Urban-Rural Differences in Oral Bisphosphonate Utilization in Ontario Following Formulary Changes and Introduction of New Formulations

by

Racquel Jandoc

A thesis submitted in conformity with the requirements for the degree of Master of Science

Graduate Department of Pharmaceutical Sciences

University of Toronto

© Copyright by Racquel Jandoc 2015
Urban-Rural Differences in Oral Bisphosphonate Utilization in Ontario Following Formulary Changes and Introduction of New Formulations

Racquel Jandoc

Master of Science

Graduate Department of Pharmaceutical Sciences

University of Toronto

2015

Abstract

Rural populations often have less access to health services and information, potentially resulting in regional variation in drug utilization. We examined urban-rural differences in oral bisphosphonate utilization in Ontario following formulary changes and introduction of new formulations. The Rurality Index of Ontario was used to define urban, nonmajor urban, and rural regions. There were 16,367,752 oral bisphosphonate claims dispensed (77% urban, 18% nonmajor urban, 5% rural) to community-dwelling seniors from 2000/01–2014/03. Trends in dispensing following formulary changes were similar between regions. When examining trends in dispensing of new formulations, urban regions had significantly faster uptake than rural regions, and consistently dispensed a higher proportion of claims. These temporal and regional differences persisted over time. Rural physicians may be less aware of new drugs, indicating a need to improve information dissemination to these areas. Further research examining whether urban-rural differences in dispensing translates into differences in clinical outcomes is warranted.
Acknowledgements

I would like to express my appreciation to many individuals, without whom this thesis would not have been possible.

First and foremost, to my supervisor, Dr. Suzanne Cadarette, thank you for providing me with this opportunity, and for your continuous patience, support, and guidance. I have accomplished much more than I thought I was capable of during my graduate training, and it is thanks to your exceptional mentorship. I am so grateful, not only for your expertise for my thesis, but also for an enriching graduate experience that has helped me grow as a student and person.

To my committee members, Dr. Muhammad Mamdani and Dr. Linda Lévesque, thank you for all of your guidance and insightful contributions. Your expert advice was invaluable to my thesis, and your comments and questions taught me to think critically about the implications of my research.

In addition, I would like to thank the Institute for Clinical Evaluative Sciences for providing access to the data and a workplace to complete my analysis. I would also like to thank Andrew Calzavara for assisting with my data and answering all of my questions about statistical coding and administrative data.

To my colleagues: Andrea Burden, Mina Tadrous, Jordan Albaum, Giulia Consiglio, Amine Amiche, Lindsay Wong, Cody Black, Sam Elbayoumi, Joann Ban, Patrick Edwards, Jennifer Youn, Erin Cicinelli, and Gina Matesic – thank you for your support, feedback, and most importantly, your friendship. The smiles, laughs, and continuous words of encouragement from you have been instrumental in helping me stay focused and motivated. I could not have asked for a better group of friends to work with.

I would like to acknowledge the funding sources that supported my graduate research: the Canadian Institutes of Health Research Drug Safety and Effectiveness Cross-disciplinary Training Program Award, University of Toronto Dean’s Fund, Ontario Ministry of Research and Innovation Early Researcher Award, and two awards from the Graduate Department of Pharmaceutical Sciences at the University of Toronto: the Clinical, Social, and Administrative Pharmacy Professors’ Award, and the Pfizer Canada Fellowship in Health Outcomes Research Award.

Last, but not least, a special thank you to my family and friends. To my parents and sister, words cannot express how grateful I am for your unconditional love and encouragement. You have always taught me to reach for the stars, and I would not be where I am today without you.

Racquel Jandoc
Table of Contents

Abstract .............................................................................................................................. ii

Acknowledgements ........................................................................................................ iii

List of Tables ..................................................................................................................... vi
List of Figures .................................................................................................................... vii
List of Appendices ........................................................................................................... viii

Abbreviations ................................................................................................................... ix

Thesis Overview ................................................................................................................. x

Chapter 1. Background ....................................................................................................... 1
1.1 Urban-Rural Differences in Health and Prescribing ....................................................... 1
1.2 Case Example: Osteoporosis Drug Utilization ............................................................... 2
1.3 The Ontario Drug Benefit Formulary ............................................................................. 3
1.4 Oral Bisphosphonate Utilization in Ontario .................................................................. 5
  1.4.1 Overall Oral Bisphosphonate Utilization ................................................................. 5
  1.4.2 New Oral Bisphosphonate Formulations ................................................................. 7
1.5 Study Objectives ......................................................................................................... 10

Chapter 2: Interrupted Time Series Analysis in Drug Utilization Research is Increasing: A Systematic Review ................................................................................................................. 13
2.1 Chapter Overview ........................................................................................................ 13
2.2 Abstract ....................................................................................................................... 14
2.3 Introduction ................................................................................................................ 15
2.4 Methods ...................................................................................................................... 17
2.5 Results ....................................................................................................................... 19
2.6 Discussion .................................................................................................................. 21

Chapter 3. Methods .......................................................................................................... 29
3.1 Chapter Overview ....................................................................................................... 29
3.2 Data Sources .............................................................................................................. 30
3.3 Defining Urban-Rural Regions .................................................................................... 30
  3.3.1 Forward Sortation Areas ..................................................................................... 30
  3.3.2 Urban Area Rural Area Codes ............................................................................. 31
  3.3.3 Statistical Classification Codes .......................................................................... 32
  3.3.4 Rurality Index of Ontario .................................................................................... 32
3.4 Outcome Measure ..................................................................................................... 33
3.5 Interrupted Time Series Analysis ................................................................................. 34
List of Tables

