Excess Heart Disease Risk Associated with Arthritis in the Canadian General Population

by

Orit Schieir

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Public Health Sciences
University of Toronto

© Copyright by Orit Schieir 2017
Excess Heart Disease Risk Associated with Arthritis in the Canadian General Population

Orit Schieir
Doctor of Philosophy
Graduate Department of Public Health Sciences
University of Toronto
2017

Abstract

This thesis, through three manuscripts, aims to clarify to what extent arthritis may be an independent risk factor for developing heart disease in the general population, underlying mechanisms of the relationship, and who may be most at-risk. Manuscript 1 is a systematic review and meta-analysis of population-based studies that estimated risk of incident myocardial infarction (MI) associated with five major types of arthritis. Results showed that MI-risk was consistently increased across arthritis and was partially explained by a higher prevalence of traditional heart disease risk factors in each type of arthritis. Knowledge gaps from the review were used to inform subsequent secondary analyses of the longitudinal Canadian National Population Health Survey (NPHS) with 16-years of follow up. Manuscript 2 used discrete time survival analysis with time-varying lagged predictors to estimate effects of arthritis on first heart disease event occurrence. Potential variations by age, sex and activity limitation were examined. Results showed that arthritis was
independently associated with incident heart disease in all women, with more marked risk in women who also reported activity limitation, but only in men with activity limitation. Manuscript 3 used a novel approach to estimate potential mediating and moderating effects of activity limitation in the arthritis-heart disease relationship and possible differences by obesity and sex/gender in the same longitudinal health survey as manuscript 2. Results showed the proportion of heart disease risk explained by activity limitation varied between men and women, and between obese and non-obese women. Activity limitation explained nearly all of the heart disease risk associated with arthritis in men (90%), most of the heart disease risk in obese women (54%), and part of the heart disease risk in non-obese women (23%). Overall, the thesis results are important for increasing public and healthcare-provider awareness of excess heart disease risk generally associated with arthritis, highlighting sex/gender disparities, and identifying activity limitation as a modifiable intermediate target for prevention strategies to help reduce heart disease risk in people with arthritis.
For Ari Scheier (1925-2017) who instilled in me a passionate curiosity for learning
Acknowledgements

I would like to thank the members of my doctoral advisory committee: Elizabeth M. Badley, Sheilah Hogg-Johnson and Richard H. Glazier. I am grateful for your thoughtful feedback, your unwavering patience, and support throughout my time in the doctoral program. As opportunities arose to participate in parallel learning and professional development, you encouraged me to do so, and this has enriched my training experience immeasurably.

Thank you, Liza, for your mentorship, for always making yourself available to help work through any study challenges throughout the process, and for sharing your remarkable depth of knowledge and expertise in musculoskeletal epidemiology. Thank you, Sheilah, for helping me hone the technical and statistical skills needed to become a better researcher. Thank you, Rick, for helping me to understand the real-world clinical applications of my theoretical work.

Last, but not least, thank you to my family. To my son Stevie, thank you for your unbridled positivity. Your smiles and laughter were contagious even when I faced the most difficult challenges. To my unborn son, thank you for symbolically and physically (from within the womb) kicking me to complete my work. Finally, thank you to my husband for supporting me and picking up the slack at home when I was too busy. Your constant love and belief in me have provided encouragement throughout this process and pushed me to be better.
Table of Contents

**Chapter 1 Introduction** ............................................................................................................. 1

1.1 Overarching Thesis Goal, Objectives & Hypotheses ....................................................... 2

1.2 Reader’s Note .......................................................................................................................... 3

1.3 References ............................................................................................................................. 4

**Chapter 2 Background Literature Review** ............................................................................. 6

2.1 Overview of Arthritis .............................................................................................................. 6

2.2 Risk Factors for Arthritis ..................................................................................................... 7

2.3 Treatment ............................................................................................................................... 8

2.4 Impact .................................................................................................................................... 8

2.5 Arthritis-Heart Disease Comorbidity .................................................................................. 9

2.5.1 Mechanisms Linking Arthritis and Heart Disease ...................................................... 10

2.6 Moderating Effects of Sex and Gender on Arthritis-Heart Disease Relationships.................. 13

2.7 Summary of Research Gaps ................................................................................................. 14

2.8 Thesis Impact .......................................................................................................................... 15

2.9 References ............................................................................................................................. 16

**Manuscript 1**

**Chapter 3 Incident Myocardial Infarction Associated with Major Types of Arthritis in the General Population: A Systematic Review and Meta-Analysis** ............................................. 33

3.1 Preface ..................................................................................................................................... 33

3.2 Abstract .................................................................................................................................. 33

3.3 Introduction ............................................................................................................................... 35

3.4 Methods .................................................................................................................................. 36
3.5 Results ........................................................................................................................................ 39
3.6 Discussion .................................................................................................................................. 41
3.7 References .................................................................................................................................. 46

Manuscript 2

Chapter 4 Sex Variations in Effects of Arthritis and Activity Limitation on First Heart Disease Event Occurrence in the Canadian General Population: Results from the Longitudinal Canadian National Population Health Survey .................................................. 66

4.1 Abstract ....................................................................................................................................... 66
4.2 Introduction .................................................................................................................................. 67
4.3 Methods ...................................................................................................................................... 69
4.4 Results ........................................................................................................................................ 73
4.5 Discussion ................................................................................................................................... 74
4.6 References ................................................................................................................................... 80
4.7 Methods Addendum ..................................................................................................................... 93
  4.7.1 Elaborated Description of Discrete-Time Survival Method ........................................... 93
  4.7.2 Data Assumptions & Treatment of Missing Values .................................................... 95

Manuscript 3

Chapter 5 Moderating and Mediating Effects of Sex, Obesity and Activity Limitation on Associations between Arthritis and Heart Disease in the General Population ........ 96

5.1 Abstract ....................................................................................................................................... 96
5.2 Introduction .................................................................................................................................. 97
5.3 Methods ...................................................................................................................................... 99
5.4 Results ........................................................................................................................................ 105
5.5 Discussion ................................................................................................................................... 106
5.6 References ................................................................................................................................... 111
Chapter 6 Discussion and Conclusions ......................................................... 126

6.1 Key Findings .......................................................................................... 126

6.1.1 Evidence that Arthritis is a Risk Factor for Heart Disease in the General
Population ..................................................................................................... 126

6.1.2 Support for Activity Limitation as a Mediator in the Arthritis-Heart
Disease Relationship .................................................................................... 127

6.1.3 Support for Sex/ Gender Differences in Arthritis-Heart Disease
Comorbidity ................................................................................................. 128

6.1.4 Contributions to Understanding Heart Disease Risk in Osteoarthritis ..... 129

6.2 Limitations and Methodological Considerations ....................................... 130

6.3 Implications ............................................................................................ 138

6.4 Conclusion .............................................................................................. 141

6.5 Student’s Role ......................................................................................... 141

6.6 References ............................................................................................. 143
List of Tables

3.1 Characteristics of included studies ................................................................. 58

3.2 Univariate meta-regression comparing relative effects of arthritis disease type and potential heterogeneity in estimated associations with incident MI ............................ 64

3.3 Sub-group analyses of population-based studies estimating risk of incident myocardial infarction associated with major types of arthritis ................................................................. 65

4.1 Characteristics of the study sample at baseline NPHS Cycle 1 (1994/1995), by self-reported diagnosed arthritis, household respondents aged 18+ free of heart disease in 1994/1995, Canada (excluding territories) ................................................................. 89

4.2 Weighted adjusted odds ratios for arthritis on first occurrence of heart disease between 1996/1997 and 2010/2011 stratified by sex, household population aged 18+ and free of heart disease in 1994/1995, Canada (excluding territories) ................................................................. 90

4.3 Weighted adjusted odds ratios for arthritis by activity limitation on first occurrence of heart disease between 1996/1997 and 2010/2011 stratified by sex, household population aged 18+ and free of heart disease in 1994/1995, Canada (excluding territories) ................. 91

4.4 Weighted adjusted odds ratios for arthritis by activity limitation on first occurrence of heart disease between 1996/1997 and 2010/2011 stratified by age and sex, household population aged 18+ and free of heart disease in 1994/1995, Canada (excluding territories) 92

5.1 Multivariable logistic regressions predicting incident heart disease and activity limitation ................................................................................................................................. 121

5.2 Direct and indirect effects of arthritis mediated through activity limitation on incident heart disease ................................................................................................................................. 122
# List of Figures

3.1 PRISMA flow diagram summarizing systematic search results ................................. 57

3.2 Forest plots of population-based studies estimating risk of incident myocardial infarction associated with major types of arthritis ................................................................. 62

4.1 NPHS participant flow diagram ................................................................................ 88

5.1 Schematic showing how the total effect of arthritis on heart disease may be decomposed into a direct and posited indirect effect mediated through activity limitation ..................... 120
List of Appendices

Appendix A Conceptual Model of Heart Disease in Arthritis

Appendix B Supplemental Material Manuscript 1: Incident Myocardial Infarction Associated with Major Types of Arthritis in the General Population: A Systematic Review and Meta-Analysis

Appendix C Supplemental Material Manuscript 2: Sex Variations in Effects of Arthritis and Activity Limitation on First Heart Disease Event Occurrence in the Canadian General Population: Results from the Longitudinal Canadian National Population Health Survey

Appendix D Supplemental Material Manuscript 3: Moderating and Mediating Effects of Sex, Obesity and Activity Limitation on Associations between Arthritis and Heart Disease in the General Population


Appendix F Description of Stata mediation macro PARAMED

Appendix G University of Toronto Research Ethics Board Approval Letter
There has been a global increase in the prevalence and impact of chronic conditions (1). This increase is projected to continue as the world population ages and rates of obesity and other common chronic disease risk factors continue to rise (2, 3). The presence of multiple chronic diseases within the same individual (multimorbidity), particularly conditions that are prevalent, disabling, and associated with excess mortality, pose the greatest health burden (4). Arthritis is currently the second and third most prevalent chronic condition reported amongst Canadian women and men, respectively, and the leading cause of physical disability in North America (5). Persons with arthritis frequently report the presence of other chronic conditions, predominantly heart disease, that are associated with poorer arthritis-related and general health outcomes (6, 7).

In Canada, arthritis with heart disease is the second most common set of conditions reported amongst persons with two and three chronic conditions (8). While arthritis and heart disease share certain risk factors (e.g. age, obesity), there is increasing evidence that arthritis may be causally related to the development of heart disease through various pathways including relationships between arthritis inflammation, arthritis medications and arthritis-attributable activity limitation with multiple traditional heart disease risk factors. Furthermore, as there are noted sex/gender differences in the prevalence, presentation, and prognosis of both diseases, further investigation of potential causal pathways according to sex/gender is warranted.
1.1 Overarching Thesis Goal, Objectives & Hypotheses

The overarching goal of this thesis is to better understand overall and intermediate relationships operating through activity limitation pathways between two of Canada’s most burdensome chronic conditions, arthritis and heart disease, as well as potential sex/gender differences in these relationships in the general population.

To accomplish this goal, a knowledge synthesis of population-based studies reporting on associations between major types of arthritis and an incident objective heart disease outcome (myocardial infarction) was first undertaken. Key findings and knowledge gaps from the synthesis were used to inform secondary analyses of longitudinal Canadian national health survey data to obtain robust estimates of effects of arthritis on first heart disease event occurrence, mediating effects of arthritis on heart disease operating through activity limitation, and sex/gender differences in these relationships.

Specific thesis objectives were to:

**Objective 1**: Perform a systematic review and meta-analysis of best-available evidence from population-based studies to quantify and compare risks for incident myocardial infarction across five major types of arthritis (rheumatoid arthritis, ankylosing spondylitis, psoriatic arthritis, gout, and osteoarthritis).

**Objective 2**: Estimate the extent to which arthritis, with and without activity limitation, is associated with time to first heart disease event occurrence in the adult Canadian general population, and examine potential sex/gender differences in these relationships.
Objective 3: Decompose total effects of arthritis on heart disease into direct and indirect effects mediated through activity limitation in the adult Canadian general population, and examine potential sex/gender differences in direct and indirect pathways.

1.2 Reader’s Note

This study is reported as a manuscript-based thesis. Chapter 2 provides a background literature review for the thesis. Chapters 3, 4, and 5 each contain a manuscript addressing the three objectives of the thesis. Elaborated methods for the manuscripts presented in Chapters 4 and 5 appear as methodological addendums at the end of each respective chapter. Chapter 6 provides a unified discussion of key findings from the three manuscripts, thesis implications, limitations and conclusions.

As sex and gender refer to separate but often related concepts (sex referring to biological and physiological factors and gender to socio-behavioural factors), “sex/gender differences” is used as common terminology throughout the thesis with the exception of the manuscript presented in Chapter 4 where only “sex differences” is used due to journal editorial policies.
1.3 References


Chapter 2
Background Literature Review

2.1 Overview of Arthritis

Arthritis consists of more than 100 conditions that affect the joints and/or surrounding tissues causing joint pain, swelling, stiffness and frequent disability (1). Current population-based prevalence estimates for arthritis are approximately 1 in 6 for Canada (16% representing approximately 4.2 million Canadians) (1) and approximately 1 in 4 for the US (23% representing over 54 million Americans) (2). Almost two-thirds of people living with arthritis are women and though arthritis prevalence increases with age, nearly 60% of people with arthritis are under the age of 65 (1, 3). A brief description of major types of arthritis most commonly affecting the adult general population is provided below.

*Osteoarthritis* (OA) typically affects the joints in the knees, hips and/or hands and is the most common chronic joint disease accounting for 60-70% of all arthritis cases. OA affects approximately 10-12% of the adult general population and over 40% of the population aged 65+ (4-7). *Gout* is characterized by recurrent acute episodes of inflammatory arthritis typically of the joints in and around the foot (though other joints can be affected) and an overproduction of uric acid in the body (hyperuricemia). Gout is the most prevalent type of inflammatory arthritis affecting 1-4% of the general population in North America and Western Europe (1, 8, 9). *Rheumatoid arthritis (RA)* is a persistent systemic joint disease typically affecting the small joints in the hands and feet (though other joints can also be affected) causing swelling and progressive joint damage; it affects 0.5-1% of the general
population (10). *Spondyloarthropathies (SpA)*, the most common of which are ankylosing spondylitis (AS), an inflammatory arthritis affecting the joints of the spine, and psoriatic arthritis (PsA), an inflammatory arthritis common in persons with psoriasis, together affect 0.1-0.5% of the adult general population (11).

### 2.2 Risk Factors for Arthritis

Non-modifiable risk factors for arthritis include genetic markers, age, sex and female hormones. Prevalence of arthritis increases with age and most types of arthritis are at least twice as common in women relative to men except for ankylosing spondylitis and gout which are more common in men relative to women (1). Gout risk, however, increases sharply in women after menopause (8). Modifiable risk factors have been associated with individual types of arthritis. Infections, smoking, and higher BMI/obesity are associated with a higher risk of developing RA (12-15). High BMI/obesity, vigorous physical activity, certain occupations, and prior joint injury can all increase stress on the joints, and have been associated with the development of OA (12, 16, 17). Hyperuricemia (sustained high levels of serum uric acid), high BMI/obesity, dietary factors (e.g. alcohol intake, meat and seafood purines, fructose), certain chronic conditions (e.g. hypertension, kidney disease, diabetes, etc.), and certain medications used to treat some of these conditions (e.g. diuretics) have been associated with increased risk for gout (8, 18, 19). AS and other SpA conditions have been most consistently linked with genetic factors, though there is mixed evidence suggesting associations with GI and other infections (20, 21). Emerging but still inconclusive evidence suggests possible associations between gut microbiomes and several types of arthritis including OA, RA, gout and SpA (22).
2.3 Treatment

Currently there is no cure for arthritis. Arthritis treatment generally focuses on relieving pain symptoms, controlling inflammation, minimizing joint damage, and improving function (23, 24). Pharmacological treatments for arthritis include use of analgesics, non-steroidal anti-inflammatory drugs (NSAIDs), coxibs, glucocorticoids (GCs), and conventional and biologic disease modifying anti-rheumatic drugs (DMARDs). While analgesics, NSAIDS, coxibs, and glucocorticoids are common treatments used across arthritis subtypes to relieve arthritis symptoms such as pain and swelling, conventional and biologic DMARDs targeting general and specific inflammatory disease processes, respectively, are mostly used to treat RA and SpA (23, 25). In addition to medications used to control pain and swelling, patients with gout may be treated with urate-lowering therapies to treat underlying hyperuricemia and/or oral colchicine (26, 27). Non-pharmacological treatments for arthritis can include joint surgeries, physical and/or occupational therapy, and self-management programs promoting positive health behaviors related to achieving and/or maintaining a healthy weight, diet, exercise, sleep, and relaxation (1, 28, 29).

2.4 Impact

Arthritis is prevalent, disabling and costly to individuals and society (7). Arthritis is currently the second and third most prevalent chronic condition reported amongst women and men in Canada, respectively (1). Arthritis symptoms including joint pain, stiffness, and swelling can have a devastating impact on an individual’s physical and emotional health, social relationships and participation, work productivity, and overall quality of life (3, 30-32). Arthritis is the leading cause of disability in North America (33); over 44% of people with
arthritis report arthritis-attributable activity limitations (2) and over 50% of working-age adults (age 25-64 years) with arthritis-attributable activity limitation report leaving the labour force (30). Annual economic costs for arthritis were estimated at 33 billion dollars in Canada (34) and over 128 billion dollars in the US (35), most of which are indirect costs due to wage losses and management of comorbidities. The prevalence and burden of arthritis is projected to continue to increase over the next 30 years with population aging, increased longevity, reduced physical activity, and rising obesity rates (34, 36, 37).

2.5 Arthritis-Heart Disease Comorbidity

Population-based estimates suggest that arthritis is associated with an approximate 2–fold increase in the odds of having another chronic condition (prevalence odds ratio (POR) 2.1, 95% CI: 1.7-2.5) (38) and half of people with arthritis have at least one other condition (39). In a US population-based study examining overall prevalence and specific patterns of multimorbidity, arthritis was part of the most prevalent dyads and triads of conditions reported in all age-sex groupings examined (41). Overall, arthritis appeared in half (15/30) of dyads and over 83% (25/30) of triads studied (41). Heart disease is one of the most prevalent co-occurring conditions with arthritis (39, 40). Persons with arthritis and heart disease report a myriad of poorer health outcomes including greater disability, depression, healthcare costs, and mortality, relative to persons with either condition alone (35, 42-44).

The co-occurrence of arthritis and heart disease may result by chance, may be due to the presence of shared risk factors, or may be causally related (45). Several cross-sectional and case-control studies support significant independent associations between arthritis and heart disease, suggesting that their co-occurrence is greater than what would be expected by
chance alone (40, 46-49). While arthritis and heart disease share some common risk factors (e.g. age, sex, smoking, obesity), there is increasing evidence that arthritis may be causally related to the development of heart disease.

2.5.1 Mechanisms Linking Arthritis and Heart Disease

Rheumatoid arthritis (RA) has been the most consistently studied type of arthritis in relation to cardiovascular morbidity and mortality (50-53) and has become recognized as an independent risk factor for heart disease similar in magnitude to diabetes (54-56). These findings have prompted the development of practice recommendations from rheumatology associations around the world aimed at reducing cardiovascular risk in patients with RA (57). Though many of the proposed mechanisms linking RA and heart disease also extend to other forms of arthritis, both overall risk for heart disease and specific cardioatherogenic mechanisms have received far less study and attention in other types of arthritis. The following section reviews potential common mechanisms across arthritis conditions that may lead to increased heart disease risk. A conceptual model summarizing common pathways for developing heart disease in persons with arthritis is presented in Appendix A.

- **Inflammation:** Systemic and chronic low-grade inflammation is common across arthritis disease types and associated with increased cardiovascular risk (58-60). Systematic inflammation via effects of pro-inflammatory molecules including acute phase reactants (e.g. c-reactive protein (CRP), erythrocyte sedimentation rate (ESR)), inflammatory cytokines (e.g. tumor necrosis factor (TNF), interleukin-6 (IL-6), interleukin-1(IL-1)), and fibrinogen contribute to the formation and progression of atherosclerosis leading to acute ischemic events (61). Systemic inflammation is more common in patients with RA,
spondyloarthopathies, and gout. Further, increased adiposity and local inflammation at affected joint site(s) in patients with osteoarthritis have been found to trigger inflammatory mediators resulting in chronic low-grade inflammation also associated with increased heart disease risk (62-66).

- **Arthritis treatment:** Arthritis medications can have variable effects on heart disease risk, with different medications increasing, decreasing, or having dual-effects on risk. The most common medications used as first-line therapy to relieve symptoms of pain and swelling across multiple types of arthritis including glucocorticoids (GCs), non-aspirin NSAIDs, and coxibs have been independently associated with increased heart disease risk (67-69). Urate-lowering therapies used to treat hyperuricemia in gout have been associated with reduced heart disease risk (70). Conventional disease modifying antirheumatic agents (DMARDs) such as methotrexate that treat generalized systemic inflammation and biologic DMARDs that target specific inflammatory pathways in RA and SpA have been associated with decreased risk (71, 72). Some of the medications above that suppress systemic inflammation may also increase lipid levels in rheumatic disease patients. Evidence to date however suggests that the cardiac benefits of suppressing inflammation outweigh the potential harms of increasing lipid levels given more modest associations between lipids and heart disease in rheumatic disease patients than in the general population (61).

- **Effect of arthritis and arthritis-attributable activity limitation on traditional heart disease risk factors:** Several traditional heart disease risk factors have been reported to be more prevalent in arthritis populations. Cross-sectional population-based estimates from the US Behavioral Risk Factor Survey (BRFSS) showed that persons aged 45 and
over with self-reported physician-diagnosed arthritis, had significantly elevated rates of activity limitation, and obesity after adjusting for socio-demographic variables (73). US population-based prevalence estimates from NHANES (49) showed that hypertension is present in approximately 40% of OA patients, smoking in 20%, diabetes in 11%, high total cholesterol in 32%, and low high-density lipoprotein cholesterol in 13%. Clinical studies have also reported associations between OA with decreased aerobic exercise, greater mean BMI, larger waist-to-hip ratios (WHR), higher systolic blood pressure, higher blood sugar, and lower mean high-density-lipoprotein cholesterol (74, 75).

Smoking, hypertension, dyslipidemia, insulin resistance, and obesity have also been reported to be highly prevalent and cluster in patients with RA (76, 77). Further, persons with increasing numbers of both markers of RA disease severity and traditional heart disease risk factors were reported to be at the highest risk of developing heart disease in a clinical sample of RA patients followed for 22 months (78).

Some traditional heart disease risk factors are also risk factors for developing certain types of arthritis (e.g. smoking, obesity) and may be more likely to precede arthritis onset. Arthritis however can also lead to negative changes in co-occurring heart disease risk factors or contribute to the onset of new traditional heart disease risk factors. For instance, systemic and chronic low-grade inflammation can affect lipid levels, insulin resistance and blood pressure, and OA and RA have been associated with a higher incidence of metabolic syndrome (61, 79-81). Importantly, more than 40% of people with any type of arthritis report activity limitation due to their arthritis associated with increased sedentary time and physical inactivity (2, 82, 83). In addition to known downstream effects of sedentary time and physical inactivity on blood sugar, blood
pressure, and weight/adiposity, arthritis attributable activity limitation may also lead to greater exposure to, or more prolonged use of GCs, non-aspirin NSAIDs, and coxibs which also negatively impact lipids, blood pressure, and/or body weight (84). Therefore arthritis $\rightarrow$ activity limitation $\rightarrow$ heart disease may represent an indirect pathway linking arthritis and heart disease.

2.6 Moderating Effects of Sex and Gender on Arthritis-Heart Disease Relationships

There are notable differences between men and women in many aspects of health including disease prevalence, symptoms, impact, access to health care, and response to therapy (85). In an effort to better understand sex and gender differences and to increase equity in healthcare, several Canadian and international healthcare organizations have issued policies calling for sex and gender-based analyses in research (86). Sex typically refers to differences in biological and physiological processes in females and males, whereas gender to differences in socially constructed roles, relationships, and behaviours. While sex and gender are separate constructs they often affect one another (87).

Sex and gender may have moderating effects on relationships between arthritis and heart disease given: a) that most types of arthritis are more prevalent in women (1); b) reports of disproportionate impacts of arthritis in women with respect to immune and inflammatory responses, radiographic progression, pain sensitivity, mental and emotional health (66, 88, 89); c) more functional limitations due to arthritis in women (90) and in obese women with arthritis in particular (91, 92); d) decreased access to and utilization of arthritis treatments in
women (93, 94); and e) sex differences in baseline risk, first manifestations and symptoms, and outcomes from heart disease (95, 96).

Sex and gender may have variable influences on relationships between different types of arthritis with heart disease and their underlying mechanisms. For instance, relative to men, women and obese women in particular, report greater functional limitations and less physical activity (91, 92) and are more susceptible to developing knee and hand OA with links to inflammation (97, 98). These disparities may potentially result in more pronounced heart disease risk in women with OA.

2.7. Summary of Research Gaps

Current knowledge regarding arthritis-heart disease comorbidity has primarily focused on heart disease risk in RA, resulting in under-recognition of possible heart disease risk in other more prevalent types of arthritis in the general population. Methodological limitations in previous studies including cross-sectional study designs, failure to exclude prevalent heart disease at study entry, and lack of adjustment, or inappropriate adjustment of covariates that may be on the causal pathway, have greatly limited the ability to draw conclusions regarding potential causal relationships between arthritis and heart disease. No studies have formally estimated the potential mediating effect of activity limitation and its relative contribution to total heart disease risk. Sex/gender differences in overall relationships and intermediate pathways have also not been explored. Large longitudinal population-based studies with long-term follow up are needed to determine whether, when, how, and in whom arthritis may increase heart disease risk in the general population.
2.8 Thesis Impact

Arthritis and heart disease are amongst the most burdensome chronic conditions in Canada and worldwide. Their independent and combined impact on individuals and society is projected to rise in coming decades with population aging and improved survival. Determining whether and to what extent arthritis increases heart disease risk is needed to improve quality of care and health outcomes of persons living with these conditions. Furthermore, determining which population sub-groups are most at-risk and identifying intermediate targets that are modifiable, can help inform prevention strategies to reduce population chronic disease rates.
## 2.9 References


Chapter 3
Incident Myocardial Infarction Associated with Major Types of Arthritis in the General Population: A Systematic Review and Meta-Analysis

Orit Schieir, Cedomir Tosevski, Richard H Glazier, Sheilah Hogg-Johnson, Elizabeth M. Badley

3.1. Preface

The present chapter includes a published systematic review and meta-analysis of best available evidence from population-based studies estimating associations between five major types of arthritis and an objective incident heart disease outcome (myocardial infarction). The review was undertaken to quantify and appraise current evidence pertaining to heart disease risk across arthritis disease types, and to identify knowledge gaps and future directions that should be addressed in planned secondary analyses of national longitudinal health survey data (presented in Chapters 4 and 5).

3.2 Abstract

Objective: To synthesize, quantify and compare risks for incident myocardial infarction (MI) across five major types of arthritis in population-based studies.

Methods: A systematic search was performed in MEDLINE, EMBASE, and CINAHL databases with additional manual/ hand searches for population-based cohort or case-control studies published in English or French between January 1980 and January 2015 with a
measure of effect and variance for associations between incident MI and five major types of arthritis: rheumatoid arthritis (RA), psoriatic arthritis (PsA), ankylosing spondylitis (AS), gout or osteoarthritis (OA), adjusted for at least age and sex. All search screening, data abstraction and quality appraisals were performed independently by two reviewers. Where appropriate, random-effects meta-analysis was used to pool results from studies with a minimum of 10 events.

