011	Epidemiology

1.6 Clinical research question 2: risk of CVD associated with different NSAIDs used in treating OA

After examining the potential mediating role of NSAIDs in the relationship between OA and increased risk of CVD, this thesis focuses on evaluating the cardiovascular safety of various NSAIDs that are used in treating OA in the real-world. Since OA is a non-curable chronic disease, health care providers frequently use NSAIDs in the treatment of OA (8). In Canada, more than a million NSAID prescriptions are written to OA patients each year (8). Data from various sources suggest that the majority of OA patients use NSAIDs (22,23). Previous research has demonstrated that NSAIDs are associated with CVD adverse events and the risk varies among different NSAIDs (25,26,48). Previous epidemiological studies investigating the risk of CVD associated with NSAIDs primarily looked into ischemic outcomes in specific patient populations (39,41,42,47–50,65–68). Consequently, data on the overall CVD safety of NSAIDs among OA patients are limited. Rahman et al. reported that OA is associated with many CVD outcomes including MI, stroke and CHF (4,16). As such, knowing the safety of NSAIDs in terms of overall CVD outcomes among OA patients is an important area of research. The second objective of this thesis is to assess the risk of CVD associated with different NSAIDs that are used to treat OA patients in the real world.

1.7 Specific thesis aims by chapter

1.7.1 Aim #1 (clinical): To disentangle the role of NSAIDs in the increased risk of CVD among OA patients.

Chapter 2 is a longitudinal study investigating the mediating role of NSAIDs in the increased risk of CVD among OA patients compared to people without OA.

1.7.2 Aim #2 (clinical): To evaluate the overall cardiovascular safety of various NSAIDs that are used in treating OA patients in the real world.

Chapter 3 is a retrospective cohort study investigating the risk of CVD associated with different NSAIDs used in the treatment of OA compared to unexposed person-time using health administrative data from BC, Canada.

1.7.3 Aim #3 (methodological): To investigate whether imputing an important study variable by multiple imputation instead of proportion-based imputation method using data from external sources would result in less biased estimates in a given relationship.

Chapter 4 is a plasmode-simulation based study investigating the OA-CVD relationship as an example in which BMI is an important confounding variable. This study used data from a population-level health survey in Canada to compare multiple imputation with proportion-based imputation in imputing unavailable BMI data using data from external sources.

1.8 Data Sources

1.8.1 Canadian health administrative databases

Administrative data are increasingly used in observational studies. There are certain advantages of this secondary data analysis. For example, the data is already collected for billing purposes, which means that using administrative data is efficient in terms of both cost and time. In most circumstances, long-term follow-up records are available at the individual level. This enables researchers to determine the temporal sequence in studies investigating the association between two diseases. Also, it offers a large sample size at the population level that better allows the study findings to represent the "real world". Several health administrative databases exist in

Canada, including provincial-level health administrative data (HAD) and population-level health survey data.

1.8.2 Data sources for chapter 2 and 3

Epidemiologic studies in chapter 2 and 3 used HAD of a previously assembled, population-based cohort (N=720,055) from British Columbia, Canada (69–73). Individuals were registered under the publicly funded BC medical services plan (MSP) as a BC resident for the fiscal years 1991/92 to 2012/13. The BC Ministry of Health and the BC Vital Statistics Agency approved access to and use of the data facilitated by Population Data BC (74) for this study. To ensure anonymity, Population Data BC removed all identifiable information including name and address. The personal health numbers were replaced with a unique serial number called study ID. Ability to link data from various sources using this study ID makes BC HAD an excellent resource for health research. The following table provides information on various data files accessed and examples of information linked at individual level in this study.

Table 1-4 British Columbia health administrative data files accessed and the information linked at individual-level in this dissertation

Data file	*Type of information*	*Use in this study*
Medical Services Plan (MSP) Payment Information File (70)	Provides information on all provincially funded health service utilization including date and type of service, practitioner and ICD-9* diagnostic code.	To create OA exposure, CVD outcome and co-morbid disease conditions.
Discharge Abstract Database (Hospital Separations) (71)	Provides information on inpatient hospitalizations including admission date, up to 25 ICD-9 or ICD-10** diagnoses codes, and separation date.	To create OA exposure, CVD outcome and co-morbid disease conditions.
Consolidation File (MSP Registration & Premium Billing) (72)	Provides information on MSP registration start and end dates, date of birth, sex and SES.	To create co-variables of age, sex, SES and MSP registration status.
Vital Statistics Deaths (73)	Provides information on date and cause of death	Follow up of study individuals ended in an event of death.
PharmaNet (69)	Provides information on all community-dispensed prescriptions, regardless of funding source including drug	To create NSAID exposure variable.

	identification number (DIN), dispensing date, quantity dispensed, and the number of days supplied	
*ICD-9: International Classification of Diseases 9th revision. **ICD-10: International Statistical Classification of Diseases and related health problems 10th revision.		

1.8.3 Data sources for chapter 4

The plasmode simulation study in chapter 4 used data from the first three cycles of the CCHS (75–77). It is a population-level health survey data that represents approximately 98% of the Canadian population and is publicly available for research. Beginning in 2001, the CCHS collected information on several variables including demographics, health status and other health determinants of approximately 130,000 respondents. The survey was repeated every two years until 2005 and annually thereafter. Data from this cross-sectional survey is primarily used for population health research.

Chapter 2: Role of Non-Steroidal Anti-Inflammatory Drugs (NSAIDs) in the association between osteoarthritis and cardiovascular diseases: A longitudinal study

2.1 Introduction

In Canada, cardiovascular diseases (CVD) have a significant disease burden in terms of both mortality and health-related quality of life. Approximately 2.4 million Canadians live with CVD, and more than 48,000 people die from CVD each year (78).

Osteoarthritis (OA) is a major musculoskeletal disorder that affects one in eight Canadian adults (8). Recent research suggests that OA is an independent risk factor for CVD (4,9,10,16,53). In a systematic review and meta-analysis of observational studies, Wang *et al.* reported that people with OA experience a 24% greater risk of CVD compared to people without OA (pooled RR: 1.24, 95% CI: 1.12 to 1.37) (9). The underlying mechanism for this higher risk of CVD among OA patients has not been studied yet (4,79).

Currently, there is no cure for OA; NSAIDs are frequently used to control the primary symptoms of OA, pain and inflammation (8). In 2007, approximately 1.2 million NSAID prescriptions were written in Canada to treat people with OA (8). The BC Osteoarthritis Survey in 2007 reported that 78% of OA patients in BC received NSAIDs (22). This finding closely matches the observations of a Spanish cross-sectional study, where 79% of OA patients received NSAIDs (23).

Findings from both observational studies and randomized clinical trials (RCTs) suggest that NSAIDs are associated with cardiovascular side effects (24–26). The substantial use of NSAIDs in OA treatment and knowledge of CVD risk associated with NSAID use warrant an investigation of the role of NSAIDs in the OA-CVD association. With respect to NSAID use, only six out of 15 observational studies that looked at the CVD risk among people with OA considered NSAID use as a confounder (9). Many experts in the field have hypothesized that NSAIDs, frequently used in OA treatment, may ultimately cause these patients to develop CVD (4,79), that is, NSAID use is believed to lay in the causal pathway of the OA-CVD relationship. Rather than considering NSAID use as merely a confounder in the investigation of the OA-CVD relationship, a "mediation analysis" appears to be a more appropriate method of inquiry. Mediation analysis using marginal structural models (MSM) is increasingly being used in epidemiological studies (51,58,62–64). Lange *et al.* proposed an algorithmic way to obtain unbiased estimates of the natural direct and indirect effects, using counterfactual principles (57). The current study is the first to evaluate the mediating role of NSAID use in the OA-CVD relationship.

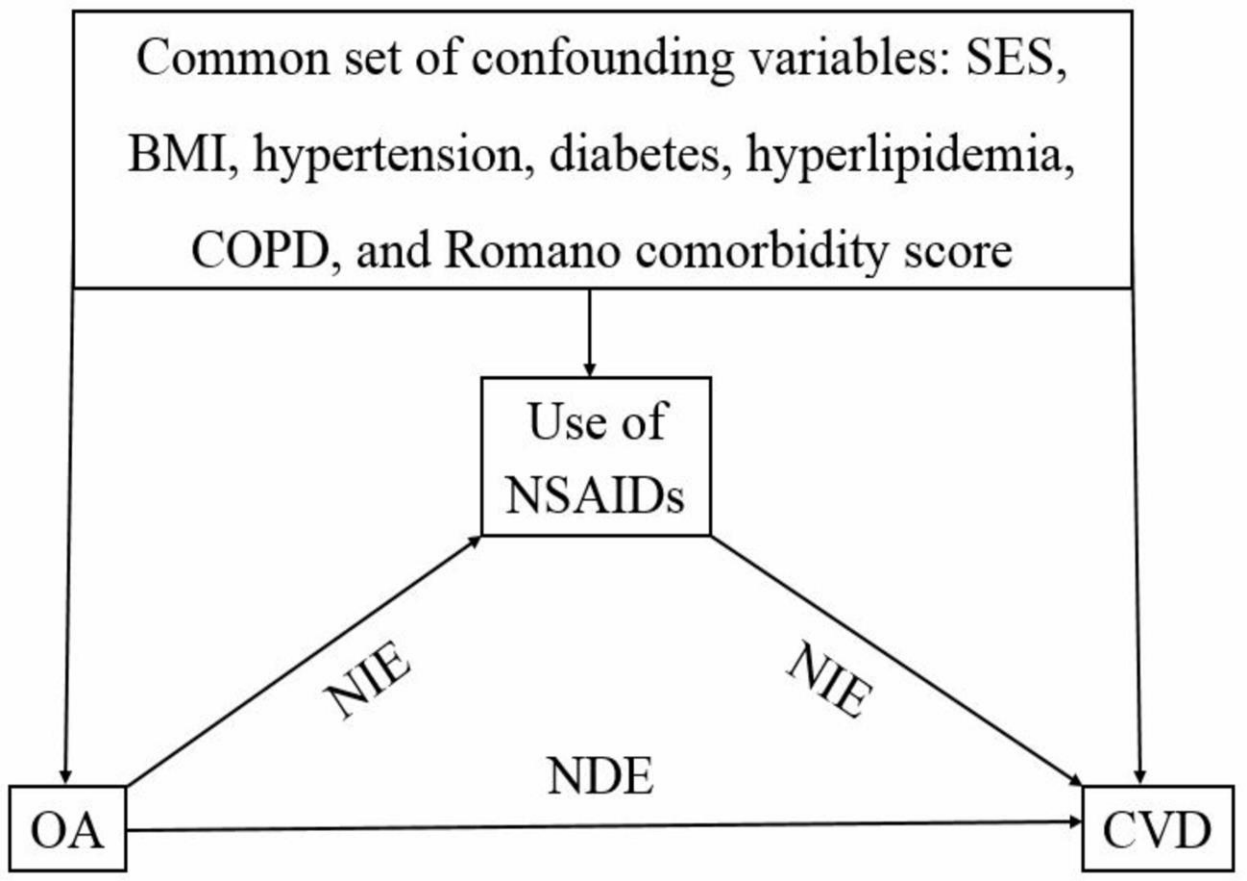


Figure 2-1 Hypothesized causal diagram of OA-CVD relationship. Here, the direct effect of OA on CVD is the primary effect of OA on CVD (OA → CVD) and the indirect effect is mediated through the use of NSAIDs (OA → NSAIDs → CVD).

2.2 Materials and Methods

2.2.1 Study population

In this longitudinal study, we analyzed a population-based cohort of 720,055 British Columbians (4,69–73). These subjects were registered in the BC Medical Services Plan (MSP) from April 01, 1991 to December 31, 2013. The cohort was assembled from the BC health administrative data (HAD) maintained by the BC Ministry of Health. Population Data BC removed all identifiable

information such as subject name and address to preserve anonymity (74). A unique serial number (study ID) replaced the personal health number of each subject. Individual-level information such as date of birth, sex, socioeconomic status (SES), billing information for any health-related consultation as well as hospital data such as date of hospital admission, diagnosis, and date of hospital separation were linked using study ID. Vital statistics deaths data was also linked at individual level. Information on all community-dispensed prescriptions, regardless of funding source was available for each subject from January 01, 1996 to December 31, 2013. For each prescription dispensed, data were available on the drug identification number (DIN), dispensing date, quantity dispensed, and the number of days supplied. This study was approved by the Behavioural Research Ethics Board at the University of British Columbia, Canada.

2.2.2 OA exposure definition

People with OA were identified using previously validated criteria using International Classification of Diseases (ICD) diagnostic codes (34). A case definition of at least two visits to a health professional, separated by at least one day, but within a two-year period, or one discharge from the hospital coded for OA (ICD-9 code of 715 or ICD-10 code of M15–M19) was adopted to identify OA patients. A visit to a health professional was defined as any service except a few procedures such as midwifery and obstetrics, anesthesia, dialysis and/or transfusion plus certain diagnostic procedures such as diagnostic radiology and ultrasound. All visits to any healthcare provider were assessed and the date of the second visit to a health professional or the date of hospital admission, whichever occurred first, was considered as the OA diagnosis date (index date). All individuals who met the OA case definition during the period of January 01, 1996 to December 31, 2008, and aged at least 20 years on the OA-index date and did not have a

history of CVD based on physician consultation and/or hospital admission before OA-diagnosis were included into the OA-exposure group. We excluded all individuals with rheumatoid arthritis (RA) from the cohort before identifying people with OA. As such, subjects with OA in our analyses never had RA diagnosis during entire study period.

2.2.3 Comparison group

A group of non-OA individuals who never had OA-related physician consultation or hospital admission and were never diagnosed with RA during the entire study period was identified. For each OA patient, all non-OA individuals who were at risk of CVD on the corresponding OA-diagnosis date were identified and then three non-OA individuals were randomly selected by matching based on exact age and sex. Once selected for an OA-case, those three matched individuals were removed from the pool of non-OA individuals and then matching was performed for the next OA-individual (without replacement).

2.2.4 Outcome variable

Composite CVD event was the primary outcome in this study. Specific CVD events such as ischemic heart disease (IHD), congestive heart failure (CHF) and stroke were the secondary outcomes. Using a definition similar to that proposed by Tonelli *et al.*, we identified primary and secondary CVD events from the discharge abstract database (hospital separations) (71), medical services plan (MSP) payment information file (70) and vital statistics deaths (73) data file using ICD-9 (410-414, 428, 430-434, 436, 438) or ICD-10 codes (I20-I25, I50, I60-I64) (80). A case definition of one hospitalization or two visits to a health professional separated by at least one day but within two years or underlying cause of death, whichever occurred first, was used to

identify CHF in administrative data. For IHD, the case definition was restricted to one hospitalization or underlying cause of death, whichever occurred first. In identifying stroke, the case definition was one hospitalization or one visit to a health professional or underlying cause of death, whichever occurred first.

2.2.5 Covariates

Older age, sex, family history of CVD, obesity, smoking, physical inactivity, unhealthy diet, stress, hypertension (high blood pressure), high cholesterol and diabetes are common risk factors for CVD (78,81–84). In this study, we used exact age on OA-index date and sex to match non-OA controls for OA-cases. SES was the neighborhood income per person equivalent that was recorded as a range of five income groups where one was the lowest and five was the highest income quintile. Adopting a similar definition as proposed by Tonelli *et al.*, we assessed the baseline history of co-morbid disease conditions such as hypertension, diabetes, hyperlipidemia and chronic obstructive pulmonary disease (COPD) using ICD-9 or ICD-10 codes (80). We used the following ICD codes to identify the various co-morbid diseases; hypertension (ICD-9 code 401 or ICD-10 code I10), diabetes (ICD-9 code 250 or ICD-10 code E11), hyperlipidemia (ICD-9 code 272 or ICD-10 code E78) and COPD (ICD-9 code 490, 492, 494, 496 or ICD-10 code J40, J43, J44, J47). We also calculated the Romano comorbidity score at baseline for each subject (85,86). The Romano score contains 19 critically important comorbidities and has been used as a risk adjustment in epidemiological studies.

2.2.6 Body mass index (BMI) imputation

The BC HAD does not include information on BMI. Previous studies have imputed BMI based on population level averages (4). In this study, we imputed BMI at the individual level by

multiple imputation technique (87) using data from Canadian Community Health Survey (CCHS) (88,89) (see Appendix A for technical details).

2.2.7 Mediator variable

Using a similar definition previously used in literature, we created a binary variable of current use of NSAIDs based on NSAID mediator index date (48,68,90). In this matched data (one individual with OA and corresponding three controls) analysis, we defined a NSAID mediator index date (based on the CVD date for an OA exposed individual), and applying that index date to that individual's control group members who were free from CVD event. An individual was said to be a current user of NSAID if the duration of the last NSAID prescription dispensed overlapped with the mediator index date (48,68,90). In a sub-group analysis, we grouped all NSAIDs into selective coxibs (Celecoxib, Lumiracoxib, Rofecoxib, Valdecoxib) only and all other conventional NSAIDs (con-NSAIDs) including diclofenac, ibuprofen and naproxen (91,92). Using similar definition, we created current use of coxibs as well as current use of con-NSAIDs for primary and secondary outcomes.

2.2.8 Statistical analysis

Study individuals were followed prospectively in the database for incident CVD. For individuals with OA, the follow-up period started from the date of their OA diagnosis (index date). For non-OA controls, follow-up period began from the index date of the corresponding OA-case for which controls were selected. Individuals were censored if they died, or emigrated or reached the study end date (December 31, 2013), whichever occurred first.

Person-years of risk for follow-up period were calculated. Covariates were selected by comparing the difference in Akaike Information Criterion (AIC) (93). If the AIC decreased by at least four points in a bivariate model based on the primary outcome, we decided to include the covariate into the model (94). We assessed the proportionality assumption by observing the log [- log (survival function)] versus log of survival time. We estimated the hazard ratio (HR) and 95% confidence interval (CI) using Cox Proportional Hazards (Cox PH) model.

We used a marginal structural model (MSM) approach based on a counterfactual framework (95) to decompose the effect of OA on increased risk of CVD into (a) NDE, i.e., the direct effect of OA on CVD that was not mediated through NSAIDs use and (b) NIE i.e. the effect of OA on CVD mediated through NSAIDs use (see Figure 2-1). Adjusting set of confounding variables were SES, BMI, hypertension, diabetes, hyperlipidemia, COPD, and Romano comorbidity score. The algorithm proposed by Lange *et al.,* allowed us to fit MSM and obtain HR estimates of natural direct and indirect effects from a weighted Cox PH model (57) (see Appendix B for technical details). The HR for the total effect (TE) was the product of HRs for NDE and NIE. The confidence interval of the TE and proportion mediated was determined by bootstrap percentile method (57,58).

2.3 Results

7,743 OA patients were identified from January 1996 to December 2008, and 23,229 non-OA controls who were at risk of CVD on the corresponding OA index date were matched based on age and sex. Table 2-1 presents the characteristics of the overall study sample (n= 30,972) by OA exposure status. Mean age of the study participants was 64.53 years. Approximately 56.04% of

the study sample was female. A higher proportion of individuals with OA were obese (29.35%) compared to non-OA individuals (19.76%). The sample was approximately equally distributed across the different socioeconomic groups. Among the four co-morbid conditions, hypertension and COPD were more common among people with OA compared to non-OA controls. Approximately 33% of OA patients had hypertension compared to 28% among non-OA controls, and 10% of OA patients had COPD compared to 7% among non-OA controls. OA patients had more comorbidities compared to the non-OA controls. The mean of Romano comorbidity score was 0.27 and 0.20 among OA patients and non-OA controls, respectively.