Table 1.1 Summary of Canadian osteoporosis practice guidelines .......................................................... 11
Table 1.2 Oral bisphosphonates currently listed on the Ontario Drug Benefit formulary .................. 12
Table 2.1 Characteristics of empirical applications of time series studies ...................................... 25
Table 3.1 Geography classifications commonly used in Ontario ....................................................... 44
Table 3.2 Rurality classifications commonly used in Ontario ............................................................ 45
Table 3.3 Variables considered in the Rurality Index of Ontario ...................................................... 46
Table 3.4 Advantages and disadvantages of interrupted time series analysis techniques ............ 47
Table 4.1a Demographic characteristics of patients dispensed an oral bisphosphonate (for initial claims with complete physician regional data) ................................................................. 59
Table 4.1b Demographic characteristics of physicians prescribing an oral bisphosphonate (for initial claims with complete physician regional data) ........................................................................ 59
Table 4.2a Results from segmented regression model: relative changes from baseline and 95% confidence intervals of proportion of alendronate/risedronate dispensed following formulary change 1a .................................................................................................................. 60
Table 4.2b Results from segmented regression model: relative changes from baseline and 95% confidence intervals of proportion of alendronate/risedronate dispensed following formulary change 2a .................................................................................................................. 60
Table 4.3 Results from regression analysis: slope estimates and 95% confidence intervals of proportion of new oral bisphosphonate formulations dispensed ........................................ 61
List of Figures

Figure 2.1 Flow diagram of systematic search results ................................................................. 26
Figure 2.2 Proportional Venn diagram of search result yield of empirical applications by search strategy, n=220 .......................................................... 27
Figure 3.1 Interrupted time series model .................................................................................... 48
Figure 4.1 Identification of eligible oral bisphosphonate (BP) claims dispensed to community-dwelling seniors in Ontario (c Refer to Appendix F for excluded formulations) .......... 62
Figure 4.2 Identification of unique patients dispensed an eligible oral bisphosphonate (BP) and physicians who dispensed an eligible oral BP claim in Ontario (c Missing patient RIO, N=3,853) ............................................................................................................... 63
Figure 4.3 Number of oral bisphosphonate claims dispensed to community-dwelling seniors in Ontario, January 2000 – March 2014 ........................................................................... 64
Figure 4.4 Number of alendronate (ALN) claims dispensed to community-dwelling seniors in Ontario, by formulation, January 2000 – March 2014 ................................................................. 64
Figure 4.5 Number of risedronate (RSD) claims dispensed to community-dwelling seniors in Ontario, by formulation, January 2000 – March 2014 ........................................................................... 65
Figure 4.6 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in Ontario following formulary changes, January 2001 – December 2008 (N=7,579,128) ......................................................................................................................... 65
Figure 4.7 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in Ontario, by region, January 2001 – December 2008 (N=7,579,128) ......................................................................................................................... 66
Figure 4.8 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in urban regions following formulary changes, January 2001 – December 2008 (N=5,848,858). ......................................................................................................................... 66
Figure 4.9 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in nonmajor urban regions following formulary changes, January 2001 – December 2008 (N=1,369,263). ......................................................................................................................... 67
Figure 4.10 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in rural regions following formulary changes, January 2001 – December 2008 (N=360,647). ......................................................................................................................... 67
Figure 4.11 Proportion of alendronate + vitamin D₃ claims of total alendronate claims dispensed to community-dwelling seniors in Ontario, by region, January 2007 – March 2014 (N=4,961,458) ......................................................................................................................... 68
Figure 4.12 Proportion of monthly risedronate claims of total risedronate claims dispensed to community-dwelling seniors in Ontario, by region, June 2009 – March 2014 (N=4,033,497) ......................................................................................................................... 68
Figure 4.13 Proportion of risedronate delayed-release claims of total risedronate claims dispensed to community-dwelling seniors in Ontario, by region, February 2012 – March 2014 (N=1,870,993) ......................................................................................................................... 69
List of Appendices

Appendix A. List of keyword terms used for MEDLINE® search ........................................... 90
Appendix B. List of identified references (methodological, reviews, and empirical papers)..... 91
Appendix C. List of references examining single institutional interventions....................... 111
Appendix D. Database information table.................................................................................. 116
Appendix E. Results for total oral bisphosphonate dispensing in Ontario, January 2000 to March 2014..................................................................................................................................... 117
Appendix F. Non-osteoporosis oral bisphosphonate formulations excluded from analysis..... 118
Appendix G. Results using patient region data........................................................................ 119
Appendix H. Results using physician region data for new oral bisphosphonate formulations, separated by brand and generic................................................................. 121
Appendix I. Full segmented model regression estimates for objective 1 ............................. 122
Appendix J. Full regression estimates for objective 2............................................................ 123
Appendix K. SAS code............................................................................................................. 124
**Abbreviations**

ALN – Alendronate

ARIMA – Autoregressive Integrated Moving Average

BMD – Bone Mineral Density

BP – Bisphosphonate

CI – Confidence Interval

FSA – Forward Sortation Area

ICES – Institute for Clinical Evaluative Sciences

IPDB – ICES Physician Database

ODB – Ontario Drug Benefit

RIO – Rurality Index of Ontario

RPDB – Registered Person Database

RSD – Risedronate

SD – Standard Deviation
Thesis Overview

This thesis is comprised of two main studies organized into five chapters. In Chapter 1, a background of urban-rural differences in health and prescribing, the drug approval process, oral bisphosphonate prescribing over time, and new oral bisphosphonate formulations available in Ontario is provided. This chapter also introduces the main objectives of this thesis and the accompanying hypotheses. Chapter 2 is the manuscript resulting from the first project: a systematic review examining the use of interrupted time series methods in drug utilization research. Chapter 3 details the methods used to conduct the second project, which was an empirical study examining urban-rural differences in oral bisphosphonate prescriptions dispensed in Ontario. The results of the empirical study are outlined in Chapter 4, and Chapter 5 discusses the major findings of this research in relation to the hypotheses and objectives, along with the study’s significance, limitations, strengths, and future considerations.
Chapter 1. Background