**Results:** We identified a total of 4,285 articles; 27 met review criteria and 25 criteria for meta-analyses. In studies adjusting for age and sex, MI-risk was significantly increased in RA (pooled RR: 1.69, 95% CI: 1.50-1.90), gout (pooled RR:1.47, 1.24-1.73), PsA (RR:1.41, 95% CI: 1.17-1.69), OA (pooled RR:1.31, 95% CI:1.01-1.71) and tended towards increased risk in AS (pooled RR: 1.24, 95% CI: 0.93-1.65). Traditional risk factors were more prevalent in all types of arthritis. MI-risk was attenuated for each type of arthritis in studies adjusting for traditional risk factors and remained significantly increased in RA, PsA and gout.

**Conclusion:** MI-risk was consistently increased in multiple types of arthritis in population-based studies, and was partially explained by a higher prevalence of traditional risk factors in all types of arthritis. Findings support more integrated cardiovascular prevention strategies for arthritis populations that target both reducing inflammation and enhancing management of traditional CV-risk factors.
3.3 Introduction

Despite reductions in mortality over time, ischemic heart disease (IHD) is a major contributor to the burden from chronic diseases and the leading cause of death world-wide (1). A number of modifiable lifestyle factors including smoking, physical inactivity, overweight/obesity, diabetes, hypertension and hyperlipidemia have long been shown to increase the risk of developing heart disease and are common targets for prevention strategies (2). More recently, inflammation has also been shown to play an integral role in the development, instability and rupture of atherosclerotic plaques leading to acute ischemic events (3, 4). This has prompted extensive research examining heart disease risk in inflammatory conditions.

Arthritis is a prevalent inflammatory joint disorder affecting the aging population characterized by joint pain, stiffness and frequent disability (5, 6). Several systematic reviews have examined whether individual types of arthritis (rheumatoid arthritis (RA) (7), ankylosing spondylitis (AS) (8), psoriatic arthritis (PsA) (9), gout (10), and osteoarthritis (OA) (11)) are independently associated with various cardiovascular outcomes, with the strongest most consistent evidence pointing towards relationships between RA and myocardial infarction (MI) (7). Previous reviews however included studies with variable study populations, sources for controls, designs, outcomes and definitions and had minimal study inclusion criteria for control of confounding. Further, focus on individual types of arthritis in each review did not allow for comparisons of IHD-risk across different types of arthritis. These limitations make it difficult for clinicians, patients and policy makers to get a clear sense of whether and to what extent IHD-risk is increased across different arthritis
populations and may result in missed opportunities for preventing or lowering IHD-risk, particularly in non-RA types of arthritis.

The present systematic review and meta-analysis was undertaken to synthesize best available evidence from population-based studies to quantify and compare risks for incident myocardial infarction in five major types of arthritis (RA, AS, PsA, gout and OA).

3.4 Methods

This review was carried out and reported according to Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) (12) and the Meta-Analysis of Observational Studies in Epidemiology (MOOSE) (13) guidelines.

**Data sources and searches:** A systematic search developed by an information specialist for population-based studies estimating associations between arthritis and incident MI was performed in MEDLINE, EMBASE and CINAHL databases. Key word and major subject headings were specified for: arthritis and for major types of arthritis (RA, AS, PsA, gout and OA), MI or acute coronary syndrome (ACS), and cohort or case-control study designs.

Search results were limited to studies published through January 2015, adults (19+) and English and French languages (sample MEDLINE search strategy provided as supplementary material). In addition to the database search, we performed manual web searches (e.g. Google scholar) and hand searched bibliographies of retrieved studies, review articles, and screened conference abstracts from 2013 and 2014 American and European rheumatology meetings (attempted to contact potential abstract authors by email twice two-weeks apart) to
identify potentially relevant studies that were accepted or in press in a peer-reviewed journal by January 2015.

**Study selection:** We included cohort and case-control studies that reported a measure of association (i.e. odds ratio, risk ratio, hazard ratio) and variability (standard error or 95% confidence interval) for effects of arthritis on incident MI or ACS defined according to established clinical criteria, physician diagnosis, validated administrative billing codes or self-reported physician diagnosis, and adjusted for at least age and sex. MI was selected as a common objective IHD outcome across studies for meta-analysis. Studies were excluded if the outcome was a composite CVD endpoint or if the study did not include a population-based comparison group. In the event of multiple publications from the same data source, we included studies if extractable estimates were reported for different types of arthritis, but only included the most recent publication if estimates were reported for the same type of arthritis.

**Data extraction and quality assessment:** The following information was abstracted from each study: country, funding, study design, data source, sample size, follow-up period, sample age range, percent female, exposure and outcome ascertainment, and measures of association with 95% confidence intervals, using a predetermined data abstraction template. Study quality was evaluated using the Newcastle-Ottawa Scale (NOS) (14). The NOS is an 8-item instrument (total score range 0–lowest to 9–highest quality) evaluating risk of bias in observational studies in relation to three domains: selection of study groups (range 0-4), comparability of study groups (range 0-2), and exposure/outcome ascertainment (range 0-3). NOS quality scores are presented as part of descriptive summaries for each study and did not influence decisions to pool studies in meta-analysis.
All search screening, data abstraction and quality appraisals was performed independently by 2 reviewers [OS, CT], with discrepancies resolved by consensus (n=3).

**Data synthesis & analysis**: In order to minimize bias that may result from combining small study effects with high variability, only studies with at least 10 events were eligible to be pooled in meta-analyses. Further, final decisions on whether or not to pool studies were based on quantitative assessments of heterogeneity described below and qualitative assessments based on the number of studies, and the consistency in direction/ magnitude of effect estimates across studies so that pooled results would be more interpretable (15).

The relative risk (RR) was selected as a common measure of association across studies. Random-effects meta-analysis with inverse variance weighting (16) was used to obtain pooled RRs and 95% confidence intervals (CI) by type of arthritis, first in studies adjusting for age and sex only, and then for studies also adjusting for at least one of the following traditional risk factors (RF): smoking, obesity/ BMI, physical activity, hyperlipidemia, diabetes, and high blood pressure. Heterogeneity among studies was assessed with Cochrane’s Q statistic and I² statistic representing the percentage of heterogeneity across studies attributable to between-study differences (17).

We used univariate random-effects meta-regression to compare associations with MI across arthritis disease types as well as examined how much arthritis disease type potentially contributed to overall heterogeneity between studies. Additional, planned subgroup analyses to examine other potential sources of heterogeneity between studies included comparisons by age, sex, inception vs. prevalent arthritis cohorts, calendar period and geographic region. Changes in pooled effect sizes after excluding each individual study were performed to
assess if any single study was strongly influencing pooled results. We assessed publication
bias graphically using a funnel plot where the natural log of the ratio of the relative risk was
plotted against its standard error, and statistically with the Egger test (18). All analyses were
performed using Stata v. 12 (StataCorp, College Station, TX).

3.5 Results

**Literature search and study selection:** Figure 3.1 summarizes study screening and
selection results. The search strategy identified 4,285 articles of which 27 met all criteria to
be included in the systematic review. Two articles reported results based on fewer than 10
outcomes (19, 20), leaving 25 articles eligible for meta-analysis (RA: 13 (21-33), AS: 3 (34-
36), PsA: 1 (28), Gout: 6 (37-42), OA: 2 (43, 44)).

Characteristics of included studies published between 1988 and 2015 are presented in Table
3.1: 26 were cohort studies (9 prospective; 17 retrospective) and 1 was a case-control study.
Study populations were from the US (8), UK (7), Sweden (5), Canada (3), China (3), and
Denmark (1) with participant follow-up ranging from 1 to 46 years. Thirteen studies included
inception cohorts and 14 prevalent cohorts with arthritis. Of the 24 studies that reported age
entry criteria, 16 had no adult age limits, and 8 were restricted to middle-aged and/or older
adult samples. Higher female to male sample ratios were reported in studies of RA and OA,
higher male to female sample ratios in studies of gout, and variable sex ratios in studies of
AS and PsA. Traditional risk-factors were more prevalent in all types of arthritis in studies
with internal population-based comparisons.
Associations between incident MI and all major types of arthritis: Overall, 22/27 (82%) studies reported higher risks for MI in arthritis groups relative to the general population. Cumulative incidence of MI based on 6,466 MIs in a combined arthritis sample of 226,962 from 25 studies with available data was 2.85% (95% CI: 2.78%-2.92%). Combined, all five types of arthritis were associated with an average 50% increased risk for MI based on 23 studies adjusted for age and sex only (Figure 3.2a), and an average 30% increased risk for MI based on 17 studies adjusted for at least one traditional risk factor (Figure 3.2b), though quantitative estimates of between-study heterogeneity were high and significant.

Subgroup analyses of incident MI by type of arthritis: We performed subgroup analyses by arthritis disease type. In studies adjusted for age and sex only, risk of incident MI was significantly increased in RA, gout, PsA, OA, and tended towards increased risk in AS (Figure 3.2a). Associations with MI were attenuated for all types of arthritis in studies that adjusted for traditional risk-factors, and remained significant for RA, gout, and PsA (Figure 3.2b). Between-study heterogeneity was high and significant in RA, gout and AS in studies adjusting for age and sex only, however heterogeneity between studies was much lower and no longer significant for RA and gout in studies with adjustment for traditional RF. Only two studies in OA and AS, respectively, included adjustment for traditional RF and the direction or magnitude of results was highly inconsistent between studies so pooling was not performed for either type of arthritis.

In meta-regression based on studies adjusting for age and sex only, estimated risks of MI in PsA, gout, AS and OA did not differ significantly from RA, and arthritis disease type was not a significant factor driving heterogeneity between studies (p=0.198) (Table 3.2). In meta-
regression analyses of studies adjusting for RF limiting to RA, gout and PsA only (OA and AS omitted due to small number of studies & inconsistency across studies), type of arthritis was a significant factor driving heterogeneity between studies (p= 0.0035) and associated risks for MI in gout were on average 20% lower than in RA (Table 3.2).

**Other subgroup and sensitivity analyses:** Results of planned subgroup analyses examining other potential sources of between-study variance including age, sex, arthritis duration, calendar period, and geographic region are presented in Table 3.3. Results of subgroup analyses from studies providing age-stratified estimates that could be harmonized across studies showed that while arthritis was associated with increased risks for incident MI across the adult age span, relative risks tended to be highest in young, then middle-aged and older adults. Among studies that provided sex-stratified estimates, pooled relative risks for MI were significantly increased in both women and men. Point estimates were consistently higher in women than in men, but confidence limits between studies overlapped and differences were inconclusive. There were no differences in effects of arthritis by duration, calendar period or region.

**3.6 Discussion**

Review results showed that risk of incident MI-risk was consistently increased across multiple major types of arthritis in population-based studies. Traditional CV RF were also more prevalent in all types of arthritis under study and explained part of the added risk for MI associated with each type of arthritis. Included studies were all of moderate to high quality based on NOS quality scores. Pooled estimates in RA and gout were based on strong and consistent evidence from multiple studies, and for PsA based on one large high quality
study. Evidence in OA and AS however was relatively sparse and there were too few studies with inconsistent results with adjustment for RF to be combined in meta-analysis. Visual examination of the funnel plot did not reveal any notable asymmetry [provided as supplementary material] and the statistical Egger test was not significant (p = 0.255) suggesting a low likelihood for publication bias.

Review results have important implications for clinical practice and for planning health services given current and projected increases in arthritis prevalence, and rising trends in obesity and other heart disease RF in the general population. First, contrary to RA where cardiovascular risk has been extensively studied and increased risk is generally well-accepted (45), risk in other types of arthritis have received far less attention and are likely under-recognized. The present study reported consistent increased risk of MI across multiple common arthritis disease types. The general public and healthcare providers should be made aware of more generalized risks for MI associated with arthritis. Second, results that traditional RF were more prevalent in all five types of arthritis under study and consistently explained part of the added risk for MI in each type, point to better management of traditional CV RF as an indirect pathway to prevent or lower MI risk in arthritis populations. This is important for patients with inflammatory arthritis as clinicians may focus on controlling inflammation as a sole means of reducing risk. Similarly, in other types of arthritis with a lower inflammatory burden, patients’ may not be perceived as higher-risk despite having several traditional RF, some of which may evolve over time secondary to arthritis. For example, disability may lead to lower physical activity or weight gain (46, 47), and certain arthritis medications may improve symptoms of inflammation but have adverse effects on weight and blood pressure (48-50). Lastly together results of meta-analyses of
studies with and without adjustment for traditional RF support common mechanisms leading to increased risk across different types of arthritis (i.e. inflammation, traditional RF). However, the relative contribution of these mechanisms may vary by type of arthritis. For example, there was a greater difference in estimated risk for MI in gout than in RA between studies that did and did not adjust for RF. The direct contribution of systemic inflammation may be higher in RA than in gout, and conversely the role of traditional RF may be greater in gout.

Many management guidelines have been developed for individual types of arthritis though few, predominantly those for RA, include recommendations for cardiovascular prevention (51). Recently, there has been a shift towards developing more integrated cardiovascular prevention strategies for inflammatory conditions. Joint recommendations for managing cardiovascular comorbidity in RA, AS and PsA were first developed by the European League Against Rheumatism (EULAR) in 2010 (52). Recommendations for managing common comorbidities in RA, PsA and psoriasis were also made by Canadian rheumatology and dermatology expert panels in 2015 (53). Gout and OA however were not included in either sets of guidance. Current results combined with increasing prevalence, high levels of OA-attributable disability, associations with obesity and physical inactivity, and prolonged treatment with NSAIDs/Coxibs (54-57), would suggest that prevention strategies recommended for other types of arthritis may help reduce risk in gout and OA as well.

Targeted education and more effective implementation of CV risk management will be important to help mitigate heart disease risk in arthritis populations. It is now recommended that rheumatologists screen / manage traditional heart disease RF given their expertise with complex arthritis treatment regimens and emerging evidence that traditional RF can affect
treatment response to certain therapies (53). However, rheumatologists may have limited
time in a standard visit to incorporate additional comprehensive cardiovascular assessments
and continued monitoring, as well as keep up with evolving evidence-based standards for
primary and secondary cardiovascular prevention. Further, given current shortages of
rheumatology specialists in several geographic regions, many types of arthritis, particularly
OA and gout are often managed outside of rheumatology (58, 59). These issues point to the
need for coordinated efforts between rheumatologists, internists, primary care and other
allied health professionals to ensure that excess cardiovascular-risk in people with arthritis is
appropriately managed.

Strengths of the present review include systematic identification and synthesis of best-
available evidence from population-based studies used to estimate effects of five prevalent
types of arthritis on MI. This review is the first to compare MI-risk across different types of
arthritis in studies with adjustment for age and sex only, and with adjustment for traditional
RF, respectively.

Limitations should also be addressed. Relatively few studies were identified for PsA, AS and
OA, and pooled risks of MI for AS and OA adjusted for traditional RF could not be
estimated due to the small number of studies and inconsistency in results between studies.
Limiting to studies published in English or French could have led to some bias in pooled
effect estimates, though tests for publication bias were not significant. While we excluded
clear duplicate publications, there could have been some overlap in study populations in
more than one study from the same country. However, pooled estimates were robust to
sensitivity analyses that eliminated individual studies (available as supplementary material).
Differences in the types and proportions of patients being treated were not examined and
could have contributed to heterogeneity. Despite differences in baseline risks and characteristics of arthritis and heart disease by age and sex (60), less than half of included studies tested for these interactions or reported stratified results. Over a third of studies did not adjust for any traditional RF and studies that did were often limited to baseline measures or lacked information on smoking, BMI/obesity and physical activity, increasing potential for residual confounding in pooled estimates. Given the common mechanisms linking arthritis and MI discussed above, risks may also extend to other less prevalent types of arthritis (e.g. systemic lupus erythematosus), inflammatory conditions other than arthritis, and to other vascular end-points but were beyond the scope of the present review and should be explored in future studies. Large longitudinal population-based studies with repeated measures examining effects of different types of arthritis, particularly OA, on IHD and other vascular end-points are needed. Stratified analyses by age and sex and mediation analyses examining direct and indirect pathways (and potential variations by type of arthritis) would help inform primary and secondary prevention strategies.

In conclusion, the present study estimated risk of incident MI associated with five major types of arthritis from population-based studies. Results showed that MI-risk was consistently increased in multiple types of arthritis, and was partially explained by a higher prevalence of traditional RF in all types of arthritis under study. Study findings support more integrated cardiovascular prevention strategies for arthritis populations that target both reducing inflammation and enhancing management of traditional CV RF.
3.7 References


44. Rahman MM, Kopec JA, Anis AH, Cibere J, Goldsmith CH. Risk of cardiovascular

45. Choy E, Ganeshalingam K, Semb AG, Szekanecz Z, Nurmohamed M.
Cardiovascular risk in rheumatoid arthritis: recent advances in the understanding of the
pivotal role of inflammation, risk predictors and the impact of treatment. Rheumatology
24907149.

46. Forman-Hoffman VL, Richardson KK, Yankey JW, Hillis SL, Wallace RB,
Wolinsky FD. Impact of functional limitations and medical comorbidity on subsequent

47. Hootman JM, Murphy LB, Helmick CG, Barbour KE. Arthritis as a potential barrier
to physical activity among adults with obesity—United States, 2007 and 2009. Morb Mortal

48. 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;330(7504):1366. doi:


Figure 3.1 PRISMA flow diagram summarizing systematic search results

Records identified through database searching (n = 4260)

Additional records identified (n = 25)

Records after duplicates removed (n = 3,397)

Title/abstract records screened (n = 3,397)

Records excluded (n = 3,140)
- Composite endpoint (39)
- Prevalent MI (17)
- No risk estimate for arthritis (41)
- No population-based controls (26)
- Narrative review/editorial, case reports, abstract, press release (104)
- Duplicate study population (3)

Full text articles assessed for eligibility (n = 257)

Total articles excluded (n = 230)

Studies included in qualitative systematic review (n = 27)

Articles excluded from meta-analysis due to fewer than 10 events (n = 2)

Studies included in quantitative meta-analyses (Total n = 25)
- Rheumatoid arthritis (13)
- Ankylosing Spondylitis (3)
- Psoriatic Arthritis (1)
- Gout = (6)
- Osteoarthritis (2)
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Funding</th>
<th>Design/ Data source</th>
<th>Study Period</th>
<th>Age range</th>
<th>Arthritis Female %</th>
<th>Arthritis Ascertainment</th>
<th>MI Type &amp; Ascertainment</th>
<th>Arthritis n</th>
<th>Arthritis # events</th>
<th>Additional Covariates</th>
<th>NOS Quality Score</th>
<th>Higher Prevalence ≥ 1 Risk Factor</th>
<th>Higher Age/sex aRR</th>
<th>Higher ≥ 1 Risk Factor aRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ogdie 2014</td>
<td>UK</td>
<td>G, A</td>
<td>RC THIN</td>
<td>1994-2010</td>
<td>18-89</td>
<td>70%</td>
<td>EHR READ codes</td>
<td>MI EHR READ Codes</td>
<td>41752</td>
<td>1032</td>
<td>Smoking, DM, HBP, HLD, DMARDs</td>
<td>4/2/3 (9)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2. Ljung 2014</td>
<td>Sweden</td>
<td>I</td>
<td>RC Population Registers</td>
<td>2001-2010</td>
<td>&gt;=18</td>
<td>76%</td>
<td>ICD 9-10 codes</td>
<td>ACS ICD-7-10</td>
<td>7704</td>
<td>221</td>
<td>COPD, DM, HBP, CBV, ATH, joint surgery, RA medications</td>
<td>3/2/3 (8)</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>4. Lindhardsen 2011</td>
<td>Denmark</td>
<td>A</td>
<td>RC Population Registers</td>
<td>1997-2006</td>
<td>&gt;=16</td>
<td>71%</td>
<td>ICD10 + DMARD Rx</td>
<td>MI ICD10</td>
<td>9921</td>
<td>265</td>
<td>CCI, SES, cardio medications</td>
<td>4/1/3 (8)</td>
<td>NA</td>
<td>Y</td>
<td>NA</td>
</tr>
<tr>
<td>5. Semb 2010</td>
<td>Sweden</td>
<td>A</td>
<td>PC AMORIS</td>
<td>1985-1996</td>
<td>&lt;20+</td>
<td>69%</td>
<td>ICD 8-10 codes</td>
<td>MI ICD 8-10</td>
<td>1779</td>
<td>214</td>
<td>None</td>
<td>4/1/3 (8)</td>
<td>Y</td>
<td>Y</td>
<td>NA</td>
</tr>
<tr>
<td>7. Sodergren 2007</td>
<td>Sweden</td>
<td>G, A</td>
<td>RC RA clinic</td>
<td>1985-2003</td>
<td>NR</td>
<td>68%</td>
<td>Rheumatologist Diagnosis (seropositive)</td>
<td>MI, MONICA definition</td>
<td>640</td>
<td>35</td>
<td>None</td>
<td>2/1/3 (6)</td>
<td>NA</td>
<td>Y</td>
<td>NA</td>
</tr>
<tr>
<td>8. Solomon 2006</td>
<td>Canada</td>
<td>NR</td>
<td>RC BCLHD</td>
<td>1999-2003</td>
<td>18+</td>
<td>71%</td>
<td>ICD 9 codes</td>
<td>MI ICD9 codes</td>
<td>25385</td>
<td>375</td>
<td>None</td>
<td>3/1/2 (6)</td>
<td>NA</td>
<td>Y</td>
<td>NA</td>
</tr>
<tr>
<td>10. Goodson 2005</td>
<td>UK</td>
<td>G</td>
<td>RC RA clinic</td>
<td>1994-2002</td>
<td>NR</td>
<td>73%</td>
<td>Rheumatologist Diagnosis</td>
<td>Hospitalized MI ICD9,10</td>
<td>1010</td>
<td>73</td>
<td>None</td>
<td>3/1/2 (6)</td>
<td>NA</td>
<td>N</td>
<td>NA</td>
</tr>
<tr>
<td>Study</td>
<td>Country</td>
<td>Funding</td>
<td>Design/ Data source</td>
<td>Study Period</td>
<td>Age range</td>
<td>Female %</td>
<td>Arthritis Ascertainment</td>
<td>MI type &amp; Ascertainment</td>
<td>Arthritis # events</td>
<td>Additional Covariates</td>
<td>NOS Quality Score</td>
<td>Higher Prevalence ≥ 1 Risk Factor</td>
<td>Higher Age/sex aRR</td>
<td>Higher ≥ 1 Risk Factor aRR</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td>---------------------</td>
<td>--------------</td>
<td>-----------</td>
<td>----------</td>
<td>------------------------</td>
<td>------------------------</td>
<td>------------------</td>
<td>---------------------</td>
<td>-----------------</td>
<td>-----------------------------</td>
<td>----------------</td>
<td>--------------------------</td>
<td></td>
</tr>
<tr>
<td>12. Turresson 2004</td>
<td>Sweden</td>
<td>A</td>
<td>RC RA clinics</td>
<td>1997-1999</td>
<td>&gt;=16</td>
<td>74%</td>
<td>1987 ACR criteria</td>
<td>MI ICD9,10 codes</td>
<td>1022</td>
<td>none</td>
<td>4/1/2 (7)</td>
<td>NA</td>
<td>Y</td>
<td>NA</td>
<td></td>
</tr>
</tbody>
</table>

Ankylosing spondylitis (AS)

| 17. Brophy 2012 | UK | G | RC EHR | 1999-2010 | >=20 | 24% | EHR READ codes | MI EHR READ codes | 1686 | none | 3/1/3 (7) | Y | N | N |

Psoriatic Arthritis (PsA)

<p>| 18. Ogdie 2014 | UK | G,A | RC THIN | 1994-2010 | 18-89 | 49% | EHR read codes | MI EHR read codes, validated | 8706 | smoking, DM, HBP, HLD, DMARD | 4/2/3 (9) | Y | Y | Y |
| 19. Li 2012 | US | G | PC NHS II | 1991-2009 | 25-60 | 100% | Self-report | Non-fatal MI Medical records | NR | Race, BMI, HLD, HBP, PA, ASA, HRT, ALC, FH, smoking, OC, VIT | 3/2/3 (8) | NR | Y | Y |</p>
<table>
<thead>
<tr>
<th>Study</th>
<th>Country</th>
<th>Funding</th>
<th>Design/ Data source</th>
<th>Study Period</th>
<th>Age range</th>
<th>Female %</th>
<th>Arthritis Ascertainment</th>
<th>MI type &amp; Ascertainment</th>
<th>Arthritis # events</th>
<th>Additional Covariates</th>
<th>NOS Quality Score</th>
<th>Higher Prevalence ≥ 1 Risk Factor</th>
<th>Higher Age/sex aRR</th>
<th>Higher ≥ 1 Risk Factor aRR</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gout</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. Kuo 2014</td>
<td>UK</td>
<td>G</td>
<td>RC CPRD</td>
<td>1997-2013</td>
<td>NR</td>
<td>28%</td>
<td>EHR READ codes</td>
<td>MI EHR READ codes</td>
<td>39111</td>
<td>1623</td>
<td>BMI, smoking, ALC</td>
<td>4/2/3 (9)</td>
<td>Y</td>
<td>Y Y Y</td>
</tr>
<tr>
<td>22. DeVera 2010</td>
<td>Canada</td>
<td>G,A</td>
<td>RC BCLHD</td>
<td>1991-2004</td>
<td>&gt;=65</td>
<td>40%</td>
<td>ICD 9 codes</td>
<td>MI ICD9 codes</td>
<td>9642</td>
<td>679</td>
<td>HBP, DM, HLD, COPD, CCI, NSAIDs, ASA, steroids, HRT, DIU</td>
<td>4/2/3 (9)</td>
<td>Y</td>
<td>Y Y</td>
</tr>
<tr>
<td>23. Choi 2007</td>
<td>USA</td>
<td>G,I</td>
<td>PC HPFUS</td>
<td>1986-1998</td>
<td>&gt;=40</td>
<td>0%</td>
<td>Medical record</td>
<td>Non-fatal MI Medical Record</td>
<td>1152</td>
<td>23</td>
<td>HBP, DM, HLD, COPD, CCI, NSAIDs, ASA, steroids, HRT, DIU</td>
<td>3/2/3 (8)</td>
<td>Y</td>
<td>MI: Y</td>
</tr>
<tr>
<td>24. Krishnan 2006</td>
<td>USA</td>
<td>I</td>
<td>PC MRFIT</td>
<td>1982-1999</td>
<td>35-57</td>
<td>0%</td>
<td>Self-reported + serum uric acid&gt;=7 mg/dl</td>
<td>MI Medical records Death certificates</td>
<td>1123</td>
<td>118</td>
<td>Centre, BMI, DM, TC, BP, smoking, ALC FH, SCR, ASA, DIU</td>
<td>3/2/3 (8)</td>
<td>Y</td>
<td>NR</td>
</tr>
<tr>
<td>26. Haugen 2013</td>
<td>USA</td>
<td>G,A</td>
<td>PC FHS</td>
<td>1990-2015</td>
<td>50-75</td>
<td>74%</td>
<td>Hand OA Radiographic (ROA) and/or symptomatic (SOA)</td>
<td>CIS/ MI Medical records Death certificates</td>
<td>726</td>
<td>18</td>
<td>BMI, lipids, DM HBP, ALC smoking, cardio medications comorbidity</td>
<td>3/2/3 (8)</td>
<td>Y</td>
<td>ROA: N SOA: Y</td>
</tr>
</tbody>
</table>
Table 3.1 Abbreviations: **RR** = relative risk, **aRR** = adjusted relative risk, **G** = Government grant, **A** = professional association /non-profit agency, **I** = Industry, **RC** = retrospective cohort, **PC** = prospective cohort, **THIN** = The Health Improvement Network, **NHIRD** = Taiwan National Health Insurance Research Database, **BCLHD** = British Columbia Longitudinal Health Database; **MRFIT** =Multiple Risk Factor Intervention Trial, **HPFUS** = Health Professional Follow-Up Study, **FHS** = Framingham Heart Study, **ERAR** = Early Rheumatoid Arthritis Register, **AMORIS** = Apolipoprotein Mortality Risk Study, **MONICA** = Multinational Monitoring of Trends and determinants in Cardiovascular diseases Study, **REP** = Rochester Epidemiology Project, **ARAMIS** = Arthritis, Rheumatism, and Aging Medical Information System, **CPRD/GPRD** =Clinical practice Research Datalink formerly called the General Practice & Research Database, **NHS** = Nurse’s Health Study, **ORALE** = Outcome of Rheumatoid Arthritis Longitudinal Evaluation, **MI** = myocardial infarction, **ACS** = acute coronary syndrome, **CHD** = coronary heart disease, **IHD** = ischemic heart disease, **CIS** = coronary insufficiency syndrome, **HBP** = high blood pressure, **DM** = diabetes mellitus, **COPD** = Chronic Obstructive Pulmonary Disorder, **HLD** = Hyperlipidemia, **CCI** = Charleston Comorbidity Index, **NSAIDs** = Non-steroidal anti-inflammatory drugs, **ASA** = aspirin, **HRT** = hormone replacement therapy, **DIU** = diuretics, **ALC** = alcohol, **SCR** = serum creatinine, **ESRD** = end-stage renal disease **NA** = not applicable, **NR** = Not Reported
**Figure 3.2** Forest Plot of Population-Based Studies Estimating Risk of Incident Myocardial Infarction Associated with Major Types of Arthritis

### A) Studies Adjusted for Age and Sex Only (N=23)

<table>
<thead>
<tr>
<th>ID</th>
<th>Study</th>
<th>RR (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>Ogdie (2015)</td>
<td>1.79 (1.57, 1.84)</td>
<td>6.21</td>
</tr>
<tr>
<td></td>
<td>Chung (2013)</td>
<td>1.70 (1.50, 1.90)</td>
<td>5.78</td>
</tr>
<tr>
<td></td>
<td>Lindhardt (2011)</td>
<td>1.69 (1.40, 1.90)</td>
<td>5.33</td>
</tr>
<tr>
<td></td>
<td>Semb (2010)</td>
<td>1.66 (1.46, 1.93)</td>
<td>5.53</td>
</tr>
<tr>
<td>OA</td>
<td>Rahman (2013)</td>
<td>1.47 (1.24, 1.73)</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>Rahm (2013)</td>
<td>1.31 (1.01, 1.71)</td>
<td>8.71</td>
</tr>
<tr>
<td></td>
<td>Overall (I-squared = 86.2%, p = 0.000)</td>
<td>1.53 (1.41, 1.66)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**NOTE**: Weights are from random effects analysis.