Table 2-1 Characteristics of study sample (n= 30,972) by OA exposure status

Variable	**Total** **N (%)**	**Individuals with OA** **N (%)**	**Individuals without OA** **N (%)**
Analytic sample	30,972 (100%)	7,743 (25%)	23,229 (75%)
CVD			
No	21,520 (69.48%)	4,860 (62.77%)	16,660 (71.72%)
Yes	9,452 (30.52%)	2,883 (37.23%)	6,569 (28.28%)
Age in years (Mean ± SD)	64.53 (12.81)	64.53 (12.81)	64.53 (12.81)
Sex (% of female)	56.04 %	56.04 %	56.04 %
BMI category			
Normal weight	10,515 (33.95%)	2,161 (27.91%)	8,354 (35.96%)
Underweight	1587 (5.12%)	256 (3.31%)	1,331 (5.73%)
Overweight	12,007 (38.76%)	3,053 (39.43%)	8,954 (38.55%)
Obese	6,863 (22.15%)	2,273 (29.35%)	4,590 (19.76%)
Socioeconomic status (SES)			
1 (Lowest SES)	5,950 (19.21%)	1,475 (19.05%)	4,475 (19.26%)
2	5,740 (18.53%)	1,409 (18.19%)	4,331 (18.64%)
3	6,954 (22.45%)	1,689 (21.81%)	5,265 (22.67%)
4	6,043 (19.52%)	1,553 (20.07%)	4,490 (19.34%)
5 (Highest SES)	6,285 (20.29%)	1,617 (20.88%)	4,668 (20.09%)

Variable	Total N (%)	Individuals with OA N (%)	Individuals without OA N (%)
Hypertension			
No	22,024 (71.11 %)	5,186 (66.98 %)	16,838 (72.48 %)
Yes	8,948 (28.89 %)	2,557 (33.02 %)	6,391 (27.52 %)
Hyperlipidemia			
No	28,675 (92.58 %)	7,219 (93.23 %)	21,456 (92.37 %)
Yes	2,297 (7.42 %)	524 (6.77 %)	1,773 (7.63 %)
Diabetes			
No	28,352 (91.54 %)	7,112 (91.86 %)	21,240 (91.44 %)
Yes	2,620 (8.46 %)	631 (8.14 %)	1,989 (8.56 %)
COPD			
No	28,472 (91.93 %)	6,944 (89.68 %)	21,528 (92.68 %)
Yes	2,500 (8.07 %)	799 (10.32 %)	1,701 (7.32 %)
Romano comorbidity score (Mean ± SD)	0.22 (0.76)	0.27 (0.83)	0.20 (0.74)
Current NSAID use			
No	30,309 (97.86%)	7,329 (94.65%)	22,980 (98.93%)
Yes	663 (2.14%)	414 (5.35%)	249 (1.07%)

There were 9,452 incident cases of primary outcome, i.e., CVDs. The mean follow-up time was approximately 9.74 (standard deviation 5.07) years and the cumulative follow-up period was

301,886 person-years. The crude incidence rate of CVD per 1,000 person-years was 38.07 and 29.05 among OA patients and non-OA controls, respectively. In the unadjusted Cox PH model, the risk of developing CVD among people with OA was 31% higher compared to people without OA (HR=1.31, 95% CI 1.25, 1.37). After adjusting for SES, BMI, hypertension, diabetes, hyperlipidemia, COPD, & Romano comorbidity score, the risk of CVD attenuated but was significantly higher among people with OA. The adjusted HR (95% CI) was 1.23 (1.17, 1.28). Among secondary outcomes, the risk was highest for CHF, followed by IHD and stroke. The adjusted HR (95% CI) was 1.42 (1.33, 1.51), 1.17 (1.10, 1.26) and 1.14 (1.07, 1.22) for CHF, IHD and stroke, respectively.

Table 2-2 Unadjusted and adjusted hazard ratios for primary and secondary outcomes

Outcome type	**Outcome**	**Unadjusted effect HR (95% CI)**	**Adjusted* effect HR (95% CI)**
Primary	CVD	1.31 (1.25, 1.37)	1.23 (1.17, 1.28)
Secondary	IHD	1.26 (1.18, 1.34)	1.17 (1.10, 1.26)
	CHF	1.51 (1.42, 1.61)	1.42 (1.33, 1.51)
	Stroke	1.22 (1.14, 1.29)	1.14 (1.07, 1.22)
* Adjusted for SES, BMI, hypertension, diabetes, hyperlipidemia, COPD, & Romano comorbidity score. Please see Appendix C for detail results of the Cox PH model.			

In the process of estimating the mediating effect, we first modeled the relationship between NSAIDs use (the mediator) with OA exposure and CVD outcome. For the primary CVD outcome, current NSAID use was significantly more prevalent among OA patients (5.35%)

compared to non-OA controls (1.07%). The odds of using NSAIDs were 5.09 times higher among people with OA compared to people without OA, adjusted OR (95% CI) was 5.09 (4.33, 5.99). The mediator was also found to be strongly associated with CVD outcome. Adjusted (HR (95% CI) was 4.14 (3.80, 4.50) after adjusting for OA, SES, BMI, hypertension, diabetes, hyperlipidemia, COPD, & Romano comorbidity score.

Finally, we implemented MSM to estimate the mediating effect. Table 2-3 presents the natural direct effect (NDE: OA→CVD) and natural indirect effect mediated through NSAIDs use (NIE: OA → NSAIDs → CVD). The 23% increased risk of CVD among people with OA was decomposed into a direct HR (95% CI) of OA of 1.13 (1.09, 1.18) and an indirect, NSAID-mediated HR (95% CI) of 1.09 (1.07, 1.09). Approximately 41% (95% CI: 33%, 51%) of the total effect of OA on the increased risk of CVD was mediated through NSAIDs use. Among the secondary outcomes, the proportion (95% CI) of total effect mediated through current NSAID use was 23% (17%, 28%) for CHF, 56% (37%, 87%) for IHD and 64% (39%, 115%) for stroke.

Table 2-3 Total, direct and mediated effect through NSAIDs use (Reference groups was no-OA)

Outcome (type)	OA Exposure status	Total effect HR (95% CI)	Natural Direct Effect (NDE) HR (95% CI)	Natural Indirect Effect (NIE) HR (95% CI)	Mediation effect (%)
CVD (primary)	Yes	1.23 (1.18, 1.28)	1.13 (1.09, 1.18)	1.09 (1.07, 1.09)	40.59
IHD (secondary)	Yes	1.17 (1.09, 1.26)	1.08 (1.01, 1.15)	1.09 (1.07, 1.11)	56.19
CHF (secondary)	Yes	1.42 (1.32, 1.51)	1.31 (1.22, 1.39)	1.08 (1.07, 1.10)	23.31
Stroke (secondary)	Yes	1.14 (1.07, 1.22)	1.06 (0.99, 1.13)	1.08 (1.07, 1.10)	63.69

Results of sub-group analysis revealed that a greater proportion of the risk of all primary and secondary outcomes were mediated through con-NSAIDs compared to selective coxibs only. Approximately 29% (21%, 38%) of the total risk of CVD was mediated through con-NSAIDs compared to 21% (15%, 29%) mediated through selective coxibs. For IHD, con-NSAIDs explained approximately 45% (26%, 79%) of the total risk compared to 32% (16%, 63%) explained by coxibs. The proportion of CHF risk mediated through con-NSAIDs was slightly higher than that of coxibs, 14% (10%, 19%) versus 12% (8%, 18%). Following similar trend, con-NSAIDs explained a higher portion of 52% (29%, 107%) of the total effect of OA on stroke. In contrast, selective coxibs explained 34% (13%, 78%).

Table 2-4 Sub-group analysis; total, direct and mediated effect through selective coxibs & con-NSAIDs use (Reference groups was no-OA)

Mediator class	Outcome (type)	OA status	Total effect HR (95% CI)	NDE HR (95% CI)	NIE HR (95% CI)	Mediation Effect (%)
Coxibs only	CVD (primary)	Yes	1.18 (1.13, 1.24)	1.14 (1.09, 1.19)	1.04 (1.03, 1.04)	21.12
	IHD (secondary)	Yes	1.12 (1.05, 1.20)	1.09 (1.02, 1.16)	1.03 (1.02, 1.05)	31.72
	CHF (secondary)	Yes	1.36 (1.28, 1.45)	1.31 (1.24, 1.39)	1.04 (1.02, 1.05)	12.09
	Stroke (secondary)	Yes	1.10 (1.03, 1.17)	1.07 (1.00, 1.14)	1.03 (1.01, 1.04)	34.25
Con-NSAIDs only	CVD (primary)	Yes	1.19 (1.14, 1.25)	1.14 (1.09, 1.19)	1.05 (1.04, 1.06)	28.62
	IHD (secondary)	Yes	1.14 (1.07, 1.22)	1.08 (1.01, 1.15)	1.06 (1.04, 1.08)	44.77
	CHF (secondary)	Yes	1.37 (1.29, 1.47)	1.31 (1.24, 1.39)	1.05 (1.03, 1.06)	14.48
	Stroke (secondary)	Yes	1.13 (1.06, 1.19)	1.07 (0.99, 1.13)	1.06 (1.04, 1.07)	52.08

2.4 Discussion

In recent years, a number of studies have reported OA as an independent risk factor for CVD (4,9,10,16,53). However, no previous study investigated the cause of the increased risk of CVD among OA patients. The necessity of evaluating the causal pathways in OA-CVD relationship has been noted in the literature (4).

The prevailing hypothesis in the OA to CVD relationship has been that OA patients frequently take NSAIDs to control their pain and inflammation and that this may lead to them developing CVD (4,79). Accordingly, we undertook this longitudinal observational cohort-based analysis to investigate the underlying mechanisms of the OA-CVD relationship. First, we re-examined the effect of OA on CVD using population-based HAD from BC, Canada. We found that the risk of developing CVD among people with OA was 23% higher compared to people without OA. Among secondary outcomes the risk of CHF was 42% higher among people with OA compared to people without OA followed by 17% greater risk of IHD and 14% greater risk of stroke.
The higher risk of CVD among OA patients observed in our study was concordant with previous research findings (4,9,10,16,53). For example, in a systematic review and meta-analysis of observational studies, people with OA were found to be at 24% greater risk of CVD compared to people without OA (9). The risk of stroke was concordant with the risk among female OA patients less than 65 years reported by Rahman *et al* (4). Overall in this study, people with OA showed relatively higher risk of stroke compared to the risk reported in previous studies. Perhaps this could be explained due to the modification of the case definition. Rahman *et al.* identified stroke cases by hospitalization only with ICD 9 or ICD 10 codes for stroke (4). In this study, we have identified stroke cases using case definition of one hospitalization or one visit to a health

professional or underlying cause of death with ICD-9 or ICD-10 code for stroke whichever occurred first (80). On the other hand, in a cohort study involving 336,906 individuals, the risk of developing incident stroke was 20%, 28%, and 41% higher among indomethacin, rofecoxib and valdecoxib users, respectively (96). Thus, it is not unlikely that increased use of NSAIDs among OA patients contribute to their increased risk of stroke.

In line with the study's hypothesis, OA patients used more NSAIDs than non-OA individuals. Since the use of NSAIDs is an independent risk factor for CVD, we investigated how much of the increased risk of CVD among OA patients was mediated through their NSAIDs use (25,26). We implemented novel methodology proposed by Lange *et al.* for evaluating mediation in a survival analysis context (57). This longitudinal study revealed that the increased risk of developing CVD among OA patients was substantially mediated (41%) through their NSAIDs use. Similar to primary CVD outcome, NSAIDs played a substantial mediating role in developing IHD, CHF, and stroke (secondary outcomes). To the best of our knowledge, this was the first study investigating the underlying mechanism of OA-CVD association. Most of the previous observational studies did not account for NSAIDs use in evaluating OA-CVD association (9,24). On the other hand, NSAIDs is proven to cause CVD. In two separate network meta-analyses, NSAIDs were found to increase CVD risk by 30% to 42% (25,26). As such, the mediating effect of NSAIDs in OA-CVD relationship observed in the study appears to be plausible.

Although the outcome under consideration in this study was CVD, two previous studies evaluated the mediating role of NSAIDs on all-cause mortality among the specific type of OA,

and their results were inconsistent. In a population-based longitudinal study, Liu *et al.* found that symptomatic knee OA increased the risk of all-cause mortality by 51% of which 76.2% was mediated through NSAIDs use (63). In a separate study, Barbour *et al.* did not find any evidence that both all-cause and CVD-specific mortality could be mediated through NSAIDs among older women with radiographic hip OA (51). These studies included specific types of OA patients, for example, symptomatic knee OA and radiographic hip OA among older women. Moreover, the outcome of both these studies was mortality. In contrast, in this study, we selected OA patients from a population-based cohort using previously validated algorithm (34). The outcome of the investigation was CVD. Meta-analysis of RCTs indicated that NSAID use is an independent risk factor for developing CVD (26). Thus it made sense to hypothesize that consuming more NSAIDs to control pain and inflammation among OA patients would lead to developing CVD. Our results differ from the literature with respect to the mediator-outcome association and can be rationalized as follows. The meta-analysis included RCTs comparing selective COX-2 inhibitors with either placebo or traditional NSAIDs or both (26). The RR of 1.42 reported in the meta-analysis by Kearney et al. (26) estimated the increased risk of vascular events associated with selective COX-2 inhibitors compared to placebo or other non-selective NSAIDs among healthy volunteers. In contrast, it is important to note that our study sample involved OA patients who use more NSAIDs than the general population. The difference in the estimated CVD risk associated with NSAID use may be due to a number of distinctions including difference in study sample (healthy volunteers vs matched OA and non-OA controls) and study outcomes (myocardial infarction, stroke or vascular death vs composite CVD events created from HAD using ICD codes).

The strengths of this study include the data that allowed the assembly of a large cohort from population-level HAD. We used previously validated algorithms to develop OA exposure, CVD outcomes and other confounding variables (34,80). The ability to exclude patients with rheumatoid arthritis to obtain an unbiased estimate of the effect size was also a strength. Finally, we estimated the mediating effect using state-of-the-art mediation analytic technique in survival context based on a counterfactual framework (57,58). This framework provided us the ability to draw interpretations from a potential causal relationship. Generally, marginal structural models are associated with certain strong assumptions including unconfoundedness, positivity, consistency and no model mis-specification (97). In our analysis for primary CVD outcome, the weights were generally well-behaved. The mean weight was 0.999 with the first quartile being 0.965 and third quartile being 1.000. Even after truncating large weights, the conclusion did not change. OA is a well-defined binary exposure variable, and we do not have any reason to suspect presence of 'multiple versions' of the exposure, and therefore not much indication that consistency assumption was violated.

Our study inherits a few limitations applicable to any observational research that use HAD. We created the mediator variable, NSAIDs use, from the prescription claim data (PharmaNet) (69) that do not contain information on over-the-counter (OTC) medication. In BC, Canada, Ibuprofen and Naproxen are available as OTC. When people buy the medication without a prescription, the NSAID use variable using prescription claim data underestimated the mediator variable. In defining the NSAID mediator variable, we used the CVD event date, both for OA individuals and non-OA controls. For non-OA controls who did not develop CVD, we assigned the CVD event date of the corresponding OA individual to check if that individual without OA

was using NSAID on that day. Since taking NSAID on censoring date without CVD does not inform about the proportion of CVD risk mediated through NSAID (primary study objective) and censoring date can vary among OA cases and non-OA controls, the mediator variable was set to zero among OA individuals who did not develop CVD (approximately 16% of the study sample). This may have resulted in overestimation of the mediating effect observed. For example, assuming everyone in the OA exposed group developed CVD, the current mediator variable definition might have missed capturing an estimated maximum of 6% of the indirect effect (41% mediating effect in 16% of the study sample). For the corresponding non-OA controls who did not develop CVD, the mediating effect of NSAID would be even less because non-OA controls were found to use substantially fewer NSAIDs than OA individuals. Although the mediation analysis done in the current context is the first of its kind, results from any observational data analysis need to be interpreted with caution. Future studies should consider more advanced causal inference methodologies, that can potentially deal with the repeatedly measured mediator variables in a longitudinal setting (98–100).

The assumptions needed to identify natural direct and indirect effects relies on the absence of unmeasured confounding (101,102). Therefore, it is important that the set of confounding variables included in this study should sufficiently control for OA exposure-CVD outcome, NSAID mediator-CVD outcome and OA exposure-NSAID mediator confounding and there is no variable in the mediator-outcome association that is affected by the OA exposure (103). One of the limitations of HAD based observational studies is that data is usually collected for regulatory purposes and often do not contain information on all variables necessary in the analysis (43). For example, BMI is a known confounder in the relationship of interest for which information is not

available in BC HAD. Measurements of other important CVD risk factors such as the family history of CVD, smoking and physical activity index are also not recorded in BC HAD. We have included all the variables that we found relevant from our extensive literature search that were available in BC HAD. Although we adopted an innovative strategy of imputing BMI, confounding may still exist (78,81–84). This unmeasured confounding may inflate the proportion of CVD risk mediated through NSAID estimated from MSM. In a sensitivity analysis (see Appendix D), we present results with the BMI variable excluded from the model and observe little change in the risk of CVD associated with the OA exposure. For example, the aHR for the primary outcome, CVD, changes from 1.23 (1.17, 1.28) to 1.25 (1.20, 1.31) with the BMI variable excluded. We observed similar changes in magnitudes (1 to 3%) for all the study outcomes when excluding BMI. Other unobserved potential confounding variables such as smoking and family history of CVD to play a substantial role in a multivariable regression model to reverse the significant risk of CVD and proportion of effect mediated through NSAID use observed in this study is unlikely.

In conclusion, our study is the first to evaluate the mediating role of NSAID use in the association between OA and CVD based on population-based HAD. Our findings suggest that a noteworthy portion of the increased risk of CVD among people with OA is mediated through their NSAIDs use. Future research is necessary using prospective follow up to confirm these findings.

Chapter 3: The cardiovascular safety of non-steroidal anti-inflammatory drugs (NSAIDs) among individuals with osteoarthritis: Findings from real-world claims data

3.1 Introduction

Osteoarthritis (OA) is known to be associated with an increased risk for cardiovascular diseases (CVD), including ischemic heart disease (IHD), congestive heart failure (CHF) and stroke (4,9,10,16,53). Furthermore, a longitudinal study investigating the mediating role of non-steroidal anti-inflammatory drugs (NSAIDs) in the observed OA-CVD association reported that a substantial proportion of the total risk of CVD among OA patients compared to non-OA controls was attributable to NSAID use (1). Additionally, the use of NSAIDs are known to be associated with cardiovascular adverse effects and the risk varies among different NSAIDs (24–26). This is particularly worrisome as there is no cure for OA and NSAIDs are widely used in the treatment of OA patients (8).

Previous observational studies investigating the CVD risk of NSAIDs were primarily focused on specific CVD events such as myocardial infarction (MI) or stroke (39,41,42,47–50,65–68). For example, in a case-control study, Levesque et al. found an elevated risk of MI among elderly adults treated with an NSAID in Quebec, Canada (48). In a separate population based case-control study, Poza *et al.* uncovered a significant association between NSAID use and stroke (50). Moreover, most of these observational studies were focused on a very specific patient population such as the elderly or people enrolled in certain health insurance plans, thus limiting

generalizability to other populations. Although a few studies have attempted to evaluate the CVD risk of NSAIDs among OA patients, their findings are not generalizable due to significant limitations such as convenience sampling of OA patients and evaluating mostly IHD risk associated only with Cox-2 inhibitors (44,104). As such, the overall cardiovascular safety of various NSAIDs among people with OA is yet unknown. Together, the substantial contributing role of NSAIDs in the increased risk of CVD among OA patients, the enormous use of NSAIDs (8,22,23) and the demonstrated varying risk for specific CVD outcomes associated with various NSAIDs warrants an investigation into the overall CVD safety of different NSAIDs used in OA treatment. The objective of this study was to evaluate the risk of CVD associated with various NSAIDs that are used in treating OA patients in the real world.

3.2 Materials and Methods

3.2.1 Study population

We used linked health administrative data (HAD) of a previously assembled population-based cohort of 720,055 individuals living in British Columbia, Canada (4,69–73). The BC Ministry of Health and the BC Vital Statistics Agency approved access to and use of the data facilitated by Population Data BC (74) for this study. At the individual level, data on demographics, billing information for physician consultations, hospital data, and prescription dispensing records were available. Additionally, information on the drug identification number (DIN), dispensing date, quantity and the number of days supplied was available for each community-dispensed prescription regardless of funding source. To be included in the cohort, an individual had to meet the following criteria: (i) diagnosed with OA between January 1997 to December 2013 using a previously validated case definition (4); (ii) at least 20 years old; (iii) no CVD diagnosis before

their OA diagnosis (34); and, (iv) not received any NSAID prescription within the last 90 days from their OA diagnosis date (105–107). Individuals diagnosed with rheumatoid arthritis (RA) were excluded from the cohort prior to identifying OA patients (44). The Behavioural Research Ethics Board at the University of British Columbia, Canada reviewed and approved this study. All inferences, opinions, and conclusions drawn in this article are those of the authors, and do not reflect the opinions or policies of the Data Steward(s).