1.1 Urban-Rural Differences in Health and Prescribing

Health status, health behaviours, and utilization of healthcare services have been shown to be associated with one’s region of residence. Universal health insurance coverage has facilitated the receipt of healthcare across Canada, yet geography remains a potential barrier to access [1]. Disparities in healthcare between urban and rural populations are evident in many chronic disease areas, with rural populations having less accessibility to health services, healthcare, and related health information [2-4]. This may translate into differences in prescribing in these regions. For instance, rural physicians in Australia perceive their prescribing to be impacted by the isolation of their practice, the remoteness of patients’ home addresses, the limited diagnostic facilities in their area, and patient access to therapy [5]. Despite the perceived limitations resulting from their location, rural physicians were found to initiate prescribing of new drugs, but the frequency of prescribing of newer therapies decreased as the practice location became more remote. This may be a result of restricted patient access to the drug due to cost or availability, inexperience of prescribers (rural areas tend to have younger doctors) [6], or less exposure and education regarding new agents. Less exposure to new agents may be a result of lower referral rates from specialists and less active pharmaceutical promotion in rural areas [5, 7, 8].

Prior research has shown prescribing practices to be influenced by pharmaceutical promotion and marketing, and that visits from pharmaceutical sales representatives often result in increased prescribing of the promoted drug [9, 10]. For example, one study indicated that the majority of physicians in obstetrics-gynecology regarded pharmaceutical representatives as having valuable information about new drugs, with one-third reporting using information from pharmaceutical representatives often, or almost always, when deciding to prescribe a new drug [11]. Promotional
efforts and aggressive marketing of new drugs has also been shown in other areas, including significant increases in prescribing of cyclooxygenase-2 inhibitors following their entry into the US market [12]. In addition, a recent study in Australia showed increased pharmaceutical promotion expenditures for risedronate prior to increased formulary access [13]. Physicians often first learn about new drugs during meetings with pharmaceutical sales representatives [10]. Other sources of new drug information include peer-reviewed literature, the media, and the advice and practice of medical colleagues [14]. The pharmaceutical industry, especially visits with sales representatives, were frequently cited as the most important source of awareness of new drugs for general physicians, as well as the most influential when deciding to prescribe new drugs [10]. General physicians have been described as passive searchers of new drug information, thus active outreach by pharmaceutical companies is the preferred source of obtaining new information as it is quick and convenient [14]. However, marketing efforts are often concentrated in urban areas, with little exposure in rural areas [8]. Thus in addition to being less aware of updated healthcare information, such as formulary changes and updated management guidelines, rural physicians may also be less aware of new drugs entering the market. This lack of awareness may potentially result in urban-rural differences in prescribing of new medications.

1.2 Case Example: Osteoporosis Drug Utilization

Few studies have examined the impact of geography and rurality on prescribing and drug dispensing, providing conflicting data that is difficult to interpret within the context of different disease demography between regions [7]. In this thesis project, urban-rural differences in dispensing were examined using osteoporosis drugs as a case example. Osteoporosis is characterized by the deterioration of bone tissue and increased susceptibility to fracture and fracture-related morbidity [15]. Over 200 million people worldwide are estimated to have osteoporosis [16] and in Canada, an estimated 16% of women and 7% of men suffer from this
disease [17]. The incidence of osteoporotic fracture increases with age, and thus the impact of osteoporosis is expected to increase worldwide as the population ages [18-20]. Fortunately, several treatments are available in Canada to prevent and treat fractures and related morbidity among men and women with osteoporosis, including bisphosphonates (oral: alendronate, etidronate, risedronate; intravenous: zoledronic acid), denosumab, raloxifene, and teriparatide [21].

Over the last fifteen years, the availability of osteoporosis drugs has changed with the introduction of new osteoporosis drugs to the market, making osteoporosis drugs an interesting case example to examine urban-rural differences in dispensing. The impact of changes in formulary accessibility to new and existing drugs may also have differential impacts in urban and rural regions. Prior studies have outlined potential barriers that may be associated with variations in osteoporosis drug use such as differential access to osteoporosis diagnostic facilities [19, 22, 23], however to our knowledge, no study has examined specific regional differences in utilization. Closer examination of urban-rural differences in osteoporosis medication dispensing and its appropriateness is therefore warranted.

1.3 The Ontario Drug Benefit Formulary

The Ontario Drug Benefit (ODB) formulary, established in 1971, is a list of approximately 3,800 drug products whose costs are mostly reimbursed by the Ministry of Health and Long-Term Care for eligible recipients [24]. A co-payment scheme was introduced to the ODB program in July 1995 to ensure the sustainability and affordability of the program, and thus all recipients are required to pay a small portion of their prescriptions. Eligible recipients include persons aged 65 years or older, those receiving professional services under the Home Care Program, residents of
long-term care facilities or Homes for Special Care, persons eligible under the Trillium Drug Program, or those receiving benefits under Ontario Works or the Ontario Disability Support Program [24].