### B) Studies adjusted for Age, Sex and at least One Traditional Risk Factors (N = 17)

<table>
<thead>
<tr>
<th>ID</th>
<th>Study</th>
<th>RR (95% CI)</th>
<th>% Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>RA</td>
<td>Ogdie (2015)</td>
<td>1.36 (1.12, 1.66)</td>
<td>6.23</td>
</tr>
<tr>
<td></td>
<td>Chung (2013)</td>
<td>1.61 (1.57, 1.69)</td>
<td>2.28</td>
</tr>
<tr>
<td></td>
<td>Maradit-Kremers (2005)</td>
<td>1.41 (1.17, 1.69)</td>
<td>4.89</td>
</tr>
<tr>
<td>Gout</td>
<td>Abbott (1988)</td>
<td>0.91 (0.65, 1.28)</td>
<td>3.01</td>
</tr>
<tr>
<td></td>
<td>Goodwin (2005)</td>
<td>1.09 (0.90, 1.31)</td>
<td>2.01</td>
</tr>
<tr>
<td>PsA</td>
<td>Ogdie (2015)</td>
<td>1.36 (1.12, 1.66)</td>
<td>6.23</td>
</tr>
<tr>
<td></td>
<td>Chou (2014)</td>
<td>1.36 (1.16, 1.59)</td>
<td>6.23</td>
</tr>
<tr>
<td></td>
<td>Chou (2014)</td>
<td>1.52 (1.37, 1.69)</td>
<td>37.35</td>
</tr>
<tr>
<td></td>
<td>Chou (2014)</td>
<td>2.00 (1.37, 2.92)</td>
<td>2.43</td>
</tr>
<tr>
<td></td>
<td>Overall (I-squared = 81.9%, p = 0.000)</td>
<td>1.32 (1.21, 1.45)</td>
<td>100.00</td>
</tr>
</tbody>
</table>

**NOTE**: Weights are from random effects analysis.
Traditional risk factors: smoking, obesity/ body mass index, physical activity, hyperlipidemia, diabetes, and high blood pressure. Fixed effects meta-analysis was used to obtain combined study effect for Ogdie 2015 from DMARD and non-DMARD groups; DeVera 2010 from female and male groups and for Haugen 2015 from radiographic and symptomatic OA groups, respectively.
### Table 3.2 Univariate meta-regression comparing relative effects of arthritis disease type and potential heterogeneity in estimated associations with incident MI

**Model 1: Studies with Adjustment for Age and Sex Only**  
(N=23)  

<table>
<thead>
<tr>
<th>Disease Type</th>
<th>OR</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid Arthritis</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Gout</td>
<td>0.87</td>
<td>0.70</td>
<td>1.09</td>
</tr>
<tr>
<td>Psoriatic Arthritis</td>
<td>0.84</td>
<td>0.55</td>
<td>1.27</td>
</tr>
<tr>
<td>Ankylosing Spondylitis</td>
<td>0.75</td>
<td>0.56</td>
<td>1.00</td>
</tr>
<tr>
<td>Osteoarthritis</td>
<td>0.78</td>
<td>0.56</td>
<td>1.08</td>
</tr>
</tbody>
</table>

Proportion of between-study variance explained Adjusted R-Squared  
20%

Knapp-Hartung joint test for all covariates  
p = 0.198

**Model 2: Studies with Adjustment for Traditional Risk-Factors (N=13*)**  

<table>
<thead>
<tr>
<th>Disease Type</th>
<th>OR</th>
<th>95% CI</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rheumatoid arthritis</td>
<td></td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Gout</td>
<td>0.79</td>
<td>0.71</td>
<td>0.89</td>
</tr>
<tr>
<td>Psoriatic Arthritis</td>
<td>0.89</td>
<td>0.68</td>
<td>1.15</td>
</tr>
</tbody>
</table>

Proportion of between-study variance explained Adjusted R-Squared  
90%

Knapp-Hartung joint test for all covariates  
0.004

* OA and AS omitted due to small number of studies & inconsistency across studies
Table 3.3 Sub-group analyses of population-based studies estimating risk of incident myocardial infarction associated with major types of arthritis

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Studies adjusting for Age and Sex Only</th>
<th>Studies Adjusting for Age, Sex and at least One Traditional Risk Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n studies</td>
<td>Random Effects RR(95% CI)</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;45</td>
<td>3</td>
<td>1.89 (1.69,2.12)</td>
</tr>
<tr>
<td>45-64</td>
<td>5</td>
<td>1.48 (1.28,1.71)</td>
</tr>
<tr>
<td>65+</td>
<td>8</td>
<td>1.28 (1.19,1.38)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>1.59 (1.39,1.83)</td>
</tr>
<tr>
<td>Male</td>
<td>11</td>
<td>1.40 (1.26,1.57)</td>
</tr>
<tr>
<td>Cohort type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inception</td>
<td>13</td>
<td>1.45 (1.31,1.60)</td>
</tr>
<tr>
<td>Prevalent</td>
<td>10</td>
<td>1.64 (1.44,1.88)</td>
</tr>
<tr>
<td>Calendar Period</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2005-2015</td>
<td>11</td>
<td>1.44 (1.30,1.60)</td>
</tr>
<tr>
<td>&lt;2005</td>
<td>12</td>
<td>1.66 (1.48,1.86)</td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North America</td>
<td>8</td>
<td>1.48 (1.25,1.76)</td>
</tr>
<tr>
<td>Europe</td>
<td>12</td>
<td>1.57 (1.40 ,177)</td>
</tr>
<tr>
<td>Asia</td>
<td>3</td>
<td>1.51 (1.25,1.83)</td>
</tr>
</tbody>
</table>

1 Traditional risk factors: smoking, obesity/ body mass index, physical activity, hyperlipidemia, diabetes, and high blood pressure
Chapter 4
Sex Variations in Effects of Arthritis and Activity Limitation on First Heart Disease Event Occurrence in the Canadian General Population: Results from the Longitudinal National Population Health Survey

Orit Schieir, Sheilah Hogg-Johnson, Richard H Glazier, Elizabeth M Badley

4.1 Abstract

Objective: To estimate sex-specific effects of arthritis and activity limitation on incident heart disease in a nationally representative, Canadian longitudinal population-based survey.

Methods: Information on sociodemographic variables, self-reported physician-diagnosed chronic conditions (including arthritis and heart disease), activity limitation, and traditional risk factors, was collected every 2 years from 1994/95 through 2010/11 as part of the longitudinal Canadian National Population Health Survey (NPHS). Deaths due to ischemic heart disease (ICD-10 codes I20-I25) and heart failure (ICD-10 codes I50.0- I50.9) were confirmed against the Canadian Vital Statistics Database. Discrete-time survival analysis stratified by sex was used to estimate effects of arthritis and activity limitation on first heart disease event occurrence.

Results: The study included 12, 591 participants with no prior history of heart disease and 1,783 incident heart disease events. After adjusting for common risk factors, arthritis was associated with a significant increased risk of incident heart disease in women (adjusted odds ratio [OR]: 1.58, confidence interval [95% CI]: 1.23-2.02). Even higher risks were reported in women with arthritis and activity limitation (OR: 2.19, 95% CI: 1.61-2.97). Arthritis was not associated with incident heart disease in men, except for when also reported with activity limitation (OR: 1.60, 95% CI: 1.14-2.26).
**Conclusion:** Women with arthritis and men with arthritis and activity limitation have significant excess risks for developing heart disease in the general population. These findings point to the need for improved access to arthritis care, cardiovascular prevention strategies particularly in women with arthritis, and directed interventions towards prevention of activity limitation.

**4.2 Introduction**

Arthritis is a frequent chronic condition affecting the aging population and a major cause of physical disability (1-3). High rates of coronary heart disease (CHD) morbidity and mortality have been reported in clinical inflammatory arthritis populations, particularly in rheumatoid arthritis (RA) (4-8). However, there is limited population-based evidence on sex-specific risks of heart disease in arthritis, and joint effects of arthritis and activity limitation by sex have not been explored. As the burden of arthritis continues to increase with the aging of the population, there is a need for improved understanding of the development of heart disease morbidity associated with arthritis in the general population.

Theoretically, any type of arthritis could increase the risk of developing heart disease through mechanisms involving systemic and local inflammation, arthritis-related pain and mobility limitations on traditional CHD-risk factors and/or effects of common arthritis medications such as non-steroidal anti-inflammatory drugs (NSAIDs) (9-12). Positive associations between physician-diagnosed arthritis and prevalent heart disease outcomes have been reported in cross-sectional population-based surveillance studies (13-15). Longitudinal population-based studies, however, are needed to confirm a causal relationship. Moreover, despite the higher prevalence of most types of arthritis in women and sex differences in the prevalence and characteristics of heart disease (16, 17), few longitudinal population-based studies report sex-specific risks of
heart disease associated with arthritis.

One longitudinal population-based study examined the risk of incident heart disease associated with all types of arthritis with findings that physician diagnosed arthritis was not associated with a significant increased risk for incident heart disease (18). This study, however, was restricted to a survivor cohort of elderly male Japanese American war veterans with a low overall incidence of heart disease and was unable to examine heart disease risks in women.

A small but growing literature has emerged examining heart disease risks associated with osteoarthritis (OA), by far the most prevalent type of arthritis affecting the general population (1, 19). Nuesch et al. (20) reported increased age and sex standardized risks for cardiovascular disease (CVD) mortality associated with symptomatic radiographic hip and/or knee OA in a UK primary care population-based study. Results from additional multivariable survival analyses of the OA patient sample from the same study showed that walking disability was significantly associated with all-cause and CVD mortality. However, persons with a history of CVD were not excluded, and analyses relating increased CVD mortality to OA were adjusted only for age and sex. Consistent with Nuesch et al. (20), Rahman et al. (21) reported significantly increased risks of incident CVD outcomes including ischemic heart disease (IHD) and heart failure (HF) in younger and older women, and older men with OA after adjusting for traditional risk factors in a Canadian health services billing claims database but lacked measures of activity limitation. Hawker et al. (22) reported that severity of OA-functional limitations and walking disability in particular were associated with serious CVD events after adjusting for multiple common risk factors in a Canadian cohort of patients aged 55+ with moderate or severe hip and knee OA. Contrary to results from the above-mentioned studies, Hoeven et al. (23) examined effects of various OA phenotypes and lower-limb disability measured at baseline on CVD incidence in a
population-based cohort aged 55+ from the Netherlands and reported that lower limb disability but not OA was associated with incident CVD. Of the 4 studies, two (22, 23) were restricted to older patient samples (20, 23), three did not report results stratified by sex (20, 22, 23) and the one that did (21) did not include a measure of activity limitation. Further, none of these studies took into account incident arthritis developing in the comparison group or possible transitions in and out of activity limitation over time.

In order to gain a clearer understanding of broad arthritis risks for developing heart disease comorbidity in the general population, the objective of the present study was to estimate sex-specific effects of arthritis and activity limitation on the development of a first-episode of heart disease in a nationally representative Canadian longitudinal population-based survey with updated measures of arthritis, activity limitation and traditional risk factors.

4.3 Methods

**Study design:** The present study was a secondary analysis of the Canadian longitudinal National Population Health Survey (NPHS). NPHS participants were selected by multi-stage cluster random sampling of Canadian households. Standardized questionnaires were administered every 2 years between 1994/1995 through 2010/2011, providing up to 9 cycles of data collection for analysis (24). NPHS response rates ranged from approximately 93% at cycle 2 to 70% at cycle 9.

**Measures:** The NPHS questionnaire included a series of items assessing the presence of long-term chronic health conditions diagnosed by a health professional. The presence of arthritis was defined as a self-reported health professional diagnosis of arthritis excluding fibromyalgia. The presence of heart disease was defined as either a self-reported health professional diagnosis of
heart disease or heart disease death with a cause of death reported as ischemic heart disease (ICD-10 codes I20.0-I25.9) or heart failure (ICD-10 codes I50.0- I50.9). Cause of death was confirmed against the 2008 Canadian Vital Statistics Death Database by Statistics Canada (25).

Other variables treated as covariates in the analysis included: sociodemographics (age, sex, education), self-reported physician-diagnosed diabetes, high blood pressure and other chronic comorbidities grouped as 0, 1, or >=2 chronic conditions, body mass index (BMI) derived from self-reports of weight in kilograms divided by height in meters squared and classified as underweight, normal weight, overweight and obese according to WHO classification standards (26), smoking classified as current, former, or never smokers, physical activity categorized as active or moderately active vs. inactive based on the sum of average daily energy expenditure during 20 leisure time activities (e.g. walking, swimming, gardening, cycling etc.) performed during the last 3 months (27, 28), activity limitation defined as any limitation at home, school, work, or with other activities due to a long-term health condition (28), and use of pain relievers in the past month.

**Statistical analysis:** The analytic cohort was restricted to participants who: 1) were 18 years or older at baseline, 2) did not have prevalent and/or a past history of heart disease at baseline and, 3) had complete information on arthritis and heart disease for ≥1 NPHS cycle.

As arthritis prevalence increases with age, arthritis status was updated at each survey cycle. The dataset was lagged so that arthritis at time t (cycles 1 through 8) predicted future incident heart disease at time t+1 (cycles 2 through 9) to strengthen temporal arguments. BMI, physical activity, diabetes, high blood pressure and use of pain relievers are risk factors for heart disease that can theoretically also be affected by exposure to arthritis. Adjustment for variables affected by exposure can inappropriately adjust away part of the effect of exposure resulting in biased
estimates of the total causal effect (29). Therefore, baseline values of these variables were included as time-invariant covariates. Updated values of age, education, smoking and other chronic comorbidities were included as time-varying covariates. Individuals were followed until a first heart disease diagnosis or heart disease death, death from causes other than heart disease, dropout, or the end of the study, whichever came first.

The NPHS, like many other national health surveys, collected information on diagnosed chronic conditions including heart disease at discrete follow-up intervals every 2 years (i.e. 2, 4, 6, 8, 10, 12, 14 and 16 years of follow up). While Cox proportional hazards models (30) are a popular choice for performing continuous time survival analysis when the exact time of event occurrence is known, they are less appropriate for analyses of time-to event outcomes that are measured intermittently with potentially multiple events recorded at the same time period. Discrete time versions of the proportional hazards model have also been developed and are commonly used in the social sciences (31, 32). Discrete time survival analysis (DTSA) methods have several advantages relevant to the present context. First, as a type of survival analysis, DTSA models timing as well as event occurrence/non-occurrence. Second, DTSA appropriately takes into account censoring. Third, DTSA easily allows for time-varying covariates. Fourth and most important, DTSA methods are well suited for carrying out survival analysis when the time scale for event occurrence is discrete (31, 32).

DTSA involves manipulation of the data into a person-period dataset where the same person’s data is split over multiple rows. Each person has a separate row of data for each time interval they are at-risk of the event occurring for the first time. Discrete proportional odds survival models can then be estimated using standard logistic regression (31, 32); we used these models in the present study to examine age- and multivariable- adjusted effects of arthritis and activity
limitation on first heart disease event occurrence.

An initial model was tested including a time-varying measure for arthritis, age, sex and study time treated as a series of indicator variables for each NPHS survey wave. A fully adjusted multivariable model was then tested by adding covariates. The same procedure was followed to test concurrent time-varying effects of arthritis and activity limitation. Interaction terms for arthritis by sex and arthritis by age were tested to examine potential variations by sex and age based on a priori hypotheses and reports from other studies (21, 33). The proportional odds assumption was verified by testing interactions for arthritis by each time indicator.

Missing values for time-varying covariates were imputed if bounded by two same responses (e.g. missing value for arthritis at cycle 2 imputed “no” if response to arthritis at cycle 1 and cycle 3 was “no”). All analyses were weighted using longitudinal sampling weights provided by Statistics Canada and the bootstrap method was used to obtain standard errors for the calculation of 95% confidence intervals to account for the complex survey design (34). All analyses were carried out using STATA v.13.

Planned sensitivity analyses examined the robustness of results when modifying the case definition for arthritis to at least 2-mentions of arthritis, to incident arthritis, and to exclude self-reported cases of rheumatoid arthritis (RA) using information on type of arthritis collected at some but not all cycles of the NPHS (cycles 4-9). The latter was done to ensure that solely strong effects of RA already known to be associated with excess heart disease risk were not driving results. We also performed analyses restricting the outcome to confirmed fatal heart disease as well as using discrete time multiple-event survival analysis estimated with a multinomial logit model to take into account competing risks of death from causes other than heart disease (32).
**Ethics:** The present study received full approval by the University of Toronto Research Ethics Board (REB).

### 4.4 Results

At baseline cycle 1 (1994/1995) 17,276 individuals were enrolled into the longitudinal NPHS and 12,591 participants met all eligibility criteria to be included in the present analysis. A participant flow diagram is presented in Figure 4.1. Baseline characteristics for the whole study sample and by arthritis are summarized in Table 4.1.

Prevalence of arthritis increased from 12.8% (95% CI: 12.1-13.5) at the start of the study in 1994/1995 to 37.7% (95% CI: 36.3-39.0) in 2008/2009. Over the study follow-up period between 1996/1997 and 2010/2011, 1,783 (Female: 959, Male: 824) first heart disease diagnoses or deaths were reported. Cumulative incidence of heart disease over the 16-year period was 22.6% (95% CI: 21.1%-24.2%) in persons who reported arthritis at least once vs. 7.9% (95% CI: 7.1%-8.8%) in persons who never reported arthritis. Over the same period, 9.2% (95% CI: 8.6-9.7) of the study sample died from causes other than heart disease and were censored at the time of death for primary analyses.

Addition of an arthritis-by-sex interaction was significant (p<0.0024) supporting a sex difference in effects of arthritis on first occurrence of heart disease. Results of DTSA analyses are presented stratified by sex. Age-adjusted and multi-adjusted odds ratios with 95% bootstrap confidence intervals indicating binary effects of arthritis on first heart disease event occurrence in the Canadian general population are summarized in Table 4.2. The age-adjusted risk of incident heart disease was significantly increased in women with arthritis relative to women without arthritis and remained significant after adjusting for all other covariates. Significant
excess risks for heart disease were also observed in women with incident arthritis. This both strengthens causal arguments and suggests that women with more recent-onset disease are also at increased risk for heart disease. Men with arthritis were not at significantly increased risk of incident heart disease relative to men without arthritis.

Modifying the exposure case-definition to 2-mentions of arthritis and excluding self-reported cases of RA did not alter findings (Table 4.2). Further, in analyses restricting the outcome to confirmed heart disease deaths in persons aged 65 or older, women but not men with arthritis were at significantly increased risk of fatal heart disease (Adjusted OR Women: 1.97 (1.07-3.60) vs. Adjusted OR Men: 0.80 (0.49-1.31). Results were also similar after repeating analyses using a discrete time competing risks model including death from causes other than heart disease as a competing event.

Results of analyses examining effects of arthritis by activity limitation are reported in Table 4.3. Arthritis, with and without activity limitation was associated with significant excess risks for heart disease in women. In men however, only arthritis in the presence of activity limitation was significantly associated with incident heart disease (Table 4.3).

In analyses stratified by age and sex, effect estimates for arthritis tended to be stronger in magnitude for younger persons (age <65) vs. older persons (≥ age 65) particularly younger women, however confidence limits overlapped and differences between groups were inconclusive (Table 4.4).

4.5 Discussion

The present study was the first to utilize longitudinal national health survey data representative of the entire Canadian general population to directly examine temporal relationships between a
broad definition of arthritis and the development of heart disease. Initially, discrete time survival analyses controlling for multiple common risk factors showed that arthritis was associated with an increased risk of developing heart disease in women but not in men. More focused analyses examining effects of a 4-level exposure definition for arthritis by activity limitation adjusting for the same common risk factors showed that arthritis again was significantly associated with increased heart disease risks in women, with even higher risks reported in women with arthritis and activity limitation. Men with arthritis who also reported activity limitation had increased heart disease risks.

Consistent with our initial findings, Kishimoto et al. (18) reported a lack of association between self-reported doctor diagnosed arthritis and incident heart disease in a sample of elderly male Japanese–American war veterans aged 71-93. The present study however extended the work of Kishimoto et al. (18) by using a representative population-based sample with younger, middle aged and older adults as well as women, and by carrying out more focused analyses examining concurrent effects of arthritis and activity limitation on heart disease event occurrence. While activity limitation was not assessed in Kishimoto et al. (18), it is reasonable to think that rates of activity limitation may have been low given the selected sample of male war veterans with lower than average rates of heart disease.

The present study used a general surveillance definition for arthritis. However, given that OA is by far the most prevalent type of arthritis, at least 10 times more prevalent than the next most common type of arthritis RA (19, 35), it's likely that the present arthritis sample was largely OA. This is further supported by similar excess risks of heart disease reported in sensitivity analyses that excluded self-reported cases of RA. Findings from the present study therefore also point to
relationships between OA, activity limitation and the development of heart disease with variations by sex.

Results from the present study are partially consistent with findings by Nuesch et al. (20) and Rahman et al. (21) who reported increased risks of CVD associated with OA. Nuesch et al. (20), however, reported similar excess risks for CVD mortality in men and women with OA. Inconsistencies between our findings and that of Nuesch et al. (20) may be due to differences in arthritis and outcome definitions, inclusion of patients with a history of CVD and/or lack of control for traditional risk factors in the Nuesch study. Rahman et al. (21) reported an increased risk of incident IHD associated with OA in all women and older men. This study however did not include a measure of activity limitation and therefore it is unknown if the older male OA sample also exhibited higher rates of activity limitation. Both Nuesch et al. (20) and Hawker et al. (22), reported that increased walking disability was associated with significant increased risks for CVD in patients with symptomatic hip and knee OA after controlling for common risk factors. Hoeven et al. (23) however, reported that lower limb disability and not OA was associated with increased risks for CVD. Inconsistencies between these three studies that combined risk estimates for men and women may partially be explained by differences in the magnitude of effects for arthritis and activity limitation by sex reported in the present study. Further, as arthritis is a major cause of activity limitation, a plausible hypothesis is that effects of arthritis on heart disease may be at least partially mediated through effects of activity limitation. It is also plausible that the degree of mediation may vary by sex.

Inflammation common across arthritis populations has been shown to promote atherosclerosis leading to the development of heart disease (9, 12) and represents a general direct pathway between arthritis and heart disease. Reports of activity limitation in persons with arthritis in the
present study may be a marker of disease severity due to inflammation that may signal to clinicians the need to intensify treatment of arthritis as well as screen and potentially initiate CVD prevention strategies. This may have particular implications for helping to identify persons with increased heart disease risks among more elderly populations where radiographic arthritis is more prevalent. In addition, arthritis is also a major cause of functional limitations (2, 3) which may act as an indirect pathway or mediator between arthritis and the development of heart disease. Functional limitations may result in reduced physical activity, increased sedentary time and changes in weight, which are modifiable risk factors for heart disease (36, 37). These effects may also exacerbate pre-existing risk factors or contribute to the development of other modifiable risk factors for heart disease such high blood pressure and diabetes further increasing heart disease risks.

The degree of mediation operating through activity limitation and/or other pathways may also vary by sex. Disproportionate impacts of arthritis in women relative to men have been reported with respect to circulating levels of proinflammatory adipokines, radiographic progression, pain, depression, and functional/mobility limitations (38-40). A greater physical and emotional burden of arthritis in women may also deter them from seeking care for and/or appropriately managing other common comorbid conditions such as diabetes, hypertension and lipid abnormalities that increase risk of heart disease (16). Work is currently underway examining these mechanistic pathways by sex.

Results from the present study have important implications for patients, clinicians, and for public health. First, study findings show excess heart disease risks associated with the broad scope of arthritis in the general population. Second, results help identify persons with arthritis, particularly women, in the general population that are at increased risk for developing heart
disease and may be in need of additional screening and initiation of CVD prevention measures. Third, findings point to optimizing treatment of arthritis and prevention of activity limitation as potential intervention targets for reducing overall chronic disease rates and improving population health. Lastly, study findings also help inform future research in arthritis comorbidity.

Strengths of the present study include a large representative population based sample with long-term follow-up, a longitudinal lagged design predicting a relatively large number of incident heart disease events among persons with no previous history of heart disease; and updated measures of arthritis, activity limitation and multiple common risk-factors.

There are certain study limitations that should be addressed. We could not examine risks of heart disease in individual types of arthritis because this information was not collected for all cycles of the survey. Information on height/weight used to calculate BMI and physician diagnosed chronic conditions was self-reported which can be less accurate than clinical measures. Self-reported height and weight used to derive BMI can tend towards underestimation, (41) however such a bias if present would likely be non-differential. A study of individually linked data from Ontario, Canada which compared reporting of acute myocardial infarction (AMI) from a similar Canadian survey to administrative data reported moderate to high agreement and validity measures (Kappa 0.48, sensitivity 0.77, specificity 0.98) (42). Furthermore, sensitivity analyses restricting the outcome to confirmed fatal heart disease provided similar directional associations as primary analyses. High agreement for diabetes and moderate to high agreement for hypertension have also been reported when comparing survey measures to administrative data (42). The case definition for arthritis used here is consistent with the definition used by the Centers for Disease Control and Prevention (CDC) for
population-based surveillance studies (43). We found that results were also robust when we modified the arthritis case definition to at least two-mentions of arthritis in sensitivity analyses. The present study did not collect specific information on NSAID use however we did include a measure of pain reliever use that included NSAIDs. Adjustment for pre-baseline NSAID use did not change relationships between OA-disability and hospitalized CVD in the Hawker study (22). Lastly, we cannot rule out the possibility of residual confounding from other unmeasured (e.g. lipids) or unknown confounding factors.