3.2.2 NSAID exposure definition

Exposure to NSAID was the independent variable in this study. We created an NSAID exposure variable using linked PharmaNet (69) data in a time-dependent fashion. Individuals were considered at risk for the duration of their NSAID prescription (66), defined as the start of the dispensing date and ending on the dispensing date in addition to the number of days supplied and a grace period of 15 days (108,109). In a sensitivity analysis, we varied the grace period from 0 and 30 days. If a new prescription of a different NSAID was dispensed during the duration of a previous prescription, we assumed that the patient had stopped taking the previous NSAID and started the newly prescribed NSAID on the dispensing date of the new prescription (45,108). If an individual received NSAIDs from different classes on the same day, the participant was excluded from the analysis (45,48,109,110). We grouped NSAIDs into overall (i.e., any NSAID use) and one of four categories: (i) coxibs, (ii) naproxen, (iii) ibuprofen, and (iv) other conventional NSAIDs (47,48,105).

3.2.3 Outcome variable

In line with the study's objective to evaluate overall CVD safety, we used a composite CVD event as the outcome (4,108). Specifically, we identified study outcome using a definition similar to the previously validated definition of identifying CVD cases from HAD (80). We used ICD-9 or ICD-10 codes to identify CVD outcomes from the discharge abstract database (hospital separations) (71), medical services plan (MSP) payment information file (70) and vital statistics deaths data file (73). The flowing table contains the list of ICD codes used in this study.

Table 3-1 List of ICD-9 and ICD-10 codes used in this study

Variable	Type	ICD-9 codes	ICD-10 codes
CVD	Outcome	410-414, 428, 430-434, 436, 438	I20-I25, I50, I60-I64
Hypertension	Baseline confounder	401	I10
Hyperlipidemia	Baseline confounder	272	E78
Diabetes mellitus	Baseline confounder	250	E11
Chronic obstructive pulmonary disease (COPD)	Baseline confounder	490, 492, 494, 496	J40, J43, J44, J47
Peptic ulcer disease (PUD)	Baseline confounder	531-533	K25-K27
Chronic kidney disease (CKD)	Baseline confounder	583, 585, 586, 592, 593	N18
Chronic liver disease (CLD)	Baseline confounder	571	K76
Smoking status (surrogate variable using ICD-9 code and ICD-10 code for personal history of nicotine dependence)	Baseline confounder	V15.82	Z87.891
Obesity (surrogate variable using ICD-9 code for	Baseline confounder	2780	E65-E68

Variable	Type	ICD-9 codes	ICD-10 codes
unspecified obesity and ICD-10 code for overweight, obesity and other hyperalimentation)			
Alcoholism (surrogate variable using ICD-9 and ICD-10 codes)	Baseline confounder	291, 303	F10, K70

3.2.4 Covariates

We identified a number of variables that are associated with NSAID use and a risk factor for CVD by extensively reviewing the literature focusing on NSAID use and CVD relationship (40,81,82,84,111). Biological variables, age in years at cohort entry was recorded as a continuous variable and sex was recorded as male or female. Socio-economic status (SES) was recorded as a categorical variable with five categories, one to five in increasing order of SES. We assessed co-morbid disease conditions including hypertension, hyperlipidemia, diabetes mellitus, chronic obstructive pulmonary disease (COPD), and peptic ulcer disease (PUD) at baseline using ICD-9 or ICD-10 codes adopting a definition similar to that proposed by Tonelli *et al.* (80). We calculated the Romano comorbidity score at baseline for each subject and recorded them as a continuous variable (85).

3.2.5 Statistical analysis

We followed the study participants prospectively from the date of cohort entry until they developed a CVD incident or were censored due to emigration, death or study end date, whichever occurred first. Using a bivariate analysis, we modeled the relationship between each of the baseline covariables with the CVD outcome. If the inclusion of covariable lowered the

Akaike Information Criterion (AIC), we included that covariable into the model (93). We used time-dependent Cox regression analyses to estimate the risk of CVD associated with NSAID use overall and with the four unique groups of NSAIDs (i.e., coxibs, naproxen, ibuprofen and other conventional NSAIDs). Hazard ratio (HR) and 95% confidence interval (CI) were the measures of association in this study. We grouped the study participants according to the NSAID dispensing status: (i) individuals who received at least one NSAID prescription ("dispensed") and, (ii) individuals who did not receive any NSAID ("non-dispensed") during the entire follow-up period. We conducted the primary analysis including the dispensed individuals to obtain an unbiased estimate of the association between NSAID use and CVD (45,48,65). In a sensitivity analysis, we included all study participants (both NSAID dispensed and non-dispensed) (112).

3.3 Results

Our study sample was comprised of 5,070 people with OA, 3,806 of which were dispensed at least one NSAID prescription (dispensed). Approximately 25% of the study sample did not receive any NSAID prescription during the entire follow-up period (non-dispensed). Table 3-2 contains detailed information on the characteristics of the study sample by NSAID dispensing status.

Overall mean age at cohort entry date was 65 years. OA individuals who did not receive any NSAID were older compared to the ones who were dispensed an NSAID, the mean age was 68 versus 64 years, respectively. Overall, approximately 55% of the study sample was female. The study sample was evenly distributed among socioeconomic groups. On average, the non-dispensed OA individuals were more likely to have more co-morbidities compared to dispensed

OA individuals; the mean of Romano co-morbidity score at baseline was 0.35 versus 0.27 respectively. Similarly, the proportion of individuals with a history of all the baseline co-morbid disease conditions except hyperlipidemia were higher among non-dispensed OA individuals. For example, the proportion of people with hypertension was 39% in non-dispensed individuals versus 31% in dispensed individuals. For diabetes it was 12% versus 9%, respectively.

Table 3-2 Characteristics of study sample by non-steroidal anti-inflammatory drugs (NSAIDs) use status during the follow up period

Variable	Total N (%)	Individuals dispensed ≥ 1 NSAID prescription N (%)	Individuals did not receive any NSAID prescription N (%)
Study sample	5,070 (100%)	3,806 (75.07%)	1,264 (24.93%)
Cardiovascular diseases (CVD)			
No	3,410 (67.26%)	2,659 (69.86%)	751 (59.41%)
Yes	1,660 (32.74%)	1,147 (30.14%)	513 (40.59%)
Age at cohort entry in years (Mean ± SD)	64.95 (13.19)	63.91 (12.86)	68.09 (13.70)
Sex (% of female)	2,767 (54.58%)	2,092 (54.97%)	675 (53.40%)
Socioeconomic status (SES)			
1 (Lowest SES)	943 (18.60%)	685 (18.00%)	258 (20.41%)
2	924 (18.22%)	676 (17.76%)	248 (19.62%)
3	1,111 (21.91%)	838 (22.02%)	273 (21.60%)
4	1,034 (20.39%)	791 (20.78%)	243 (19.22%)
5 (Highest SES)	1,058 (20.87%)	816 (21.44%)	242 (19.15%)
Hypertension			
No	3,387 (66.80%)	2,622 (68.89%)	765 (60.52%)
Yes	1,683 (33.20%)	1,184 (31.11%)	499 (39.48%)
Hyperlipidemia			
No	4,691 (92.52%)	3,516 (92.38%)	1,175 (92.96%)
Yes	379 (7.48%)	290 (7.62%)	89 (7.04%)
Diabetes			
No	4,588 (90.49%)	3,479 (91.41%)	1,109 (87.74%)

Variable	Total N (%)	Individuals dispensed ≥ 1 NSAID prescription N (%)	Individuals did not receive any NSAID prescription N (%)
Yes	482 (9.51%)	327 (8.59%)	155 (12.26%)
Chronic obstructive pulmonary disease (COPD)			
No	4,541 (89.57%)	3,424 (89.96%)	1,117 (88.37%)
Yes	529 (10.43%)	382 (10.04%)	147 (11.63%)
Peptic ulcer disease (PUD)			
No	4,862 (95.90%)	3,656 (86.06%)	1,206 (95.41%)
Yes	208 (4.10%)	150 (3.94%)	58 (4.59%)
Romano comorbidity score (Mean ± SD)	0.29 (0.86)	0.27 (0.83)	0.35 (0.92)

There were 1,660 individuals who developed incident CVD over the study period of which 523 (31.50%) were diagnosed with ischemic heart disease, 532 (32.04%) were diagnosed with congestive heart failure (CHF) and 605 (36.44%) were diagnosed with stroke. Appendix E presents detail distribution of CVD outcomes by NSAID dispensing status. The proportion of individuals who developed CVD was substantially higher among non-dispensed individuals compared to dispensed individuals, 41% versus 30%.

Results from the primary analysis indicated that the risk of CVD associated with exposure to overall NSAID was 48% higher compared to unexposed person-time. After adjusting for age, sex, SES, COPD, diabetes, hypertension, hyperlipidemia, PUD and Romano comorbidity score, the adjusted HR and 95% CI from the time-dependent Cox regression model was 1.48 (1.27, 1.73). Compared to unexposed person-time, coxibs and naproxen were associated with similar risk of CVD followed by other conventional NSAIDs and ibuprofen, adjusted HR (95% CI) was

1.58 (1.24, 2.01), 1.58 (1.11, 2.24), 1.39 (1.10, 1.75) and 1.36 (0.75, 2.47) respectively (Table 3-3).

Table 3-3 Risk of CVD associated with various NSAIDs used in the treatment of OA

Exposure to NSAID	HR and 95% CI	
	Unadjusted	Adjusted*
Unexposed person-time	Reference	Reference
Overall NSAID	1.37 (1.17, 1.59)	1.48 (1.27, 1.73)
Coxibs	1.48 (1.16, 1.88)	1.58 (1.24, 2.01)
Naproxen	1.29 (0.91, 1.83)	1.58 (1.11, 2.24)
Ibuprofen	1.39 (0.77, 2.52)	1.36 (0.75, 2.47)
Other conventional NSAIDs	1.20 (0.95, 1.52)	1.39 (1.10, 1.75)

CVD: Cardiovascular disease

NSAIDs: Non-Steroidal Anti-Inflammatory Drugs

OA: Osteoarthritis

*Adjusted for age, sex, socioeconomic status, chronic obstructive pulmonary, diabetes, hypertension, hyperlipidemia, peptic ulcer disease and Romano comorbidity score

Results from the sensitivity analysis including all OA individuals in the study sample indicated that the CVD risk associated with exposure to overall NSAID was 16% higher compared to

unexposed person-time, the adjusted HR (95% CI) was 1.16 (0.99, 1.34). Although the directionality of CVD risk associated with various NSAID groups was similar to that observed in the primary analysis, the magnitude of association was substantially less. For example, when non-dispensed OA individuals were included in the analysis, the risk of CVD associated with exposure to coxibs versus unexposed person-time was only 24% compared to the 58% increased risk estimated from the primary analysis. We observed similar attenuation in the CVD risk associated with exposure to other groups of NSAIDs. The following table contains detailed results from the sensitivity analysis.

Table 3-4 Results from sensitivity analysis: effect of including OA individuals who were not dispensed an NSAID during follow up period.

Exposure to NSAID	Analysis	
	Primary	Sensitivity
Unexposed person-time	Reference	Reference
Overall NSAID	1.48 (1.27, 1.73)	1.16 (0.99, 1.34)
Coxib	1.58 (1.24, 2.01)	1.24 (0.98, 1.58)
Naproxen	1.58 (1.11, 2.24)	1.18 (0.84, 1.67)
Ibuprofen	1.36 (0.75, 2.47)	1.05 (0.58, 1.91)
Conventional NSAID	1.39 (1.10, 1.75)	1.09 (0.87, 1.37)

3.4 Discussion

This study is the first retrospective cohort study using BC HAD that has investigated the overall CVD safety of various NSAIDs used to treat OA in the real world. Our findings indicate that the

risk of CVD associated with exposure to NSAIDs was significantly higher compared to unexposed person-time among OA patients who were dispensed at least one NSAID prescription during the follow up period. Considering exposure to different classes of NSAIDs compared to unexposed person-time, we found that coxibs and naproxen were associated with similar CVD risk which was higher than that of other conventional NSAIDs including ibuprofen.

In general, findings from previous research indicate that NSAID use increases the risk of CVD. Thus, our primary finding of a 48% increased risk of CVD associated with NSAID use is concordant with the literature. Previous studies have estimated the CVD risk associate with NSAID use either by comparing with non-NSAID use or by directly comparing exposure to different types of NSAIDs. For example, in a meta-analysis of studies based on Canadian and European HADs comparing the risk of acute MI with NSAID use, Bally *et al.* reported that the odds of MI were higher among current NSAID users compared to non-users (defined by no NSAID within last 365 days from index date) (41). The OR (95% CI) was 1.24 (0.91, 1.82), 1.58 (1.07, 2.17), 1.53 (1.07, 2.33), 1.48 (1.00, 2.26) and 1.50 (1.06, 2.04) for celecoxib, refecoxib, naproxen, ibuprofen and diclofenac, respectively (41). In a nested case-control study using data from UK GPRD, Andersohn *et al.* observed that the current use of coxibs and diclofenac significantly increased the risk of acute MI compared to no NSAID use within one year from the MI event date (47). The increased risk of MI associated with celecoxib and diclofenac was 56% and 37%, respectively (47).

In a cohort study involving older adults with arthritis, Solomon *et al.* found that coxibs were associated with a 28% higher risk of CVD events when compared to non-selective NSAIDs (105). In a sub-group analysis of Prospective Randomized Evaluation of Celecoxib Integrated

Safety versus Ibuprofen Or Naproxen (PRECISION) trial, Solomon et al. (113) compared the cardiovascular, gastrointestinal and renal safety of celecoxib with ibuprofen and naproxen among OA patients. The risk of composite CVD outcome consisting of non-fatal MI or stroke and cardiovascular death was not different among OA patients exposed to celecoxib and naproxen (113). Our study finding of similar CVD risk associated with exposure to coxibs and naproxen is consistent with this finding. The minor differences in the estimated CVD risk among different studies is likely due to the differences in study design (cohort/RCT/case-control), the outcome studied (composite CVD versus MI or stroke), study population and the way NSAID exposure variable was defined based on the study objective. For example, PRECISION is a non-inferiority trial comparing risk of adverse drug reactions of coxibs with two commonly used NSAIDs, ibuprofen and naproxen. The intent-to-treat (ITT) analysis censored at 30 months (113). In contrast, the time-dependent NSAID exposure in this study estimated the CVD risk associated with various NSAIDs used to treat OA compared to non-exposure person-time over a long follow-up period. Almost all previous observational studies investigating NSAID-CVD association included individuals who received at least one NSAID prescription within a pre-defined calendar period (45,48,65). Our primary analysis involving OA individuals who were dispensed at least one NSAID prescription during the follow up period is comparable to existing literature.

It has been noted that older individuals with multiple co-morbidities are usually advised not to take NSAIDs because of the risk of bleeding or other contraindications (112). This selective under-use of drugs by elderly patients may lead to an artificial protective association in NSAID users (112). Accordingly, in a study investigating the relationship between drug treatment and

mortality among older individuals, Glynn *et al.* found that NSAIDs were associated with a 26% reduction in death rates compared to individuals who did not fill a NSAID prescription within the last 120 days prior to hospitalization (112). The authors, however, concluded that the mortality benefit observed was an artificial association resulting from the reduced prescription of NSAID among older individuals (112). In this study, we observed that the OA individuals who did not receive any NSAID during follow up (non-dispensed) were older (68 versus 64 years of age) and had higher co-morbidities compared to the OA individuals who received NSAIDs (dispensed). Among dispensed individuals, the majority received their first NSAID prescription within two years from the follow-up start date, the mean time to receive the first NSAID prescription was 1.93 years. On the other hand, individuals in the non-dispensed group were followed up for an average of six years and were not dispensed any NSAID prescription. Including the non-dispensed individuals who are intrinsically at higher risk for CVD into the analyses would systematically contribute only to the unexposed person-time leading to attenuation of the true NSAID-CVD association. In line with the previous study findings by Glynn *et al.* (112), our sensitivity analysis including the non-dispensed individuals resulted in substantial decrease in the CVD risk associated with NSAID use among individuals with OA. The directionality of CVD risk associated with various NSAIDs remained similar.

This study has several strengths. We used population-level HAD from British Columbia, Canada and we modeled exposure to NSAID in a time-dependent fashion. This enabled us to account for the effect of all NSAID prescriptions dispensed during the entire follow-up period on the CVD outcome. A grace period, usually three to five times of the plasma half-lives of the drug molecule, is recommended to separate the treatment from other treatments (106). The half-lives

of NSAIDs are usually short to medium, ranging between 1 to 14 hours (107). Thus, the grace period of 15 days used in creating the NSAID exposure variable was more than adequate and consistent with the existing literature (108,109). Change in grace period from 0 and 30 days did not result in substantial variation in the risk of CVD associated with exposure to NSAID. Also, the directionality of higher risk among the unique NSAID groups was similar for all three grace periods. Another strength of this study is that, unlike previously published studies in which a specific CVD event (e.g., MI or stroke) was identified based on hospital admission, we identified the composite CVD outcome using a case definition previously validated by Tonelli *et al.*, who used HAD from Alberta, Canada (80). As such, findings from this study provide new knowledge on the overall CVD safety of different NSAIDs that are prescribed to people with OA in the real world. For individuals who received different classes of NSAIDs on the same day, it was not possible to know which NSAID the individual took or if they had taken both; thus, we excluded those individuals (45,48,109,110). We also excluded individuals diagnosed with RA (44). This makes the study findings more relevant to the treatment of OA patients with NSAIDs.

Similar to other studies using a linked health administrative database to investigate CVD risk with NSAID use, one of the limitations of our study is that the information on over-the-counter (OTC) NSAID was not available in BC HAD (4). Although, ibuprofen and naproxen are available in Canada as OTC medication, prescription use of NSAIDs usually involves much higher doses. It is also assumed that people who use prescription NSAIDs use less OTC than others. We created the NSAID exposure variable based on prescription dispensing records (PharmaNet data). So, the study findings are more relevant to prescription use of NSAIDs than OTC use. Another point noted is that in the time-dependent NSAID exposure variable, an

individual was considered exposed to NSAID for the durations of NSAID prescriptions and unexposed to NSAIDs for the rest of follow-up period. The individuals were also allowed to switch between NSAIDs. Our models were not designed to make direct comparisons among different NSAIDs. Another limitation includes the unavailability of information on confounding variables. HADs are usually collected for administrative purposes. As such, information on all confounding variables is not always available (105). For example, obesity and smoking are two important CVD risk factors for which information is not available in BC HAD. We created three surrogate variables for smoking (44), obesity (5) and alcoholism (5) using ICD codes. Additionally, we assessed a few other baseline co-variables, such as chronic kidney disease and chronic liver disease. We also assessed the use of low-dose aspirin within the last 120 days of cohort entry date (114). Overall, the proportion of individuals in the surrogate variables for obesity, smoking (history of tobacco use) and alcoholism were extremely low, ranging from 0.61% to 1.58%. Only a small proportion of individuals in the study sample had chronic kidney (4.02%) and liver disease (1.60%) at baseline. None of these variables contributed to the model and therefore were not adjusted for in the final model.

In conclusion, we estimated the overall cardiovascular safety and more specifically, the safety profile of different NSAIDs used in treating OA in the real world. We expect that this new knowledge on treatment choices with NSAID in terms of overall CVD safety will help improve clinical management of OA patients.