Drug products undergo an extensive review process in order to be listed on the ODB formulary. The first step is national approval of the drug product by Health Canada, followed by review and subsequent recommendations by the ODB program. For national approval, drug manufacturers must submit their application for their product to the Therapeutic Products Directorate of Health Canada, which approves the product after extensive review of clinical evidence for the safety and efficacy of the drug [25]. Once a drug has been approved for sale within Canada by the directorate, Health Canada issues a: 1) “Notice of Compliance,” indicating that the product meets regulation standards [26], and 2) a unique “Drug Identification Number” for the product [27].

Approval by Health Canada does not mean automatic coverage through the ODB. Thus the second step to ODB coverage, following national approval by Health Canada, is for drug companies to file a submission to the ODB Program. The submission is reviewed by the province’s expert advisory committee: the Committee to Evaluate Drugs (formerly known as the Drug Quality and Therapeutics Committee); who then make a recommendation as to whether the drug should be listed on the ODB formulary [24, 28]. As of 2002, however, drug manufacturers first apply to the Canadian Agency for Drugs and Technologies in Health and have their drugs reviewed through the Common Drug Review prior to submitting their application to the Committee to Evaluate Drugs [28, 29].

Final ODB coverage decisions are made by the Executive Officer of the Ontario Public Drug Program. Decisions are based upon the Committee to Evaluate Drugs’ recommendations,
recommendations from other advisory bodies, product listing agreements with drug companies, drug program budgets, public interests, and government priorities [28]. Drug products may be listed on the ODB formulary as “General Benefit,” “Limited Use,” “Conditional Listing,” or may be approved as part of the “Exceptional Access Program” (previously known as the Individual Clinical Review, or Section 8 of the ODB Act). Each listing status has varying degrees of restriction to ODB-eligible recipients. General Benefit has no restrictions, i.e., ODB-covered drugs are openly available. Limited Use and Conditional Listing are more restrictive, covering new and existing drug products for patients who meet specific conditions. Drugs listed under the Exceptional Access Program also have restricted access and are only eligible on a case by case basis [24]. The ODB formulary listing status of a drug may change over time as more evidence becomes available. The potential impact of ODB formulary changes on oral bisphosphonate utilization in urban and rural regions is the primary focus of this thesis.

1.4 Oral Bisphosphonate Utilization in Ontario

1.4.1 Overall Oral Bisphosphonate Utilization

Oral bisphosphonates (alendronate, etidronate, and risedronate) are the primary drugs used to treat osteoporosis and prevent fractures, with 99% of osteoporosis medication users receiving an oral bisphosphonate as initial treatment [30]. Since oral bisphosphonates are considered the cornerstone of osteoporosis treatment and comprise the majority of osteoporosis pharmacotherapy prescribed, this thesis focuses on the utilization of oral bisphosphonates as the case example.

Alendronate and risedronate have proven efficacy for prevention of both vertebral and nonvertebral fractures, whereas etidronate has proven efficacy for prevention of vertebral
fractures only [15]. Consequently, the 2002 Canadian osteoporosis practice guidelines recommended alendronate and risedronate as first-line therapy, and etidronate as second-line therapy [31]. The current 2010 osteoporosis guidelines continue to recommend alendronate and risedronate as first-line therapy, and etidronate as second-line therapy [21]. Table 1.1 briefly summarizes guideline recommendations for osteoporosis management over time. Despite the stronger evidence of benefit with alendronate and risedronate compared to etidronate, etidronate has been available as General Benefit since 1996 and alendronate and risedronate had restrictive coverage until 2007 [30]. Table 1.2 summarizes oral bisphosphonate dates of Notice of Compliance and ODB formulary listing statuses.

Utilization practices of oral bisphosphonates have changed over time in response to formulary listing status changes. One study examined the number of incident users of osteoporosis medications in Ontario from fiscal years 1995/1996 to 2008/2009 and showed large increases in oral bisphosphonate utilization from 1996 to 2000. Etidronate was most commonly prescribed to new users as it was the only oral bisphosphonate listed as General Benefit on the ODB formulary at this time [30]. Beginning in the 2000/2001 fiscal year, there was a shift in utilization from etidronate to alendronate and risedronate over the remaining duration of the study. Etidronate comprised the majority of prescriptions for new bisphosphonate users until the year 2000 and since then, the number of new bisphosphonate users prescribed alendronate or risedronate has been increasing [30].

The changes in drug utilization beginning in the 2000/2001 fiscal year coincided with changes in drug listing status on the ODB formulary which made alendronate (2000) and risedronate (2001) more accessible. As described in Table 1.2, alendronate and risedronate were initially restricted to patients who failed, experienced side effects, or were intolerant to etidronate. The Limited Use
criteria changed in April 2003, becoming less restrictive with broader indications. In addition to the above conditions, alendronate and risedronate were now also available to patients who met at least two of the following three criteria: 1) age > 75 years, 2) prior osteoporotic fracture, and/or 3) a bone mineral density (BMD) T-score < -3.0 SD. A third formulary change in 2007 removed all restrictions for alendronate and risedronate, listing both formulations as General Benefit, thereby increasing their availability to all ODB-eligible recipients [30]. Prior studies have identified formularies to be an effective means of influencing prescribing [32-36]. Indeed, when physicians in British Columbia and Ontario were not restricted by a drug formulary, alendronate and risedronate were the preferred therapy for new osteoporosis treatment, yet etidronate was still being dispensed [30]. As previously described, rural physicians may be less aware of health information, such as formulary updates, and therefore it is possible that etidronate utilization may be occurring at a disproportional rate in rural regions compared to urban regions.