In conclusion, the present study examined sex variations in effects of arthritis and activity limitation on the development of heart disease comorbidity in the Canadian general population. Results showed that arthritis was associated with increased risks of developing heart disease in women, and in men with arthritis who also reported activity limitation. This provides a new impetus for more longitudinal population based research examining mediational pathways involving arthritis, activity limitation and heart disease, and potential variations by sex.
4.6 References


Potential participants assessed for study eligibility
NPHS longitudinal household sample cycle 1 1994/1995
(n = 17,276)

Excluded (n = 4685)
1. Ineligible
   • Age <18 years at baseline cycle 1 1994/1995 = 3,159
   • Prevalent (and/or a prior history of) heart disease at baseline cycle 1 1994/1995 = 746

2. Eligible but insufficient data for analysis
   • Total non-response for arthritis and/or heart disease = 780

Final Analytic Sample = 12,591
Female n = 6863, Male n=5728
Table 4.1 Characteristics of the study sample at baseline NPHS Cycle 1 (1994/1995), by self-reported diagnosed arthritis, household respondents aged 18+ free of heart disease in 1994/1995, Canada (excluding territories)*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total Sample weighted % or weighted mean (sd)</th>
<th>No Arthritis weighted % or weighted mean (sd)</th>
<th>Arthritis weighted % or weighted mean (sd)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>12,591</td>
<td>10,633</td>
<td>1,958</td>
</tr>
<tr>
<td>Age (mean, sd)</td>
<td>43.0 (0.2)</td>
<td>40.7 (0.2)</td>
<td>58.3 (0.5)</td>
</tr>
<tr>
<td>Sex - Female</td>
<td>51.0</td>
<td>49.0</td>
<td>64.0</td>
</tr>
<tr>
<td>≥ High School Degree</td>
<td>75.4</td>
<td>77.9</td>
<td>58.6</td>
</tr>
<tr>
<td>High blood pressure</td>
<td>9.5</td>
<td>6.5</td>
<td>22.6</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2.9</td>
<td>2.1</td>
<td>8.1</td>
</tr>
<tr>
<td>Body mass index classification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight (BMI&lt;18.5)</td>
<td>2.6</td>
<td>2.8</td>
<td>1.3</td>
</tr>
<tr>
<td>Normal (BMI 18.5-24.9)</td>
<td>49.5</td>
<td>51.2</td>
<td>37.9</td>
</tr>
<tr>
<td>Overweight (BMI 25-29.9)</td>
<td>35.2</td>
<td>34.4</td>
<td>40.4</td>
</tr>
<tr>
<td>Obese (BMI 30+)</td>
<td>12.7</td>
<td>11.6</td>
<td>20.4</td>
</tr>
<tr>
<td>Physical activity - Inactive</td>
<td>60.6</td>
<td>60.1</td>
<td>63.8</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never smoker</td>
<td>38.8</td>
<td>39.1</td>
<td>36.4</td>
</tr>
<tr>
<td>Past smoker</td>
<td>30.0</td>
<td>29.1</td>
<td>36.7</td>
</tr>
<tr>
<td>Current smoker</td>
<td>31.2</td>
<td>31.8</td>
<td>26.9</td>
</tr>
<tr>
<td>Other chronic conditions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>65.6</td>
<td>69.0</td>
<td>42.5</td>
</tr>
<tr>
<td>1</td>
<td>25.3</td>
<td>24.4</td>
<td>31.4</td>
</tr>
<tr>
<td>≥2</td>
<td>9.1</td>
<td>6.6</td>
<td>26.1</td>
</tr>
<tr>
<td>Activity limitation</td>
<td>15.0</td>
<td>10.9</td>
<td>42.6</td>
</tr>
<tr>
<td>Use of pain relievers past month</td>
<td>62.4</td>
<td>60.3</td>
<td>77.2</td>
</tr>
</tbody>
</table>

* Values are the weighted percentage unless indicated otherwise. NPHS = National Population Health Survey; BMI = body mass index.
**Table 4.2** Weighted adjusted odds ratios for arthritis on first occurrence of heart disease\(^1\) between 1996/1997 and 2010/2011 stratified by sex, household population aged 18+ and free of heart disease in 1994/1995, Canada (excluding territories)*

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th></th>
<th>Male</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age adjusted OR (95% CI)</td>
<td>Fully adjusted OR (95% CI)(^\dagger)</td>
<td>Age adjusted OR (95% CI)</td>
<td>Fully adjusted OR (95% CI)(^\dagger)</td>
</tr>
<tr>
<td>Arthritis</td>
<td>1.81 (1.45-2.27)</td>
<td>1.58 (1.23-2.02)</td>
<td>1.11 (0.89-1.40)</td>
<td>1.01 (0.79-1.29)</td>
</tr>
<tr>
<td>Arthritis (2 mentions of arthritis)</td>
<td>1.64 (1.33-2.02)</td>
<td>1.39 (1.11-1.75)</td>
<td>1.11 (0.88-1.40)</td>
<td>1.00 (0.78-1.28)</td>
</tr>
<tr>
<td>Arthritis (excluding RA(^3))</td>
<td>1.92 (1.51-2.44)</td>
<td>1.60 (1.22-2.10)</td>
<td>1.26 (0.98-1.61)</td>
<td>1.12 (0.85-1.46)</td>
</tr>
<tr>
<td>Incident arthritis (N=9932)</td>
<td>1.73 (1.32-2.27)</td>
<td>1.65 (1.22-2.22)</td>
<td>1.29 (0.94-1.76)</td>
<td>1.20 (0.88-1.63)</td>
</tr>
</tbody>
</table>

* First occurrence of heart disease was defined as either a self-reported health professional diagnosis of heart disease or death, with cause of death recorded as ischemic heart disease or heart failure. ORs = odds ratios; 95% CI = 95% confidence interval; RA = rheumatoid arthritis.

\(^\dagger\) Fully adjusted analyses were adjusted for age, education, high blood pressure, diabetes mellitus, body mass index classification, smoking, physical activity, other chronic conditions, and use of pain relievers.
Table 4.3 Weighted adjusted odds ratios for arthritis by activity limitation on first occurrence of heart disease between 1996/1997 and 2010/2011 stratified by sex, household population aged 18+ and free of heart disease in 1994/1995, Canada (excluding territories)*

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th>Male</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age adjusted OR (95% CI)</td>
<td>Fully adjusted OR (95% CI)</td>
</tr>
<tr>
<td>No Arthritis,</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>No activity limitation</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity limitation,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No arthritis</td>
<td>2.12 (1.53-2.94)</td>
<td>1.69 (1.16-2.45)</td>
</tr>
<tr>
<td>Arthritis,</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No activity limitation</td>
<td>1.64 (1.26-2.14)</td>
<td>1.51 (1.12-2.02)</td>
</tr>
<tr>
<td>Arthritis &amp;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity limitation</td>
<td>2.79 (2.14-3.63)</td>
<td>2.19 (1.61-2.97)</td>
</tr>
</tbody>
</table>

* ORs = odds ratios; 95% CI = 95% confidence interval.

† Fully adjusted analyses were adjusted for age, education, high blood pressure, diabetes mellitus, body mass index classification, smoking, physical activity, other chronic conditions, and use of pain relievers.
Table 4.4 Weighted adjusted odds ratios for arthritis by activity limitation on first occurrence of heart disease\(^1\) between 1996/1997 and 2010/2011 stratified by age and sex, household population aged 18+ and free of heart disease in 1994/1995, Canada (excluding territories)*

<table>
<thead>
<tr>
<th></th>
<th>Female</th>
<th></th>
<th>Male</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Age&lt;65</td>
<td>Age&gt;=65</td>
<td>Age&lt;65</td>
<td>Age&gt;=65</td>
</tr>
<tr>
<td>Arthritis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Arthritis</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Arthritis</td>
<td>1.64 (1.13-2.38)</td>
<td>1.49 (1.08-2.06)</td>
<td>1.04 (0.70-1.56)</td>
<td>1.00 (0.73-1.36)</td>
</tr>
<tr>
<td>Arthritis by Activity Limitation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No Arthritis, No activity limitation</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
<td>Reference</td>
</tr>
<tr>
<td>Activity limitation, No arthritis</td>
<td>1.98 (1.17-3.33)</td>
<td>1.17 (0.70-1.98)</td>
<td>2.28 (1.36-3.76)</td>
<td>1.75 (1.05-2.90)</td>
</tr>
<tr>
<td>Arthritis, No activity limitation</td>
<td>1.79 (1.16-2.78)</td>
<td>1.25 (0.86-1.82)</td>
<td>0.98 (0.63-1.54)</td>
<td>0.93 (0.64-1.36)</td>
</tr>
<tr>
<td>Arthritis &amp; Activity limitation</td>
<td>2.14 (1.34-3.40)</td>
<td>2.04 (1.36-3.05)</td>
<td>1.70 (0.94-3.06)</td>
<td>1.54 (1.01-2.35)</td>
</tr>
</tbody>
</table>

* ORs = odds ratios; 95% CI = 95% confidence interval.

\(^\dagger\) Adjusted for age, education, high blood pressure, diabetes mellitus, body mass index classification, smoking, physical activity, other chronic conditions, and use of pain relievers.
4.7 Methods Addendum

4.7.1 Elaborated Description of Discrete-Time Survival Method

Different survival analysis methods are preferred when study time is continuous vs. discrete (event occurrence is measured at discrete follow-up periods). Like continuous time survival analysis, discrete-time survival analysis (DTSA) models both if and when an event occurred taking into account censoring. However, DTSA is a more appropriate choice for analyzing event occurrence in longitudinal panel data where repeat assessments occur at discrete follow-up periods and multiple events may be recorded at the same time (“ties”). DTSA also easily allows for time-varying covariates in the model (31, 32).

The discrete-time hazard is the conditional probability individual i will experience the event of interest for the first time in a specific time period j given that it did not occur before that time period (equation 1)

$$h(t_{ij}) = \Pr[T_i = j | T_i \geq j] \quad (1)$$

where $h_{ij}$ is the hazard probability for individual i at time period j, and Ti is a discrete random variable representing the time period j when individual i experiences the event of interest (31).

The first step to performing DTSA is to create a person-period dataset where the same person appears over as many rows of data as there are time intervals at-risk of the event occurring for the first time for that person. DTSA can then be fit using either logistic or complimentary log-log (cloglog) regression estimated using maximum likelihood. Though the two models differ in their underlying assumptions of the timing of event occurrence (the logistic model assumes that event occurrence is naturally discrete, whereas the cloglog model is more flexible for modelling
event occurrence that is truly continuous but measured at discrete time periods), estimates from
the 2 models are very similar when the outcome is rare (31, 32). Given the logistic model is more
widely used in the literature, and target audiences are often already familiar with logistic
regression, all DTSA analyses in the present study were estimated with logistic regression based
on equation 2

$$\text{logit } h(t_{ij}) = (\alpha_1 D_{1ij} + \alpha_2 D_{2ij} + \cdots + \alpha_j D_{jij}) + (\beta_1 Z_{1ij} + \beta_2 Z_{2ij} + \cdots + \beta_p Z_{pij})$$  (2)

where $\alpha_1, \alpha_2, \ldots, \alpha_j$ are intercept parameters for j time indicators $D_{1ij}, D_{2ij}, \ldots D_{jij}$ represented as
separate dummy variables for each time period and $\beta_1, \beta_2, \ldots, \beta_p$ represent the slope parameters
for each predictor (31, 32).

Model assumptions were verified as follows: the linearity assumption by testing higher order
polynomials for continuous variables; the additivity assumption by testing interaction terms for
arthritis by other covariates in the model for which there was theoretical support for possible
interaction effects (age, education, race, BMI, inactivity, activity limitation); the proportional
odds assumption by testing whether the addition of interaction terms for arthritis with each time
indicator resulted in a significant improvement in model fit using a likelihood ratio test, and the
no unobserved heterogeneity assumption was verified by repeating analyses with a multilevel
logistic regression model assuming a normal distributed random effect (31, 32).

Extensions of discrete time survival analysis can also be performed to take into account
potential competing risks (31, 44). Competing risks refers to the concept that persons are
simultaneously at-risk for multiple different events and competing event occurrence can preclude
the person from experiencing the outcome event (e.g. death from other causes before heart
disease event occurrence). A multinomial regression model with a three-level outcome (censored, heart disease event occurrence, death from causes other than heart disease) was performed to take into account competing risks of death (44).

### 4.7.2 Data Assumptions & Treatment of Missing Values

For all discrete-time survival analyses, heart disease event occurrence was treated as a single non-repeatable event. Timing of heart disease event occurrence was defined as the first NPHS survey wave where a person self-reported a healthcare provider diagnosis of heart disease or the survey wave with a corresponding cause of death, whichever came first. It was assumed that timing of arthritis onset was the first survey wave that arthritis was reported to be present. Therefore, any missing values for arthritis before the survey wave arthritis was first reported were re-coded as 0 (not present). Further, given that arthritis is a chronic condition, arthritis was treated as an absorbing state such that once a person reported “yes” to having arthritis, it was assumed that arthritis remained present at each survey wave through to the end the person’s observation period. As persons may enter in and out of states of activity limitation, missing values for activity limitation were not imputed. Missing values for categorical time-varying covariates were carried forward from the last completed survey wave. Missing values for age were carried forward from the previous cycle adding 2 years for every forward cycle.
Chapter 5
Moderating and Mediating Effects of Sex, Obesity and Activity Limitation on Associations between Arthritis and Heart Disease in the General Population

Prepared for submission to Arthritis Care Res
Orit Schieir, Sheilah Hogg-Johnson, Richard H Glazier, Elizabeth M Badley

5.1 Abstract

Objective: To estimate mediating and moderating effects of activity limitation, sex and obesity on overall associations between arthritis and heart disease in the general population.

Methods: Data were from the longitudinal Canadian National Population Health Survey (NPHS). The NPHS collected information on sociodemographic variables, self-reported physician-diagnosed chronic conditions including arthritis and heart disease, activity limitation, and lifestyle/health behaviors every 2 years from 1994/95 through 2010/11. Cause of death from ischemic heart disease and heart failure was confirmed against the Canadian Vital Statistics Database. Mediation analyses were used to estimate sex-specific direct and indirect effects of arthritis via activity limitation on heart disease, controlling for multiple baseline covariates.

Results: The sample included 11,655 adults without prevalent heart disease at baseline or the first follow up cycle and 1442 subsequent first heart disease events. In men, only indirect effects of arthritis via activity limitation on heart disease were significant (OR: 1.09, 95% CI: 1.05-1.14, % mediated: 90%). In women, there were significant interactions for effects of arthritis with obesity. Indirect effects of arthritis through activity limitation explained a greater proportion of
total heart disease risk in obese (OR: 1.18, 95% CI: 1.06 - 1.31, % mediated: 54%) vs. non-obese women (OR: 1.08, 95% CI: 1.04 - 1.12, % mediated: 23%).

**Conclusion:** Activity limitation explained part of the increased heart disease risk associated with arthritis in the general population and is a modifiable target for prevention strategies to reduce heart disease risk in persons with arthritis, particularly in obese women with arthritis.

**5.2 Introduction**

While rheumatoid arthritis (RA) is recognized as an independent risk factor for heart disease (1, 2), a recent meta-analysis of population-based studies has shown that this heart disease risk also extends to multiple other prevalent types of arthritis in the general population (3). Inflammation is a common feature of arthritis that has been shown to have pro-atherogenic effects (1, 4) and is one potential direct mechanism linking arthritis and heart disease. However, inflammation alone does not fully explain added heart disease risk associated with arthritis (3, 5), highlighting the role of other mechanisms to account for total excess risk. The latter may be particularly relevant to explain associated heart disease risk in osteoarthritis (OA) (6-8) and other prevalent types of arthritis with a relatively lower inflammatory load than RA. Given current and projected increases in both the prevalence and burden of arthritis (9, 10), it’s important to understand how arthritis may lead to the development of heart disease in the general population and who may be most at-risk.

Arthritis frequently causes functional limitations that affect an individual’s ability to carry out daily activities. Results from a recent US population-based study showed that of the 54.4 million Americans who reported having doctor-diagnosed arthritis, 23.7 million (44%) also reported
having activity limitation attributable to their arthritis (11). Functional limitations, particularly mobility limitations, have been independently associated with increased heart disease risk in people with arthritis (7, 12, 13). These findings suggest that activity limitation may be a potential indirect pathway (a mediator) linking arthritis and heart disease. A visual decomposition of the total effect of arthritis on heart disease into direct and proposed indirect effects mediated through activity limitation is presented in Figure 5.1.

In addition to these potential direct and indirect pathways, several studies have reported effect modification by sex, where heart disease risk associated with arthritis is greater for women than in men (6, 8, 14-17). Activity limitation is also reported more often in women than in men with arthritis, and more often in obese vs. non-obese women (18-22). A previous longitudinal population-based study performed by our group (17) examining independent and concurrent effects of arthritis and activity limitation on incident heart disease controlling for many common confounders showed that arthritis was significantly associated with increased heart disease risk in women, and the relative risk was higher in women who also reported activity limitation than in those who did not. However, only men with arthritis who also reported activity limitation had increased heart disease risk. Given these reported variations in associations between arthritis and heart disease, it is plausible that activity limitation may both mediate and moderate associations between arthritis and heart disease.

No studies have examined mediating and moderating effects of activity limitation, sex and obesity on relationships between arthritis and heart disease. The objectives of the present study were therefore to build on previous work by: 1) decomposing total effects of arthritis on heart disease into direct and indirect effects mediated through activity limitation while simultaneously allowing for possible interactions between arthritis and activity limitation and, 2) examining
potential variations in these direct and indirect pathways by sex and obesity, in a large longitudinal population-based cohort. We hypothesized that activity limitation would partially mediate associations between arthritis and heart disease in both men and women, and that indirect effects of activity limitation would be greater for obese women with arthritis relative to non-obese women.

5.3 Methods

**Study design:** This was a longitudinal population-based study that utilized data from the Canadian National Population Health Survey (NPHS). The NPHS is a longitudinal panel survey developed, administered and housed by Statistics Canada. Survey participants were selected by multi-stage cluster random sampling of households from the 10 Canadian provinces in 1994/95 (excluding people living on Indian Reserves, Canadian Forces Bases and some remote areas in Quebec and Ontario) and were followed prospectively through to 2010/2011 when the survey ended (23). Participants completed standardized in-depth interviews on several health-related measures including socio-demographic factors, self-reported doctor-diagnosed chronic conditions, and lifestyle/health behaviors at baseline and again every 2-years throughout the follow up period providing up to 9 cycles of data per person (response rates ranged from 93% to 70% at each cycle). The present study was limited to adult participants (age 18+) who had at least one available measure of arthritis, activity limitation and heart disease, respectively, for analysis. Participants were excluded if they reported prevalent heart disease at either of the first two NPHS visits. The present study was approved by the University of Toronto Research Ethics Board (REB).

**Measures**
**Exposure:** A time-varying binary indicator for arthritis (present/absent) was ascertained as part of the chronic conditions section of the NPHS questionnaire. Participants were re-asked at each survey cycle two years apart to self-report if they had doctor diagnosed arthritis, excluding fibromyalgia. The questionnaire was worded as follows: “Now I would like to ask about certain chronic health conditions that you may have. We are interested in long-term conditions that have lasted, or are expected to last, 6 months or more and that have been diagnosed by a health professional…Do you have arthritis, excluding fibromyalgia?”

**Mediator:** A time-varying binary indicator of activity limitation (present/absent) was ascertained as part of the health limitations section of the questionnaire. Participants were re-asked about limitations in daily activities due to long-term physical or mental health conditions at work, home, school or other at each survey cycle. The questionnaire was worded as follows: “The next few questions deal with any health limitations which affect daily activities. Because of a long-term physical or mental condition or a health problem, are you limited in the kind or amount of activity:” and was followed by 4 separate queries … “at home”, “…at school”, “…at work”, “…other activities” (24).

**Outcome:** Heart disease was defined as a first fatal or non-fatal heart disease event. Non-fatal heart disease was captured by re-asking patients to self-report if they had doctor-diagnosed heart disease including heart attack, angina or heart failure at each survey cycle. Fatal heart disease was ascertained from death records with a cause of death listed as ischemic heart disease (ICD-10 codes I20.0-I25.9) or heart failure (ICD-10 codes I50.0-150.9). Cause of death was confirmed against the 2008 Canadian Vital Statistics Death Database (25).
**Confounders:** We included all available baseline measures of potential confounders related to associations between arthritis and heart disease (exposure-outcome relationship), arthritis and activity limitation (exposure-mediator relationship), and, activity limitation and heart disease (mediator-outcome relationship) as recommended when performing mediation analyses (26). *Age in years* was included as a continuous variable, *education* (≤ high school degree vs. any post-secondary education) and *race* (white vs. non-white) were collected as part of the demographic interview and included as binary variables. Self-reported doctor-diagnosed *diabetes* and *high blood pressure*, respectively, were ascertained as part of the chronic conditions section of the questionnaire (described above) and were modelled as separate variables to control for potential confounding effects of pre-existing cases of these two conditions which are more common in arthritis populations and are independent risk factors for heart disease. The sum of positive responses to other self-reported doctor-diagnosed chronic conditions was treated as a single count variable and grouped as 0, 1, or >=2 chronic conditions. Underweight, healthy weight, overweight and obese weight classes were defined using standard WHO BMI cut-points for the adult general population and calculated from self-reported height and weight (27). Current and past history of *cigarette smoking* was self-reported and grouped into current, former, or never smokers. Physical inactivity is an independent risk factor for heart disease that is also associated with activity limitation. We therefore included a baseline measure of physical activity measured 2 years prior to the first assessment of activity limitation. *Physical activity* was categorized as active or moderately active vs. inactive based on the sum of average daily energy expenditure during 20 leisure time activities (e.g. walking, swimming, gardening, cycling etc..) performed during the last 3 months (24, 28). Alcohol intake was self-reported as the number of drinks per
day categorized as 0, 1-2 drinks per day and >2 drinks per day to coincide with dietary health guidelines. Use of pain relievers was self-reported (y/n) for the past month.

**Statistical analysis:** A popular method for performing mediation analyses is the product of coefficients method (“Baron & Kenny approach”) (29). In this approach two regressions are performed. Model 1: outcome is regressed on the exposure, mediator and covariates. Model 2: Mediator is regressed on the exposure and covariates. The direct effect is the effect of exposure controlling for the mediator in the outcome model. The indirect effect is the product of the effect of the mediator on the outcome in Model 1 and the effect of the exposure on the mediator in Model 2. The total effect is the sum of the direct and indirect effects. This approach however, was developed for use with continuous mediators and outcomes, and assumes no exposure mediator interactions.

This approach is not ideal in the current study context given that: 1) due to the longitudinal panel nature of the NPHS, heart disease is a discrete-censored outcome measured at fixed periods two years-apart; 2) activity limitation (mediator) is a binary measure, and 3) there could be possible interactions between arthritis and activity limitation (exposure-mediator interactions).

Prevention researchers have recently developed a novel approach for performing mediation analyses appropriate for use with longitudinal health surveys with discrete censored outcomes, that combines discrete time survival analysis (DTSA) with classical mediation methods described above (30). DTSA has many advantages including that it appropriately handles censored data, is specially geared for modelling event occurrence measured at discrete points in time and easily allows for time-varying covariates. In brief, DTSA models the hazard function for a target event as a cumulative set of conditional probabilities that an event occurred for the
first time in a specific time period given that it did not occur before and can be estimated with standard logistic regression (31, 32). In the context of mediation, DTSA would be used to fit Model 1 with the target event outcome regressed on a set of time specific intercepts, exposure, mediator, and confounders using logistic regression. Standard regression is then used to fit the Model 2 with the mediator as the outcome regressed on the exposure and covariates, and direct and indirect effects are estimated through the product of coefficients method described above. This approach however, still does not allow for binary mediators and assumes no exposure-mediator interactions.

Causal inference researchers have developed very flexible regression-based extensions of the product of coefficients method for estimating total, direct and indirect effects allowing for various types of non-linear mediators and outcomes and exposure-mediator interactions, provided that stringent assumptions related to no unmeasured confounding in any of the tested pathways are met (26, 33, 34). Plug-ins for performing causal mediation analyses with standard software have also been made available (34). These software plug-ins allow Model 1 to be fit with either linear, logistic, log-linear, poisson or negative binomial regression, and Model 2 to be fit with linear or logistic regression.

Given that discrete time survival analysis can be fit with logistic regression, we applied a combined approach similar to Fairchild et al. (30) but rather combined discrete time survival analysis with causal mediation methods to estimate direct and indirect effects based on formulas for binary mediator and outcomes (26, 34, 35).

As arthritis, activity limitation and heart disease all increase over time as a function of age, we included repeated measures of these 3 variables. Further, to ensure correct chronological
ordering in the estimated mediated effect, we lagged the dataset so that measures of arthritis (time t: NPHS cycles 1-7) would precede measures of activity limitation (time+1: NPHS cycles 2-8), and both arthritis and activity limitation would precede measures of heart disease (t+2: NPHS cycles 3-9). All covariates were measured at a single time point at baseline, to avoid biases that may result from adjustment for covariate changes over time secondary to arthritis (36).

We used the Stata mediation program plug-in PARAMED to fit all regression models and estimate causal natural direct effects (DE), indirect effects (IE) and total effects on the odds ratio scale along with their respective standard errors and 95% confidence intervals (CI). Logistic regression was used to fit Model 1 with heart disease as the outcome regressed on separate time specific intercepts corresponding to each NPHS cycle, arthritis, activity limitation, and covariates. Logistic regression was also used to fit Model 2 with activity limitation (mediator) as the outcome regressed on time, arthritis and covariates. We examined possible effect modification of arthritis by obesity by adding interaction terms to Model 1 and Model 2.

Standard errors and 95% confidence intervals (CI) for estimated direct and indirect effects were derived using the delta method. As bootstrap methods may be preferred in contexts with non-linear outcomes (34) we also provide re-estimated bootstrap 95% CIs based on 500 sample replications in a supplemental appendix. Lastly, we calculated the proportion mediated which here refers to the proportion of the total effect of arthritis on heart disease operating through indirect effects of activity limitation as $\frac{OR_{DE}(OR_{IE}-1)}{(OR_{DE} \times OR_{IE}-1)}$ (26, 33).

Sensitivity analyses included repeating the analysis above re-estimating direct and indirect effects after allowing for an exposure-mediator interaction for arthritis with activity limitation (34). All analyses were carried out in Stata v. 13 and v.14.
5.4 Results

The starting sample of the NPHS in 1994/95 was 17,276. Persons aged <18 years (n=3,159) and prevalent cases of heart disease reported in the first 2 NPHS cycles (n=1,087) were excluded. Of the 13,030 potentially eligible participants, 1,375 (11%) were excluded due to complete missing data on arthritis, activity limitation or heart disease, leaving a final analytic sample of 11,655 (Females=6,373, Males=5,282) participants. Between 1998/1999 and 2010/2011, 1442 (Females=776, Males=666) first heart disease diagnoses or deaths were reported. Prevalence of arthritis increased from 13% to 37% and activity limitation from 11% to 25% over the study period.

Interactions for arthritis with BMI class were not significant in men. In women, there was a significant interaction for arthritis with obese weight class in Model 1 with heart disease as the outcome (p=0.036). The same interaction was borderline significant in Model 2 with activity limitation (mediator) as the outcome (p=0.09). We therefore present all regression and estimated direct and indirect effects stratified by sex, and additional sub-analyses for non-obese and obese women.

Table 5.1 summarizes regression results from Model 1 and Model 2 for each subgroup. In men, arthritis was not associated with heart disease when adjusting for activity limitation and all covariates [OR: 1.01, 95% CI: 0.83-1.23]. There were however, significant positive associations between arthritis and activity limitation [OR: 3.14, 95% CI: 2.89-3.42] and between activity limitation and heart disease controlling for arthritis and all covariates [OR: 1.61, 95% CI: 1.32-1.96]. In analyses including all women, there was a significant effect of arthritis on heart disease controlling for activity limitation and all covariates [OR: 1.33, 95% CI:1.11-1.59]. There were
also significant positive associations between arthritis and activity limitation [OR: 2.78, 95% CI: 2.60-2.98] and between activity limitation and heart disease controlling for arthritis and all covariates [OR:1.64, 95% CI: 1.39-1.94]. In stratified analyses of women by obesity status, associations between arthritis and heart disease, arthritis and activity limitation, and activity limitation and heart disease were all significant in non-obese women. However, in obese women, arthritis was more strongly associated with activity limitation than in non-obese women and only associations between arthritis and activity limitation, and activity limitation and heart disease, respectively, were significant.