Chapter 4: Using external data to estimate unmeasured confounders: A plasmode simulation study comparing alternative approaches to impute body mass index in a study of the relationship between osteoarthritis and cardiovascular disease

4.1 Introduction

Administrative databases are increasingly being used to undertake epidemiological research. As these data are primarily collected for administrative purposes, such as physician billings or hospital utilization tracking, researchers typically do not have any say over the type of information collected. Hence, administrative databases often do not include information on all the variables necessary to answer specific research questions. Specifically, data on potential confounders is often not collected. Consequently, not being able to adjust for potential confounders is a major challenge while using administrative databases (43). For example, being obese or overweight and the related construct of body mass index (BMI) are an important risk factor for many diseases (78,81,82,84,115). However, BMI information is usually not included in administrative databases, including those in British Columbia (BC). To overcome this limitation, in a longitudinal study investigating the association between osteoarthritis (OA) and cardiovascular diseases (CVD) using BC administrative databases, Rahman et al. imputed BMI categories for all study participants using data from another source, namely the Canadian Community Health Survey (CCHS) (4). The imputation was done randomly in accordance with the proportions observed among individuals in CCHS who were grouped based on the OA exposure, CVD outcome and demographic variables (4). In a separate BC population-based study investigating the association between rheumatoid arthritis (RA) and diabetes mellitus,

Schmidt et al. used a similar imputation based on proportions of obesity among RA patients versus the general population (5). However, none of the studies attempted to measure the extent of potential bias or confounding resulting from BMI being imputed using the proportion-based imputation method (4,5).

A limitation of assigning a BMI category to an individual randomly according to a population-level proportion, as described in studies mentioned above (4,5), is that this does not account for the imprecision associated with the imputation technique employed and may introduce bias (87,116). Moreover, such imputation of BMI category of individuals in study data from administrative databases using population-level proportion does not account for individual-level heterogeneity. The resulting imputed values of BMI may be unrealistic for study individuals (e.g., may contradict the characteristics recorded in the administrative data). If, instead, we impute BMI values that are generated based on individual level covariates that are predictive of BMI, those BMI values would be much more realistic and consistent with individual's characteristics. Rubin et al. introduced the multiple imputation technique that has the potential to overcome these limitations of proportion-based imputation because of the following two essential features (117). First, multiple imputation accounts for the imprecision involved with the imputation, thus lessens bias and provides valid statistical inferences (87,116). Second, a multivariable multiple imputation model uses variables recorded at the individual level, and therefore, imputes the missing BMI value from a series of plausible values, had it not been missing (118).

The primary objective of this study was to investigate whether imputing an important study variable taking information from external data sources using multiple imputation instead of proportion-based imputation would result in less biased estimates of a given relationship. In the current context, we investigated the relationship between OA and increased risk of CVD as an example of a relationship in which BMI is an important confounding variable for which information is not available in administrative databases and could be imputed from an alternative source, such as a survey database. We hypothesize that imputing BMI at the individual level using the multiple imputation (117) provides more realistic BMI values and introduces less bias in the estimated model coefficients compared to coefficients from a model in which BMI categories are imputed using proportion-based imputation (4,5).

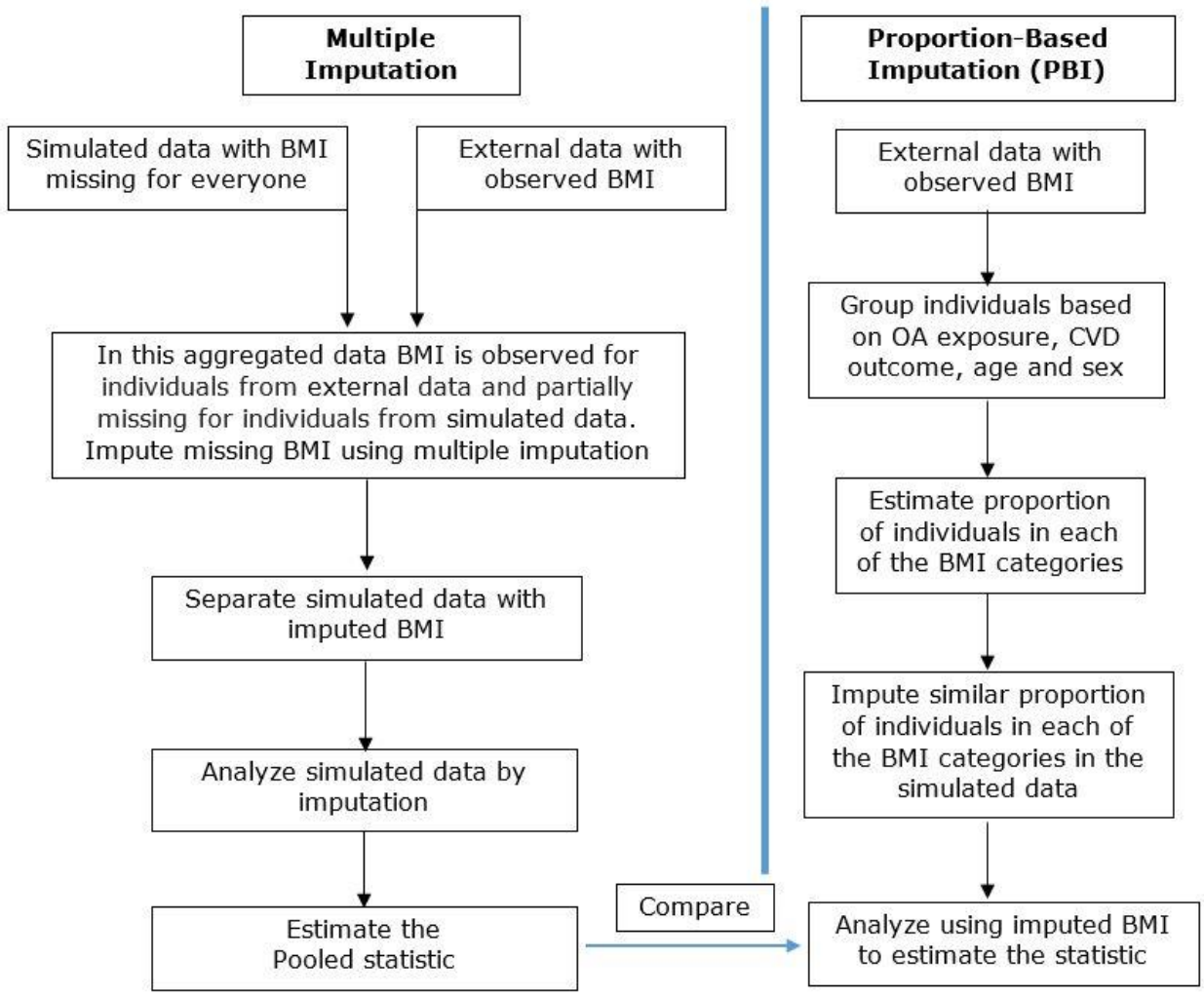


Figure 4-1 Conceptual framework to compare multiple imputation with PBI in imputing BMI variable missing for everyone in a study data using information from an external database

4.2 Materials and Methods

4.2.1 Data sources

4.2.1.1 Study dataset

To adequately assess the effect on bias and uncertainty associated with imputation of missing information on BMI, particularly in administrative data, we needed a dataset in which BMI is at least partially recorded at the individual level. Since administrative databases do not contain information on BMI, prospectively collecting BMI information for a large sample from administrative data would be both time and cost restrictive. Instead, we used data from the

CCHS, a large national health survey data representing approximately 98% of the Canadian population. After starting in 2001, the CCHS was repeated every two years until 2005 collecting information on a large number of variables from approximately 130,000 respondents. We created a study dataset from CCHS cycle 3.1 (2005) (henceforth, Data.1) (77,119,120). Specifically, Data.1 had complete information (no missing values) on OA exposure, CVD outcome and potential confounding variables including BMI. We created a dichotomous explanatory variable of OA using responses of two CCHS questions (121). The first question, administered to all respondents, was "Do you have arthritis or rheumatism?" and the second question, which was only asked of respondents who answered yes to the first question, was "What type of arthritis?". As respondents were not asked this second question in CCHS cycles after 3.1, formed the rationale for using CCHS cycle 3.1 for our study. The dichotomous outcome variable for CVD was obtained directly from a survey question 'Do you have heart disease?' that was asked of all respondents. The BMI was recorded as a categorical variable with underweight, normal weight, over weight and obese categories. Responses such as 'Don't Know', 'Refusal' or 'Not Stated' were excluded.

We then created 500 simulated datasets (Data.2) from the Data.1 using plasmode simulation, a previously used approach to assess biases in epidemiologic studies using administrative databases (122). Ethics approval for this study using publicly available CCHS data was covered by item 7.10.3 in University of British Columbia's Policy #89: Research and Other Studies Involving Human Subjects (2) and Article 2.2 in of the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans (TCPS 2) (3).

4.2.1.2 Simulated datasets using plasmode simulation

At first step we fitted a multivariable logistic regression model using Data.1 (N=84,452). In this model the exposure was OA, outcome was CVD, and age, sex, level of education, household income level, physical activity index, smoking status, diabetes, hypertension and BMI category were entered as the confounding variables. We then used resampling with replacement from Data.1. We did not modify the OA exposure status and any of the co-variables during sampling (122). As such, the associations among the study variables remained unchanged in the sampled populations. We used the following outcome generation model:

$$\text{Logit}\,[\text{Prob}(Y = 1)] = \beta_0 + \beta_1 Z + \beta_2 X \quad \ldots\ldots\ldots\ldots \qquad \text{(eq. 1)}$$

where Y is the outcome (e.g., CVD), beta0 is the intercept, beta1 is the effect of exposure (OA), Z is the exposure indicator (whether the respondent has OA or not), X is the covariate matrix, beta2 is a vector that includes effects associated with covariates (X = age, sex, level of education, household income level, physical activity index, smoking status, diabetes, hypertension and BMI category). We replaced the Odds Ratio (OR) with 1.60 (estimated from the Data.1 in the first step). As such, we knew the true OR of 1.60 for all the simulated datasets upfront. We applied this outcome-generating model to the OA exposure and covariate data sampled with replacement from the study dataset. Finally, we created a binary CVD outcome status for each subject using the probability of outcome obtained from the outcome-generating model. We created the simulated datasets using this plasmode simulation technique starting from the sampling of individuals (122). After each iteration, we monitored the average of the estimated OR that was stabilized at 35^{th} iteration and did not change further. We created 500 simulated datasets (Data.2) (n=75,000).

4.2.1.3 Setting missing information for BMI

We set the BMI variable missing for everyone in copies of the 500 simulated datasets (Data.3). This Data.3 mimics administrative data in which BMI is not recorded and missing for everyone in the database.

4.2.1.4 External survey data

Similar to previously published proportion-based imputation method (4), we created a large dataset (Data.4) by compiling data from CCHS cycles 1.1 (2001) (75) and 2.1 (2003) (76). This Data.4 served as the external survey data in both multiple imputation and proportion-based imputation methods. Table 4-1 summarizes the different datasets that are used in this study.

Table 4-1 Description of various datasets used to compare multiple imputation approach with proportion-based imputation method in imputing BMI variable missing for everyone in a database

Dataset	Description	Number of participants (N)	Comments
Data.1	Study dataset created from CCHS cycle 3.1 (2005)	84,452	Complete information on all variables including BMI
Data.2	500 simulated datasets created from study dataset using plasmode simulation	75,000	Complete information on all variables including BMI
Data.3	Copies of 500 simulated datasets but setting BMI as missing for everyone in the datasets	75,000	Missing information for BMI for everyone in the dataset Complete information on other variables
Data.4	External survey data created by compiling data from CCHS cycles 1.1 (2001) and 2.1 (2003)	149,810	Complete information on all variables including BMI

4.2.2 Multiple imputation

In multiple imputation, multiple copies, usually three to five, of complete datasets are created by imputing the missing value (123). Each of the complete datasets is then analyzed separately using an appropriate statistical method. Finally, the estimates from each of the complete datasets are combined using Rubin's rules to produce a single, pooled estimate (87). The type of distribution under which a missing value will be imputed is an important consideration in selecting the multiple imputation model (124). Unlike Markov Chain Monte Carlo method which assumes a joint multivariate normal distribution among all variables entered into the imputation model, the fully conditional specification method uses a separate conditional distribution to impute the missing variable (124). In addition, fully conditional specification method is advantageous because it allows selecting the imputation model based on the type of the missing variable. In this study, we aggregated data by setting Data.3 under Data.4. In this aggregated data BMI was observed for individuals from Data.4 and missing for individuals from Data.3. We implemented multiple imputation (with a number of imputations = 5) by PROC MI in SAS (version 9.4) using fully conditional specification logistic regression (118). We used information on age, sex, OA and CVD in the imputation model.

4.2.3 Proportion-based imputation

We grouped individuals in Data.4 based on OA, CVD, 10-year age category and sex. Within each group we calculated the proportions of individuals in each of the four BMI categories. Finally, individuals in Data.3 were grouped based on OA, CVD, 10-year age category and sex; and then similar proportions of BMI categories were imputed.

4.2.4 Analytic approach

We carried out the analysis in two steps. First, we implemented multiple imputation and proportion-based imputation to impute the missing BMI variable in Data.3. We then analyzed the imputed datasets and compared results with the known values estimated from Data.2. Based on the recommendations provided by Ratitch et al. (116) and Liu et al. (118) in evaluating the performance of imputation methods, we compared the proportion of individuals in each of the four BMI categories, and the ORs estimated from the multivariable logistic model. In this model, CVD outcome was regressed on OA exposure adjusting for age, sex, physical activity index, level of education, household income level, smoking status, diabetes, hypertension and BMI category.

To evaluate the performance of the plasmode simulation approach adopted in this study, we first estimated the proportion of individuals in each of the four BMI categories within each of the 500 simulated datasets (Data.2). Then we calculated the average of the proportions in each BMI category and compared that with the proportions observed in Data.1. We also fitted the multivariable logistic regression model in each of the 500 simulated datasets (Data.2) separately and calculated the average of the 500 ORs. The 95% confidence interval (CI) of the averaged OR was calculated by the percentile method. We compared this average OR from Data.2 with the OR estimated from Data.1.

In multiple imputation method, we created five complete datasets for each of the 500 simulated datasets (Data.3) by imputing the missing BMI. We fitted the multivariable logistic regression model in each of the complete datasets separately and combined the ORs using Rubin’s rule to

obtain 500 pooled ORs. Finally, the average of the 500 pooled ORs was compared with the average of the 500 ORs estimated from Data.2. Rubin's rule in pooling the ORs accounted for the uncertainty associated with the imputed BMI (125). This method takes input of point estimates and standard errors from multiple imputed datasets and then generate a pooled estimate with an overall confidence interval. Rubin's rule assumes that the estimated statistics are approximately normally distributed. Similar to the method proposed by Ratitch et al. (116), we applied log transformation to normalize the ORs. After combining, the pooled OR was back-transformed to its original log scale (116).

After imputing BMI using proportion-based imputation method, we estimated the average of the proportions in each BMI category observed in 500 simulated datasets (Data.3). We also analyzed the imputed datasets separately using the multivariable logistic regression model and calculated the average of the 500 ORs and 95% CI.

4.3 Results

The study data (Data.1) contained 84,452 survey respondents including 11,489 respondents with OA exposure and 4,963 respondents with CVD outcome. Table 4-2 presents the characteristics of the overall study sample by OA exposure status. The proportion of females among individuals with OA was substantially higher compared to individuals without OA, 71% versus 50%. The proportion of individuals with OA was consistently higher among all groups over 50 years of age. For example, in the age group of 60-69 years, the proportion of individuals was 27% and 11% among people with OA and without OA, respectively. The prevalence of obesity was higher among individuals with OA. A substantial proportion (58%) of individuals were physically

inactive among people with OA compared to non-OA individuals (48%). The prevalence of co-morbid disease conditions, such as diabetes and high blood pressure, was significantly higher among people with OA. Individuals without OA appeared to have higher education and income level compared to people with OA.

Table 4-2 Characteristics of study sample (n= 84,452) for the study data (Data.1) created using data from CCHS cycle 3.1 (2005) according to osteoarthritis status

Variable	N (%)	With osteoarthritis	Without osteoarthritis
Study sample	84,452 (100%)	11,489 (13.60%)	72,963 (86.40%)
CVD status			
No	79,489 (94.12%)	9,750 (84.86%)	69,739 (95.58%)
Yes	4,963 (5.86%)	1,739 (15.14%)	3,224 (4.42%)
Sex			
Female	44,455 (52.64%)	8,171 (71.12%)	36,284 (49.73%)
Male	39,997 (47.36%)	3,318 (28.88%)	36,679 (50.27%)
Age category			
20-29 years	13,518 (16.00%)	150 (1.31%)	13,368 (18.32%)
30-39 years	16,726 (19.81%)	413 (3.59%)	16,313 (22.36%)
40-49 years	16,233 (19.22%)	1,091 (9.49%)	15,142 (20.75%)
50-59 years	15,401 (18.23%)	2,712 (23.61%)	12,689 (17.39%)
60-69 years	11,483 (13.59%)	3,094 (26.93%)	8,389 (11.49%)
70-79 years	7,556 (8.95%)	2,644 (23.01%)	4,912 (6.73%)
≥ 80 years	3,535 (4.19%)	1,385 (12.06%)	2,150 (2.95%)
BMI category			
Underweight	1,887 (2.23%)	231 (2.01%)	1,656 (2.27%)
Normal weight	37,868 (44.84%)	4,127 (35.92%)	33,741 (46.24%)
Overweight	29,660 (35.12%)	4,208 (36.63%)	25,452 (34.88%)
Obese	15,037 (17.81%)	2,923 (25.44%)	12,114 (16.60%)
Physical activity index			
Inactive	41,890 (49.60%)	6,621 (57.63%)	35,269 (48.34%)
Moderately active	22,121 (26.19%)	2,789 (24.28%)	19,332 (26.49%)
Active	20,441 (24.21%)	2,079 (18.10%)	18,362 (25.17%)
Highest level of education			
Less than secondary	15,730 (18.62%)	3,628 (31.59%)	12,102 (16.59%)
Secondary graduate	12,973 (15.36%)	1,658 (14.43%)	11,315 (15.51%)
Some post- secondary	6,545 (7.75%)	794 (6.91%)	5,751 (7.88%)
College or university degree	49,204 (58.26%)	5,409 (47.08%)	43,795 (60.02%)

Variable	N (%)	With osteoarthritis	Without osteoarthritis
Total household income			
Less than $30,000	22,384 (26.51%)	5,112 (44.49%)	17,272 (23.67%)
$30,000 to $49,999	19,166 (22.69%)	2,739 (23.84%)	16,427 (22.51%)
$50,000 to $79,999	21,505 (25.46%)	2,158 (18.78%)	19,347 (26.52%)
$80,000 or more	21,397 (25.34%)	1,480 (12.88%)	19,917 (27.29%)
Smoking			
Never smoked	26,094 (30.90%)	3,386 (29.47%)	22,708 (31.12%)
Former occasional	13,196 (15.62%)	1,613 (14.04%)	11,583 (15.88%)
Former daily	23,653 (28.00%)	4,250 (36.99%)	19,403 (26.59%)
Current occasional	4,211 (4.99%)	312 (2.72%)	3,899 (5.34%)
Daily smoker	17,298 (20.48%)	1,928 (16.78%)	15,370 (21.07%)
Diabetes			
No	79,465 (94.10%)	10,075 (87.69%)	69,390 (95.10%)
Yes	4,987 (5.90%)	1,414 (12.31%)	3,573 (4.90%)
High blood pressure			
No	68,711 (81.36%)	6,989 (60.83%)	61,722 (84.59%)
Yes	15,741 (18.64%)	4,500 (39.17%)	11,241 (15.41%)

Table 4-3 presents the proportion of individuals in each of the four BMI categories. Plasmode simulation appeared to produce simulated data that closely resembling the study data. The proportion of individuals in each of the BMI categories were comparable between Data.1 and Data.2; 2% versus 2% in underweight, 45% versus 46% in the normal weight, 35% versus 35% in the overweight and 18% versus 17% in the obese category. After imputing BMI categories missing for everyone in Data.3, both multiple imputation and proportion-based imputation methods underestimated the proportion of individuals in normal weight category (Table 4-3). Compared to the known proportion of 46% normal weight individuals in Data.2, the proportion was only 38% when imputed using proportion-based imputation. In contrast, multiple imputation produced a less biased proportion of 41% in this category. Proportion-based imputation substantially overestimated the proportion of obese individuals (25%) compared to the known proportion of 17% in Data.2. Whereas the proportion of obese individuals was similar, 17%

versus 17%, when BMI category was imputed using multiple imputation. Among the overweight individuals, although proportion-based imputation produced a proportion similar to that was observed in Data.2, multiple imputation overestimated the proportion by 3%.