1.4.2 New Oral Bisphosphonate Formulations

New oral bisphosphonate formulations have been developed that are marketed as advantageous over prior therapies, yet their impact on oral bisphosphonate use in Ontario has not yet been examined. These new formulations include: 1) alendronate + vitamin D₃ (January 2007), 2) monthly risedronate (June 2009), and 3) risedronate delayed-release (February 2012). Although these new formulations have additional advantages compared to existing therapies, there is currently no evidence demonstrating improved fracture outcomes.

1.4.2.1 Alendronate + Vitamin D₃

Alendronate + vitamin D₃ is a new weekly bisphosphonate that combines the antiresorptive effect of alendronate with the benefits of cholecalciferol (vitamin D₃). It was developed by Merck Canada Inc. under the brand name Fosavance™, and has been available on the ODB
formulary as General Benefit since January 2007 [37]. The first generic form of alendronate + vitamin D₃ (Teva-alendronate/cholecalciferol) was listed as General Benefit in January 2014 [37]. Vitamin D deficiency is common worldwide, especially among osteoporotic women [38]. Vitamin D is associated with increased bone density, reduced bone modeling, decreased risk of falls, as well as decreased fracture risk and thus vitamin D, along with calcium is often recommended to osteoporotic patients [38, 39]. Indeed, current osteoporosis guidelines recommend a minimum of 10 µg (400 IU) of vitamin D for healthy adults at low risk of vitamin D deficiency, and 20 µg (800 IU) of vitamin D daily for adults over the age of 50 with moderate risk of vitamin D deficiency [21]. Daily vitamin D and calcium supplements are available to patients as over-the-counter products, however adherence to both vitamin D and calcium is poor [40]. The novelty of the alendronate + vitamin D₃ combination tablet is that it overcomes the barriers to self-supplementation with vitamin D, thereby decreasing the inconvenience of buying vitamin D separately, or remembering to take them daily, as well as providing some supplemental vitamin D that many osteoporotic patients are lacking at no extra cost [39]. Prior research has shown that the combination tablet has equivalent antiresorptive effects compared to alendronate alone [41], as well as improved vitamin D status [41, 42]. Although there currently is no evidence demonstrating improved fracture outcomes with the new alendronate + vitamin D₃ formulation compared to alendronate alone, this combination tablet has been shown to be more effective at increasing BMD and reducing bone turnover compared to standard therapy, and therefore may potentially be advantageous over existing formulations of alendronate [42].

1.4.2.2 Monthly Risedronate

Patients often prefer weekly dosing regimens over daily intervals, and weekly dosing has consistently been associated with improved adherence compared to daily dosing [43, 44]. It has been suggested that monthly dosing may improve adherence over weekly bisphosphonates due to
the convenience and practicality of even less frequent dosing intervals [45], and thus a new monthly risedronate formulation was developed by Warner Chilcott under the brand name Actonel™. Monthly risedronate was listed on the ODB as General Benefit in June 2009 [24]. Generic monthly risedronate formulations have been available as General Benefit since June 2012. Prior research has shown comparable efficacy and safety between monthly and daily risedronate for the treatment of osteoporosis [46, 47]. Monthly risedronate therefore provides a new treatment option for patients who prefer a less frequent dosing regimen than existing daily or weekly intervals.

1.4.2.3 Risedronate Delayed-Release

Warner Chilcott’s Actonel DR™, or risedronate delayed-release, tablets are a new weekly formulation of risedronate that has been listed as General Benefit since February 2012. Oral bisphosphonates are available in a number of dosing formulations for patient convenience, yet adherence is suboptimal [48]. Strict administration guidelines of bisphosphonates have been reported as a major barrier to continuing treatment [49]. For instance, oral bisphosphonates must be administered with plain water following an overnight fast and 30-60 minutes prior to eating or drinking [50]. One study identified that 56% of patients did not comply with at least one of the instructions regarding administration of a bisphosphonate [51]. Risedronate delayed-release is a new bisphosphonate formulation which contains a chelating compound that eliminates the need to fast prior to administration. Risedronate delayed-release taken before or after a meal has been shown to have similar clinical effectiveness as daily risedronate. As such, this novel formulation may increase convenience without compromising efficacy or safety [52], potentially becoming a preferred treatment option among patients and physicians.
1.5 Study Objectives

The two main objectives of this thesis are outlined below, and were studied in the context of osteoporosis, using oral bisphosphonate medications as a case example. This thesis sought to:

1) Assess urban-rural differences in oral bisphosphonate dispensing in Ontario following formulary changes using interrupted time series analysis.

**Hypothesis:** Rural physicians will respond more slowly to drug formulary changes that made alendronate and risedronate more accessible to the public, as rural physicians may be less aware of this new information.

2) Examine urban-rural differences in the dispensing of new oral bisphosphonate formulations upon entry into the market compared to existing oral bisphosphonate formulations.

**Hypothesis:** Rural regions will have a slower uptake of new oral bisphosphonate formulations, as pharmaceutical marketing is often concentrated in urban regions thus rural physicians may be less aware of these new products.