Table 5.2 summarizes estimated direct and indirect effects of arthritis mediated by activity limitation on heart disease. In men, only indirect effects of arthritis were significant and explained 90% of total excess heart disease risk associated with arthritis. In analyses of all women, both direct and indirect effects of arthritis were significant and activity limitation explained 29% of total excess heart disease risk. Non-obese women had significant direct and indirect effects of arthritis with activity limitation explaining 23% of total heart disease risk, while obese women only had significant indirect effects with activity limitation explaining a relatively greater proportion (54%) of total excess heart disease risk associated with arthritis.

Additional interaction terms for arthritis with activity limitation in heart disease outcome models were not significant in men (p=0.713) or women (p=0.102) and results remained unchanged in sensitivity analyses that re-estimated direct and indirect effects specifying interactions between arthritis and heart disease (data not shown).
5.5 Discussion

Results from the present longitudinal population-based study showed that activity limitation explained part of the increased heart disease risk associated with arthritis. Further, the proportion of heart disease risk explained by activity limitation differed for men compared to women and for obese compared to non-obese women. In men, activity limitation almost entirely explained any increase in heart disease risk associated with arthritis. In obese women, most of the excess heart disease risk associated with arthritis was explained by indirect effects of activity limitation. In non-obese women however, while activity limitation significantly explained part of the heart disease risk associated with arthritis, most of the risk operated through mechanisms other than activity limitation.

The present study is a direct extension of a previous study examining individual and concurrent effects of arthritis and activity limitations on heart disease risk in the same longitudinal population-bases survey (17). In the previous study however, arthritis and activity limitation were measured at the same time and temporal precedence between arthritis and activity limitation needed to assess mediation could not be determined. We only identified one other relevant study that estimated overall relationship of radiographic hip OA (RHOA) with cardiovascular mortality and 5 potential mediators in a prospective cohort study of elderly women 65 years and older (37). In this study, the indirect effect of RHOA through physical function assessed by 6m walk test was the only borderline significant mediator (HR = 1.06, 95% CI: 1.00-1.13, p=0.07) explaining 25% of overall CVD mortality associated with RHOA. Other mediators examined included physical activity, disability, hip pain and NSAID use. There are considerable differences between this and the present study in terms of study populations,
outcomes, mediator definitions and time points. In addition, the authors noted that a portion of the sample underwent total hip replacement (THR) and sensitivity analyses that excluded women who underwent THR showed more robust associations between RHOA with mortality suggesting a protective benefit of THR on mortality. Though not examined, it is also likely that THR had a beneficial impact on the mediators examined as well.

Results from the current study are consistent with evidence from multiple studies reporting associations supporting different parts of the pathways examined here. These include reported associations between arthritis pain and disability with activity limitation (9, 11), and associations between activity limitation and increased heart disease risk through associations with physical inactivity, sedentary lifestyle, social participation restriction and greater psychological distress (38, 39). Possible explanations for stronger associations between arthritis and heart disease in women vs. men may include reports of disproportionate impacts of arthritis, and higher use of NSAIDs/coxibs in women (19, 40, 41). Possible explanation for stronger mediated effects of activity limitation in obese vs. non-obese women may include reports of stronger associations of arthritis with disability and physical inactivity in obese women (21, 42, 43).

Findings from the present study have important implications for public health and for clinical practice given that: 1) heart disease is a leading cause of morbidity and mortality, with an even higher estimated global burden of disease in 2010 than previously reported in 1990 (44, 45); 2) prevalence of arthritis is increasing with more recent birth cohorts having a higher prevalence of arthritis at younger ages than previous generations; 3) arthritis disease burden is also increasing with reports of nearly 20% increases in arthritis attributable activity limitation from 2002 to 2015; and, 4) obesity a major risk factor for both arthritis and heart disease that is also associated with disability, continues to rise in the general population, with particular increases in the
population aged 65+ with arthritis (46). Activity limitation is modifiable and study results point to it being an important intermediate target for prevention strategies to help mitigate heart disease risk in a growing arthritis population, particularly for obese woman with arthritis who may benefit most from such interventions.

Strengths of the present study include use of novel methods to test mediating and moderating mechanisms underlying the relationship between arthritis and heart disease in a large longitudinal population-based sample with updated measures of arthritis, activity limitation and numerous confounding variables. To our knowledge this is the first study to provide evidence that part of the heart disease risk associated with arthritis is explained by an indirect effect of arthritis on activity limitation and that the magnitude of the mediated effect varies by sex and obesity.

This study also has limitations that should be considered. First, the case definition for arthritis is self-reported and captures all forms of arthritis (other than fibromyalgia). The magnitude of direct and indirect effects of arthritis through activity limitation may differ in different types of arthritis. Given that OA is by far the most prevalent type of arthritis in the general population (47), results are very likely to reflect heart disease mechanisms in OA. If so, results here would support a previously under-recognized indirect pathway linking OA and heart disease with important implications for prevention strategies. Second, the activity limitation measure is not specific to arthritis, though we controlled for possible confounding effects of other chronic conditions in the multivariable models and independent associations between arthritis and activity limitation were quite high (range OR: 2.71 -3.26). Third, Statistics Canada recommends the use of weights to appropriately account for the complex survey design in the NPHS, however the mediation software package used to estimate direct and indirect effects did not allow for
inclusion of survey weights. We did however re-run all regression models using survey weights and parameter estimates for arthritis, activity limitation and all confounders were very similar in weighted and unweighted regression models. Another limitation of the mediation software was that standard logistic regression used to estimate adjusted associations between arthritis and activity limitation in Model 2 does not account for within person clustering over time. Ignoring clustering of observations may have resulted in somewhat smaller estimated standard errors used to calculate part of indirect or mediated effect of arthritis. We examined this potential bias by separately re-running Model 2 using multilevel modelling that appropriately accounts for clustered observations and found that associations between arthritis and activity limitation remained significant and in the same direction as when estimated with logistic regression, though as expected standard errors were larger in multi-level models. Lastly, mediation analysis assumes that there are no omitted confounders in any pathway for estimated direct and indirect effects to have valid causal interpretations. While we did include numerous measured confounding factors identified for each causal pathway between arthritis, activity limitation and heart disease, we cannot rule out possible effects of unmeasured (e.g. lipids) or unknown confounders on estimated direct and indirect effects. None the less, we provide strong support for mediation by demonstrating multi-adjusted associations between arthritis and activity limitation and activity limitation with heart disease furthering current understanding of the development of arthritis heart disease comorbidity in the general population.

In conclusion, results of this large longitudinal population-based study showed that activity limitation explained part of the increased heart disease risk associated with arthritis. These findings suggest that optimized treatments of pain and prevention of disability should be prioritized not only to improve arthritis outcomes but as prevention strategies to reduce heart
disease risk in the general population with arthritis, particularly in obese women with arthritis.
5.6 References


17. Schieir O, Hogg-Johnson S, Glazier RH, Badley EM. Sex Variations in the Effects of Arthritis and Activity Limitation on First Heart Disease Event Occurrence in the Canadian


**Figure 5.1** Schematic showing how the total effect of arthritis on heart disease may be decomposed into a direct and posited indirect effect mediated through activity limitation.

- **c** represents the total effect of arthritis on heart disease; **c’** the direct effect of arthritis on heart disease controlling for activity limitation; **a** the effect of arthritis on activity limitation; and, **b** the effect of activity limitation on heart disease controlling for arthritis. Combined **a** and **b** represent the indirect effect of arthritis mediated through activity limitation. Note that for simplicity, confounders are not depicted but all associations are assumed to be controlled for all relevant confounders.
Table 5.1 Multivariable logistic regressions predicting incident heart disease and activity limitation

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>All</td>
<td>Non-Obese</td>
<td>Obese</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1: Heart</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Disease as the</td>
<td>Adjusted</td>
<td>95% CI</td>
<td>Adjusted</td>
<td>95% CI</td>
<td>Adjusted</td>
<td>95% CI</td>
<td>Adjusted</td>
</tr>
<tr>
<td>outcome</td>
<td>OR</td>
<td></td>
<td>OR</td>
<td></td>
<td>OR</td>
<td></td>
<td>OR</td>
</tr>
<tr>
<td>Arthritis(c')</td>
<td>1.01</td>
<td>0.83</td>
<td>1.23</td>
<td>1.33</td>
<td>1.11</td>
<td>1.59</td>
<td>1.37</td>
</tr>
<tr>
<td>Activity limitation(b)</td>
<td>1.61</td>
<td>1.32</td>
<td>1.96</td>
<td>1.64</td>
<td>1.39</td>
<td>1.94</td>
<td>1.57</td>
</tr>
<tr>
<td>Model 2: Activity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Limitation (mediator)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>as the outcome</td>
<td>Arthritis(a)</td>
<td>3.14</td>
<td>2.89</td>
<td>3.42</td>
<td>2.78</td>
<td>2.60</td>
<td>2.98</td>
</tr>
</tbody>
</table>

Models 1 and Model 2 both included separate indicator variables for time reflecting each NPHS cycle and were adjusted for the following confounders: age, race, education, high blood pressure, diabetes, BMI class, physical inactivity, smoking, alcohol intake, sum of non-cardiovascular comorbidities, and use of pain relievers. $c'$ represents the effect of arthritis on heart disease controlling for activity limitation and all confounders; $b$ the effect of activity limitation on heart disease controlling for arthritis and all confounders; and, $a$ the effect of arthritis on activity limitation controlling for all confounders.
Table 5.2 Direct and indirect effects of arthritis mediated through activity limitation on incident heart disease

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td></td>
<td>Adjusted OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Direct effect</td>
<td>1.01</td>
<td>0.83</td>
</tr>
<tr>
<td>Indirect effect</td>
<td>1.09</td>
<td>1.05</td>
</tr>
<tr>
<td>Total Effect</td>
<td>1.11</td>
<td>0.91</td>
</tr>
<tr>
<td>% Mediated</td>
<td>90%</td>
<td>29%</td>
</tr>
</tbody>
</table>

The direct effect represents the effect of arthritis on heart disease that operates through pathways other than activity limitation. The indirect effect represents the effect of arthritis on heart disease that operates through activity limitation pathways. The % mediated represents the proportion of the overall association between arthritis and heart disease that is explained by the indirect effect of activity limitation. 95% confidence intervals were estimated using the delta method.
5.7 Methods Addendum: Elaborated Description of Mediation Approach

There were several important methodological considerations that affected the choice of mediation method that could be used in the present context. These included a binary mediator, a discrete time to event outcome, time varying covariates needed to incorporate updated measures of arthritis and activity limitation that increase over time with age, and flexibility to allow for potential exposure-mediator interactions of arthritis with activity limitation.

As summarized in the manuscript above, the classical product of coefficients (“Baron & Kenny”) method for performing mediation is of limited use when either the mediator or outcome are non-linear and there may be exposure mediator interactions (26, 34). Researchers have therefore sought to extend this classical approach for studying mediating mechanisms in broader research contexts. Fairchild et al. showed that discrete time-survival analysis (DTSA) could be combined with the product of coefficients method to evaluate underlying mechanisms of event occurrence measured at discrete time periods (30). The rationale for utilizing DTSA is that it is specially geared for analyses of event occurrence measured at discrete time intervals, which is often how longitudinal data is collected. Further, DTSA is highly flexible and can easily handle time-varying covariates). Underlying equations for regression-based (using logistic regression) and structural-equation modelling (SEM)-based applications are detailed in Fairchild et al. (30). Though this combined approach provides a useful framework for performing mediation analysis with discrete time-to-event outcomes, the SEM-based application has only been integrated into more specialized SEM software (MPLUS) and does yet allow for binary mediators. Further, both the regression-based and SEM-based applications assume no exposure-mediator interactions.
As there were no pre-existing mediation methods that could address all of the methodological
constraints of the present study, I explored alternate options where DTSA could be integrated
into other mediation methods.

Mediation methods have also recently been re-conceptualized by causal inference researchers
based on the counterfactual framework reviewed in detail elsewhere (48, 49). General
regression-based approaches have since been developed for estimating these counterfactual
direct and indirect effects with a variety of linear and nonlinear outcomes and mediators allowing
for exposure mediator interactions provided that the following transparent but often untestable
assumptions about confounding are met (26, 34, 50, 51):

1. No exposure outcome confounding
2. No mediator outcome confounding
3. No exposure mediator confounding
4. No mediator-outcome confounding affected by exposure

A detailed explanation of the different formulae available for performing causal mediation
analysis is presented in Appendix E.

In brief, in the context of a binary mediator and binary outcome, direct and indirect effects
allowing for exposure mediator interactions can be derived from two logistic regression models
shown below

\[
\logit\{P(Y = 1|A = a, M = m, C = c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a * m + \theta_4 c
\]  (1)

\[
\logit\{P(M = 1|A = a, C = c)\} = \beta_0 + \beta_1 a + \beta_2 c
\]  (2)

Direct and indirect effects on the odds ratio scale can then be obtained by plugging in parameter
estimates from the above two regressions into the formulae below

\[
\begin{align*}
OR^{DE} & \cong \frac{\exp(\theta_1 a)\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta_1^* c)\}}{\exp(\theta_1 a^*)\{1 + \exp(\theta_2 + \theta_3 a^* + \beta_0 + \beta_1 a^* + \beta_1^* c)\}} \\
OR^{IE} & \cong \frac{\{1 + \exp(\beta_0 + \beta_1 a^* + \beta_1^* c)\}\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_1^* c)\}}{\{1 + \exp(\beta_0 + \beta_1 a + \beta_1^* c)\}\{1 + \exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a^* + \beta_1^* c)\}}.
\end{align*}
\]

Standard errors along with 95% confidence intervals for direct and indirect effects can also be obtained using either the delta method or bootstrap method.

The authors provide plug-ins for carrying out causal mediation analysis in standard software packages (34). Software plug-ins fit the outcome and mediator regression models and provide estimated total, direct and indirect effects along with their standard errors and 95% confidence intervals using either the delta method or bootstrap method. A detailed description of the Stata mediation macro PARAMED used in the present study is presented in Appendix F.

As the approach above uses logistic regression to estimate the outcome model and DTSA is also estimated using logistic regression, we integrated DTSA in to the above causal mediation method. We created a person period dataset in the usual way described in Chapter 5. The Stata mediation macro PARAMED was then used to fit a DTSA model for the outcome heart disease with logistic regression and a second logistic regression model with activity limitation (mediator) as the outcome. Direct and indirect effects were estimated based on results of these two regression models using the formulae above.
Chapter 6
Discussion and Conclusions

6.1 Key Findings

6.1.1 Evidence that Arthritis is a Risk Factor for Heart Disease in the General Population

Manuscript one presented in Chapter 3 was the first systematic review and meta-analysis to quantify and compare incident MI-risk across the five most prevalent types of arthritis in population-based studies. Results showed that MI-risk was increased across multiple arthritis disease types and that overall arthritis was associated with an approximate 50% increase in age/sex-adjusted, and 30% increase in multi-adjusted (age/sex + at least one other traditional heart disease risk factor) risk of MI in population-based studies of moderate to high quality. The prevalence of traditional heart disease risk factors was also consistently higher in each type of arthritis than in the general population. Though there was significant heterogeneity in pooled estimates for overall effects of arthritis, sensitivity analyses using meta-regression suggested that age/sex adjusted risk of MI did not differ significantly across arthritis disease types, but that multi-adjusted associations with MI were on average 20% higher in RA than in gout, pointing to potentially similar total risk of MI across arthritis disease types but conceivable differences in direct vs. indirect mechanisms. Though the review sought to identify evidence of MI-risk in five common types of arthritis, more than half of identified studies were in RA, and there was a relative dearth of high quality evidence that could be identified in OA, which is at least 10 times more prevalent in the general population than RA (1). This, in addition to limitations of included studies noted in Chapter 3 (e.g. inconsistent confounder adjustment, common lack of gender/sex stratified analyses, etc.), and other conflicting published reports regarding relationships between
arthritis, activity limitation and various cardiovascular outcomes (2-6) supported the need for additional analyses providing more robust population-based estimates of heart disease risk associated with arthritis (+/- activity limitation) stratified by sex.

Manuscript two presented in Chapter 4 utilized 16 years of longitudinal data from the Canadian National Population Health Survey (NPHS). A weighted sample representing over 19 million adult Canadians was used to estimate sex-specific effects of arthritis on first-ever heart disease event occurrence. After adjusting for multiple common risk factors, results of initial discrete time-survival analysis with time-varying lagged predictors showed that women but not men with arthritis had a significant increased risk of heart disease. Moreover, results were robust to several sensitivity analyses modifying the case definition of arthritis (excess heart disease risk in women range 39%-65% depending on the arthritis case definition used, Chapter 4, Table 4.2). More detailed analyses examining concurrent effects of arthritis and activity limitation showed that all women with arthritis, and women with arthritis and activity limitation in particular, were at significantly increased risk of developing heart disease. Contrary to women however, all men with activity limitation, with and without arthritis, had higher heart disease risk.

Together, manuscripts one and two support arthritis (not just RA) as an independent risk factor for developing heart disease in the general population, particularly in women.

6.1.2 Support for Activity Limitation as a Mediator in the Arthritis-Heart Disease Relationship

Manuscript two presented in Chapter 4 showed that activity limitation alone and in combination with arthritis significantly increases heart disease risk in the general population. These results have important implications for risk stratification. However, as arthritis and activity limitation
were both assessed at the same time, temporal ordering between arthritis and activity limitation needed to test for mediation could not be assessed. Manuscript three presented in Chapter 5 extended the work in manuscript two and was the first study to formally examine a complete proposed indirect pathway between arthritis, activity limitation and heart disease specifying appropriate chronological precedence between each variable arthritis, activity limitation and heart disease using novel mediation methods. Results after controlling for multiple covariates in each path showed that indirect effects of arthritis via activity limitation on heart disease were significant, with differences in relative contributions of the indirect effect to the total effect in men vs. women and in non-obese women vs. obese women. The proportion of total heart disease risk explained by activity limitation was greater in men and in obese women with arthritis than in non-obese women with arthritis, respectively.

Results are important in terms of highlighting the role of non-inflammatory mechanisms leading to increased heart disease risk in persons with arthritis. As activity limitation is potentially modifiable (7, 8) it can be targeted for intervention to help reduce heart disease morbidity associated with arthritis in the general populations. Results also suggest that such interventions may potentially have even more benefit in terms of reducing heart disease risk in obese women with arthritis.

6.1.3 Support for Sex/ Gender Differences in Arthritis-Heart Disease Comorbidity

The thesis manuscripts provide consistent evidence of sex/gender disparities in heart disease risk associated with arthritis. Manuscripts two and three examined one potential source for the disparity by examining effects of activity limitation in relation to overall and intermediate effects of arthritis on heart disease. Results showed that in men, total heart disease risk associated with
arthritis was only marginally increased and almost entirely explained by activity limitation, while in women total heart disease risk associated with arthritis was much higher and explained by activity limitation as well as other pathways, with notable differences in obese vs. non-obese women.

Results of the present study add to increasing evidence showing a greater burden of arthritis in women vs. men. In addition to higher prevalence and severity, more pain, worse function, poorer emotional health and more social participation restriction (9-11), results point to a higher burden of heart disease comorbidity in women vs. men. Study findings call for more careful cardiovascular screening and follow-up of women with arthritis and further research to better understand sex and gender-related pathways between arthritis and heart disease, respectively.

6.1.4 Contributions to Understanding Heart Disease Risk in Osteoarthritis

OA is the most common form of arthritis. Population-based estimates suggest that upwards of 3 out of 5 people with arthritis have OA (12-14). Therefore, present study findings are highly likely to reflect overall and intermediate associations between OA and heart disease. If so, results presented in manuscripts two and three contribute to the relatively scant and conflicting literature that has led to a hitherto limited understanding of heart disease risk in OA. For instance, the systematic review presented in Chapter 3 only identified two population-based studies estimating risk of incident MI associated with OA (4, 15). Another recent systematic review that included population-based studies examining associations between OA and a broad scope of CVD outcomes only identified 10 prospective studies (16) several of which reported conflicting results. Studies included in this review reported consistent associations between lower limb OA (knee and/or hip) and CVD in female samples, inconsistent associations between radiographic
hand OA and CVD, and often null associations for OA (any type) and CVD when restricted to male samples or when single measures of association were reported from mixed sex samples. Results from the present study supporting a mediating path between arthritis, activity limitation and heart disease and effect modification by sex help reconcile these seemingly inconsistent studies. Moreover, results potentially help identify OA subgroups at greater risk of heart disease in the general population, namely women and those with activity limitation, as well as one of the potential mechanisms contributing to this excess risk.

6.2 Limitations and Methodological Considerations

Justification of study design and data source: The primary focus of this thesis is to better understand overall associations, mechanisms, and potential sex/gender differences in arthritis-heart disease relationships in the general population. Given the need to understand temporal relationships between arthritis and heart disease (and intermediate pathways), the long latency for, and relatively rare occurrence of incident cases of heart disease in the general population, longitudinal population-based studies with long-term follow-up and multiple assessment points are needed. Analysis of secondary data sources including administrative billing databases, electronic health records (EHR) and longitudinal health surveys can be a faster and less expensive option than primary data sources for carrying out these types of analyses. Administrative healthcare billing databases, however, are less well-suited to the present research context given that they: 1) often have limited look-back periods or require additional linkages to other data sources to determine prior history of heart disease needed to exclude prevalent disease cases at baseline; 2) can lack generalizability if limited to a sub-set of the population (i.e. elderly or fraction covered by health or drug benefit plan); 3) may be more vulnerable to under-
ascertainment of arthritis cases given the episodic nature of arthritis, general under-treatment of arthritis in the general population, and high comorbidity rates that may lead to under-assignment of arthritis billing codes in persons with arthritis and more salient conditions; 4) often lack measures of important covariates (e.g. SES, smoking, activity limitation etc.) (17-19). EHRs are often very large requiring added processing power, can be difficult to clean and extract usable data given that different providers may enter diagnoses and relevant clinical data in different fields, and often lack important covariate information (20). Longitudinal national health surveys on the other hand, collect repeat measures of a myriad of health-related variables according to standardized practices in a large sample weighted to represent an entire population.

The present study was therefore a secondary analysis of the longitudinal Canadian National Population Health Survey (NPHS). The NPHS is an extremely rich data source with several important advantages relevant to the proposed study including: long-term follow-up, multiple assessment points; standardized data collection methods; detailed information on self-reported physician-diagnosed chronic conditions, deaths and cause of death verified against the Canadian Vital Statistics Database, as well as updated measures of multiple important covariates.

Limitations of the NPHS include reliance on self-report for information on chronic conditions, activity limitation and height/weight used to calculate BMI, and lack of, or limited information on lipids/ hypercholesterolemia, insulin resistance, type of arthritis, some arthritis and heart disease treatments.

Selection bias: Common threats of selection bias in systematic reviews and meta-analyses relate to the comprehensiveness and rigour of the process used to identify and include studies in the review, size and quality of included studies, and likelihood of publication bias (21, 22). The systematic review and meta-analysis presented in chapter 3 employed several strategies to
minimize such biases, including a detailed search strategy developed by an information specialist
performed in three databases, supplemental grey literature searches, use of 2 reviewers to carry
out all study screening, selection, appraisal and extraction of information (with conflicts resolved
by consensus), criteria for study size to be included in meta-analysis and use of both graphical
and statistical methods to assess publication bias. There may however, have still been some
selection bias if there were additional relevant studies published in languages other than English
or French that were not included in the review. Moreover, while graphical and statistical methods
for assessing publication bias were not significant for overall effect estimates of arthritis, there
may have been a threat of publication bias in qualitative and/or quantitative summaries of
individual types of arthritis particularly in PsA, AS and OA where very few studies were
identified, and of those identified none were small negative studies.

Common sources of selection bias in cohort studies stem from a lack of comparability in exposed
and unexposed groups, and if non-response or attrition over time result in systematic differences
between persons included vs. excluded from the study sample (23). In analyses of the NPHS,
people from the same target population were categorized as having or not having arthritis.
Further, only 5.8% of the NPHS sample used in manuscript two, and 10.5% of the sample used
in manuscript three had total non-response for either arthritis, activity limitation and/or heart
disease and were excluded from all analyses. In sensitivity analyses comparing demographic
characteristics of total non-responders to at least partial responders included in analyses, there
were no differences with respect to sex or race, though total non-responders were slightly older
and were more likely to have not completed high school (p<0.05). Survival analysis methods
were used to appropriately account for censoring in the analytic sample over time. Further, the
likelihood of potential systematic differences in attrition between exposure groups was likely low.
given that median follow up time did not differ between arthritis vs. non-arthritis groups, and estimates for arthritis on heart disease were robust to sensitivity analyses accounting for potential competing risks of death from other causes reported in manuscript two presented in Chapter 4. Proportions of missingness for time-invariant covariates measured at baseline were relatively low ranging from 0-6%. Variables closer to the top of this range were physical inactivity, BMI, use of pain medications and alcohol consumption. Given the relatively low percent of missing values no imputation was performed. This small reduction in sample size may have affected the precision of study estimates. In addition, up to 20% of time-varying covariates were missing at any given measurement cycle. In order to maximize data available for each analysis, single imputation methods were performed (described in detail in the methods addendum to Chapter 4). More sophisticated multiple imputation methods are generally preferred over single imputed methods for addressing missing data, however the current imputation procedure was guided by reasonable theoretical assumptions and was overall more feasible given the many other study complexities.

**Information bias:** NPHS measures were often self-reported and are vulnerable to misclassification (either under or over-reporting) because of recall-related and social desirability biases (24). As the NPHS is a general health survey collected prospectively with no a priori focus on arthritis or heart disease, in general misclassification due to self-report would likely be non-differential between arthritis and non-arthritis groups.

More detailed discussion of the impacts of self-report measures of exposure, outcome, mediator and covariates is provided below. Arthritis status was ascertained by self-reported physician diagnosis of arthritis excluding fibromyalgia. This case definition was similar to the general surveillance definition used by the Centers for Disease Control (CDC) for population-based
surveillance studies and shown to be valid (25). Another, validation study reported an estimated 71% sensitivity and specificity, respectively, for self-reported physician-diagnosed arthritis in a senior center population age 50+ (26). As symptom based conditions are captured more frequently in surveys than in medical records, misclassification in self-reported arthritis is more likely to result in a higher estimated prevalence of arthritis. If the arthritis group included a proportion of healthy persons without arthritis, then the bias in reported associations between arthritis and heart disease, and arthritis and activity limitation would likely tend towards underestimation and true associations may be even higher than those reported in the present study. Even if the inverse scenario was more likely such that healthy persons with arthritis were misclassified as not having arthritis, the bias in reported associations would also likely tend towards underestimation of true associations. Theoretically, reported associations in the present study could be biased towards overestimation if persons without arthritis but with a higher predisposition for heart disease (e.g. greater genetic susceptibility, higher prevalence of traditional heart disease risk factors etc….) were more likely to be misclassified as having arthritis, though there is no substantive reason this would occur in the present study context.

The non-fatal heart disease outcome was also ascertained by self-reported physician diagnosis. A validation study of NHANES (n=1,666) comparing data on self-reported chronic conditions to hospital records found that self-reports of ischaemic heart disease (IHD) and heart attack were accurate in > 80 % of cases (27). A study of individually linked data from Ontario which compared reporting of acute myocardial infarction (AMI) from a similar Canadian survey to administrative data reported moderate to high agreement and validity measures (Kappa 0.48, sensitivity 0.77, specificity 0.98) (28). It should also be noted that mortality from IHD and heart failure (HF) part of the primary heart disease outcome was confirmed against the 2008 Canadian
vital statistics database. Therefore, information on cause of death was missing from 2009-2011 resulting in some under-ascertainment of the heart disease outcome, likely in the range of 50-75 deaths when extrapolating from the distribution of deaths reported from 1994-2008.