Table 4-3 Comparison of proportion of individuals in each of the four BMI categories after imputing BMI for everyone in the simulated data with the BMI categories originally recorded by Statistics Canada in Canadian Community Health Survey cycle 3.1 (2005) data

BMI category	Data.1 (%)	Data.2 (%)	Data.3 (%)	
			BMI imputed using multiple imputation	BMI imputed using proportion-based imputation
Under weight	2.23	2.26	2.54	2.00
Normal weight	44.84	45.98	41.45	38.35
Over weight	35.12	34.93	38.66	34.77
Obese	17.81	16.83	17.34	24.87

Data.1: Study data created from CCHS cycle 3.1 (2005).
Data.2: Simulated data created from study data using Plasmode simulation.
Data.3: Copy of simulated data in which BMI was missing for everyone in the dataset.

Table 4-4 presents the adjusted ORs estimated using multivariable logistic regression. The odds of having CVD among people with OA was 1.60 times higher than that of among non-OA controls. The adjusted OR (95% CI) averaged after analyzing all the plasmode simulated datasets (Data.2) was 1.59 (1.36, 1.82), closely resembling the known OR of 1.60 estimated from Data.1. The BMI imputed using multiple imputation produced a less biased estimate of the OA-CVD association compared to the proportion-based imputation. The adjusted OR (95% CI) was 1.62 (1.39, 1.86). In contrast, BMI category imputed by proportion-based imputation resulted in an

overestimate of the OA-CVD association, the adjusted OR (95% CI) was 1.66 (1.41, 1.90), much higher than that of observed in Data.2.

Table 4-4 Comparing adjusted* ORs estimated from multivariable logistic regression models

Data		Data.1	Data.2	Data.3	Data.3
Missing BMI category		None	None	BMI missing for everyone in the data	BMI missing for everyone in the data
Imputation method used		None	None	Multiple imputation	Proportion-based imputation
Adjusted OR*	Without BMI	1.62 (1.51, 1.73)	1.63 (1.19, 2.23)	1.63 (1.19, 2.23)	1.63 (1.19, 2.23)
	Including BMI	1.60 (1.49, 1.71)	1.59 (1.36, 1.82)	1.62 (1.39, 1.86)	1.66 (1.41, 1.90)

* adjusted for age, sex, physical activity index, education and income level, smoking status, diabetes and hypertension.

Data.1: Study data created from CCHS cycle 3.1 (2005).

Data.2: Simulated data created from study data using plasmode simulation.

Data.3: Copy of simulated data in which BMI was missing for everyone in the dataset.

In a sensitivity analysis, we compared the existing multiple imputation model with a large multiple imputation model. In the later model we used information on other covariables including the level of education, household income level, physical activity index, smoking status, diabetes and hypertension in addition to the age, sex, OA and CVD. Both the existing and large multiple imputation models produced similar proportions in each of the four BMI categories. In multivariable logistic regressions, BMI categories imputed by existing and large multiple

imputation models resulted in similar point estimates of OA-CVD association. The adjusted OR (95% CI) was found to be 1.62 (1.39, 1.86) and 1.62 (1.38, 1.85) when imputed BMI from the existing and large multiple imputation models were entered into the multivariable logistic regression models, respectively.

4.4 Discussion

In the current work, we found that multiple imputation approach performed better than the proportion-based imputation method that has been previously used in imputing important variable in the studies based on administrative database. Although a few observational studies attempted to account for missing BMI by additionally accessing survey data, the effect of the imputed BMI in estimating an unbiased measure of exposure-outcome association remained unknown (4,5). To the best of our knowledge, this is the first plasmode simulation-based study comparing multiple imputation with proportion-based imputation in imputing BMI category, a confounding variable in the association between OA and CVD. After imputing BMI category missing for everyone in the simulated data using information from external population-level survey data, compared to proportion-based imputation multiple imputation produced proportions closer to the known proportions across the BMI categories except overweight individuals. Also, BMI imputed using multiple imputation resulted in substantially less bias in the OA-CVD association. In contrast, the proportion-based imputation overestimated the association compared to the known point estimate.

Multiple imputation is superior to proportion-based imputation method from the logistical aspect. In proportion-based imputation individuals in both external survey data and study sample are grouped using exposure, outcome, age and sex variables. It is possible that the study sample may

contain a group of individuals that were not observed in survey data. Consequently, there would be no known proportion of BMI category from survey data to be assigned to that particular group of individuals in the study sample. For example, we grouped individuals in both external survey data (Data.4) and simulated data (Data.3) using OA exposure, CVD outcome, age and sex variables. In this study, the external data did not represent men aged 20 to 30 years who had both OA and CVD. In contrast, simulated datasets contained some individuals from that group (0.03% to 0.06%). When BMI was imputed using proportion-based imputation, the model assigned everyone in this group to obese category since there was no known proportion for the first three BMI categories of underweight, normal weight and overweight. In a sensitivity analysis, we excluded these individuals from the simulated data for whom information was not available in the external data and the results did not change much. This result is not surprising as the proportion of individuals excluded were negligible, approximately 0.05% of the simulated datasets. In addition, assigning population level proportions in simulated data may result in a fractional number of individuals in each BMI category. Since an individual can be part of only one of the four mutually exclusive BMI categories, fractions of individuals were rounded to the closest full number. This rounding up in each BMI category within a number of groups, for example, 56 in this study, may also contribute to the variation in overall proportions in each BMI category. In contrast, the fully conditional specification method of multiple imputation used in this study is advantageous because it enables one to fit an appropriate imputation model when the multivariable distribution cannot be drawn for the missing values (126). This has been reflected in imputing the BMI for men aged 20 to 30 years who had both OA and CVD. Multiple imputation imputed missing BMI in Data.3 with very similar proportions that was observed in corresponding simulated data.

Although BMI imputed using both of the imputation methods overestimated the adjusted OR for the OA-CVD association, compared to proportion-based imputation multiple imputation introduced much less bias in the point estimate. Perhaps this can be explained by the inter-relationship among the numerous variables that are recorded at the individual level. Franklin et al. reported that the variables recorded for each individual have a unique covariance structure that can serve as a proxy for unmeasured confounders and be used to eliminate bias (122). It is more likely that the BMI category imputed by multiple imputation using individual-level information would fit in the existing co-variate structure better than the BMI category imputed using population-level proportions. Consequently, the BMI category from multiple imputation method would perform more realistically in the multivariable outcome model, such as multivariable logistic regression employed in this analysis. The estimated ORs was not substantially different from each other after considering the overlap of the confidence intervals.

Other studies have evaluated multiple imputation as a tool for imputing missing values. In a plasmode simulation-based study, Liu et al. found that multiple imputation performed better than inverse probability weighting in handling missing data in marginal structural models (126). In a separate study using data from a case-cohort study, Muhlenbruch et al. concluded that multiple imputation was a valid method for imputing 90% missingness of a continuous variable of waist circumference (127). Our findings of the superiority of multiple imputation over proportion-based imputation were consistent with these study findings.

BMI is an important confounding variable in OA-CVD association (78,81,82,84,115). Although our objective was to compare the imputation methods in imputing missing BMI, we also looked

into the contribution of the imputed BMI into the outcome models. There is a notion that the variables used to impute the BMI variable already contain the information into the outcome model. As such, we wanted to see if the imputed BMI variable contributes to the model when added along with other co-variables. The effect of adjusting for BMI was small (1% to 4%). This appeared to be reasonable because BMI was entered as a covariable into a multivariable logistic regression model. It was expected that one covariable would not make a substantial change on the estimated OR. Our findings indicated that multiple imputation was superior because the effect of adjusting for BMI imputed using proportion based imputation was in the unexpected direction compared to that observed in both study data and simulated datasets.

There are a number of strengths in our study. We used publicly available data from a Canadian population-level health survey. Plasmode simulation enabled us to create 500 large (N=75,000) datasets to test the hypothesis in multiple data settings. In addition, plasmode simulation technique has the advantage of informing the true effect size set out during the data generation step that is useful in estimating the bias. One of the limitations of this study is that the plasmode simulation takes inspiration from a particular dataset to preserve realistic settings. We can be confident that the methods proposed here performs well for CCHS data under the settings we have evaluated. Future research using a variety of datasets is necessary to confirm the study findings. Another point to be noted is the cross-sectional nature of the health survey data. In longitudinal studies, individuals are prospectively followed for a long period of time. It is possible that the BMI of an individual may change over time. Imputation of BMI using survey data, either by multiple imputation or proportion-based imputation, cannot address this change in

BMI over time. Under such circumstances, the imputed BMI can potentially serve as a proxy for the baseline BMI.

In conclusion, our simulation study showed that multiple imputation approach introduced less bias in the association between OA and CVD than conventional proportion-based imputation when BMI was imputed for everyone in the data using information from external survey data. A simple multiple imputation model including OA exposure, CVD outcome and demographic variables of age and sex adequately imputed BMI that performed realistically in estimating the OA-CVD association. Researchers often face the challenge of missing variable. In imputing a study variable that is not recorded in a study data, multiple imputation is advantageous over imputation method based on population level proportions.

Chapter 5: Conclusion

This thesis focuses on and reveals the role of NSAIDs in the increased risk of CVD among people with OA. This concluding chapter presents the key findings from each of the three separate but related investigations and discusses the overall implications of the thesis and the research context. It also discusses the strengths and limitations of the research and provides some recommendations for future work in this area.

5.1 Key findings

The work undertaken in this thesis suggests that NSAID use by people living with OA affects the causal pathway that explains the OA-CVD relationship. Chapter 2's longitudinal study using health administrative data (HAD) from British Columbia (BC), Canada addressed a gap in the literature by examining the mediating role of NSAIDs in the increased risk of CVD among OA patients. The key study finding was that 41% of the increased risk of CVD among OA patients was mediated through their NSAID use. The implementation of a marginal structural model (57) in this study supports this interpretation of a causal relationship between NSAID use and CVD among OA patients.

A subsequent step in better understanding the influence of NSAIDs on CVD risk among OA patients involved identifying OA patients who received NSAIDs and evaluating whether this NSAID use resulted in adverse CVD outcomes. The retrospective cohort study in Chapter 3 used BC HAD to evaluate the risk of incident CVD events associated with different NSAIDs that are used in treating OA in the real-world settings. Findings indicate that exposure to NSAIDs was associated with a 48% increased risk of CVD that was independent of age, sex, SES, Romano

comorbidity score, and specific co-morbidities including COPD, diabetes, hypertension, hyperlipidemia, and peptic ulcer disease. In addition, when exposure to different groups of NSAID was compared with unexposed person-time, the risk of CVD varied among the different NSAIDs that are used to treat OA. Relative to unexposed person-time, Coxibs and naproxen may increase CVD risk more than conventional NSAIDs including ibuprofen. The study presented a comprehensive evaluation of the CVD adverse effect of NSAID use in treating OA.

The methodological approach deployed to investigate the above questions required considering the best available approaches available to impute/supplement data not typically available in HAD. Recent HAD-based epidemiological studies in the area of arthritis research imputed BMI for all study participants using a proportion-based imputation method. However, the potential bias of this imputation method had not been evaluated prior to this thesis. Thus the simulation-based study in Chapter 4 addressed an important knowledge gap by comparing the multiple imputation approach with the proportion-based imputation method. The study findings showed that BMI imputed using multiple imputation introduced less bias when evaluating the OA-CVD association than BMI imputed using the proportion-based imputation method. Our study also proposes a valid approach for imputing BMI for all members of a study cohort defined in an administrative dataset by accessing additional data from external sources.

5.2 Significance and implications of the research

A wide literature was reviewed to rationalize the research questions addressed in this thesis and to identify appropriate study designs and analytic approaches. The thesis thus contributes to a

number of research fields, namely rheumatology, pharmacoepidemiology, and health services research.

The first study to evaluate the mediating effect of NSAIDs in the OA-CVD relationship, Chapter 2 contributes to the rheumatology literature by describing how NSAIDs influenced CVD outcomes among OA patients from BC, Canada. People with OA were found to use substantially more NSAIDs than the non-OA control population, which may lead them to develop CVD. This evidence has important clinical implications. By demonstrating the substantial mediating role of NSAIDs in the increased risk of CVD among OA patients compared to non-OA controls, results highlight the importance of monitoring for cardiovascular adverse effects among OA patients. Kopec et al. reported a 3.3% decline in the proportion of new OA patients who were dispensed an NSAID prescription between 2000 and 2014 (128). Although it appears that NSAIDs are not used as first-line therapy in OA treatment, health care providers should still explain this increased risk of CVD to OA patients being treated with NSAIDs and may encourage them to try alternative treatment approaches, including exercise and physiotherapy. In addition to its significance for clinical practice, Chapter 2 also contributes to the pharmacoepidemiology literature on the application of marginal structural models in HAD-based observational studies in the field of OA research (51,63).

The retrospective cohort study in Chapter 3, which evaluated the risks of composite CVD adverse effects associated with NSAID treatment, contributes to rheumatology literature by describing the overall cardiovascular safety of different NSAIDs among OA patients. This study

also contributes to the pharmacoepidemiologic literature and adds to the evidence base describing the cardiovascular risks associated with exposure to NSAIDs.

The plasmode simulation-based study in Chapter 4, which compared two imputation approaches in imputing BMI, contributes to rheumatology literature by proposing a valid imputation approach to impute important study variables not recorded in observational data. Chapter 4 also contributes to pharmacoepidemiology literature on the application of plasmode simulation in evaluating bias in studies based on secondary data analysis (122,129–131). This study also adds to the evidence base for the superiority of the multiple imputation approach over other conventional imputation techniques (126,127).

5.3 Strengths and limitations of the research

The independent manuscripts in chapters 2 to 4 describe study-specific strengths and limitations. This section examines the strengths and limitations of the thesis as a whole. This discussion emphasizes the epidemiological studies conducted in chapters 2 and 3.

The application of novel methodologies in addressing all three research questions is a strength of this thesis. For example, marginal structural modelling was used in chapter 2 to break down the total effect of OA exposure on CVD outcomes, separating the direct effect of OA and indirect effect of OA mediated through NSAID use. Unlike conventional difference of coefficients approach in mediation analysis, the natural direct and indirect effects were estimated from a single weighted Cox regression model using the algorithm proposed by Lange et al. (57). This framework enables one to make interpretations of a potential causal relationship. In addition,

matching non-OA controls without replacement to each of the OA cases enriched the comparison group without duplication of control individuals in the study sample. Moreover, the OA cases (34), CVD outcomes and co-morbid disease conditions (80) were identified using case definitions previously validated using Canadian health administrative data, from British Columbia and Alberta, respectively. Rahman et al. reported that the administrative case-definition used in this study to identify OA cases had high validity with a higher than 5 positive likelihood ratio (34,132). The specificity was 91-100% and the positive predictive values (PPVs) ranged between 85 to 100% depending on reference standard (34,132). Tonelli et al. reported high validity of the case definition used to identify the CVD outcomes in this study having at least 70% PPV and sensitivity compared to the gold standard (80). Another strength is that we identified OA cases who did not have CVD prior to their OA diagnosis and then prospectively followed them for incident CVD outcomes. This nullified the possibility of reverse causality bias (34).

The time-dependent Cox regression analysis deployed in the retrospective cohort study design in chapter 3 is a methodological strength in this research area. Most of the previously published observational studies investigating the risk of CVD, namely myocardial infarction or stroke, used a case-control study design with odds ratio (OR) as the point estimate. Although the case-control study design is a valid approach and widely used in health research, it is somewhat difficult to interpret the OR in general terms. In contrast, the Cox regression analysis approach produces hazard ratio (HR) as the point estimate. Compared to OR, HR is straightforward to interpret as a direct measure of the CVD risk associated with the exposure to NSAID. Furthermore, the time-

dependent Cox regression analysis accounted for all prescriptions for various NSAIDs that were dispensed over the follow-up period.

The plasmode simulation approach implemented in chapter 4 was both time- and cost-efficient in comparing the two imputation approaches in multiple datasets. Since the health administrative datasets in Canada do not record information on BMI, prospectively collecting BMI information on large samples from multiple datasets may not be viable. Although computationally intense, plasmode simulation is capable of producing multiple datasets, for example 500 in chapter 4. A strength of plasmode simulation is that it is capable of keeping the data structure in the simulated datasets similar to the original data (122). Also, multiple imputation is well-accepted in the literature as a valid approach in handling missing data (123).

All research work in this thesis involved secondary analysis of population-level administrative data. The epidemiological studies in chapters 2 and 3 were based on health administrative data of a large population-based cohort from British Columbia. The ability to establish linkage between various datasets, for example medical services plan (MSP) payment information file, discharge abstract database (hospital separations), PharmaNet, consolidation file (MSP registration & premium billing) and vital statistics deaths data files makes BC HAD very valuable for pharmacoepidemiologic research. The long-term follow-up of individuals in the cohort makes the dataset suitable for longitudinal studies. Finally, findings from HAD-based studies are usually considered to have high external validity and therefore generalizable at the population-level. In chapter 4, the data for the simulation was drawn from the large sample size available in CCHS, another Canadian population-level health survey.

The limitation of any research based on HAD is also applicable to this thesis. In administrative databases data are primarily collected for billing purposes. As a result, research conducted using HAD faces several limitations. In pharmacoepidemiologic studies when a medication is available over-the-counter (OTC), using information from prescription claim records may result in underestimation of the drug exposure. In Canada, NSAIDs specifically ibuprofen and naproxen are available OTC. As such, the mediating effect of NSAID estimated in chapter 2 and the CVD risk associated with exposure to NSAIDs estimated in chapter 3 are conservative estimates. Worldwide, HADs are extensively used in pharmacoepidemiologic studies involving NSAIDs (24,45,48,105,108). In chapters 2 and 3, the strictest administrative case definition for OA was used to overcome the limitation of uncertainty around identifying the OA cases from HAD (34). The CVD outcomes were also identified using administrative case definitions previously validated by Tonelli et al. (80).

One of the limitations of the CCHS data used in chapter 4 is that the OA exposure and CVD outcomes were self-reported (88). However, in that study, in which the OA-CVD relationship was taken as an example to compare two imputation methods, the results of the imputed datasets were compared with the known values from the corresponding simulated datasets. Thus, the self-report of the OA exposure and CVD outcome did not have any impact on addressing the research question. Although discussed in detail in independent manuscripts, it is also important to note the limitation of potential unmeasured confounding in HAD-based epidemiological studies. The unavailability of a required variable is a challenge that researchers often face in adjusting for confounders in HAD-based studies (4,108). With all these limitations, both the BC HAD and

CCHS offer cost- and time-efficient options for population-level health research and have been extensively used in epidemiological research. This thesis contributes to the growing body of Canadian health research using administrative data.

5.4 Future research and recommendations

This thesis reexamined the OA-CVD relationship and estimated the effect of NSAID use, an important mediating factor, using currently available data [i.e., which do not include information on several important variables]. Future research should aim to address the limitations of current research. Future studies prospectively collecting information on OTC NSAID use, BMI, physical activity level, smoking status and family history of CVD will help in controlling for measured and unmeasured confounding. Other causal pathways namely decreased physical activity and chronic inflammation have also been hypothesized for the OA-CVD relationship (4,19–21). When information is available on all relevant study variables, future studies can estimate mediation through multiple mediator variables (63).

In Chapter 3, in order to ensure a sufficient number of study outcomes within each NSAID exposure group, different doses of NSAIDs were not differentiated, as in previous studies (108–110). Future prospective studies may investigate how the effect on CVD depends on both the duration and dose of NSAIDs. In addition, methodological work should be done to investigate how the effect of NSAIDs on CVD risk depends on the definition of the NSAID variable. This may be done using two different approaches: 1) within a single study design (e.g., longitudinal) current NSAID use may be compared to NSAID use defined as a time dependent variable; 2)

across study designs (e.g., nested case-control versus longitudinal) using the same study population the impact of each NSAID variable definition can be explored.