To assess urban-rural differences in oral bisphosphonate utilization following formulary changes, the following two studies were conducted: 1) A systematic literature review to identify the use and reporting of interrupted time series methods used to evaluate the effects of healthcare interventions on drug utilization (*Chapter 2*); and 2) An interrupted time series analysis to determine the impact of formulary changes on dispensing, stratified by region (*Chapters 3-5*). The systematic review informed analysis for study 2. To examine urban-rural differences in the dispensing of new oral bisphosphonates, the number of claims dispensed for each new formulation was identified and compared between regions (*Chapters 3-5*).
Table 1.1 Summary of Canadian osteoporosis practice guidelines

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Goals of management</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Confirm or rule out osteoporosis</td>
<td>• Treatment of low BMD</td>
<td>• Fracture risk assessment</td>
<td>• Prevention of fractures and consequences</td>
<td></td>
</tr>
<tr>
<td>• Prevention of fracture</td>
<td>• Prevention of fracture</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Indications for treatment</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>• Low BMD (T-score &lt; -2.5)</td>
<td>• Low BMD (T-score &lt; -2.5) in combination with:</td>
<td>• Low BMD (T-score &lt; -2.5)</td>
<td>• 10 year fracture risk &gt; 20%</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Prior fragility fracture</td>
<td>• Fragility fractures after age 40</td>
<td>• Prior fragility fracture of hip or spine</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Age 65+ years</td>
<td>• Age 65+ years</td>
<td>• One or more fragility fractures</td>
<td></td>
</tr>
<tr>
<td></td>
<td>• Family history of osteoporosis</td>
<td>• Family history of fractures</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Systemic use of glucocorticoids for more than 3 months</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Recommended pharmacotherapy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First-line</strong></td>
<td>• Estrogen</td>
<td>• Alendronate</td>
<td>• Alendronate</td>
<td>• Alendronate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Raloxifene</td>
<td>• Raloxifene</td>
<td>• Denosumab</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Risedronate</td>
<td>• Risedronate</td>
<td>• Raloxifene</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>• Teriparatide&lt;sup&gt;a&lt;/sup&gt;</td>
<td>• Risedronate</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>• Zoledronic acid</td>
</tr>
<tr>
<td><strong>Second-line</strong></td>
<td>• Alendronate</td>
<td>• Calcitonin</td>
<td>• Calcitonin</td>
<td>• Calcitonin&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>• Etidronate</td>
<td>• Estrogen</td>
<td>• Estrogen</td>
<td>• Etidronate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Etidronate</td>
<td>• Etidronate</td>
<td>• Teriparatide&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

BMD: Bone mineral density

<sup>a</sup> Currently not available on the ODB

<sup>b</sup> No longer authorized for sale within Canada as of October 2013 due to increased risk of cancer associated with its prolonged use [55]
Table 1.2 Oral bisphosphonates currently listed on the Ontario Drug Benefit formulary

<table>
<thead>
<tr>
<th>Druga</th>
<th>Regimen</th>
<th>Dose</th>
<th>Notice of Compliance</th>
<th>Limited Use 1</th>
<th>Limited Use 2</th>
<th>General Benefitb</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Generic Daily</td>
<td>10mg</td>
<td>Apr 2003</td>
<td>Sep 2003d</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Generic Weekly</td>
<td>70mg</td>
<td>May 2005</td>
<td>Jul 2005d</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alendronate + Vitamin D3</td>
<td>Brand Weekly</td>
<td>70mg &amp; 70mcg</td>
<td>Feb 2006</td>
<td></td>
<td></td>
<td>Jan 2007</td>
</tr>
<tr>
<td></td>
<td>Generic Weekly</td>
<td>70mg &amp; 70mcg</td>
<td>Mar 2013</td>
<td></td>
<td></td>
<td>Jan 2014</td>
</tr>
<tr>
<td>Alendronate + Vitamin D3</td>
<td>Brand Weekly</td>
<td>70mg &amp; 140mcg</td>
<td>Aug 2008</td>
<td></td>
<td></td>
<td>Jun 2009</td>
</tr>
<tr>
<td></td>
<td>Generic Weekly</td>
<td>70mg &amp; 140mcg</td>
<td>Mar 2013</td>
<td></td>
<td></td>
<td>Jan 2014</td>
</tr>
<tr>
<td>Etidronate + Calcium Carbonate</td>
<td>Brand Cyclical</td>
<td>400mg/500mg tab</td>
<td>Jul 1995</td>
<td></td>
<td></td>
<td>Oct 1996</td>
</tr>
<tr>
<td></td>
<td>Generic Cyclical</td>
<td>400mg/500mg tab</td>
<td>Apr 2008</td>
<td></td>
<td></td>
<td>Jun 2008</td>
</tr>
<tr>
<td></td>
<td>Generic Daily</td>
<td>5mg</td>
<td>Jan 2010</td>
<td></td>
<td>Apr 2010</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Generic Weekly</td>
<td>35mg</td>
<td>Jan 2010</td>
<td></td>
<td>Apr 2010</td>
<td></td>
</tr>
<tr>
<td>Risedronate Delayed-release</td>
<td>Brand Weekly</td>
<td>35mg</td>
<td>July 2011</td>
<td></td>
<td></td>
<td>Feb 2012</td>
</tr>
<tr>
<td></td>
<td>Generic Weekly</td>
<td>35mg</td>
<td>n/a</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Risedronate</td>
<td>Brand Monthly</td>
<td>150mg</td>
<td>Sep 2008</td>
<td></td>
<td></td>
<td>Jun 2009</td>
</tr>
<tr>
<td></td>
<td>Generic Monthly</td>
<td>150mg</td>
<td>Dec 2011</td>
<td></td>
<td></td>
<td>Jun 2012</td>
</tr>
</tbody>
</table>

Information summarized from the Ontario Drug Benefit formulary [24] and Cadarette et al [30]