The activity limitation variable used in analyses was a general flag for self-reported restrictions in a person’s daily activities at work, home, school or other, due to a long-term chronic health condition. As the measure was not specific to arthritis, it is possible that there may have been greater misclassification of activity limitation in persons with arthritis who also had other disabling chronic conditions. Multivariable analysis however was used to control for effects of specific individual conditions as well as a composite comorbidity score, and independent associations between arthritis and activity limitation were very strong.

Information on height and weight used to derive BMI was self-reported. Relative to BMI based on objectively measured height and weight, BMI based on self-reported height and weight has been shown to be biased towards underestimation (29). This underestimation has also been shown to be greater for obese vs. normal or underweight groups, resulting in inflated associations between obesity measures derived from self-reported vs. measured height and weight (30). As obesity is a strong risk factor for arthritis, and arthritis is more common in women, it is plausible that measurement error in self-reported BMI in the present study could partially explain stronger associations between arthritis and heart disease in women. Self-reported physician diagnosed diabetes has been shown to be valid, although self-reported physician diagnosed high blood pressure is reported to have more moderate sensitivity with bias tending towards underestimation (31). Self-reported cigarette smoking status and physical activity have been shown to have high accuracy (32, 33).
**Confounding:** Uncontrolled confounding is a common source of bias in observational studies. Though the present study included measures of, and statistically adjusted for multiple important confounders in the focal relationships examined, residual confounding from included covariates measured with error (e.g. BMI), and/or from other possible known (e.g. lipids, insulin resistance, depression/ other mental health conditions, specific arthritis and heart disease treatments) and unknown omitted confounders cannot be ruled out. Mediation results presented in chapter 5 are based on strong underlying assumptions related to no uncontrolled confounding in any examined pathway. These results should therefore be interpreted as strong support for a mediated pathway between arthritis, activity limitation and heart disease rather than definitively causal.

**Generalizability:** The present study used general surveillance measures for arthritis and heart disease for the general population. It is highly likely that reported associations for arthritis reflect associations for OA and other prevalent forms of arthritis with ischemic heart disease and heart failure, given the wording and conditions listed as part of the heart disease questions in the survey, and that sensitivity analyses limiting to confirmed IHD and HF deaths were similar to main results, though this cannot be confirmed.

The target population of the NPHS was household residents in ten Canadian provinces in 1994/1995, excluding Territories, persons living on Indian Reserves and Crown Lands, residents of health institutions, full-time members of the Canadian military bases & some remote areas in Ontario and Quebec. Statistics Canada used multi-stage cluster random sampling (where geographic clusters then household dwellings within each cluster were selected) to obtain the NPHS survey sample in Cycle 1 (1995/95). Survey attrition rates due to non-response over time were relatively low ranging from 5-9% for each subsequent survey cycle (34). To account for the complex survey design and attrition over time, Statistics Canada provides data users with
longitudinal survey weights and bootstrap replication weights generated from 500 replications. When weighted analyses are performed with appropriate survey software, analysis results are generalizable to the target Canadian population in 1994/95. All analyses for manuscript two in Chapter 4 were weighted and estimated with appropriate Stata survey suite procedures.

When survey and bootstrap weights are not applied in NPHS analyses, consequences of the complex survey design is potentially underrepresenting persons in larger households (parents and children) and over-representing persons in smaller households (singles and elderly) (35). Further, clustering typically results in increased variability, and ignoring clustering can result in underestimated standard errors and bias in hypothesis tests. The effect on point estimates and standard errors may range from negligible to fairly large depending on the research question and modeling procedures. In general, the bias is usually greater for estimating prevalence and other summary statistics and relatively minimal for estimating measures of association (36-38). The mediation procedure utilized did not support use of survey weights, therefore results for mediation analyses presented in Chapter 5 were unweighted. Sensitivity analyses re-estimating the same multivariable models used in the calculation of total, direct and indirect effects (i.e. Model 1 and Model 2) with survey weighted logistic regression showed that measures of association for arthritis and activity limitation with heart disease, and for arthritis with activity limitation, were all quite similar with no substantive differences in the direction, magnitude or significance of relationships but that as expected, weighted standard errors were slightly larger than unweighted. For instance, in women, unweighted odds ratios (OR) and standard errors (SE) for associations between arthritis and heart disease controlling for activity limitation and covariates were OR =1.33, SE=0.12 vs. weighted OR=1.29, SE=0.16. Given the similarity in weighted and unweighted results and focus on a mechanistic relationship, it would be reasonable
to assume that results are still generalizable to the target Canadian population.

**Ethical considerations:** NPHS participants provided written informed consent upon enrolling in the NPHS survey study. As a secondary analysis of the NPHS, the present study did not require any new data collection or pose any new harms. The present study was submitted and approved by the University of Toronto Research Ethics Board after undergoing delegated review (Appendix G).

### 6.3 Implications

**Policy:** The aging of the population, improvements in survival and increase in common chronic disease risk factors such as obesity and physical inactivity has resulted in a greater proportion of the population living longer with multiple chronic conditions (39). Arthritis is highly prevalent, its individual and societal impact have increased in recent decades, and is projected to continue to increase in the next 30 years (1, 40). Half of people with arthritis also have at least one other chronic condition. Arthritis-heart disease is amongst the most common pattern of multimorbidity in the Canadian population across all age groups (41). Arthritis commonly affects joints in the lower limbs causing impairments in mobility and increased disability. This in turn makes it more challenging to be physically active (42) needed to protect against and/or help manage other chronic conditions including heart disease, diabetes, obesity and high blood pressure.

Present results supporting arthritis in general (and likely OA) as a risk factor for heart disease, and activity limitation as a potentially significant intermediate target have important implications for policies related to primary and secondary prevention of heart disease comorbidity in the general population. Results support more intensified efforts aimed at arthritis prevention
including broader promotion of physical activity and healthy weight guidelines in the general population. Other possible strategies could also include public health messaging and incentives to increase communication between care providers and patients to debunk common misconceptions about arthritis being a common part of aging that cannot be prevented or treated, and that physical activity should be avoided to prevent worsening of arthritis symptoms (43-45). Secondary prevention efforts are also needed to minimize heart disease risk in persons living with arthritis. These could include dissemination of tailored education regarding arthritis-heart disease risk to multiple arthritis care providers such as specialists, primary care, and allied health professionals emphasizing the need for effective pain management strategies and prevention of disability, along with broader health care coverage of allied health visits, self-management and exercise programs. Further, given current and projected increases in unmet needs for care in a growing faction of the population with arthritis (17, 46), implementation of newer models of care (47) that focus on improving referral systems, increasing training, communication, and shared care between different arthritis care providers can not only lead to better arthritis outcomes, but also reductions in the overall chronic disease burden in the general population.

**Practice:** Many clinical practice guidelines have been developed to increase awareness and provide guidance for managing heart disease risk in patients with RA (48). Recommendations generally include early tight control of inflammation, screening and appropriate initiation of cardiovascular prevention using risk factor thresholds for the general population and exercising caution with prolonged use of NSAIDs, coxibs and corticosteroids particularly in patients with multiple traditional heart disease risk factors. Furthermore, physical activity interventions have been shown to decrease pain and improve function by almost 40% (7) and chronic disease self-management programs have been shown to improve patient confidence in the management of
their condition and to reduce pain, fatigue, and depression (49). Unfortunately, both are underutilized by arthritis patients. Physician behavior including encouraging patients to take part in these programs and writing prescriptions for exercise have been shown to increase adherence (50). Evidence from the present study would support medical associations and training programs disseminating more inclusive information and recommendations about managing excess cardiovascular risk across arthritis. Such recommendations could be similar to those already developed for patients with RA with added guidance for increasing physician recommendations to patients of available self-management and physical activity programs, and more careful screening and monitoring of female arthritis patients.

**Research:** There are still many unanswered questions regarding arthritis heart disease comorbidity that should be addressed in future studies. First, evidence from intervention studies examining effects of increased cardiac screening and/or of specific interventions targeting disability on reducing heart disease are lacking. Second, results from the present study provide indirect evidence linking OA and heart disease. Given the prevalence and impact of OA on disability in the general population, more direct evidence of heart disease risk in OA populations is needed, particularly studies in hand OA where overall relationships and mechanisms are less clear. Third, the present study only examined one potential mediating path linking arthritis and heart disease. Investigation of other intermediate pathways and how multiple pathways interact with one another to increase heart disease risk is needed to further inform prevention strategies and improve quality of care. For instance, depression and other mental health conditions are commonly associated with arthritis and heart disease as well as with multiple factors posited to be on the causal pathway between arthritis and heart disease including activity limitation, pain sensitivity, physical inactivity, use of pain medications and treatment adherence (39, 41, 51).
Analyses of these complex mechanistic pathways and interrelationships between mediators would require further development of current regression-based approaches for performing mediation that can appropriately handle time-varying measures, multiple and non-linear mediators. Lastly, present evidence for sex/gender disparities in heart disease risk associated with arthritis call for more frequent inclusion of sex/gender based analyses in arthritis comorbidity studies. Further investigation of potential sex-specific biologic mediators (e.g. hormones, immune system, inflammatory markers/ adipokines) and gender differences in complex mediated pathways highlighted above can help inform pharmacological and non-pharmacological treatments for heart disease in women that may lead to improved outcomes.

6.4 Conclusion

Results from the thesis offer several additional insights into relationships between arthritis and heart disease in the general population by providing strong evidence for arthritis in general being a risk factor for heart disease, activity limitation being an intermediate path in the overall relationship, and sex/gender differences in overall and intermediate relationships. Together results point to more inclusive and integrated CVD prevention strategies (e.g. improving timely access to arthritis care, optimizing management strategies for controlling inflammation, pain and preventing disability) to mitigate added heart disease risk across arthritis, and prioritization of arthritis prevention as a means of reducing chronic disease rates in the general population.

6.5 Student’s role

I was involved in all phases of the research including: refining the research questions, developing the study protocol, gaining access to the data, data analysis, interpretation of results, manuscript/
thesis preparation and dissemination of study findings. While completing the dissertation I acquired expertise and practical skills in advanced analysis methods including longitudinal analyses of complex surveys, discrete-time survival analysis and mediation analysis.
6.6 References


GCs = glucocorticoids, NSAIDs = non-steroid anti-inflammatory drugs (excluding aspirin),
coxibs = cyclooxygenase inhibitors

Traditional heart disease risk factors that are modifiable include: smoking, BMI/obesity, insulin resistance/ diabetes, high blood pressure and physical inactivity
Appendix B
Chapter 3 Supplemental Material Manuscript 1

Sample MEDLINE Search Strategy

MEDLINE (January 1980-January 2015)

1 Arthritis/
2 arthriti*.mp.
3 polyarthriti*.mp.
4 arthrochondriti*.mp.
5 arthrosynovitis.mp.
6 oligoarthriti*.mp.
7 section 31.mp.
8 joint inflammation*.mp.
9 ARTHRITIS, PSORIATIC/
10 (psoria* adj3 arthropath*).mp.
11 (psoria* adj3 rheumatism).mp.
12 arthritic psoriasis.mp.
13 Arthritis, Rheumatoid/
14 caplan syndrome/
15 felty's syndrome/
16 rheumatoid nodule/
17 exp rheumatoid factor/
18 (arthrit* adj2 rheum*).mp.
19 (caplan* adj2 syndrome*).mp.
20 (felty* adj2 syndrome*).mp.
21 rheumatoid.mp.
22 inflammatory arthritis.tw.
23 Gout/
24 gout?.mp.
25 Arthritis, Gouty/
26 Osteoarthritis/
27 degenerative arthritides.mp.
28 osteoarthr*.mp.
29 (degenerative adj2 joint disease*).mp.
30 arthros?s.mp.
31 hip osteoarthritis/
32 coxartheros?s.mp.
33 coxarthros?s.mp.
34 malum coxae senilis.mp.
35 OSTEOARTHRITIS, KNEE/
36 gonarthros?s.mp.
37 Osteoarthritis, Spine/
38 spondylosis.mp.
39 (barre lieou adj2 syndrome).mp.
40 brachialgia paraesthetica nocturna.mp.
41 (neri barre adj2 syndrome).mp.
42 (spinal adj2 osteophytosis).mp.
43 posterior cervical sympathetic syndrome.mp.
44 spondylarthros?s.mp.
45 spondyloarthros?s.mp.
46 vertebral artery syndrome.mp.
47 periarthritis/
48 periarthrit*.mp.
49 SACROILIITIS/
50 sacroiliitis.mp.
51 sacroiliitide*.mp.
52 sacroileitis.mp.
53 Spondylarthritis/
54 spondylarthrit*.mp.
55 Spondylarthropathies/
56 spondylarthropath*.mp.
57 (spondylo* adj2 arthropath*).mp.
58 Spondylitis, Ankylosing/
59 (ankylosing adj2 spondyl*).mp.
60 (bechterew* adj2 disease*).mp.
61 (rheumatoid* adj2 spondylitis).mp.
62 marie struempell* disease*.mp.
63 exp myocardial infarction/
64 myocardial infarct$.tw.
65 heart attack$.tw.
66 heart infarct$.tw.
67 (coronary adj3 syndrome$).tw.
68 acute coronary.tw.
69 coronary thrombosis/
70 coronary thrombosis.tw.
71 ami.tw.
72 Prospective Studies/
73 prospectiv$.mp.
74 longitudinal.mp.
75 retrospectiv$.mp.
76 Retrospective Studies/
77 risk:.mp.
78 incidence$.mp.
79 exp cohort studies/
80 between group:.tw.
81 cohort$.tw.
82 controlled clinical trial.pt.
83 epidemiologic methods/
84 limit 83 to yr="1971 - 1988"
85 exp case-control studies/
86 (case$ and control$).tw.
87 limit 83 to yr=1966-1989
88 or/1-62
89 or/63-71
90 or/72-82,84
91 or/85-87
92 90 or 91
93 88 and 89
94 limit 93 to (english or french)
95 limit 94 to "all child (0 to 18 years)"
96 limit 94 to "all adult (19 plus years)"
97 95 and 96
98 95 not 97
99 94 not 98
100 limit 99 to yr="1980 -Current"
101 90 and 100
102 91 and 100
103 101 or 102
104 102 not 101
105 from 101 keep 1-388
106 from 104 keep 1-9
Funnel Plot of Population-Based Studies Reporting Risk of Incident Myocardial Infarction Associated with Major Types of Arthritis.

1a) Studies adjusting for Age and Sex only

1b) Studies adjusting for Age, Sex and at least one Traditional Risk Factor

Traditional risk factors: smoking, obesity/ body mass index, physical activity, hyperlipidemia, diabetes, and high blood pressure.
Sensitivity Analyses Examining Impact of Omitting Individual Studies on Pooled Effect of Major Types of Arthritis on incident MI

<table>
<thead>
<tr>
<th>Study omitted</th>
<th>RR</th>
<th>95% CI</th>
<th>RR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ogdie</td>
<td>1.53</td>
<td>1.42, 1.66</td>
<td>1.32</td>
<td>1.20, 1.45</td>
</tr>
<tr>
<td>Ljung</td>
<td>1.54</td>
<td>1.42, 1.67</td>
<td>1.31</td>
<td>1.20, 1.44</td>
</tr>
<tr>
<td>Chung</td>
<td>1.52</td>
<td>1.41, 1.65</td>
<td>1.32</td>
<td>1.20, 1.45</td>
</tr>
<tr>
<td>Lindhardsen</td>
<td>1.53</td>
<td>1.41, 1.67</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Semb</td>
<td>1.55</td>
<td>1.42, 1.68</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Holmqvist</td>
<td>1.55</td>
<td>1.42, 1.68</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Sodergen</td>
<td>1.56</td>
<td>1.44, 1.69</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Solomon</td>
<td>1.54</td>
<td>1.42, 1.67</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Maradit-Kremers</td>
<td>1.53</td>
<td>1.41, 1.66</td>
<td>1.32</td>
<td>1.20, 1.45</td>
</tr>
<tr>
<td>Fischer</td>
<td>-</td>
<td>-</td>
<td>1.34</td>
<td>1.21, 1.47</td>
</tr>
<tr>
<td>Goodson</td>
<td>1.53</td>
<td>1.41, 1.66</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Turresson</td>
<td>1.54</td>
<td>1.41, 1.68</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Solomon</td>
<td>1.52</td>
<td>1.40, 1.65</td>
<td>1.35</td>
<td>1.23, 1.47</td>
</tr>
<tr>
<td>Evers</td>
<td>1.51</td>
<td>1.40, 1.64</td>
<td>1.31</td>
<td>1.20, 1.44</td>
</tr>
<tr>
<td>Chou</td>
<td>1.54</td>
<td>1.43, 1.67</td>
<td>1.31</td>
<td>1.19, 1.43</td>
</tr>
<tr>
<td>Brophy</td>
<td>1.53</td>
<td>1.40, 1.67</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Ogdie</td>
<td>1.56</td>
<td>1.44, 1.69</td>
<td>1.33</td>
<td>1.20, 1.47</td>
</tr>
<tr>
<td>Kuo</td>
<td>1.52</td>
<td>1.40, 1.65</td>
<td>1.35</td>
<td>1.21, 1.50</td>
</tr>
<tr>
<td>Kuo</td>
<td>1.51</td>
<td>1.39, 1.63</td>
<td>1.30</td>
<td>1.19, 1.43</td>
</tr>
<tr>
<td>DeVeria</td>
<td>1.51</td>
<td>1.39, 1.63</td>
<td>1.33</td>
<td>1.21, 1.46</td>
</tr>
<tr>
<td>Choi</td>
<td>1.53</td>
<td>1.41, 1.65</td>
<td>1.28</td>
<td>1.18, 1.39</td>
</tr>
<tr>
<td>Krishnan</td>
<td>-</td>
<td>-</td>
<td>1.35</td>
<td>1.24, 1.47</td>
</tr>
<tr>
<td>Abbott</td>
<td>1.53</td>
<td>1.41, 1.66</td>
<td>1.31</td>
<td>1.20, 1.43</td>
</tr>
<tr>
<td>Haugen</td>
<td>1.53</td>
<td>1.41, 1.66</td>
<td>1.32</td>
<td>1.21, 1.45</td>
</tr>
<tr>
<td>Rahman</td>
<td>1.53</td>
<td>1.41, 1.66</td>
<td>1.32</td>
<td>1.21, 1.45</td>
</tr>
<tr>
<td>Combined</td>
<td>1.53</td>
<td>1.41, 1.66</td>
<td>1.32</td>
<td>1.21, 1.45</td>
</tr>
</tbody>
</table>

1 Traditional risk factors: smoking, obesity/body mass index, physical activity, hyperlipidemia, diabetes, and high blood pressure.
Appendix C  
Supplemental Material Manuscript 2

Whole Sample Fully Adjusted Discrete Time Survival Models Showing Interaction for Arthritis by Sex

<table>
<thead>
<tr>
<th>Variables</th>
<th>OR</th>
<th>Bootstrap 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthritis</td>
<td>1.517</td>
<td>1.204</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.912</td>
</tr>
<tr>
<td>Male Sex</td>
<td>1.869</td>
<td>1.470</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.377</td>
</tr>
<tr>
<td>Arthritis x Male sex</td>
<td>0.690</td>
<td>0.500</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.952</td>
</tr>
<tr>
<td>Age, years</td>
<td>1.060</td>
<td>1.054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.066</td>
</tr>
<tr>
<td>&lt; High School degree</td>
<td>0.881</td>
<td>0.753</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.030</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2.151</td>
<td>1.587</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.916</td>
</tr>
<tr>
<td>High blood pressure</td>
<td>1.604</td>
<td>1.328</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.936</td>
</tr>
<tr>
<td>Inactive</td>
<td>1.178</td>
<td>1.019</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.362</td>
</tr>
<tr>
<td>BMI class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal weight</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>1.014</td>
<td>0.643</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.600</td>
</tr>
<tr>
<td>Overweight</td>
<td>1.212</td>
<td>1.015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.446</td>
</tr>
<tr>
<td>Obese</td>
<td>1.420</td>
<td>1.148</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.758</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None smoker</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>1.404</td>
<td>1.161</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.697</td>
</tr>
<tr>
<td>Past</td>
<td>1.048</td>
<td>0.895</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.228</td>
</tr>
<tr>
<td>No of comorbidities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.412</td>
<td>1.191</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.674</td>
</tr>
<tr>
<td>2 or more</td>
<td>1.561</td>
<td>1.295</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.881</td>
</tr>
<tr>
<td>Use of pain relievers</td>
<td>1.203</td>
<td>1.039</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.393</td>
</tr>
</tbody>
</table>
Whole Sample Fully Adjusted Discrete Time Survival Models Showing Interaction for Arthritis by Sex and Age

<table>
<thead>
<tr>
<th>Variables</th>
<th>OR</th>
<th>95% Bootstrap</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthritis</td>
<td>1.641</td>
<td>1.243</td>
</tr>
<tr>
<td>Male Sex</td>
<td>1.877</td>
<td>1.475</td>
</tr>
<tr>
<td>Arthritis x Male sex</td>
<td>0.683</td>
<td>0.497</td>
</tr>
<tr>
<td>Arthritis x Age 65</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 1</td>
<td>0.905</td>
<td>0.669</td>
</tr>
<tr>
<td>1 1</td>
<td>0.787</td>
<td>0.613</td>
</tr>
<tr>
<td>Age, years</td>
<td>1.064</td>
<td>1.055</td>
</tr>
<tr>
<td>&lt; High School degree</td>
<td>0.876</td>
<td>0.749</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2.142</td>
<td>1.576</td>
</tr>
<tr>
<td>High blood pressure</td>
<td>1.608</td>
<td>1.331</td>
</tr>
<tr>
<td>Inactive</td>
<td>1.176</td>
<td>1.017</td>
</tr>
<tr>
<td>BMI class</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Underweight</td>
<td>1.015</td>
<td>0.644</td>
</tr>
<tr>
<td>Overweight</td>
<td>1.210</td>
<td>1.014</td>
</tr>
<tr>
<td>Obese</td>
<td>1.413</td>
<td>1.142</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None smoker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Current</td>
<td>1.392</td>
<td>1.151</td>
</tr>
<tr>
<td>Past</td>
<td>1.044</td>
<td>0.892</td>
</tr>
<tr>
<td>No of comorbidities</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.408</td>
<td>1.188</td>
</tr>
<tr>
<td>2 or more</td>
<td>1.559</td>
<td>1.295</td>
</tr>
<tr>
<td>Use of pain relievers</td>
<td>1.204</td>
<td>1.040</td>
</tr>
</tbody>
</table>
Sex-Stratified Fully Adjusted Discrete Time Survival Models Listing All Included Variable Effect Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>OR</th>
<th>Bootstrap 95% CI</th>
<th>OR</th>
<th>Bootstrap 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arthritis</td>
<td>1.575</td>
<td>1.230 2.016</td>
<td>1.006</td>
<td>0.006 0.014</td>
</tr>
<tr>
<td>Age, years</td>
<td>1.059</td>
<td>1.051 1.066</td>
<td>1.062</td>
<td>1.053 1.071</td>
</tr>
<tr>
<td>&lt; High School degree</td>
<td>0.852</td>
<td>0.686 1.059</td>
<td>0.907</td>
<td>0.722 1.140</td>
</tr>
<tr>
<td>Diabetes</td>
<td>2.288</td>
<td>1.503 3.481</td>
<td>2.068</td>
<td>1.353 3.160</td>
</tr>
<tr>
<td>High blood pressure</td>
<td>1.463</td>
<td>1.141 1.877</td>
<td>1.785</td>
<td>1.386 2.300</td>
</tr>
<tr>
<td>Inactive</td>
<td>1.185</td>
<td>0.975 1.441</td>
<td>1.181</td>
<td>0.946 1.474</td>
</tr>
<tr>
<td>BMI class</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Normal weight</td>
<td></td>
<td>Reference</td>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td>Underweight</td>
<td>0.821</td>
<td>0.488 1.383</td>
<td>1.817</td>
<td>0.672 4.911</td>
</tr>
<tr>
<td>Overweight</td>
<td>1.065</td>
<td>0.834 1.360</td>
<td>1.382</td>
<td>1.086 1.757</td>
</tr>
<tr>
<td>Obese</td>
<td>1.155</td>
<td>0.879 1.518</td>
<td>1.790</td>
<td>1.279 2.505</td>
</tr>
<tr>
<td>Smoking</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None smoker</td>
<td></td>
<td>Reference</td>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td>Current</td>
<td>1.305</td>
<td>0.995 1.712</td>
<td>1.455</td>
<td>1.075 1.969</td>
</tr>
<tr>
<td>Past</td>
<td>1.051</td>
<td>0.847 1.303</td>
<td>1.020</td>
<td>0.781 1.333</td>
</tr>
<tr>
<td>No of comorbidities</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td></td>
<td>Reference</td>
<td></td>
<td>Reference</td>
</tr>
<tr>
<td>1</td>
<td>1.485</td>
<td>1.137 1.941</td>
<td>1.364</td>
<td>1.062 1.750</td>
</tr>
<tr>
<td>2 or more</td>
<td>1.690</td>
<td>1.308 2.182</td>
<td>1.446</td>
<td>1.094 1.912</td>
</tr>
<tr>
<td>Use of pain relievers</td>
<td>1.222</td>
<td>0.988 1.511</td>
<td>1.184</td>
<td>0.960 1.461</td>
</tr>
</tbody>
</table>
# Appendix D

## Supplemental Material Manuscript 3

Model 1 Detailed Logistic Regression Output with all Covariates

| Model 1: | Outcome | Men | | Women | | | | | |
| --- | --- | --- | | | | | | | | |
| | Heart | All | | All | | Non-Obese (BMI<30) | | Obese (BMI 30+) | |
| Disease | | Adjusted OR | 95% CI | Adjusted OR | 95% CI | Adjusted OR | 95% CI | Adjusted OR | 95% CI |
| Arthritis | 1.01 | 0.83 | 1.23 | 1.33 | 1.11 | 1.59 | 1.37 | 1.13 | 1.67 | 1.18 | 0.77 | 1.80 |
| Activity Limitation | 1.61 | 1.32 | 1.96 | 1.64 | 1.39 | 1.94 | 1.57 | 1.30 | 1.89 | 1.97 | 1.35 | 2.87 |
| Age centered | 1.06 | 1.06 | 1.07 | 1.06 | 1.05 | 1.07 | 1.06 | 1.05 | 1.07 | 1.05 | 1.04 | 1.07 |
| Non-white | 0.85 | 0.56 | 1.30 | 0.84 | 0.52 | 1.35 | 0.80 | 0.47 | 1.36 | 0.88 | 0.27 | 2.91 |
| <= HS education | 0.90 | 0.75 | 1.08 | 0.91 | 0.77 | 1.07 | 0.90 | 0.75 | 1.08 | 0.93 | 0.64 | 1.37 |
| High blood pressure | 1.96 | 1.58 | 2.44 | 1.42 | 1.18 | 1.72 | 1.42 | 1.15 | 1.76 | 1.50 | 1.00 | 2.25 |
| Diabetes | 1.90 | 1.34 | 2.68 | 2.19 | 1.64 | 2.93 | 2.53 | 1.78 | 3.59 | 1.65 | 0.97 | 2.81 |
| Inactive | 1.09 | 0.91 | 1.30 | 0.98 | 0.83 | 1.16 | 1.01 | 0.84 | 1.21 | 0.88 | 0.60 | 1.31 |
| BMI Class | | | | | | | | | | | | |
| Underweight vs. healthy weight | 1.40 | 0.55 | 3.61 | 1.00 | 0.60 | 1.67 | | | | | | |
| Overweight vs. healthy weight | 1.20 | 0.99 | 1.46 | 1.09 | 0.92 | 1.30 | | | | | | |
| Obese vs. healthy weight | 1.49 | 1.16 | 1.93 | 1.14 | 0.92 | 1.42 | | | | | | |
| Smoking | | | | | | | | | | | | |
| Current smoker vs. never smoker | 1.68 | 1.32 | 2.12 | 1.46 | 1.18 | 1.80 | 1.29 | 1.02 | 1.63 | 2.44 | 1.48 | 4.04 |
| Past smoker vs. never smoker | 1.11 | 0.89 | 1.39 | 1.27 | 1.06 | 1.53 | 1.18 | 0.96 | 1.44 | 1.77 | 1.17 | 2.67 |
| Pain medication past month | 1.01 | 0.85 | 1.21 | 1.23 | 1.03 | 1.47 | 1.21 | 1.00 | 1.48 | 1.21 | 0.79 | 1.84 |
| Number of Comorbid Conditions | | | | | | | | | | | | |
| 1 vs. 0 | 1.22 | 1.01 | 1.48 | 1.24 | 1.03 | 1.48 | 1.24 | 1.02 | 1.52 | 1.31 | 0.86 | 1.98 |
| 2+ vs. 0 | 1.27 | 0.96 | 1.68 | 1.49 | 1.21 | 1.84 | 1.59 | 1.25 | 2.01 | 1.18 | 0.73 | 1.91 |
### Model 1 Detailed Logistic Regression Output with all Covariates Continued