In the simulation-based study in chapter 4, comparing two imputation approaches was based on CCHS data. There are, however, other useful health survey data available in Canada, for example, the National Population Health Survey (NPHS) (133) and the Canadian Health Measures Survey (CHMS) (134). Future studies using data from various sources are necessary to confirm our findings.

5.5 Conclusion

This thesis suggests that NSAIDs play a substantial role in the development of CVD among people with OA. Furthermore, it suggests that not all NSAIDs used in treating OA in the real world are the same in terms of cardiovascular safety. After accounting for all prescriptions dispensed to individuals with OA, Coxibs and naproxen appeared to increase the risk of CVD more than conventional NSAIDs including ibuprofen. The knowledge gained has the potential to contribute to the better care of OA patients, and to help inform OA-related policies in BC and Canada.

References

1. Atiquzzaman M, Karim ME, Kopec J, Wong H, Anis AH. Role of Nonsteroidal Antiinflammatory Drugs in the Association Between Osteoarthritis and Cardiovascular Diseases: A Longitudinal Study. Arthritis Rheumatol. 2019;71(11):1835–43.

2. Board of Governors. Policy # 89: Research Involving Human Participants [Internet]. University of British Columbia; 2012 [cited 2014 Oct 14]. Available from: http://universitycounsel.ubc.ca/files/2012/06/policy89.pdf

3. Government of Canada IAP on RE. Interagency Advisory Panel on Research Ethics [Internet]. 2016 [cited 2018 Sep 30]. Available from: http://www.pre.ethics.gc.ca/eng/policy-politique/initiatives/tcps2-eptc2/chapter2-chapitre2/

4. Rahman MM, Kopec JA, Anis AH, Cibere J, Goldsmith CH. Risk of Cardiovascular Disease in Patients With Osteoarthritis: A Prospective Longitudinal Study: Cardiovascular Disease in Osteoarthritis Patients. Arthritis Care Res. 2013 Dec;65(12):1951–8.

5. Schmidt TJ. Cardiovascular disease prevention in rheumatoid arthritis : three population-based studies in British Columbia [Internet]. University of British Columbia; 2016 [cited 2017 Nov 30]. Available from: https://open.library.ubc.ca/cIRcle/collections/ubctheses/24/items/1.0305710

6. Wittenauer R, Smith L, Aden K. Priority Medicines for Europe and the World "A Public Health Approach to Innovation." Backgr Pap. 2013;31.

7. What is Osteoarthritis? [Internet]. arthritis.org. [cited 2019 Mar 6]. Available from: http://www.arthritis.org/about-arthritis/types/osteoarthritis/what-is-osteoarthritis.php

8. Government of Canada PHA of C. Life with Arthritis in Canada : A personal and public health challenge - Chapter Seven: Arthritis-related prescription medication use - Public Health Agency of Canada [Internet]. 2010 [cited 2016 Mar 21]. Available from: http://www.phac-aspc.gc.ca/cd-mc/arthritis-arthrite/lwaic-vaaac-10/9-eng.php

9. Wang H, Bai J, He B, Hu X, Liu D. Osteoarthritis and the risk of cardiovascular disease: a meta-analysis of observational studies. Sci Rep [Internet]. 2016 Dec [cited 2017 Sep 6];6(1). Available from: http://www.nature.com/articles/srep39672

10. Veronese N, Trevisan C, De Rui M, Bolzetta F, Maggi S, Zambon S, et al. Association of Osteoarthritis With Increased Risk of Cardiovascular Diseases in the Elderly: Findings From the Progetto Veneto Anziano Study Cohort. Arthritis Rheumatol. 2016 May 1;68(5):1136–44.

11. Haugen IK, Ramachandran VS, Misra D, Neogi T, Niu J, Yang T, et al. Hand osteoarthritis in relation to mortality and incidence of cardiovascular disease: data from the Framingham Heart Study. Ann Rheum Dis. 2015 Jan;74(1):74–81.

12. Tsuboi M, Hasegawa Y, Matsuyama Y, Suzuki S, Suzuki K, Imagama S. Do musculoskeletal degenerative diseases affect mortality and cause of death after 10 years in Japan? J Bone Miner Metab. 2011 Mar 1;29(2):217–23.

13. Nuesch E, Dieppe P, Reichenbach S, Williams S, Iff S, Juni P. All cause and disease specific mortality in patients with knee or hip osteoarthritis: population based cohort study. BMJ. 2011 Mar 8;342(mar08 2):d1165–d1165.

14. Haara M, Manninen P, Kroger H, Arokoski J, Karkkainen A, Knekt P, et al. Osteoarthritis of finger joints in Finns aged 30 or over: prevalence, determinants, and association with mortality. Ann Rheum Dis. 2003 Feb;62(2):151–8.

15. Kluzek S, Sanchez-Santos MT, Leyland KM, Judge A, Spector TD, Hart D, et al. Painful knee but not hand osteoarthritis is an independent predictor of mortality over 23 years follow-up of a population-based cohort of middle-aged women. Ann Rheum Dis. 2016 Oct 1;75(10):1749–56.

16. Rahman MM, Kopec JA, Cibere J, Goldsmith CH, Anis AH. The relationship between osteoarthritis and cardiovascular disease in a population health survey: a cross-sectional study. BMJ Open. 2013;3(5):e002624.

17. Kadam UT, Jordan K, Croft PR. Clinical comorbidity in patients with osteoarthritis: a case-control study of general practice consulters in England and Wales. Ann Rheum Dis. 2004 Apr 1;63(4):408–14.

18. Jonsson H, Helgadottir GP, Aspelund T, Eiriksdottir G, Sigurdsson S, Ingvarsson T, et al. Hand osteoarthritis in older women is associated with carotid and coronary atherosclerosis: the AGES Reykjavik study. Ann Rheum Dis. 2009 Nov 1;68(11):1696–700.

19. Fernandes GS, Valdes AM. Cardiovascular disease and osteoarthritis: common pathways and patient outcomes. Eur J Clin Invest. 2015 Apr 1;45(4):405–14.

20. Swift A. Osteoarthritis 1: Physiology, risk factors and causes of pain. [Review]. Nurs Times. 2012 Feb;108(7):12–5.

21. Osteoarthritis Symptoms [Internet]. arthritis.org. [cited 2019 Mar 6]. Available from: http://www.arthritis.org/about-arthritis/types/osteoarthritis/symptoms.php

22. Li Linda, Kopec Jacek, Cibere Jolanda, Esdaile John. British Columbia Osteoarthritis Survey [Internet]. British Columbia, Canada: A joint initiative of British Columbia Ministry of Health, The Arthritis Society, BC & Yukon Division, Arthritis Research

Centre of Canada; 2008 Jan [cited 2015 Jul 30]. Available from: http://arthritis.rehab.med.ubc.ca/files/2011/08/BCOASurvey.pdf

23. Lanas A, Garcia-Tell G, Armada B, Oteo-Alvaro A. Prescription patterns and appropriateness of NSAID therapy according to gastrointestinal risk and cardiovascular history in patients with diagnoses of osteoarthritis. BMC Med. 2011 Apr 14;9:38.

24. McGettigan P, Henry D. Cardiovascular Risk with Non-Steroidal Anti-Inflammatory Drugs: Systematic Review of Population-Based Controlled Observational Studies. PLoS Med [Internet]. 2011 Sep [cited 2016 Mar 21];8(9). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3181230/

25. Trelle S, Reichenbach S, Wandel S, Hildebrand P, Tschannen B, Villiger PM, et al. Cardiovascular safety of non-steroidal anti-inflammatory drugs: network meta-analysis. BMJ. 2011 Jan 11;342:c7086.

26. Kearney PM, Baigent C, Godwin J, Halls H, Emberson JR, Patrono C. Do selective cyclo-oxygenase-2 inhibitors and traditional non-steroidal anti-inflammatory drugs increase the risk of atherothrombosis? Meta-analysis of randomised trials. BMJ. 2006 Jun 3;332(7553):1302–8.

27. Sellam J, Berenbaum F. The role of synovitis in pathophysiology and clinical symptoms of osteoarthritis. Nat Rev Rheumatol. 2010 Nov;6(11):625–35.

28. Anti-Inflammatory Medications | Arthritis Society [Internet]. [cited 2019 May 19]. Available from: https://arthritis.ca/treatment/medication/anti-inflammatory-medications

29. CalmX some as, Artist W an E, Director F, producer, Creator VGC, inventors freelance writer for some 18 years S specialized in writing about, et al. Hippocrates Discovered Aspirin (He Just Didn't Know It) [Internet]. ThoughtCo. [cited 2019 May 19]. Available from: https://www.thoughtco.com/history-of-aspirin-4072562

30. Rao P, Knaus EE. Evolution of Nonsteroidal Anti-Inflammatory Drugs (NSAIDs): Cyclooxygenase (COX) Inhibition and Beyond. J Pharm Pharm Sci. 2008 Sep 20;11(2):81.

31. Vane JR. Inhibition of Prostaglandin Synthesis as a Mechanism of Action for Aspirin-like Drugs. Nature New Biol. 1971 Jun;231(25):232.

32. Government of Canada HC. Drug Product Database Online Query [Internet]. 2012 [cited 2019 Jul 11]. Available from: https://health-products.canada.ca/dpd-bdpp/index-eng.jsp

33. Bauman AE. Updating the evidence that physical activity is good for health: an epidemiological review 2000–2003. J Sci Med Sport. 2004 Apr 1;7(1, Supplement 1):6–19.

34. Rahman MM. The epidemiology of osteoarthritis and its association with cardiovascular disease and diabetes. University of British Columbia; 2014.

35. Macera CA, Hootman JM, Sniezek JE. Major public health benefits of physical activity. Arthritis Care Res. 2003;49(1):122–8.

36. Pasceri Vincenzo, Yeh Edward T. H. A Tale of Two Diseases. Circulation. 1999 Nov 23;100(21):2124–6.

37. Liuzzo G, Goronzy JJ, Yang H, Kopecky SL, Holmes DR, Frye RL, et al. Monoclonal T-cell proliferation and plaque instability in acute coronary syndromes. Circulation. 2000 Jun 27;101(25):2883–8.

38. Choi HK, Hernán MA, Seeger JD, Robins JM, Wolfe F. Methotrexate and mortality in patients with rheumatoid arthritis: a prospective study. The Lancet. 2002 Apr 6;359(9313):1173–7.

39. Varas-Lorenzo C, Riera-Guardia N, Calingaert B, Castellsague J, Salvo F, Nicotra F, et al. Myocardial infarction and individual nonsteroidal anti-inflammatory drugs meta-analysis of observational studies. Pharmacoepidemiol Drug Saf. 2013 Jun;22(6):559–70.

40. Singh G, Wu O, Langhorne P, Madhok R. Risk of acute myocardial infarction with nonselective non-steroidal anti-inflammatory drugs: a meta-analysis. Arthritis Res Ther. 2006;8(5):R153.

41. Bally M, Dendukuri N, Rich B, Nadeau L, Helin-Salmivaara A, Garbe E, et al. Risk of acute myocardial infarction with NSAIDs in real world use: bayesian meta-analysis of individual patient data. BMJ. 2017 May 9;357:j1909.

42. Varas-Lorenzo C, Riera-Guardia N, Calingaert B, Castellsague J, Pariente A, Scotti L, et al. Stroke risk and NSAIDs: a systematic review of observational studies. Pharmacoepidemiol Drug Saf. 2011 Dec 1;20(12):1225–36.

43. Solomon DH, Rassen JA, Glynn RJ, Lee J, Levin R, Schneeweiss S. THe comparative safety of analgesics in older adults with arthritis. Arch Intern Med. 2010 Dec 13;170(22):1968–78.

44. Cunnington M, Webb D, Qizilbash N, Blum D, Mander A, Funk MJ, et al. Risk of ischaemic cardiovascular events from selective cyclooxygenase-2 inhibitors in osteoarthritis. Pharmacoepidemiol Drug Saf. 2008 Jun 1;17(6):601–8.

45. Rahme E, Nedjar H. Risks and benefits of COX-2 inhibitors vs non-selective NSAIDs: does their cardiovascular risk exceed their gastrointestinal benefit? A retrospective cohort study. Rheumatology. 2007 Mar 1;46(3):435–8.

46. Mamdani M, Juurlink DN, Lee DS, Rochon PA, Kopp A, Naglie G, et al. Cyclo-oxygenase-2 inhibitors versus non-selective non-steroidal anti-inflammatory drugs and

congestive heart failure outcomes in elderly patients: a population-based cohort study. The Lancet. 2004 May 29;363(9423):1751–6.

47. Andersohn F, Suissa S, Garbe E. Use of First- and Second-Generation Cyclooxygenase-2–Selective Nonsteroidal Antiinflammatory Drugs and Risk of Acute Myocardial Infarction. Circulation [Internet]. 2006 Apr 25 [cited 2018 Oct 23]; Available from: https://www.ahajournals.org/doi/abs/10.1161/CIRCULATIONAHA.105.602425?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dpubmed

48. Lévesque LE, Brophy JM, Zhang B. The Risk for Myocardial Infarction with Cyclooxygenase-2 Inhibitors: A Population Study of Elderly Adults. Ann Intern Med. 2005 Apr 5;142(7):481.

49. Helin-Salmivaara A, Virtanen A, Vesalainen R, Grönroos JM, Klaukka T, Idänpään-Heikkilä JE, et al. NSAID use and the risk of hospitalization for first myocardial infarction in the general population: a nationwide case–control study from Finland. Eur Heart J. 2006 Jul 2;27(14):1657–63.

50. García-Poza P, Abajo FJ de, Gil MJ, Chacón A, Bryant V, García-Rodríguez LA. Risk of ischemic stroke associated with non-steroidal anti-inflammatory drugs and paracetamol: a population-based case-control study. J Thromb Haemost. 2015 May 1;13(5):708–18.

51. Barbour KE, Lui L-Y, Nevitt MC, Murphy LB, Helmick CG, Theis KA, et al. Hip Osteoarthritis and the Risk of All-Cause and Disease-Specific Mortality in Older Women: A Population-Based Cohort Study. Arthritis Rheumatol. 2015 Jul;67(7):1798–805.

52. Hoeven TA, Leening MJG, Bindels PJ, Castaño-Betancourt M, Meurs JB van, Franco OH, et al. Disability and not osteoarthritis predicts cardiovascular disease: a prospective population-based cohort study. Ann Rheum Dis. 2015 Apr 1;74(4):752–6.

53. Ong KL, Wu BJ, Cheung BMY, Barter PJ, Rye K-A. Arthritis: its prevalence, risk factors, and association with cardiovascular diseases in the United States, 1999 to 2008. Ann Epidemiol. 2013 Feb 1;23(2):80–6.

54. Nielen MM, van Sijl AM, Peters MJ, Verheij RA, Schellevis FG, Nurmohamed MT. Cardiovascular disease prevalence in patients with inflammatory arthritis, diabetes mellitus and osteoarthritis: a cross-sectional study in primary care. BMC Musculoskelet Disord. 2012 Aug 21;13:150.

55. Kishimoto M, Masaki H. Arthritis as a Risk Factor for Incident Coronary Heart Disease in Elderly Japanese-American Males. :6.

56. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol. 1986 Dec;51(6):1173–82.

57. Lange T, Vansteelandt S, Bekaert M. A Simple Unified Approach for Estimating Natural Direct and Indirect Effects. Am J Epidemiol. 2012 Aug 1;176(3):190–5.

58. Rochon J, du Bois A, Lange T. Mediation analysis of the relationship between institutional research activity and patient survival. BMC Med Res Methodol. 2014;14(1):9.

59. Cole SR, Hernán MA. Fallibility in estimating direct effects. Int J Epidemiol. 2002 Feb 1;31(1):163–5.

60. Kaufman JS, MacLehose RF, Kaufman S. A further critique of the analytic strategy of adjusting for covariates to identify biologic mediation. Epidemiol Perspect Innov EPI. 2004 Oct 8;1:4.

61. VanderWeele TJ. Causal mediation analysis with survival data. Epidemiol Camb Mass. 2011 Jul;22(4):582–5.

62. Lange T, Hansen JV. Direct and Indirect Effects in a Survival Context: Epidemiology. 2011 Jul;22(4):575–81.

63. Liu Q, Niu J, Li H, Ke Y, Li R, Zhang Y, et al. Knee Symptomatic Osteoarthritis, Walking Disability, NSAIDs Use and All-cause Mortality: Population-based Wuchuan Osteoarthritis Study. Sci Rep [Internet]. 2017 Dec [cited 2017 Oct 23];7(1). Available from: http://www.nature.com/articles/s41598-017-03110-3

64. Nordahl H, Lange T, Osler M, Diderichsen F, Andersen I, Prescott E, et al. Education and Cause-specific Mortality: The Mediating Role of Differential Exposure and Vulnerability to Behavioral Risk Factors. Epidemiology. 2014 May;25(3):389–96.

65. Solomon DH, Avorn J, Stürmer T, Glynn RJ, Mogun H, Schneeweiss S. Cardiovascular outcomes in new users of coxibs and nonsteroidal antiinflammatory drugs: High-risk subgroups and time course of risk. Arthritis Rheum. 2006 May 1;54(5):1378–89.

66. Schjerning Olsen A-M, Fosbøl EL, Lindhardsen J, Folke F, Charlot M, Selmer C, et al. Duration of Treatment With Nonsteroidal Anti-Inflammatory Drugs and Impact on Risk of Death and Recurrent Myocardial Infarction in Patients With Prior Myocardial Infarction: A Nationwide Cohort Study. Circulation. 2011 May 24;123(20):2226–35.

67. Bak S, Andersen M, Tsiropoulos I, Garcia Rodriguez LA, Hallas J, Christensen K, et al. Risk of stroke associated with nonsteroidal anti-inflammatory drugs: a nested case-control study. Stroke. 2003 Feb;34(2):379–86.

68. Fischer LM, Schlienger RG, Matter CM, Jick H, Meier CR. Current Use of Nonsteroidal Antiinflammatory Drugs and the Risk of Acute Myocardial Infarction. Pharmacother J Hum Pharmacol Drug Ther. 2005 Apr 1;25(4):503–10.

69. PharmaNet | www.popdata.bc.ca [Internet]. [cited 2019 Jul 11]. Available from: https://www.popdata.bc.ca/data/health/PharmaNet

70. Medical Services Plan (MSP) Payment Information File | www.popdata.bc.ca [Internet]. [cited 2019 Jul 11]. Available from: https://www.popdata.bc.ca/data/health/msp

71. Discharge Abstracts Database (Hospital Separations file) | www.popdata.bc.ca [Internet]. [cited 2019 Jul 11]. Available from: https://www.popdata.bc.ca/data/health/dad

72. Consolidation file | www.popdata.bc.ca [Internet]. [cited 2019 Jul 11]. Available from: https://www.popdata.bc.ca/data/population/consolidationfile

73. Vital Statistics Deaths | www.popdata.bc.ca [Internet]. [cited 2019 Jul 11]. Available from: https://www.popdata.bc.ca/data/population/vsdeaths

74. About PopData | www.popdata.bc.ca [Internet]. [cited 2017 Oct 24]. Available from: https://www.popdata.bc.ca/aboutus

75. Government of Canada SC. Canadian Community Health Survey (CCHS) Cycle 1.1 [Internet]. 2007 [cited 2018 Sep 30]. Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=3359

76. Government of Canada SC. Canadian Community Health Survey (CCHS) Cycle 2.1 [Internet]. 2007 [cited 2018 Sep 30]. Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=4995

77. Government of Canada SC. Canadian Community Health Survey (CCHS) Cycle 3.1 [Internet]. 2007 [cited 2018 Sep 30]. Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=22642

78. Canada PHA of, Canada PHA of. Heart disease – heart health [Internet]. aem. 2013 [cited 2017 May 17]. Available from: https://www.canada.ca/en/public-health/services/diseases/heart-disease-heart-health.html

79. Link Between OA and CVD Risk [Internet]. arthritis.org. [cited 2017 Oct 23]. Available from: http://www.arthritis.org/living-with-arthritis/comorbidities/heart-disease/osteoarthritis-ups-cvd-risk.php

80. Tonelli M, Wiebe N, Fortin M, Guthrie B, Hemmelgarn BR, James MT, et al. Methods for identifying 30 chronic conditions: application to administrative data. BMC Med Inform Decis Mak. 2015;15:31.