BMD: Bone mineral density; ODB: Ontario Drug Benefit; SD: Standard deviation

*a Only includes the Notice of Compliance and ODB-listing date of the first brand and/or generic product of each formulation

*b Openly available without restriction

*c Only available for treatment of osteoporosis in patients who failed* etidronate, experienced side effects, or have contraindications to etidronate

*d Only available for treatment of osteoporosis in patients who failed* etidronate, experienced side effects, or have contraindications to etidronate; or meet at least two of the following conditions: age >75 years, prior osteoporotic fracture, BMD T-score ≤ -3.0 SD

*e Failed: continued loss of BMD of >3% after two years of therapy, or a new fracture after one year of therapy
Chapter 2: Interrupted Time Series Analysis in Drug Utilization Research is Increasing: A Systematic Review

2.1 Chapter Overview

This chapter is a manuscript that has been submitted with revisions for considered publication in the *Journal of Clinical Epidemiology*, and is currently under review. This manuscript is the first of two projects conducted for this thesis and includes the systematic review that was completed to identify the use and reporting of interrupted time series methods in drug utilization research, and to inform the empirical time series application presented in Chapters 3-5.

Authors:
Racquel Jandoc\(^1\), Andrea M. Burden\(^1\), Muhammad Mamdani\(^{1,2,3}\), Linda E. Lévesque\(^4\), Suzanne M. Cadarette\(^1\)

Affiliations:
\(^1\) Leslie Dan Faculty of Pharmacy, University of Toronto, Toronto ON, Canada
\(^2\) Applied Health Research Center, La Ka Shing Knowledge Institute, St. Michael’s Hospital, Toronto ON, Canada
\(^3\) Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto ON, Canada
\(^4\) Department of Public Health Sciences, Queen’s University, Kingston ON, Canada

Status:
Manuscript submitted for publication in *J Clin Epidemiol*; submitted with revisions October 6, 2014 – under review.
2.2 Abstract

Objective: To describe the use and reporting of interrupted time series methods in drug utilization research.

Study Design and Setting: We completed a systematic search of MEDLINE®, Web of Science®, and reference lists to identify English-language articles through to December 2013 that employed interrupted time series methods in drug utilization research. We tabulated the number of studies by publication year, and summarized methodological detail.

Results: We identified 220 eligible empirical applications since 1984. Only 17 (8%) were published before 2000, and 90 (41%) were published since 2010. Segmented regression was the most commonly applied interrupted time series method (67%). Most studies assessed drug policy changes (51%, n=112); 22% (n=48) examined the impact of new evidence, 18% (n=39) examined safety advisories, and 16% (n=35) examined quality improvement interventions. Autocorrelation was considered in 66% of studies, 31% reported adjusting for seasonality, and 15% accounted for non-stationarity.

Conclusions: Use of interrupted time series methods in drug utilization research has increased, particularly in recent years. Despite methodological recommendations, there is large variation in reporting of analytic methods. Developing methodological and reporting standards for interrupted time series analysis is important to improve its application in drug utilization research.
2.3 Introduction

Interrupted time series analysis is the strongest and most commonly used quasi-experimental design to assess the impact of an intervention when a randomized controlled trial is not feasible [56-61]. This method has been applied in a variety of disciplines, and was first introduced to the field of health services research in 1981 to evaluate the impact of regionalized perinatal care [62]. Interrupted time series methods use aggregate data collected over equally spaced time intervals before and after an intervention, with the key assumption that data trends prior to the intervention can be extrapolated to predict trends had the intervention not occurred [61]. Routinely maintained pharmacy and medical databases provide rich data sources to apply interrupted time series methods [61].

Several methodological issues need to be considered when completing an interrupted time series analysis. First, given the serial nature of the design; autocorrelation, non-stationarity, and seasonality need to be considered [61, 63]. Autocorrelation refers to the serial dependence of outcome measure error terms. For example, prescription patterns closer to each other may be more similar than those further apart [57, 60, 61]. The presence of autocorrelation can be assessed using the Ljung-Box chi-square statistic [64] or Durbin-Watson statistic [57, 60, 61, 65], and corrected for if necessary. Non-stationary data exhibit an underlying trend that is unrelated to the intervention. For example, the use of a drug commonly increases once it enters the market [63]. Non-stationary can be tested using the augmented Dickey-Fuller test [66]. Seasonality represents regular seasonal fluctuations in the outcome, for example, use of medications to treat influenza. When present, terms for seasonality (e.g., months) should be included in the model [57, 61]. Failing to account for autocorrelation, non-stationarity, and seasonality may lead to biased results.
The number of data points available for analysis, and the number of observations within each data point, are important when using interrupted time series analysis. Although there is no gold standard, it is generally agreed that more data points and observations are better. Depending on the minimum effect size and the amount of variation, a minimum of nine data points pre-, post-, and when applicable, between interventions [56, 67]; and at least 100 observations per data point is encouraged [61]. A larger number of observations in each data point provides more stable estimates and thus reduces the variability and outliers within a time series analysis. Data point outliers that are explainable, such as a sudden peak in drug dispensing in anticipation of a drug restriction policy, should be controlled for using an indicator term [61]. Outliers that result from random variation can be treated as regular data points [61]. A larger number of data points also permits more stable estimates for forecasting pre-intervention trends had the intervention not occurred. In general however, caution should be used when forecasting beyond the data points observed in interrupted time series analyses. Another caveat when conducting interrupted time series methods relates to possible outcome measure ceiling or floor effects. For example, when studying the impact of an intervention in improving the proportion of patients treated with a drug, the outcome has a natural ceiling of 100%, and thus depending on the initial level of measurement, minimal change in the outcome may be observed [68]. Authors must consider ceiling and floor effects when designing their study and interpreting results.