<table>
<thead>
<tr>
<th>Alcohol Intake</th>
<th>All</th>
<th>95% CI</th>
<th>All</th>
<th>95% CI</th>
<th>All</th>
<th>95% CI</th>
<th>All</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2 drinks per day vs. abstainer</td>
<td>Men</td>
<td>Adjusted OR</td>
<td>0.74</td>
<td>0.61</td>
<td>0.91</td>
<td>0.69</td>
<td>0.52</td>
<td>0.91</td>
</tr>
<tr>
<td>2+ drinks per day vs. abstainer</td>
<td>Women</td>
<td>Adjusted OR</td>
<td>0.72</td>
<td>0.49</td>
<td>1.06</td>
<td>1.11</td>
<td>0.69</td>
<td>1.79</td>
</tr>
</tbody>
</table>
## Model 2 Detailed Logistic Regression Output with all Covariates

| Model 2: Outcome Activity Limitation | Men | | | Women |
|-------------------------------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
|                                     | All | All | Non-Obese (BMI <30) | Obese (BMI 30+) |
|                                     | Adjusted OR | 95% CI | Adjusted OR | 95% CI | Adjusted OR | 95% CI | Adjusted OR | 95% CI |
| Arthritis                           | 3.14 | 2.89 | 3.42 | 3.15 | 2.89 | 3.42 | 3.26 | 2.75 | 3.86 |
| Age_centered                       | 1.01 | 1.00 | 1.01 | 1.01 | 1.01 | 1.02 | 1.01 | 1.00 | 1.02 |
| Non-white                           | 0.92 | 0.78 | 1.08 | 0.86 | 0.75 | 1.00 | 0.86 | 0.73 | 1.01 |
| <= HS education                     | 1.00 | 0.92 | 1.09 | 0.98 | 0.91 | 1.05 | 0.96 | 0.89 | 1.04 |
| High blood pressure                 | 1.15 | 1.00 | 1.31 | 1.12 | 1.02 | 1.24 | 1.13 | 1.01 | 1.26 |
| Diabetes                            | 1.84 | 1.46 | 2.32 | 1.65 | 1.38 | 1.97 | 1.58 | 1.26 | 1.98 |
| Inactive                            | 1.03 | 0.95 | 1.11 | 1.09 | 1.03 | 1.16 | 1.10 | 1.03 | 1.17 |
| BMI Class                           |     |     |     |     |     |     |     |     |     |
| Underweight vs. healthy weight      | 1.92 | 1.29 | 2.86 | 1.03 | 0.86 | 1.23 |     |     |     |
| Overweight vs. healthy weight       | 0.88 | 0.81 | 0.96 | 1.03 | 0.97 | 1.11 |     |     |     |
| Obese vs. healthy weight            | 1.17 | 1.04 | 1.31 | 1.16 | 1.06 | 1.26 |     |     |     |
| Smoking                             |     |     |     |     |     |     |     |     |     |
| Current smoker vs. never smoker     | 1.59 | 1.44 | 1.75 | 1.31 | 1.22 | 1.42 | 1.31 | 1.21 | 1.42 |
| Past smoker vs. never smoker        | 1.19 | 1.08 | 1.31 | 1.00 | 0.93 | 1.08 | 1.00 | 0.92 | 1.09 |
| Pain medication past month          | 1.23 | 1.14 | 1.33 | 1.30 | 1.21 | 1.39 | 1.27 | 1.17 | 1.37 |
| Number of Comorbid Conditions       |     |     |     |     |     |     |     |     |     |
| 1 vs. 0                             | 2.12 | 1.95 | 2.30 | 1.84 | 1.72 | 1.97 | 1.85 | 1.72 | 1.99 |
| 2+ vs. 0                            | 3.91 | 3.48 | 4.40 | 3.41 | 3.14 | 3.71 | 3.46 | 3.15 | 3.80 |

Note: The table above shows the adjusted odds ratios (OR) and 95% confidence intervals (CI) for various factors associated with activity limitation, including Arthritis, Age_centered, Non-white, <= HS education, High blood pressure, Diabetes, Inactive, and BMI Class. The data is divided into all, Non-Obese (BMI <30), and Obese (BMI 30+) categories.
## Model 2 Detailed Logistic Regression Output with all Covariates Continued

<table>
<thead>
<tr>
<th>Model 2: Outcome Activity Limitation</th>
<th>Men</th>
<th>Women</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>All</td>
</tr>
<tr>
<td>Alcohol Intake</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1-2 drinks per day vs. abstainer</td>
<td>0.86</td>
<td>0.93</td>
</tr>
<tr>
<td></td>
<td>0.79</td>
<td>1.02</td>
</tr>
<tr>
<td>2+ drinks per day vs. abstainer</td>
<td>0.83</td>
<td>0.84</td>
</tr>
<tr>
<td></td>
<td>0.72</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Adjusted OR and 95% CI for alcohol intake.
Sensitivity Analysis Direct and Indirect Effects of Arthritis Mediated through Activity Limitation on Incident Heart Disease with 95% confidence intervals estimated using bootstrap methods based on 500 replications

<table>
<thead>
<tr>
<th></th>
<th>Men</th>
<th></th>
<th>Women</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>All</td>
<td>Non-Obese</td>
<td>Obese</td>
</tr>
<tr>
<td></td>
<td>Adjusted OR</td>
<td>95% CI</td>
<td>Adjusted OR</td>
<td>95% CI</td>
</tr>
<tr>
<td>Direct effect</td>
<td>1.01</td>
<td>0.82</td>
<td>1.21</td>
<td>1.33</td>
</tr>
<tr>
<td>Indirect effect</td>
<td>1.09</td>
<td>1.05</td>
<td>1.14</td>
<td>1.10</td>
</tr>
<tr>
<td>Total Effect</td>
<td>1.11</td>
<td>0.91</td>
<td>1.32</td>
<td>1.45</td>
</tr>
<tr>
<td>% Mediated</td>
<td>90%</td>
<td></td>
<td>29%</td>
<td></td>
</tr>
</tbody>
</table>
Appendix E

Online Appendix "Mediation analysis allowing for exposure-mediator interactions and causal interpretation: theoretical assumptions and implementation with SAS and SPSS macros"

Linda Valeri and Tyler J. VanderWeele

We let $Y_a$ and $M_a$ denote respectively the values of the outcome and mediator that would have been observed had the exposure $A$ been set to level $a$. We let $Y_{am}$ denote the value of the outcome that would have been observed had the exposure, $A$, and mediator, $M$, been set to levels $a$ and $m$, respectively.

The average controlled direct effect comparing exposure level $a$ to $a^*$ and fixing the mediator to level $m$ is defined by $CDE_{a,a^*}(m) = E[Y_{am} - Y_{a^*m}]$. The average natural direct effect is then defined by $NDE_{a,a^*}(a^*) = E[Y_{aM_a} - Y_{a^*M_a}]$. The average natural indirect effect can be defined as $NIE_{a,a^*}(a) = E[Y_{aM_a} - Y_{aM_{a^*}}]$, which compares the effect of the mediator at levels $M_a$ and $M_{a^*}$ on the outcome when exposure $A$ is set to $a$. Controlled direct effects and natural direct and indirect effects within strata of $C = c$ are then defined by:

$$CDE_{a,a^*|c}(m) = E[Y_{am} - Y_{a^*m}|c], \quad NDE_{a,a^*|c}(a^*) = E[Y_{aM_a} - Y_{a^*M_a}|c] \quad \text{and} \quad NIE_{a,a^*|c}(a) = E[Y_{aM_a} - Y_{aM_{a^*}}|c]$$

For a dichotomous outcome the total effect on the odds ratio scale conditional on $C = c$ is given by $OR_{a,a^*|c}^{TE} = \frac{P(Y_{am}=1|c)/(1-P(Y_{am}=1|c))}{P(Y_{a^*m}=1|c)/(1-P(Y_{a^*m}=1|c))}$. The controlled direct effect on the odds ratio scale is given by $OR_{a,a^*|c}^{CDE}(m) = \frac{P(Y_{am}=1|c)/(1-P(Y_{am}=1|c))}{P(Y_{a^*m}=1|c)/(1-P(Y_{a^*m}=1|c))}$. The natural direct effect on the odds ratio scale conditional on $C = c$ is given by $OR_{a,a^*|c}^{NDE}(a^*) = \frac{P(Y_{aM_a}=1|c)/(1-P(Y_{aM_a}=1|c))}{P(Y_{a^*M_{a^*}}=1|c)/(1-P(Y_{a^*M_{a^*}}=1|c))}$. The natural indirect effect on the odds ratio scale conditional on $C = c$ is given by $OR_{a,a^*|c}^{NIE}(a) = \frac{P(Y_{aM_a}=1|c)/(1-P(Y_{aM_a}=1|c))}{P(Y_{aM_{a^*}}=1|c)/(1-P(Y_{aM_{a^*}}=1|c))}$.

As discussed in the text, identification assumptions (i)-(iv) will suffice to identify these direct and indirect effects. If we let $X \perp Y|Z$ denote that $X$ is independent of $Y$ conditional on $Z$ then these four identification assumptions can be expressed formally in terms of counterfactual independence as (i) $Y_{am} \perp A|C$, (ii) $Y_{am} \perp M|\{A,C\}$, (iii) $M_a \perp A|C$, and (iv) $Y_{am} \perp M_{a^*}|C$. Assumptions (i) and (ii) suffice to identify controlled direct ef-
fects; assumptions (i)-(iv) suffice to identify natural direct and indirect effects (Pearl, 2001; VanderWeele and Vansteelandt, 2009). The intuitive interpretation of these assumptions as described in the text follows from the theory of causal diagrams (Pearl, 2001). Alternative identification assumptions have also been proposed (Imai 2010a; Hafeman and VanderWeele, 2011). However, it has been shown that the intuitive graphical interpretation of these alternative assumptions are in fact equivalent (Shpitser and VanderWeele, 2011). Technical examples can be constructed where one set of identification assumptions holds and another does not, but on a causal diagram corresponding to a set of non-parametric structural equations, whenever one set of the assumptions among those in VanderWeele and Vansteelandt (2009), Imai (2010a), and Hafeman and VanderWeele (2011) holds, the others will also.

1 Continuous Mediator and Outcome

Effects using regression

Suppose that both the mediator and the outcome are continuous and that the following models fit the observed data:

\[ E(M|A = a, C = c) = \beta_0 + \beta_1 a + \beta_2 c \] (1)

\[ E(Y|A = a, M = m, C = c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a \times m + \theta_4 c \] (2)

If the covariates \( C \) satisfied the no-unmeasured confounding assumptions (i)-(iv) above, then the average controlled direct effect and the average natural direct and indirect effects were derived by VanderWeele and Vansteelandt, 2009.

In particular, if the regression models (1) and (2) are correctly specified and assumptions of no unmeasured confounding of exposure-outcome relationship (i) and no unmeasured confounding of the mediator-outcome relationship (ii) hold, then we could compute the con-
trolled direct effect as follows:

\[
CDE = E[Y_{am} - Y_{a^*m}|C = c] \\
= E[Y|C = c, A = a, M = m] - E[Y|C = c, A = a^*, M = m] \\
= (\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a_m + \theta_4 c) - (\theta_0 + \theta_1 a^* + \theta_2 m + \theta_3 a^* m + \theta_4 c) \\
= (\theta_1 a + \theta_3 a_m - \theta_1 a^* - \theta_3 a^* m) \\
= \theta_1 (a - a^*) + \theta_3 m(a - a^*).
\]

If the regression models (1) and (2) are correctly specified and assumptions (i) and (ii) together with two additional assumptions of (iii) no unmeasured confounding of the exposure-mediator relationship and (iv) that there is no mediator-outcome confounder that is affected by the exposure hold, then we could compute the natural direct effects by:

\[
NDE = E[Y_{aM_{a^*}} - Y_{a^*M_{a^*}}|C = c] \\
= \sum_m \{E[Y|C = c, A = a, M = m] - E[Y|C = c, A = a^*, M = m]\} \times P(M = m|C = c, A = a^*) \\
= \sum_m \{(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 am + \theta_4 c) - (\theta_0 + \theta_1 a^* + \theta_2 m + \theta_3 a^* m + \theta_4 c)\} \times P(M = m|C = c, A = a^*) \\
= \sum_m \{(\theta_1 a + \theta_2 m + \theta_3 am) - (\theta_1 a^* + \theta_2 m + \theta_3 a^* m)\} \times P(M = m|C = c, A = a^*) \\
= \{(\theta_1 a + \theta_2 E[M|A = a^*, C = c] + \theta_3 a E[M|A = a^*, C = c]) - (\theta_1 a^* + \theta_2 E[M|A = a^*, C = c] + \theta_3 a^* E[M|A = a^*, C = c])\} \\
= \{(\theta_1 a + \theta_2 (\beta_0 + \beta_1 a + \beta_2 c) + \theta_3 a(\beta_0 + \beta_1 a + \beta_2 c) - (\theta_1 a^* + \theta_2 (\beta_0 + \beta_1 a + \beta_2 c) + \theta_3 a^*(\beta_0 + \beta_1 a + \beta_2 c))\} \\
= \{(\theta_1 a - a^*) + \theta_3 a(\beta_0 + \beta_1 a + \beta_2 c) - (\theta_1 a^* + \theta_3 a^*(\beta_0 + \beta_1 a + \beta_2 c))\} \\
= (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a + \theta_3 \beta_2 c)(a - a^*).
\]

Moreover under the same assumptions we can compute the natural indirect effects by:
\[ NIE = E[Y_{aM_a} - Y_{aM_a^*} | C = c] \]
\[ = \sum_m E[Y | C = c, A = a, M = m] \times P(M = m | C = c, A = a) - \sum_m E[Y | C = c, A = a, M = m] \times P(M = m | C = c, A = a^*) \]
\[ = \sum_m (\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \theta_4_c) \times P(M = m | C = c, A = a) - \sum_m (\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \theta_4_c) \times P(M = m | C = c, A = a^*) \]
\[ = (\theta_0 + \theta_1 a + \theta_2 E[M | A = a, C = c] + \theta_3 a E[M | A = a, C = c] + \theta_4_c) - (\theta_0 + \theta_1 a + \theta_2 E[M | A = a^*, C = c] + \theta_4_c) \]
\[ = (\theta_1 a + \theta_2 (\beta_0 + \beta_1 a + \beta_2 c) + \theta_3 a (\beta_0 + \beta_1 a + \beta_2 c)) - (\theta_1 a^* + \theta_2 (\beta_0 + \beta_1 a^* + \beta_2^c) + \theta_3 a^* (\beta_0 + \beta_1 a^* + \beta_2 c)) \]
\[ = (\theta_2 \beta_1 + \theta_3 \beta_1 a) (a - a^*). \]

If the regression models (1) and (2) are correctly specified and assumptions (i) and (ii) hold, then we could compute the total effect by:

\[ TE = E[Y_a - Y_{a^*} | C = c] \]
\[ = E[Y_{a,M(a)} - Y_{a^*,M(a^*)} | C = c] + E[Y_{a,M(a)} - Y_{a,M(a)} | C = c] \]
\[ = (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 c + \theta_2 \beta_1 + \theta_3 \beta_1 a) (a - a^*). \]

Finally if the regression models (1) and (2) are correctly specified and assumptions (i)-(iv) hold then we could compute the proportion mediated by:

\[ PM = \frac{E[Y_{aM_a} - Y_{aM_a^*} | C = c]}{E[Y_a - Y_{a^*} | C = c]} \]
\[ = \frac{\theta_2 \beta_1 + \theta_3 \beta_1 a}{\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 c + \theta_2 \beta_1 + \theta_3 \beta_1 a}. \]

\[ \text{Standard errors} \]
Suppose that model (1) and (2) have been fit using standard linear regression software and that the resulting estimates \( \hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)' \) and \( \hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_4)' \) have covariance matrices \( \Sigma_\beta \) and \( \Sigma_\theta \). Then the covariance matrix of \( (\hat{\beta}', \hat{\theta}') \) is

\[
\Sigma = \begin{bmatrix}
\Sigma_\beta & 0 \\
0 & \Sigma_\theta
\end{bmatrix}
\]

Standard errors of the controlled and natural direct and indirect effects can be obtained (using the delta method) as

\[
\sqrt{\Gamma \Sigma \Gamma'} |a - a^*|
\]

with \( \Gamma = (0, 0, 0', 0, 1, 0, m, 0') \) for the controlled direct effect, \( \Gamma = (\theta_3, \theta_3a^*, \theta_3c', 0, 1, 0, \beta_0 + \beta_1a^* + \beta'_2c, 0') \) for the pure natural direct effect (same expression holds for the total natural direct effect upon substituting \( a \) and \( a^* \)), \( \Gamma = (0, \theta_2 + \theta_3a, 0', 0, \beta_0, \beta_1a, 0') \) for the total natural indirect effect (the same expression holds for the pure natural indirect effect upon substituting \( a \) and \( a^* \)), \( \Gamma = (\theta_3, \theta_3(a + a^*) + \theta_2, \theta_3c', 0, 1, \beta_0 + \beta_1(a + a^*) + \beta'_2c, 0') \) for the total effect and for the proportion mediated \( \Gamma = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8) \) where

\[
d_1 = -\theta_3 \frac{\theta_2 \beta_1 + \theta_3 \beta_1 a}{(\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta'_2c + \theta_2 \beta_1 + \theta_3 \beta_1 a)^2}
\]

\[
d_2 = \frac{(\theta_2 + \theta_3a)(-(\theta_2 \beta_1 + \theta_3 \beta_1 a) + (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta'_2c + \theta_2 \beta_1 + \theta_3 \beta_1 a)) - \theta_3 a^*}{(\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta'_2c + \theta_2 \beta_1 + \theta_3 \beta_1 a)^2}
\]

\[
d_3 = -\frac{\theta_3 c'(\theta_2 \beta_1 + \theta_3 \beta_1 a)}{(\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta'_2c + \theta_2 \beta_1 + \theta_3 \beta_1 a)^2}
\]

\[
d_4 = 0
\]
2 Continuous Mediator and Binary Outcome

Effects using regression

Suppose that the mediator is continuous and the outcome is binary and is rare. Suppose that the following models fit the observed data:

\[ E(M|A = a, C = c) = \beta_0 + \beta_1 a + \beta_2 c \]  \hspace{1cm} (3)

\[ \text{logit}\{P(Y = 1|A = a, M = m, C = c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a \times m + \theta_4 c \]  \hspace{1cm} (4)

and that the error term in the regression model for M is normally distributed with mean 0 and variance \(\sigma^2\). If the regression models (3) and (4) are correctly specified and assumptions (i) and (ii) hold then the conditional controlled direct effect on the odds ratio scale would be given by (VanderWeele and Vansteelandt, 2010):

\[ d_5 = -\frac{\theta_2 \beta_1 + \theta_3 \beta_1 a}{(\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 c + \theta_2 \beta_1 + \theta_3 \beta_1 a)^2} \]

\[ d_6 = \frac{\beta_1 (-\theta_2 \beta_1 + \theta_3 \beta_1 a) + (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 c + \theta_2 \beta_1 + \theta_3 \beta_1 a)}{(\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 c + \theta_2 \beta_1 + \theta_3 \beta_1 a)^2} \]

\[ d_7 = \frac{\beta_1 a (\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 c + \theta_2 \beta_1 + \theta_3 \beta_1 a) - (\beta_0 + \beta_1 (a + a^*) + \beta_2 c)(\theta_2 \beta_1 + \theta_3 \beta_1 a)}{(\theta_1 + \theta_3 \beta_0 + \theta_3 \beta_1 a^* + \theta_3 \beta_2 c + \theta_2 \beta_1 + \theta_3 \beta_1 a)^2} \]

\[ d_8 = 0 \]
\[
OR^{CDE} = \frac{P(Y_{am}=1|c)/(1-P(Y_{am}=1|c))}{P(Y_{am}=1|a, m, c)/(1-P(Y_{am}=1|a, m, c))} \\
= \frac{P(Y=1|a, m, c)/(1-P(Y=1|a, m, c))}{P(Y=1|a^*, m, c)/(1-P(Y=1|a^*, m, c))} \\
= \frac{\exp[\theta_0+\theta_1a+\theta_2m+\theta_3am+\theta_4c]}{\exp[\theta_0+\theta_1a^*+\theta_2m+\theta_3am+\theta_4c]} \\
= \exp[(\theta_1 + \theta_3m)(a - a^*)].
\]

If the regression models (3) and (4) are correctly specified and assumptions (i)-(iv) hold, the outcome \(Y\) is rare, and the error term for linear regression model (1) is normally distributed and has constant variance \(\sigma^2\), then we could compute the natural direct effects by:

\[
OR^{NDE} = \exp[\log\left\{\frac{P(Y_{am}=1|c)/(1-P(Y_{am}=1|c))}{P(Y_{am}=1|m, c)/(1-P(Y_{am}=1|m, c))}\right\}] \\
= \exp[\log\{P(Y_{am} = 1|c)\} - \log\{P(Y_{am} = 1|m, c)\}] \\
\sim \exp[\theta_0 + \theta_1a + \theta_4c + (\theta_2 + \theta_3a)\{\beta_0 + \beta_1a + \beta_2c\} + \frac{1}{2}(\theta_2 + \theta_3a)^2\sigma^2 - \{\theta_0 + \theta_1a + \theta_4c + (\theta_2 + \theta_3a)\}^2\sigma^2] \\
= \exp[\{\theta_1 + \theta_3(\beta_0 + \beta_1a + \beta_2c + \theta_2\sigma^2)\}(a - a^*) + 0.5\theta_3^2\sigma^2(a^2 - a^*^2)].
\]

If the regression models (3) and (4) are correctly specified and assumptions (i)-(iv) hold, the outcome \(Y\) is rare, and the error term for linear regression model (1) is normally distributed and has constant variance \(\sigma^2\), then we could compute the natural indirect effects by:

\[
OR^{NIE} = \exp[\log\left\{\frac{P(Y_{am}=1|c)/(1-P(Y_{am}=1|c))}{P(Y_{am}=1|m, c)/(1-P(Y_{am}=1|m, c))}\right\}] \\
= \exp[\log\{P(Y_{am} = 1|c)\} - \log\{P(Y_{am} = 1|m, c)\}] \\
\sim \exp[\theta_0 + \theta_1a + \theta_4c + (\theta_2 + \theta_3a)\{\beta_0 + \beta_1a + \beta_2c\} + \frac{1}{2}(\theta_2 + \theta_3a)^2\sigma^2 - \{\theta_0 + \theta_1a + \theta_4c + (\theta_2 + \theta_3a)\}^2\sigma^2] \\
= \exp[(\theta_2\beta_1 + \theta_3\beta_1a)(a - a^*)].
\]

If the regression models (3) and (4) are correctly specified and assumptions (i)-(iv) hold,
the outcome $Y$ is rare, and the error term for linear regression model (3) is normally distributed and has constant variance $\sigma^2$, then we could compute the total effects by:

$$OR^{TE} = \exp[\log\left\{ \frac{P(Y_{a,M_a^*} = 1|c)}{1-P(Y_{a,M_a^*} = 1|c)} \right\} \times \exp[\log\left\{ \frac{P(Y_{a,M_a} = 1|c)}{1-P(Y_{a,M_a} = 1|c)} \right\}]$$

$$= E[Y_{a,M_a^*} - Y_{a^*,M_a^*}|C = c] \times E[Y_{a,M_a} - Y_{a^*,M_a}|C = c]$$

$$= \exp[(\theta_1 + \theta_3\beta_0 + \theta_3\beta_1a^* + \theta_3\beta_2c + \theta_2\beta_1 + \theta_3\beta_1a + \theta_3\theta_2\sigma^2)(a - a^*) + 0.5\theta_3^2\sigma^2(a^2 - a^{*2})].$$

If the regression models (3) and (4) are correctly specified and assumptions (i)-(iv) hold then we can compute the proportion mediated by:

$$PM = OR^{NDE} \times (OR^{NIE} - 1)/(OR^{NDE} \times OR^{NIE} - 1).$$

These expressions apply also if the outcome is not rare and log-linear rather than logistic models are fit to the outcome model; the direct and indirect effect will have now an interpretation on the risk ratio scale rather than on the odds ratio scale.

These expressions apply also if the outcome is a count variable. In particular if $Y \sim Pois(\lambda)$ for $\lambda = \exp\{\theta_0 + \theta_1a + \theta_2m + \theta_3a \ast m + \theta'_4c\}$ the outcome regression can be defined as:

$$\log\{E(Y|A = a, M = m, C = c)\} = \theta_0 + \theta_1a + \theta_2m + \theta_3a \ast m + \theta'_4c$$

The natural direct effect for binary outcome on the risk ratio scale coincides with the natural direct effect for poisson count outcome since:

$$RR^{NDE} = \exp[\log\left\{ \frac{E(Y_{a,M_a}|C)}{E(Y_{a^*,M_a^*}|C)} \right\}]$$

The same argument holds for the natural indirect effect. Finally, the argument can be
extended to the case in which the count outcome is modeled with a negative binomial distribution. This is the case since the negative binomial distribution can be represented as an over-dispersed poisson and the mean of the two models coincide.