81. Cardiac Health Foundation Of Canada [Internet]. [cited 2017 May 25]. Available from: http://cardiachealth.ca/index.php?o=cardiac+rehabilitation&type=view_article&id=24

82. What Are Coronary Heart Disease Risk Factors? - NHLBI, NIH [Internet]. [cited 2017 May 25]. Available from: https://www.nhlbi.nih.gov/health/health-topics/topics/hd

83. Risk factors [Internet]. [cited 2017 May 25]. Available from: https://www.bhf.org.uk/heart-health/risk-factors

84. Cardiovascular disease risk factors | World Heart Federation [Internet]. [cited 2017 May 25]. Available from: http://www.world-heart-federation.org/press/fact-sheets/cardiovascular-disease-risk-factors/

85. Romano PS, Roos LL, Jollis JG. Presentation adapting a clinical comorbidity index for use with ICD-9-CM administrative data: differing perspectives. J Clin Epidemiol. 1993;46(10):1075–1079.

86. Romano PS, Roost LL, Jollis JG. Further evidence concerning the use of a clinical comorbidity index with ICD-9-CM administrative data. J Clin Epidemiol. 1993;46(10):1085–1090.

87. Sterne JAC, White IR, Carlin JB, Spratt M, Royston P, Kenward MG, et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. BMJ. 2009 Jun 29;338:b2393.

88. Government of Canada SC. Canadian Community Health Survey (CCHS) [Internet]. 2005 [cited 2014 Oct 10]. Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SurvId=1630&InstaId=22642&SDDS=3226

89. Government of Canada SC. Canadian Community Health Survey (CCHS) [Internet]. 2007 [cited 2017 Oct 24]. Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&Id=4995

90. Cheetham TC, Graham DJ, Campen D, Hui R, Spence M, Levy G, et al. Myocardial Infarction and Its Association with the Use of Nonselective NSAIDs: A Nested Case-Control and Time-to-Event Analysis. Perm J. 2008;12(1):16–22.

91. Aviña-Zubieta JA, Abrahamowicz M, Vera D, A M, Choi HK, Sayre EC, et al. Immediate and past cumulative effects of oral glucocorticoids on the risk of acute myocardial infarction in rheumatoid arthritis: a population-based study. Rheumatology. 2013 Jan 1;52(1):68–75.

92. Aviña-Zubieta JA, Abrahamowicz M, Choi HK, Rahman MM, Sylvestre M-P, Esdaile JM, et al. Risk of cerebrovascular disease associated with the use of glucocorticoids in patients with incident rheumatoid arthritis: a population-based study. Ann Rheum Dis. 2011 Jun 1;70(6):990–5.

93. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Control. 1974 Dec;19(6):716–23.

94. Abrahamowicz M, Bartlett G, Tamblyn R, du Berger R. Modeling cumulative dose and exposure duration provided insights regarding the associations between benzodiazepines and injuries. J Clin Epidemiol. 2006 Apr;59(4):393–403.

95. Pearl J. Causality: Models, Reasoning and Inference. 2nd ed. New York, NY, USA: Cambridge University Press; 2009.

96. Roumie CL, Mitchel EF, Kaltenbach L, Arbogast PG, Gideon P, Griffin MR. Nonaspirin NSAIDs, Cyclooxygenase 2 Inhibitors, and the Risk for Stroke. Stroke. 2008 Jul 1;39(7):2037–45.

97. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. Am J Epidemiol. 2008 Sep 15;168(6):656–64.

98. Zheng W, Laan M van der. Causal Mediation in a Survival Setting with Time-Dependent Mediators. UC Berkeley Div Biostat Work Pap Ser [Internet]. 2012 Jun 28; Available from: https://biostats.bepress.com/ucbbiostat/paper295

99. Lin S-H, Young J, Logan R, Tchetgen Tchetgen EJ, VanderWeele TJ. Parametric mediational g-formula approach to mediation analysis with time-varying exposures, mediators, and confounders. Epidemiol Camb Mass. 2017 Mar;28(2):266–74.

100. Lin S-H, Young JG, Logan R, VanderWeele TJ. Mediation analysis for a survival outcome with time-varying exposures, mediators, and confounders. Stat Med. 2017 Nov 20;36(26):4153–66.

101. Pearl J. The Causal Mediation Formula—A Guide to the Assessment of Pathways and Mechanisms. Prev Sci. 2012 Aug 1;13(4):426–36.

102. Nguyen QC, Osypuk TL, Schmidt NM, Glymour MM, Tchetgen Tchetgen EJ. Practical Guidance for Conducting Mediation Analysis With Multiple Mediators Using Inverse Odds Ratio Weighting. Am J Epidemiol. 2015 Mar 1;181(5):349–56.

103. VanderWeele T. Explanation in Causal Inference: Methods for Mediation and Interaction. Oxford University Press; 2015. 729 p.

104. Lanas A, Tornero J, Zamorano JL. Assessment of gastrointestinal and cardiovascular risk in patients with osteoarthritis who require NSAIDs: the LOGICA study. Ann Rheum Dis. 2010 Aug 1;69(8):1453–8.

105. Solomon DH, Rassen JA, Glynn RJ, Lee J, Levin R, Schneeweiss S. The Comparative Safety of Analgesics in Older Adults With Arthritis. Arch Intern Med. 2010 Dec 13;170(22):1968–78.

106. Kaushal N, Singh S, Gulati M, Vaidya Y, Kaushik M. Study of regulatory requirements for the conduct of bioequivalence studies in US, Europe, Canada, India, ASEAN and SADC countries: Impact on generic drug substitution. J Appl Pharm Sci. 2016;206–22.

107. Day RO, McLachlan AJ, Graham GG, Williams KM. Pharmacokinetics of Nonsteroidal Anti-Inflammatory Drugs in Synovial Fluid: Clin Pharmacokinet. 1999;36(3):191–210.

108. Solomon DH, Glynn RJ, Rothman KJ, Schneeweiss S, Setoguchi S, Mogun H, et al. Subgroup Analyses to Determine Cardiovascular Risk Associated With Nonsteroidal Antiinflammatory Drugs and Coxibs in Specific Patient Groups. Arthritis Rheum. 2008 Aug 15;59(8):1097–104.

109. Solomon DH, Rassen JA, Glynn RJ, Lee J, Levin R, Schneeweiss S. The Comparative Safety of Analgesics in Older Adults With Arthritis. Arch Intern Med. 2010 Dec 13;170(22):1968–78.

110. Huang W-F, Hsiao F-Y, Wen Y-W, Tsai Y-W. Cardiovascular events associated with the use of four nonselective NSAIDs (Etodolac, Nabumetone, Ibuprofen, or Naproxen) Versus a Cyclooxygenase-2 Inhibitor (Celecoxib): A population-based analysis in taiwanese adults. Clin Ther. 2006 Nov 1;28(11):1827–36.

111. Canada PHA of, Canada PHA of. Heart disease – heart health [Internet]. aem. 2013 [cited 2017 May 17]. Available from: https://www.canada.ca/en/public-health/services/diseases/heart-disease-heart-health.html

112. Paradoxical Relations of Drug Treatment with Mortality in Ol... : Epidemiology [Internet]. [cited 2018 Nov 7]. Available from: https://journals.lww.com/epidem/Fulltext/2001/11000/Paradoxical_Relations_of_Drug_Treatment_with.17.aspx

113. Solomon DH, Husni ME, Wolski KE, Wisniewski LM, Borer JS, Graham DY, et al. Differences in Safety of Nonsteroidal Antiinflammatory Drugs in Patients With Osteoarthritis and Patients With Rheumatoid Arthritis. Arthritis Rheumatol. 2018;70(4):537–46.

114. Hippisley-Cox J, Coupland C. Risk of myocardial infarction in patients taking cyclo-oxygenase-2 inhibitors or conventional non-steroidal anti-inflammatory drugs: population based nested case-control analysis. BMJ. 2005 Jun 9;330(7504):1366.

115. Risk factors [Internet]. [cited 2018 Mar 13]. Available from: https://www.bhf.org.uk/heart-health/risk-factors

116. Ratitch B, Lipkovich I. Combining Analysis Results from Multiply Imputed Categorical Data. :19.

117. Little RJA, Rubin DB. Statistical analysis with missing data. New York: Wiley; 1987. 278 p. (Wiley series in probability and mathematical statistics. Applied probability and statistics).

118. Liu Y, De A. Multiple Imputation by Fully Conditional Specification for Dealing with Missing Data in a Large Epidemiologic Study. Int J Stat Med Res. 2015;4(3):287–95.

119. Government of Canada SC. Canadian Community Health Survey (CCHS) [Internet]. 2005 [cited 2014 Oct 13]. Available from:

http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SurvId=1630&InstaId=22642&SDDS=3226

120. Statistics Canada. CCHS 3.1 Public Use Microdata File User Guide [Internet]. 2006 [cited 2014 Oct 14]. Available from: http://abacus.library.ubc.ca/bitstream/10573/41781/21/Guide_E.Pdf

121. Statistics Canada. CCHS 3.1 Public Use Microdata File Data Dictionary [Internet]. 2006 [cited 2014 Oct 14]. Available from: http://abacus.library.ubc.ca/handle/10573/41781

122. Franklin JM, Schneeweiss S, Polinski JM, Rassen JA. Plasmode simulation for the evaluation of pharmacoepidemiologic methods in complex healthcare databases. Comput Stat Data Anal. 2014 Apr;72:219–26.

123. Rubin DB. Multiple imputation for nonresponse in surveys. New York: Wiley; 1987. 258 p. (Wiley series in probability and mathematical statistics. Applied probability and statistics).

124. Multiple Imputation in SAS Part 1 [Internet]. IDRE Stats. [cited 2017 Nov 15]. Available from: https://stats.idre.ucla.edu/sas/seminars/multiple-imputation-in-sas/mi_new_1/

125. Rubin DB. Frontmatter. In: Multiple Imputation for Nonresponse in Surveys [Internet]. John Wiley & Sons, Inc.; 1987 [cited 2015 Aug 12]. p. i–xxix. Available from: http://onlinelibrary.wiley.com/doi/10.1002/9780470316696.fmatter/summary

126. Liu S, Chrysanthopoulou SA, Chang Q, Hunnicutt JN, Lapane KL. Missing Data in Marginal Structural Models: A Plasmode Simulation Study Comparing Multiple Imputation and Inverse Probability Weighting. Med Care [Internet]. 2019 Jan 16 [cited 2019 Jan 28];Publish Ahead of Print. Available from: https://insights.ovid.com/crossref?an=00005650-900000000-98494

127. Mühlenbruch K, Kuxhaus O, di Giuseppe R, Boeing H, Weikert C, Schulze MB. Multiple imputation was a valid approach to estimate absolute risk from a prediction model based on case–cohort data. J Clin Epidemiol. 2017 Apr 1;84:130–41.

128. Trends in Prescribing of NSAIDs and Opioids Among Osteoarthritis Patients in British Columbia, Canada, 1998-2014 [Internet]. ACR Meeting Abstracts. [cited 2019 Dec 12]. Available from: https://acrabstracts.org/abstract/trends-in-prescribing-of-nsaids-and-opioids-among-osteoarthritis-patients-in-british-columbia-canada-1998-2014/

129. Franklin JM, Eddings W, Glynn RJ, Schneeweiss S. Regularized Regression Versus the High-Dimensional Propensity Score for Confounding Adjustment in Secondary Database Analyses. Am J Epidemiol. 2015 Oct 1;182(7):651–9.

130. Fewell Z, Davey Smith G, Sterne JAC. The Impact of Residual and Unmeasured Confounding in Epidemiologic Studies: A Simulation Study. Am J Epidemiol. 2007 Sep 15;166(6):646–55.

131. Vaughan LK, Divers J, Padilla MA, Redden DT, Tiwari HK, Pomp D, et al. The use of plasmodes as a supplement to simulations: A simple example evaluating individual admixture estimation methodologies. Comput Stat Data Anal. 2009 Mar;53(5):1755–66.

132. Rahman MM, Kopec JA, Goldsmith CH, Anis AH, Cibere J. Validation of Administrative Osteoarthritis Diagnosis Using a Clinical and Radiological Population-Based Cohort. Int J Rheumatol [Internet]. 2016 [cited 2019 Dec 12];2016. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5227164/

133. Government of Canada SC. National Population Health Survey: Health Institutions Component, Longitudinal (NPHS) [Internet]. 2007 [cited 2016 Mar 21]. Available from: http://www23.statcan.gc.ca/imdb/p2SV.pl?Function=getSurvey&SDDS=5003

134. Canadian Health Measures Survey (CHMS) Data User Guide: Cycle 1 [Internet]. [cited 2016 Mar 21]. Available from: http://www23.statcan.gc.ca/imdb-bmdi/document/5071_D2_T1_V1-eng.htm

135. Sinharay S, Stern HS, Russell D. The use of multiple imputation for the analysis of missing data. Psychol Methods. 2001;6(4):317.

136. Vittinghoff E, Glidden DV, Shiboski SC, McCulloch CE. Survival Analysis. In: Regression Methods in Biostatistics [Internet]. Springer, Boston, MA; 2012 [cited 2017 Nov 15]. p. 203–59. (Statistics for Biology and Health). Available from: https://link.springer.com/chapter/10.1007/978-1-4614-1353-0_6

Appendices

Appendix A BMI Missing value imputation in chapter 2

In 1978 Rubin et al. first proposed the method of multiple imputation to handle the issue of missing data (123). In recent years multiple imputation method is increasingly used in imputing the missing data because of the advancement in statistical computation (135).

Multiple imputation method has three steps. In step one, multiple imputation creates a number (equivalent to the desired number of imputations set by the researcher) of complete datasets in which the missing value of one variable is replaced by the imputed values based on observed data of a set of other variables. One major advantage of multiple imputation is that it accounts for the uncertainty associated with the imputation procedure (87).

In step two, each imputed dataset is analyzed separately using any standard statistical methods. The association between the exposure and the outcome would vary between the imputed datasets because of variation in the imputation of the missing values. In step three, the results from multiple datasets are combined using Rubin's rules (87).

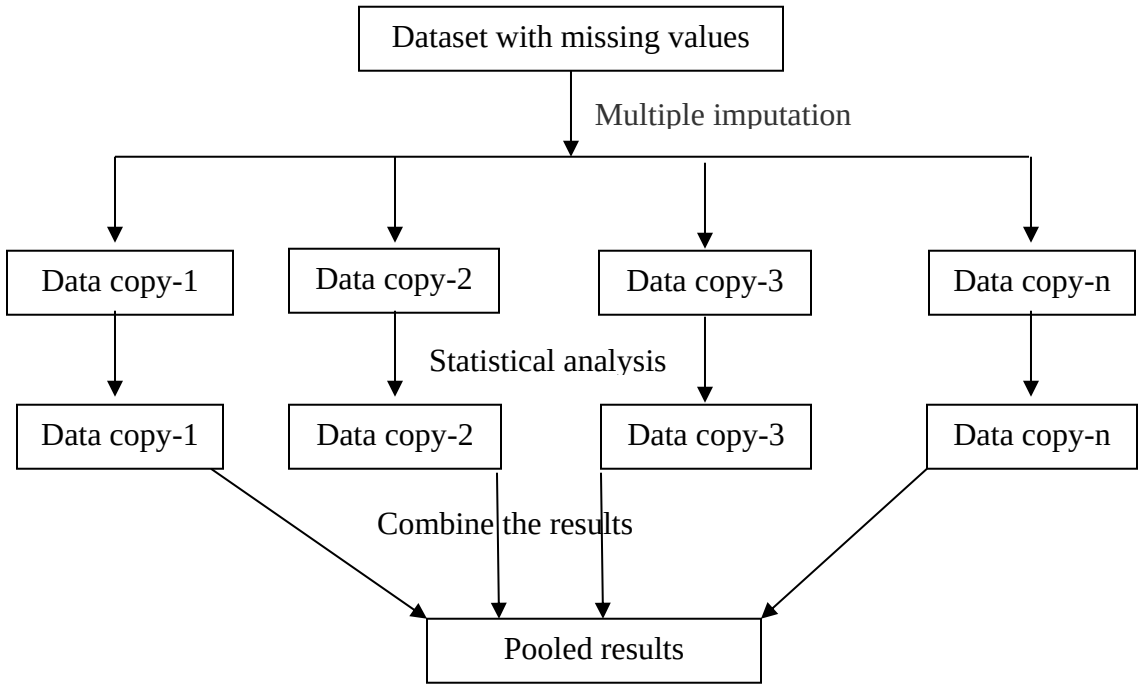


Figure 5-1 Illustration of multiple imputation

A.1 Assumptions in multiple imputation

There are certain assumptions in multiple imputation (117)-

- Missing completely at random (MACR): When the probability of missing of a variable is not related to the variable itself as well as any other variable in the dataset.
- Missing at random (MAR): When the probability of missing of a variable does not depend on the value of the variable itself after controlling for other observed variables.

Health administrative data in British Columbia are primarily collected for billing purposes and does not record information on body mass index. Since the information is not available for anyone in the population, we assume that the missingness of BMI is not related to the BMI value

itself. Similarly, the variables recorded in the BC HAD is not related to the missingness of BMI (not recorded as this variable was not necessary for administrative purpose). Thus, we assume that the BMI is missing completely at random in the BC HAD.

A.2 Methods of multiple imputation

The selection of imputation method in multiple imputation depends on the missingness pattern (monotone/non-monotone/arbitrary) and type of imputed variable (continuous or categorical) (124).

Missing pattern

In monotone missing pattern the missing data is situated at the end when the data read from left to right. In other words, when a variable is missing for an individual in the dataset, it is assumed that all subsequent variables are missing for that individual. In arbitrary missing data the missingness of a variable is scattered among full data values (136).

Table 5-1 Illustration of monotone missing pattern

Observation	Variable 1	Variable 2	Variable 3
1	O	O	O
2	O	O	M
3	O	M	M
O = observed and M = missing			

Table 5-2 Illustration of arbitrary missing pattern

Observation	Variable 1	Variable 2	Variable 3
1	O	O	O
2	O	M	O
3	M	M	O
O = observed and M = missing			

Methods

The type of distribution under which the missing value will be imputed is an important consideration in multiple imputation (124). Markov Chain Monte Carlo (MCMC) method assumes that variables entered into the imputation model have a joint multivariate normal distribution. The MCMC based algorithm fills the missing value of a variable by drawing from a conditional distribution (multivariate normal is assumed) based on the other observed data. This method may produce biased estimates when the sample size is relatively small and the proportion of missing data is high. On the other hand fully conditional specification method does not assume a joint distribution. Instead, this method uses a separate conditional distribution for each imputed variable (124).

A.3 Description of multiple imputation method used in this study

CCHS data contains information on demographics, health status and health determinants including BMI in a large sample of the Canadian population (89,119). A detailed description of the survey design, sampling frame and interviewing procedures can be obtained from Statistics Canada. First, we created a reference dataset using CCHS cycles 2.1 and 3.1 datasets including all survey respondents aged at least 20 years. We created a binary variable of OA and CVD using self-reported responses to questions related to the corresponding disease status. Age was recorded as 10-year age-group, and sex was recorded as male/female. Statistics Canada calculated BMI by dividing self-reported weight in kilograms by self-reported height in meters squared and recorded as a continuous variable. Information on all variables was observed for everyone in the reference dataset.

We grouped the individuals in our analytic data from BC HAD into 10-year age groups. We then set the analytic data (N=30,972) under the reference dataset created from the CCHS data (N=166,381). In this combined dataset (N= 197,353) all the variables were observed for everyone except BMI score of the individuals from our analytic data created from BC HAD. So in the final dataset, BMI was missing for approximately 16% of the individuals. Finally, assuming an arbitrary missing pattern we imputed BMI score using FCS regression method. Age, sex, OA and CVD variables were entered into the multiple imputation model. Finally, we categorized the BMI score into four categories that were defined as underweight (<18.5 kg/m2), normal weight (18.5–24.9 kg/m2), overweight (25.0–29.9 kg/m2), and obese (>30.0 kg/m2).

Appendix B Mediation analysis in chapter 2 using marginal structural model (MSM)

B.1 Marginal structural model

Marginal structural model (MSM) is based on the counterfactual framework, i.e., the marginal expectation of a counterfactual outcome is modeled (57). We will explain the concept using the following example-

Let,

A = binary exposure, a random variable. Small letter case 'a' denotes observed exposure.

A* = counterfactual exposure [A* = | A - 1 |], is the exposure status that we do not observe.

Y = outcome

M = the mediator

C = common set of confounding variables

In this example, the nested counterfactuals, Ya*;Ma denotes the outcome that we would have observed if the exposure was set to counterfactual exposure, a* and the mediator was set to the value it would have taken if the exposure was set to a (57).

Similarly, the nested counterfactuals, Ya;Ma* denotes the outcome that we would have observed if the exposure was set to observed exposure, a and the mediator was set to the value it would have taken if the exposure was set to counterfactual exposure, a* (57).

The nested counterfactuals, Ya;Ma* can be modeled as (57)-

$$E[Ya;Ma^*] = \beta_0 + \beta_1 a + \beta_2 a^* \quad \text{----------(equation B.1)}$$

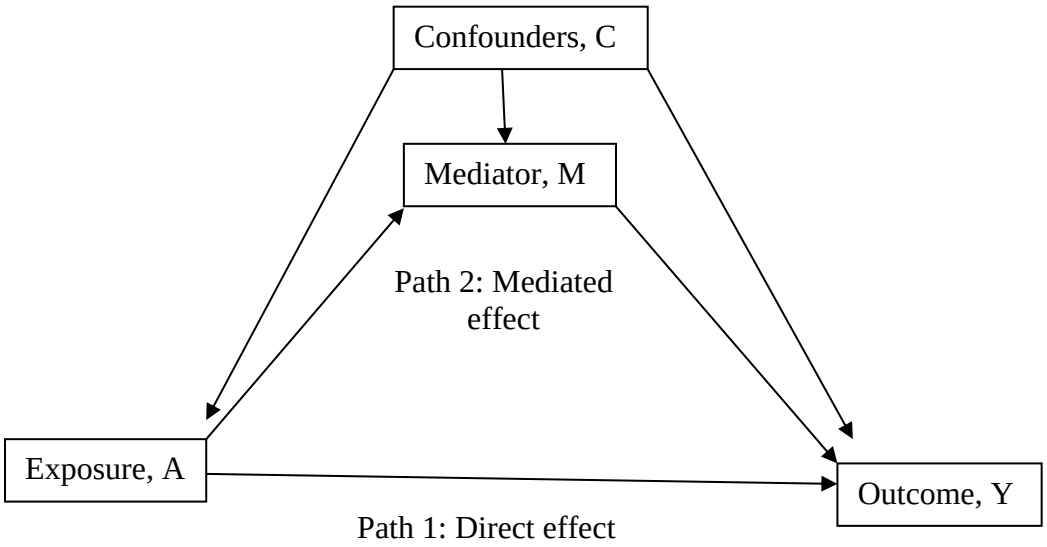


Figure 5-2 Illustration of direct and indirect effects

In this model the exposure is included twice, the observed exposure (a) and the counterfactual exposure (a*) to establish that it works through 2 distinct causal pathways (57)-

The $\beta_1(A - A^*)$ estimates the natural direct effect (NDE) and $\beta_2(A - A^*)$ estimates the natural indirect effect (NIE). The sum of the NDE and NIE provides the estimate for total effect.

The generalized linear MSMs can be written as follows (57)-

$$g (E [Ya;Ma^*]) = \beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 a \times a^* \quad \text{----------(equation B.2)}$$

In this equation, 'g' is the link function specifying the outcome model, logistic regression, for example, and β_3 is the interaction term that be included, if necessary. If there is no interaction present, i.e., $\beta_3 = 0$ and 'g' is the logit link, then-

Odds ratio (OR) for NDE can be obtained by exponentiating the $\beta_1(a - a^*)$

OR for NIE can be obtained by exponentiating the $\beta_2(a - a^*)$

The product of the NDE odds ratio and NIE odds ratio gives the total effect (57).

In case of survival analysis involving time-to-event outcome, Lange et al. proposed that unbiased estimates of natural direct and indirect effects can be obtained from weighted cox regression model using a duplicated data set (58). In Cox PH model, the hazard function related to counterfactual survival time can be estimated from the following equation (57)-

$\lambda(t) = \lambda_0(t) \exp(\beta_0 + \beta_1 a + \beta_2 a^* + \beta_3 a \times a^*)$ ----------(equation B.3)

Here, λ_0 is the unspecified baseline hazards. Similar to the logistic regression model, the hazard ratios for A and A* are the estimates of the NDE and NIE, respectively and the total effect (HR) is the product of hazard ratios for NDE and NIE.

B.2 MSM implementation strategy

NDE and NIE can be estimated from weighted Cox regression model using a duplicated data set (58). In the first replicated data in this study, OA* took the original value of the OA exposure status. In the second replicated data OA* took the opposite ('counterfactual') value of the OA exposure, i.e., OA* = | OA – 1 |.Below are the steps that were implemented to calculate the weight and hazard ratios for OA and OA* that serve as estimates for the natural direct effect and indirect effect, respectively.

Step 1: In this step, a new variable, OA*, was created that represented the counterfactual of the value of the OA exposure status. Each observation in the analytic data was replicated twice (once

OA* = 1 and once with OA* = 0). When the newly created variable OA* = OA, then the temporary mediator variable, mtemp was set to the observed NSAID use, a binary variable. As a result, mtemp was missing for all newly created observations, but equal to the observed NSAID use for the actual observations. Using this mtemp as the mediator-outcome (NSAID use yes/No), the model fitted the observed data and then fitted probabilities were computed for both actual and "created" observations.

Conceptually-

	Study ID	OA	OA*	M (temp.)	CVD	C
Copy 1	1	0	1	.	1	
	2	0	1	.	0	
	3	1	1	1	1	
	4	0	1	.	0	
	5	1	1	0	1	

	Study ID	OA	OA*	M (temp.)	CVD	C
Copy 2	1	0	0	1	1	
	2	0	0	0	0	
	3	1	0	.	1	
	4	0	0	0	0	
	5	1	0	.	1	

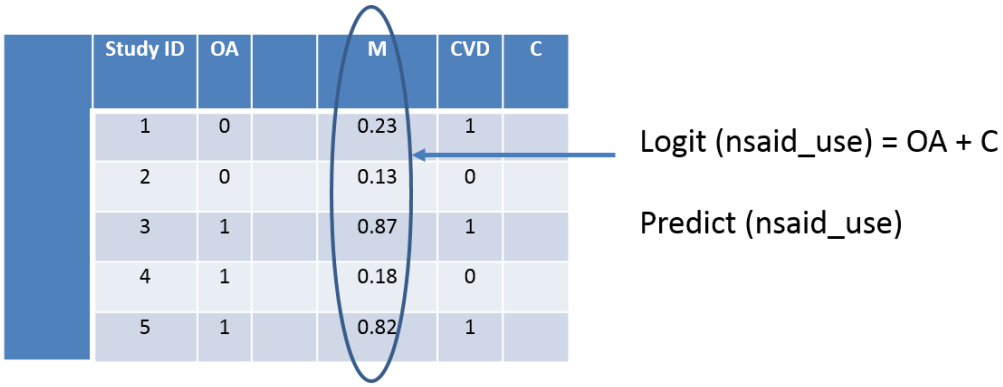
SAS code-

```
data MSM;
set MSM;

OA* = 1;
mtemp =.;
if OA = OA* then mtemp = nsaid_use;
output;

OA* = 0;
mtemp =.;
if OA = OA* then mtemp = nsaid_use;
output;
run;
```

Step 2: In this step we fitted a multivariable logistic regression model for the mediator, mtemp (nsaid_use), conditioning on OA exposure status and common set of confounding variables (C) and computed the predicted values.

Since we needed to calculate the predicted values by changing the value of OA exposure status to the counterfactual (OA*), we created a copy of the OA exposure variable (atemp) and used that in the model.

Conceptually-



SAS code-

```
data MSM;
set MSM;
atemp = OA;
run;
proc logistic data = MSM;
class
atemp (param = ref ref = "0")
hypertension (param = ref ref = "0")
diabetes (param = ref ref = "0")
hyperlipidemia (param = ref ref = "0")
COPD (param = ref ref = "0")
SES_ category (param = ref ref = "5")
bmi_category (param = ref ref = "Normal weight")
mtemp (param = ref ref = "0") ;
```

```
model mtemp = atemp hypertension diabetes hyperlipidemia COPD SES_cat bmi_category
Romano_comorbidity_score / link = glogit;
score out = MSM;
run;


data MSM;
set MSM;
if nsaid_use = 0 then weight_direct = P_0; * direct weight because atemp was set to OA;
if nsaid_use = 1 then weight_direct = P_1; * direct weight because atemp was set to OA;
run;
```

Step 3: Similar to step 2, we set the variable, atemp to OA* and calculated predicted values.

Conceptually-

SAS code-

```
data MSM;
set MSM;
atemp = OA*;
run;
proc logistic data = MSM;
class
atemp (param = ref ref = “0”)
hypertension (param = ref ref = “0”)
diabetes (param = ref ref = “0”)
hyperlipidemia (param = ref ref = “0”)
COPD (param = ref ref = “0”)
SES_ category (param = ref ref = “5”)
bmi_category (param = ref ref = “Normal weight”)
mtemp (param = ref ref = “0”) ;
model mtemp = atemp hypertension diabetes hyperlipidemia COPD SES_cat bmi_category
Romano_comorbidity_score / link = glogit;
score out = MSM;
run;


data MSM;
set MSM;
```

```
if nsaid_use = 0 then weight_indirect = P_0; * indirect weight because atemp was set to OA*;
if nsaid_use = 1 then weight_indirect = P_1; * indirect weight because atemp was set to OA*;
run;
```

Step 4: Finally the weights were computed based on the following equation-

$$Wi = \frac{[P\ (nsaid\_use = nsaid\_usei\ l\ OA = OA^*i,\ C=Ci\ )]}{[P\ (nsaid\_use = NSAID\ usei\ l\ OA = OAi,\ C=Ci\ )]}$$

---------(equation B.4)

Conceptually-

	Study ID		OA*	M	CVD	C
	1		0	0.19	1	
	2		0	0.14	0	
	3		1	0.83	1	
	4		0	0.17	0	
	5		1	0.82	1	

Weight (W) = Indirect wt./ direct wt.

Weight (W) = [P (nsaid_use = nsaid_usei l OA = OA*I, C=Ci)] / [P (nsaid_use = nsaid_usei l OA = OAi, C=Ci)]

	Study ID	OA		M	CVD	C
	1	0		0.23	1	
	2	0		0.13	0	
	3	1		0.87	1	
	4	1		0.18	0	
	5	1		0.82	1	

SAS code-

```
data MSM;
set MSM;
w = weight_indirect / weight_direct;
run;
```

Step 5: In this step, we fitted a weighted Cox regression model to estimate the NDE and NIE adjusting for hypertension, diabetes, hyperlipidemia, COPD, BMI, SES and Romano comorbidity score.

Conceptually-

$h(\text{time to CVD}) = h0(t) \exp(\beta_1 OA + \beta_2 OA^* + \ldots + \beta n\, C)$ (weighted by W) -(equation B.5)

- The hazard ratios for OA and OA* are the estimate of the NDE and NIE, respectively.
- The total effect of OA on CVD is the product of hazard ratios for OA (NDE) and OA* (NDE).

SAS code-

```
proc phreg data = MSM_CVD covsandwich;
class
OA (param = ref ref = “0”)
OA* (param = ref ref = “0”)
hypertension (param = ref ref = “0”)
diabetes (param = ref ref = “0”)
hyperlipidemia (param = ref ref = “0”)
COPD (param = ref ref = “0”)
SES_ category (param = ref ref = “5”)
bmi_category (param = ref ref = “Normal weight”)
cvd;
model year_cvd * cvd (0) = OA OA* hypertension diabetes hyperlipidemia COPD
SES_category bmi_category Romano_comorbidity_score  / r1 ties = efron;
WEIGHT w;
ID studyid;
run;
```

Appendix C Detailed analysis results in chapter 2

Table 5-3 Multivariable Cox regression analysis for primary CVD outcome

Variable	HR (95% CI)
OA	
No	Reference
Yes	1.23 (1.17, 1.28)
Hypertension	
No	Reference
Yes	1.75 (1.68, 1.83)
Diabetes	
No	Reference
Yes	1.56 (1.46, 1.66)
Hyperlipidemia	
No	Reference
Yes	0.82 (0.76, 0.89)
COPD	
No	Reference
Yes	1.39 (1.31, 1.49)
BMI	
Normal weight	Reference
Under weight	0.94 (0.85, 1.04)
Over weight	1.07 (1.02, 1.13)
Obese	1.18 (1.10, 1.24)
SES	
5 (Highest)	Reference
4	1.00 (0.95, 1.08)
3	1.09 (1.03, 1.17)
2	1.16 (1.09, 1.24)
1 (Lowest)	1.24 (1.16, 1.32)
Romano comorbidity score	1.12 (1.09, 1.15)

Table 5-4 Multivariable Cox regression analysis for secondary outcome, IHD

Variable	HR (95% CI)
OA	
No	Reference
Yes	1.17 (1.10, 1.26)
Hypertension	
No	Reference
Yes	1.69(1.59,1.80)
Diabetes	
No	Reference
Yes	1.81 (1.66,1.98)
Hyperlipidemia	
No	Reference
Yes	0.81 (0.72, 0.92)
COPD	
No	Reference
Yes	1.33 (1.20, 1.48)
BMI	
Normal weight	Reference
Under weight	0.91 (0.78, 1.06)
Over weight	1.07 (0.99, 1.15)
Obese	1.22 (1.13, 1.32)
SES	
5 (Highest)	Reference
4	0.99 (0.91, 1.10)
3	1.07 (0.98, 1.18)
2	1.10 (1.00, 1.22)
1 (Lowest)	1.14 (1.04, 1.26)
Romano comorbidity score	1.09 (1.05, 1.13)

Table 5-5 Multivariable Cox regression analysis for secondary outcome, CHF

Variable	HR (95% CI)
OA	
No	Reference
Yes	1.42 (1.33, 1.51)
Hypertension	
No	Reference
Yes	1.89 (1.78, 2.01)
Diabetes	
No	Reference
Yes	1.72 (1.57, 1.88)
Hyperlipidemia	
No	Reference
Yes	0.66 (0.58, 0.76)
COPD	
No	Reference
Yes	1.59 (1.45, 1.75)
BMI	
Normal weight	Reference
Under weight	0.98 (0.84, 1.13)
Over weight	1.04 (0.97, 1.11)
Obese	1.05 (0.97, 1.14)
SES	
5 (Highest)	Reference
4	1.08 (0.97,1.19)
3	1.13 (1.03, 1.25)
2	1.29 (1.17, 1.42)
1 (Lowest)	1.46 (1.33, 1.60)
Romano comorbidity score	1.13 (1.09, 1.17)

Table 5-6 Multivariable Cox regression analysis for secondary outcome, stroke

Variable	HR (95% CI)
OA	
No	Reference
Yes	1.14 (1.07, 1.22)
Hypertension	
No	Reference
Yes	1.63 (1.53, 1.73)
Diabetes	
No	Reference
Yes	1.48 (1.35, 1.63)
Hyperlipidemia	
No	Reference
Yes	0.91 (0.82, 1.02)
COPD	
No	Reference
Yes	1.28 (1.15, 1.41)
BMI	
Normal weight	Reference
Under weight	0.91 (0.79, 1.06)
Over weight	1.07 (0.99, 1.14)
Obese	1.15 (1.06,1.24)
SES	
5 (Highest)	Reference
4	0.99 (0.90, 1.09)
3	1.07 (0.98, 1.17)
2	1.13 (1.03, 1.24)
1 (Lowest)	1.18 (1.08, 1.29)
Romano comorbidity score	1.12 (1.08, 1.16)

Appendix D Effect of imputed BMI variable (Sensitivity analysis) in chapter 2

Body mass index (BMI) is a risk factor for CVD (78,81,82,84,115). BMI is a confounding variable in this study investigating OA-CVD relationship. In recent years researchers have attempted to account for the BMI in the analysis for which information was not available in the HAD. For example, in a longitudinal study using BC HAD, Rahman et al. imputed BMI categories of individuals in HAD (4). The imputation was done in accordance with the proportions observed among individuals in population-based health survey data who were grouped based on the osteoarthritis (OA) exposure, CVD outcome, and demographic variables (age and sex) (4). In a sensitivity analysis of a separate population-based study using BC HAD, Schmidt et al. used proportions of obesity among rheumatoid arthritis (RA) patients as well as general population observed in two separate survey datasets (5).

Assigning a BMI category to an individual in BC HAD similar to population-level proportion observed in survey data has few major limitations; (a) this does not account for the uncertainty associated with the imputation technique (b) such imputation does not account for the individual level data and (c) the resulting imputed values of BMIs may be unrealistic for individual patients (87,116).

We proposed an improved method compared to the currently published proportion based imputation method to account for BMI in studies involving HAD. In this study, we imputed BMI using multiple imputation to overcome these limitations of the conventional proportion based imputation method (117).

In this sensitivity analysis, we excluded the BMI variable from the model. We observed very little change in the risk of CVD associated with the OA exposure. For example, the aHR for the

primary outcome, CVD, was 1.25 (1.20, 1.31) compared to 1.23 (1.17, 1.28) with BMI into the model. There was only one to three percent change in the risk with or without BMI into the model for all study outcomes. Since BMI was entered as a covariable into a multivariable model, it is unlikely that one covariable would make a substantial change on the estimated risk. It appeared that the imputed BMI variable had very little potential to introduce bias in this analysis.

Table 5-7 Effect of imputed BMI variable on the hazard ratios for primary and secondary outcomes

Outcome type	Outcome	Unadjusted effect HR (95% CI)	Adjusted* effect HR (95% CI)	
			With BMI	Without BMI
Primary	CVD	1.31 (1.25, 1.37)	1.23 (1.17, 1.28)	1.25 (1.20, 1.31,)
Secondary	IHD	1.26 (1.18, 1.34)	1.17 (1.10, 1.26)	1.20 (1.13, 1.29)
	CHF	1.51 (1.42, 1.61)	1.42 (1.33, 1.51)	1.43 (1.34, 1.52)
	Stroke	1.22 (1.14, 1.29)	1.14 (1.07, 1.22)	1.16 (1.09, 1.24)
* Adjusted for SES, hypertension, diabetes, hyperlipidemia, COPD, & Romano comorbidity score.				

Using contingency tables, we investigated the crude relationship of the imputed BMI with the OA exposure and CVD outcome separately in our analytic data. The proportion of obese individuals was higher among people with OA compared to non-OA individuals, 29.35% versus 19.76%. Furthermore, we have used a logistic regression model for investigating the association between OA exposure and imputed BMI. After adjusting for other baseline co-variables including hypertension, diabetes, hyperlipidemia, COPD, Romano co-morbidity score and SES,

the adjusted OR and 95% CI was 1.93 (1.80, 2.07) for being obese. On the other hand, the risk of CVD was 20% higher among obese people compared to people with normal weight. The adjusted hazard ratio (95% CI) was 1.20 (1.15, 1.28). We observed that the imputed BMI in our analytic data was strongly associated with the OA exposure and posed a substantial risk for CVD outcome. Our findings are consistent with the evidence that obesity is an independent risk factor for both OA and CVD, therefore justifying this variable as a confounder in the OA-CVD relationship. Although there was no substantial change in the point estimate, the inclusion of this BMI variable in the outcome model decreased the AIC substantially (by more than four points).

Appendix E Distribution of study outcome

Table 5-8 Distribution of CVD outcomes by NSAID dispensing status

Distribution of study outcome	Total N (%)	Individuals dispensed ≥ 1 NSAID prescription N (%)	Individuals did not receive any NSAID prescription N (%)
Cardiovascular diseases (CVD)	1,660	1,147	513
Ischemic heart disease (IHD)	523 (31.50%)	355 (30.95%)	168 (32.75%)
Congestive heart failure (CHF)	532 (32.04%)	375 (32.69%)	157 (30.60%)
Stroke	605 (36.44%)	417 (36.36%)	188 (36.65%)