**Standard errors**

We now consider standard errors for the controlled direct effect and natural direct and indirect effect odds ratios. Suppose that model (4) has been fit using standard logistic regression software and that model (3) has been fit using standard linear regression software. Suppose furthermore that the resulting estimates $\hat{\beta} = (\beta_0, \beta_1, \beta_2')$, $\hat{\theta} = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4')$ and $\hat{\sigma}^2$ of $\sigma$ have covariance matrices $\Sigma_{\beta}$ and $\Sigma_{\theta}$. Then the covariance matrix of $(\hat{\beta}', \hat{\theta}', \hat{\sigma}^2)$ is

$$
\Sigma = \begin{bmatrix}
\Sigma_{\beta} & 0 & 0 \\
0 & \Sigma_{\theta} & 0 \\
0 & 0 & \Sigma_{\sigma^2}
\end{bmatrix}
$$

Standard errors of the controlled and natural direct and indirect effects can be obtained (using the delta method) as

$$\sqrt{\Gamma \Sigma \Gamma^T |a - a^*|}$$

with $\Gamma = (0, 0, 0', 0, 1, 0, m, 0', 0)$ for the log of controlled direct effect odds ratio, $\Gamma = (\theta_3, \theta_3 a^*, \theta_3 c', 0, 1, \theta_3 \sigma^2, \beta_0 + \beta_1 a^* + \beta_2 c + \theta_2 \sigma^2 + \theta_3 \sigma^2(a + a^*), 0', \theta_2 \theta_3 + 0.5 \theta_3^2(a + a^*))$ for the log pure natural direct effect odds ratio (same expression holds for the total natural direct effect upon substituting $a$ and $a^*$), $\Gamma = (0, \theta_2 + \theta_3 a, 0', 0, 0, \beta_1, \beta_1 a, 0', 0)$ for the log of total natural indirect effect (the same expression holds for the pure natural indirect effect upon substituting $a$ and $a^*$), $\Gamma = (\theta_3, \theta_3(a + a^*) + \theta_2, \theta_3 c', 0, 1, \theta_3 \sigma^2 + \beta_1, \beta_0 + \beta_1(a + a^*) + \beta_2 c + \theta_2 \sigma^2 + \theta_3 \sigma^2(a^2 - a^*), 0', 0.5 \theta_3^2(a^2 - a^{*2})$) for the logarithm of the total effect.
3 Binary Mediator and Continuous Outcome

Effects using regression

Suppose that the outcome is continuous, the mediator is binary and that the following models fit the observed data:

\[
\text{logit}\{P(M = 1|A = a, C = c)\} = \beta_0 + \beta_1 a + \beta_2 c
\]  

(5)

\[
E(Y|A = a, M = m, C = c) = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a \cdot m + \theta_4 c
\]  

(6)

In particular, if the regression models (5) and (6) are correctly specified and assumptions (i) and (ii) hold then we could compute the average controlled direct effect as in section 1

If the regression models (5) and (6) are correctly specified and assumptions (i)-(iv) hold then we could compute the average natural direct effects by:

\[
NDE = E[Y_{aM_a} - Y_{a^*M_{a^*}}|C = c] = \sum_m \{E[Y|C = c, A = a, M = m] - E[Y|C = c, A = a^*, M = m]\} \times P(M = m|C = c, A = a^*)
\]

\[
= \sum_m \{E[Y|C = c, A = a, M = m] - E[Y|C = c, A = a^*, M = m]\} \times P(M = m|C = c, A = a^*)
\]

\[
= \{\theta_1(a - a^*)\} + \{\theta_3(a - a^*)\} \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}{1 + \exp[\beta_0 + \beta_1 a^* + \beta_2 c]}.
\]

If the regression models (5) and (6) are correctly specified and assumptions (i)-(iv) hold then we could compute the average natural indirect effects by:

\[
NIE = E[Y_{aM_a} - Y_{aM_{a^*}}|C = c] = \sum_m E[Y|C = c, A = a, M = m] \times P(M = m|C = c, A = a) - \sum_m E[Y|C = c, A = a, M = m] \times P(M = m|C = c, A = a^*)
\]
m| × P(M = m|C = c, A = a*)

= \sum_m(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \theta_4 c) \times P(M = m|C = c, A = a) - \sum_m(\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a m + \theta_4 c) \times P(M = m|C = c, A = a^*)

= (\theta_2 + \theta_3 a)\{E[M|A = a, C = c] - E[M|A = a^*, C = c]\}

= (\theta_2 + \theta_3 a)\{\frac{\exp[\beta_0 + \beta_1 a + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a + \beta_2 c]} - \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}\}.

If the regression models (5) and (6) are correctly specified and assumptions (i)-(iv) hold then we could compute the total effect by:

TE = E[Y_a - Y_{a^*}|C = c]

= E[Y_{aM_a} - Y_{a^*M_{a^*}}|C = c] + E[Y_{aM_a} - Y_{a^*M_{a^*}}|C = c]

= (\theta_1(a - a^*)) + (\theta_3(a - a^*))\{\frac{\exp[\beta_0 + \beta_1 a + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a + \beta_2 c]} + (\theta_2 + \theta_3 a)\{\frac{\exp[\beta_0 + \beta_1 a + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a + \beta_2 c]} - \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}\}.

If the regression models (5) and (6) are correctly specified and assumptions (i)-(iv) hold then we could compute the proportion mediated by:

PM = \frac{E[Y_{aM_a} - Y_{a^*M_{a^*}}|C = c]}{E[Y_a - Y_{a^*}|C = c]}

= \frac{(\theta_2 + \theta_3 a)\{\frac{\exp[\beta_0 + \beta_1 a + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a + \beta_2 c]} - \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}\}}{(\theta_2 + \theta_3 a)\{\frac{\exp[\beta_0 + \beta_1 a + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a + \beta_2 c]} - \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}\} + (\theta_3(a - a^*))\{\frac{\exp[\beta_0 + \beta_1 a + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a + \beta_2 c]} - \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}{1+\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}\}}.

**Standard errors**

Suppose that model (6) have been fit using standard linear regression software and that model (5) have been fit using standard logistic regression. The resulting estimates are \(\hat{\beta}\) of \(\beta = (\beta_0, \beta_1, \beta_2)\)' and \(\hat{\theta}\) of \(\theta = (\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)\)' have covariance matrices \(\Sigma_\beta\) and \(\Sigma_\theta\). Then
the covariance matrix of \((\hat{\beta}', \hat{\theta}')\) is

\[
\Sigma = \begin{bmatrix} \Sigma_{\beta} & 0 \\ 0 & \Sigma_{\theta} \end{bmatrix}
\]

Standard errors of the controlled and natural direct can be obtained (using the delta method) as

\[
\sqrt{\Gamma \Sigma \Gamma'} |a - a^*|
\]

with \(\Gamma = (0, 0, 0', 0, 1, 0, m, 0')\) for the controlled direct effect, \(\Gamma = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8)\) for the pure natural direct effect (same expression holds for the total natural direct effect upon substituting \(a\) and \(a^*\)), where

\[
d_1 = \frac{\theta_3 \exp[\beta_0 + \beta_1 a^* + \beta_2' c](1 + \exp[\beta_0 + \beta_1 a^* + \beta_2' c]) - \theta_3 \{\exp[\beta_0 + \beta_1 a^* + \beta_2' c]\}^2}{(1 + \exp[\beta_0 + \beta_1 a^* + \beta_2' c])^2}
\]

\[
d_2 = \frac{\theta_3 a^* \exp[\beta_0 + \beta_1 a^* + \beta_2' c](1 + \exp[\beta_0 + \beta_1 a^* + \beta_2' c]) - \{\exp[\beta_0 + \beta_1 a^* + \beta_2' c]\}^2}{(1 + \exp[\beta_0 + \beta_1 a^* + \beta_2' c])^2}
\]

\[
d_3 = \frac{\theta_3 c' \exp[\beta_0 + \beta_1 a^* + \beta_2' c](1 + \exp[\beta_0 + \beta_1 a^* + \beta_2' c]) - \{\exp[\beta_0 + \beta_1 a^* + \beta_2' c]\}^2}{(1 + \exp[\beta_0 + \beta_1 a^* + \beta_2' c])^2}
\]

\[
d_4 = 0
\]

\[
d_5 = 1
\]

\[
d_6 = 0
\]

\[
d_7 = \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2' c]}{1 + \exp[\beta_0 + \beta_1 a^* + \beta_2' c]}
\]

\[
d_8 = 0'
\]

Standard errors of the natural indirect can be obtained (using the delta method) as

\[
\sqrt{\Gamma \Sigma \Gamma'}
\]
For the natural indirect effect (the same expression holds for the pure natural indirect effect upon substituting a and \(a^*\)) let

\[
A = \frac{\exp[\beta_0 + \beta_1a + \beta_2c] \{1 + \exp[\beta_0 + \beta_1a + \beta_2c]\} - \{\exp[\beta_0 + \beta_1a + \beta_2c]\}^2}{\{1 + \exp[\beta_0 + \beta_1a + \beta_2c]\}^2}
\]

\[
B = \frac{\exp[\beta_0 + \beta_1a^* + \beta_2c] \{1 + \exp[\beta_0 + \beta_1a^* + \beta_2c]\} - \{\exp[\beta_0 + \beta_1a^* + \beta_2c]\}^2}{\{1 + \exp[\beta_0 + \beta_1a^* + \beta_2c]\}^2}
\]

\[
K = \frac{\exp[\beta_0 + \beta_1a + \beta_2c]}{\{1 + \exp[\beta_0 + \beta_1a + \beta_2c]\}}
\]

\[
D = \frac{\exp[\beta_0 + \beta_1a^* + \beta_2c]}{\{1 + \exp[\beta_0 + \beta_1a^* + \beta_2c]\}}
\]

and

\[
\Gamma = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8), \text{ where}
\]

\[
d_1 = \{\theta_2 + \theta_3a\} [A - B]
\]

\[
d_2 = \{\theta_2 + \theta_3a\} [aA - a^*B]
\]

\[
d_3 = \{\theta_2 + \theta_3a\} c \{A - B\}
\]

\[
d_4 = 0
\]

\[
d_5 = 0
\]

\[
d_6 = K - D
\]

\[
d_7 = a[K - D]
\]

\[
d_8 = 0
\]

Standard errors of the controlled and total effect and percentage mediated can be obtained
(using the delta method) as

\[ \sqrt{\Gamma \Sigma \Gamma’} \]

let

\[ A = \frac{\exp[\beta_0 + \beta_1 a + \beta_2 c](1 + \exp[\beta_0 + \beta_1 a + \beta_2 c]) - \{\exp[\beta_0 + \beta_1 a + \beta_2 c]\}^2}{1 + \exp[\beta_0 + \beta_1 a + \beta_2 c]^2} \]

\[ B = \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2 c](1 + \exp[\beta_0 + \beta_1 a^* + \beta_2 c]) - \{\exp[\beta_0 + \beta_1 a^* + \beta_2 c]\}^2}{1 + \exp[\beta_0 + \beta_1 a^* + \beta_2 c]^2} \]

\[ K = \frac{\exp[\beta_0 + \beta_1 a + \beta_2 c]}{1 + \exp[\beta_0 + \beta_1 a + \beta_2 c]} \]

\[ D = \frac{\exp[\beta_0 + \beta_1 a^* + \beta_2 c]}{1 + \exp[\beta_0 + \beta_1 a^* + \beta_2 c]} \]

for the total effect \( \Gamma = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8) \), where

\[ d_1 = \theta_3(a - a^*)B + (\theta_2 + \theta_3 a)(A - B) \]

\[ d_2 = a^*\theta_3(a - a^*)B + (\theta_2 + \theta_3 a)(aA - a^*B) \]

\[ d_3 = c^*\theta_3(a - a^*)B + (\theta_2 + \theta_3 a)(A - B) \]

\[ d_4 = 0 \]

\[ d_5 = a - a^* \]

\[ d_6 = K - D \]

\[ d_7 = (a - a^*)D + a[K - D] \]

\[ d_8 = 0 \]
and for the proportion mediated $\Gamma = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8)$ where

$$d_1 = \frac{[(\theta_2 + \theta_3 a)(A - B)] - [\{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D\} - [\{(\theta_2 + \theta_3 a)(A - B)\} + (a - a^*)\theta_1 + \theta_3 D\}]^2}{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D]^2}$$

$$d_2 = \frac{[(\theta_2 + \theta_3 a)(A - a^* B)] - [\{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D\} - [\{(\theta_2 + \theta_3 a)(A - a^* B)\} + (a - a^*)\theta_1 + \theta_3 D\}]^2}{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D]^2}$$

$$d_3 = \frac{[(\theta_2 + \theta_3 a)c'(A - B)] - [\{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D\} - c' \{(\theta_2 + \theta_3 a)(A - B)\} + (a - a^*)\theta_1 + \theta_3 D\}]^2}{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D]^2}$$

$$d_4 = 0$$

$$d_5 = \frac{(a - a^*)(\theta_2 + \theta_3 a)[K - D]}{[(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D]^2}$$

$$d_6 = \frac{a[K - D]\{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D\} - [K - D]\{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D\}]^2}{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D]^2}$$

$$d_7 = \frac{[K - D]\{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D\} - a[K - D\{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D\}]{\{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D\}]^2}{(\theta_2 + \theta_3 a)(K - D) + (a - a^*)\theta_1 + \theta_3 D]^2}$$

$$d_8 = 0.$$

### 4 Binary Mediator and Binary Outcome

**Effects using regression**

Suppose that both the outcome and the mediator are binary and that the following models fit the observed data:

$$\logit\{P(M = 1|A = a, C = c)\} = \beta_0 + \beta_1 a + \beta_2 c$$  \hspace{1cm} (7)

$$\logit\{P(Y = 1|A = a, M = m, C = c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a * m + \theta_4 c$$  \hspace{1cm} (8)

If the regression models (7) and (8) are correctly specified and assumptions (i) and (ii) hold then we can compute the controlled direct effect odds ratio as the case in which the mediator is continuous and the outcome is binary.
If the regression models (7) and (8) are correctly specified and assumptions (i)-(iv) hold and the outcome Y is rare, then we could compute the average natural direct effects by:

\[
OR^{NDE} = \exp[\log\left\{ \frac{P(Y_{aM*}=1|c)/(1-P(Y_{aM*}=1|c))}{P(Y_{aM*}=1|c)/(1-P(Y_{aM*}=1|c))} \right\}]
\]

\[
= \exp[\log\{P(Y_{aM*} = 1|c)\} - \logit\{P(Y_{aM*} = 1|c)\}]
\]

\[
\sim \exp[\log\left\{ \frac{\exp(\theta_0 + \theta_1 a + \theta_2 c) + \exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)}{1 + \exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)} \right\} - \log\left\{ \frac{\exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)}{1 + \exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)} \right\}]
\]

\[
= \left\{ \frac{\exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)}{\exp(\theta_0 + \theta_1 a + \theta_2 c)} \right\}.
\]

If the regression models (7) and (8) are correctly specified and assumptions (i)-(iv) hold and the outcome Y is rare, then we could compute the average natural indirect effects by:

\[
OR^{NIE} = \exp[\log\left\{ \frac{P(Y_{aM*}=1|c)/(1-P(Y_{aM*}=1|c))}{P(Y_{aM*}=1|c)/(1-P(Y_{aM*}=1|c))} \right\}]
\]

\[
= \exp[\log\{P(Y_{aM*} = 1|c)\} - \logit\{P(Y_{aM*} = 1|c)\}]
\]

\[
\sim \exp[\log\left\{ \frac{\exp(\theta_0 + \theta_1 a + \theta_2 c) + \exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)}{1 + \exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)} \right\} - \log\left\{ \frac{\exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)}{1 + \exp(\theta_0 + \theta_1 a + \theta_2 c + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)} \right\}]
\]

\[
= \left\{ \frac{[1+\exp(\beta_0 + \beta_1 a + \beta_2 c)][1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)]}{[1+\exp(\theta_0 + \theta_1 a + \theta_2 c)][1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)]} \right\}.
\]

If the regression models (7) and (8) are correctly specified and assumptions (i)-(iv) hold, the outcome Y is rare, then we could compute the total effects by:

\[
OR^{TE} = \exp[\log\left\{ \frac{P(Y_{aM*}=1|c)/(1-P(Y_{aM*}=1|c))}{P(Y_{aM*}=1|c)/(1-P(Y_{aM*}=1|c))} \right\} \times \exp[\log\left\{ \frac{P(Y_{aM*}=1|c)/(1-P(Y_{aM*}=1|c))}{P(Y_{aM*}=1|c)/(1-P(Y_{aM*}=1|c))} \right\}]
\]

\[
= \left\{ \frac{\exp(\theta_1 a)(1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c))}{\exp(\theta_1 a)[1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)]} \right\} \times \left\{ \frac{[1+\exp(\beta_0 + \beta_1 a + \beta_2 c)][1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)]}{[1+\exp(\theta_0 + \theta_1 a + \theta_2 c)][1+\exp(\theta_2 + \theta_3 a + \beta_0 + \beta_1 a + \beta_2 c)]} \right\}.
\]

If the regression models (7) and (8) are correctly specified and assumptions (i)-(iv) hold then we can compute the proportion mediated by:
\[ PM = OR^{NDE} \times (OR^{NIE} - 1)/(OR^{NDE} \times OR^{NIE} - 1). \]

These expressions apply also if the outcome is not rare and log-linear rather than logistic models are fit to the outcome model; the direct and indirect effect will have now an interpretation on the risk ratio scale rather than on the odds ratio scale.

These expressions apply also if the outcome is a count variable. In particular if \( Y \sim Poi(\lambda) \) for \( \lambda = \exp\{\theta_0 + \theta_1 a + \theta_2 m + \theta_3 a \times m + \theta'_4 c\} \) the outcome regression can be defined as:

\[
\log\{E(Y|A = a, M = m, C = c)\} = \theta_0 + \theta_1 a + \theta_2 m + \theta_3 a \times m + \theta'_4 c
\]

The natural direct effect for binary outcome on the risk ratio scale coincides with the natural direct effect for poisson count outcome since:

\[
RR^{NDE} = \exp[\log\left\{\frac{E(Y_{aM\ast}|c)}{E(Y_{a\ast M\ast}|c)}\right\}]
\]

The same argument holds for the natural indirect effect. Finally, the argument can be extended to the case in which the count outcome is modeled with a negative binomial distribution. This is the case since the negative binomial distribution can be represented as an over-dispersed poisson and the mean of the two models coincide.

**Standard Errors:**

Suppose that model (7) and (8) have been fit using standard logistic regression software and that the resulting estimates \( \hat{\beta} \) of \( \beta = (\beta_0, \beta_1, \beta_2)' \) and \( \hat{\theta} \) of \( \theta = (\theta_0, \theta_1, \theta_2, \theta_3, \theta'_4)' \) have covariance matrices \( \Sigma_{\beta} \) and \( \Sigma_{\theta} \). Then the covariance matrix of \( (\hat{\beta}', \hat{\theta}') \) is
\[ \Sigma = \begin{bmatrix} \Sigma_\beta & 0 \\ 0 & \Sigma_\theta \end{bmatrix} \]

Standard errors of the controlled and natural direct and indirect effects can be obtained (using the delta method) as

\[ \sqrt{\Gamma \Sigma \Gamma'} \]

with \( \Gamma = (0, 0, 0', \beta - \beta^*, 0, m(\beta - \beta^*), 0') \) for the controlled direct effect, \( \Gamma = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8) \) for the logarithm of the pure natural direct effect (same expression holds for the logarithm of the total natural direct effect upon substituting \( a \) and \( a^* \)), where let

\[ A = \frac{\exp[\theta_2 + \theta_3a + \beta_0 + \beta_1a^* + \beta_2'c]}{1 + \exp[\theta_2 + \theta_3a + \beta_0 + \beta_1a^* + \beta_2'c]} \]

\[ B = \frac{\exp[\theta_2 + \theta_3a^* + \beta_0 + \beta_1a^* + \beta_2'c]}{1 + \exp[\theta_2 + \theta_3a^* + \beta_0 + \beta_1a^* + \beta_2'c]} \]

and

\[ d_1 = A - B \]

\[ d_2 = a^*(A - B) \]

\[ d_3 = c'(A - B) \]

\[ d_4 = 0 \]

\[ d_5 = (a - a^*) \]

\[ d_6 = A - B \]

\[ d_7 = aA - a^*B \]
\[
d_8 = 0'
\]

for the logarithm of the natural indirect effect (the same expression holds for the pure natural indirect effect upon substituting \(a\) and \(a^*\)) let

\[
A = \frac{\exp[\theta_2 + \theta_3a + \beta_0 + \beta_1a + \beta_2'c]}{1 + \exp[\theta_2 + \theta_3a + \beta_0 + \beta_1a + \beta_2'c]}
\]

\[
B = \frac{\exp[\theta_2 + \theta_3a + \beta_0 + \beta_1a^* + \beta_2'c]}{1 + \exp[\theta_2 + \theta_3a + \beta_0 + \beta_1a^* + \beta_2'c]}
\]

\[
K = \frac{\exp[\beta_0 + \beta_1a + \beta_2'c]}{1 + \exp[\beta_0 + \beta_1a + \beta_2'c]}
\]

\[
D = \frac{\exp[\beta_0 + \beta_1a^* + \beta_2'c]}{1 + \exp[\beta_0 + \beta_1a^* + \beta_2'c]}
\]

and

\[
\Gamma = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8)
\]

where

\[
d_1 = (D + A) - (K + B)
\]

\[
d_2 = a^*[D - B] + a[A - K]
\]

\[
d_3 = c'[(D + A) - (K + B)]
\]

\[
d_4 = 0
\]

\[
d_5 = 0
\]

\[
d_6 = A - B
\]

\[
d_7 = a[A - B]
\]

\[
d_8 = 0'
\]
Standard errors of the logarithm of the total effect and percentage mediated can be obtained (using the delta method) as

\[ \sqrt{\Gamma \Sigma \Gamma'} \]

Let \( d_i(\log(pnde)) \) and \( d_i(\log(tnie)) \) for \( i = 1, \ldots, 8 \), the gamma elements derived for the logarithm of the pure natural direct effect and the total natural indirect effect respectively. For the total effect \( \Gamma = (d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8) \), where

\[
\begin{align*}
    d_1 &= d_1(\log(pnde)) + d_1(\log(tnie)) \\
    d_2 &= d_2(\log(pnde)) + d_2(\log(tnie)) \\
    d_3 &= d_3(\log(pnde)) + d_3(\log(tnie)) \\
    d_4 &= d_4(\log(pnde)) + d_4(\log(tnie)) \\
    d_5 &= d_5(\log(pnde)) + d_5(\log(tnie)) \\
    d_6 &= d_6(\log(pnde)) + d_6(\log(tnie)) \\
    d_7 &= d_7(\log(pnde)) + d_7(\log(tnie)) \\
    d_8 &= d_8(\log(pnde)) + d_8(\log(tnie))
\end{align*}
\]
Title

`paramed -- causal mediation analysis using parametric regression models`

Syntax

```
paramed varname, avar(varname) mvar(varname) a0(real) a1(real) m(real) yreg(string) mreg(string) [cvars(varlist) interaction casecontrol fulloutput c(numlist) bootstrap reps(integer 200) level(cilevel) seed(passthru)]
```

`varname` - this specifies the outcome variable.

`avar(varname)` - this specifies the treatment (exposure) variable.

`mvar(varname)` - this specifies the mediator variable.

`a0(real)` - this specifies the baseline level of the treatment (exposure).

`a1(real)` - this specifies the alternative treatment (exposure) level.

`m(real)` - this specifies the level of mediator at which the controlled direct effect is to be estimated. If there is no treatment (exposure)-mediator interaction the controlled direct effect is the same at all levels of the mediator and so an arbitrary value can be chosen.

`yreg(string)` - this specifies the types of regression to be implemented for the outcome variable. This can be either `linear`, `logistic`, `loglinear`, `Poisson` or `Negative binomial`.

`mreg(string)` - this specifies the types of regression to be implemented for the mediator. This can be either `linear` or `logistic`.

Description

`paramed` performs causal mediation analysis using parametric regression models. Two models are estimated: a model for the mediator conditional on treatment (exposure) and covariates (if specified), and a model for the outcome conditional on treatment (exposure), the mediator and covariates (if specified). It extends statistical mediation analysis (widely known as Baron and Kenny procedure) to allow for the presence of treatment (exposure)-mediator interactions in the outcome regression.
model using counterfactual definitions of direct and indirect effects.

**paramed** allows continuous, binary or count outcomes, and continuous or binary mediators, and requires the user to specify an appropriate form for the regression models.

**paramed** provides estimates of the controlled direct effect, the natural direct effect, the natural indirect effect and the total effect with standard errors and confidence intervals derived using the delta method by default, with a bootstrap option also available.

**Options**

**cvars(varlist)** - this option specifies the list of covariates to be included in the analysis. Categorical variables need to be coded as a series of dummy variables before being entered as covariates.

**interaction** - this specifies whether a treatment (exposure)-mediator interaction is present (no interaction if this option is omitted).

**fulloutput** - this option specifies the output mode, which can be either *reduced* or *full*. The reduced output is the default option (if this option is omitted). The results matrix contains the controlled direct effect, natural direct effect, natural indirect effect and total effect. When the *full* option is specified, both conditional effects and effects evaluated at the mean covariate levels are shown.

**c(numlist)** - this option is used when the output option is *full*. When the output mode is *full*, fixed values must be provided for the covariates at which conditional effects are computed (the number of values must correspond to the number of covariates).

**casecontrol** - this option is used for implementing mediation analysis when data arise from a case-control design, provided the outcome in the population is rare. If this option is omitted, the data will not be treated as from a case-control design.

**bootstrap** - this specifies whether a bootstrap should be run to compute bias-corrected bootstrap confidence intervals. No bootstrap will be run if this option is omitted.

**reps(integer 200)** - this specifies the number of replications for bootstrap. The default is 200.

**level(cilevel)** - this specifies the confidence level for bootstrap. If this option is omitted, the current default level obtained from c(level) (from creturn list) will be used.

**seed(passthru)** - this specifies the seed for bootstrap. If this option is omitted, a random seed will be used.
Example

. use paramed_example.dta

Continuous outcome, continuous mediator, a binary treatment coded 0 and 1, two covariates, no interaction between treatment and mediator, delta method standard errors

. paramed y_cont, avar(treat) mvar(m_cont) cvars(var1 var2) a0(0) a1(1) m(1) yreg(linear) mreg(linear)

Continuous outcome, binary mediator, a binary treatment coded 0 and 1, two covariates, include an interaction between treatment and mediator, bootstrap standard errors with default bootstrap settings

. paramed y_cont, avar(treat) mvar(m_bin) cvars(var1 var2) a0(0) a1(1) m(1) yreg(linear) mreg(logistic) inter boot

Binary outcome, binary mediator, a binary treatment coded 0 and 1, no covariates, bootstrap standard errors with 1000 replications and fixing the seed to 1234

. paramed y_bin, avar(treat) mvar(m_bin) a0(0) a1(1) m(1) yreg(logistic) mreg(logistic) boot reps(1000) seed(1234)

Count outcome with a Poisson model, binary mediator, a binary treatment coded 0 and 1, two covariates, no interaction between treatment and mediator, bootstrap standard errors with 1000 replications and fixing the seed to 1234

. paramed y_poisson, avar(treat) mvar(m_bin) cvars(var1 var2) a0(0) a1(1) m(1) yreg(poisson) mreg(logistic) boot reps(1000) seed(1234)

Saved results

`paramed` saves the following results in `e()`:

Matrices

- `e(effects)` matrix containing direct, indirect and total effect estimates, standard errors, p-value and 95% confidence intervals

Authors
Further reading


Acknowledgments

This work was supported by the UK Medical Research Council Methodology Research Programme (Grant number: G0900678) and a UK Medical Research Council Career Development Award in Biostatistics (Grant number: G0802418).

The command is based on the MEDIATION macros in SAS and SPSS by Linda Valeri and Tyler VanderWeele.

Also see

Help: [R] regress, [R] logit, [R] glm
Appendix G

UNIVERSITY OF TORONTO
OFFICE OF THE VICE PRESIDENT, RESEARCH

PROTOCOL REFERENCE # 28610
April 3, 2013
Dr. Elizabeth Badley
DALLA LANA SCHOOL OF PUBLIC HEALTH
FACULTY OF MEDICINE
Mrs. Orit Schieir
DALLA LANA SCHOOL OF PUBLIC HEALTH
FACULTY OF MEDICINE

Dear Dr. Badley and Mrs. Orit Schieir,

Re: Your research protocol entitled, "Incidence of coronary heart disease associated with Arthritis: A Canadian population-based cohort study"

We are writing to advise you that the Health Sciences Research Ethics Board (REB) has granted approval to the above-named research protocol under the REB's delegated review process. Your protocol has been approved for a period of one year and ongoing research under this protocol must be renewed prior to the expiry date.

Any changes to the approved protocol or consent materials must be reviewed and approved through the amendment process prior to its implementation. Any adverse or unanticipated events in the research should be reported to the Office of Research Ethics as soon as possible.

Please ensure that you submit an Annual Renewal Form or a Study Completion Report 15 to 30 days prior to the expiry date of your current ethics approval. Note that annual renewals for studies cannot be accepted more than 30 days prior to the date of expiry.

If your research is funded by a third party, please contact the assigned Research Funding Officer in Research Services to ensure that your funds are released.

Best wishes for the successful completion of your research.

Yours sincerely,

Judith Friedland, Ph.D.
REB Chair

Daniel Gyewu
REB Manager