A clear intervention time point helps to identify pre- and post-intervention data points, yet if intervention effects are gradual or delayed, then a lag period may be considered [61]. Lagged intervention effects can be accounted for by excluding the lag period from the analysis, or modeling the lag period as a separate segment in the time series [61]. Here, graphical figures displaying the results of interrupted time series analysis are particularly useful. Even without statistical output, figures allow readers to visually examine baseline trends, the time point at
which the intervention occurred, and the impact of the intervention [60, 61]. All interrupted time series studies should therefore include graphical display to facilitate interpretation of study results.

The main threat to validity in interrupted time series analysis relates to time-varying confounding, such as changes in outcome coding, co-interventions, or changes in the population under study [56, 60, 61, 69]. These threats need to be considered at the individual study level and require intimate knowledge of the data and healthcare utilization trends. The use of a comparison outcome in the same population, or a comparison group using the same outcome in a group not exposed to the intervention, helps to alleviate concerns related to time-varying confounding [60, 61]. Indeed, an advantage of interrupted time series analysis is the ease in stratifying results by different groups [60].

Interrupted time series analysis has been applied in a variety of disciplines, however its use to study the impact of healthcare interventions on drug utilization has not been well described. The purpose of our study was to describe the use and reporting of interrupted time series methods in drug utilization research.

2.4 Methods

We completed a systematic MEDLINE® keyword and Web of Science® citation search to identify all English-language articles that employed interrupted time series methods to study drug utilization in humans. Empirical applications that examined the impact of interventions at the population level, including: drug policy changes, new evidence in the form of guideline changes or major publications, quality improvement interventions, and government or media
safety advisories; on prescription drug utilization were eligible. We defined drug utilization as the number or proportion of: drugs dispensed, or patients dispensed a drug or meeting an adherence target. Systematic reviews, methodological contributions, letters to the editor, and conference abstracts were excluded since the focus was on use and reporting of empirical applications. We also excluded single institution studies so we could focus on methods used at a population level that may be more generalizable.

We first searched MEDLINE® from inception (1946) to December 2013 with keyword terms related to time series analysis and drug utilization (Appendix A). We then used Web of Science® to perform a citation search of methodological papers identified in the keyword search [60, 61, 70], and a commonly cited paper [62]. Finally, we manually searched reference lists from all methodological contributions (Appendix B.1-10), review papers (Appendix B.11-14), and eligible empirical applications (Appendix B.15-204) identified in the keyword and citation searches to identify any additional empirical applications. Two authors (RJ and AMB) independently completed each search and reviewed articles for eligibility. Discrepancies were resolved through discussion with a third author (SMC). A proportional Venn diagram was created to illustrate the number of empirical applications identified by each search strategy. The number of empirical applications was then plotted by publication year.

We abstracted the following characteristics for each application: intervention(s) of interest; primary data source, and methodological detail (time intervals, outcome measure, interrupted time series methods used, and methodological considerations reported). As described above, there are several methodological considerations in interrupted time series analysis, and we have taken care to abstract whether authors reported these; however, we were unable to evaluate aspects that would require access to each study’s raw data. Thus methodological considerations
abstracted included reporting of: autocorrelation, non-stationarity, and seasonality; use of a comparison group; clearly defined time points; number of pre- and post-intervention points; outliers; forecasting; and absolute and/or relative changes with confidence intervals or standard errors. Additional considerations abstracted included the use of lag periods, sensitivity analysis, and graphical figures to display results. One author (RJ) abstracted all data and a second author (AMB) verified all abstracted data. All methodological considerations were summarized using descriptive statistics.

2.5 Results

Of 1917 unique articles identified, 10 were methodological contributions (Appendix B.1-10), 4 were review papers (Appendix B.11-14), and 220 were eligible empirical applications (Appendix B.15-234, Figure 2.1).

Each search strategy proved important, with 52 (24%) empirical applications identified solely by the keyword search, 33 (15%) identified solely by the citation search, 30 (14%) identified only by the reference list search; and only 35 (16%) identified by all three search strategies (Figure 2.2). Most segmented regression papers (92 of 133, 69%) were identified by the citation search, whereas most autoregressive integrated moving average (ARIMA) papers (26 of 30, 84%) were identified by the keyword search. One eligible paper that did not appear in our original search was identified by a reviewer during the peer-review process.

The first empirical application was published in 1984, yet relatively few (n=17, 8%) were published before the year 2000 (Figure 2.3). Since 2000, use has increased with an average of 15
applications published per year. Forty-one percent (n=90) have been published in the last four years, with a high of 31 articles published in 2013.

Table 2.1 summarizes the characteristics of the 220 empirical applications identified, of which 92% utilized administrative pharmacy databases. Policy changes were the most common interventions evaluated (51%), followed by new evidence (22%), safety advisories (18%), and quality improvement interventions (16%). Seventy-one percent examined prescriptions dispensed (22% number, 35%, proportion and 14% standard dose) as the primary outcome measure, and 29% used the number or proportion of patients dispensed the drug of interest or meeting an adherence target. Most applications examined drug utilization over monthly (76%) or quarterly (14%) intervals. Of the 200 papers (91%) reporting detailed methods, segmented regression (67%), ARIMA models (16%), and linear regression (11%) were the most commonly applied analyses. Other analytical methods included generalized estimating equations, logistic, nonlinear, and Poisson models. Fifty percent (n=67) of papers using segmented regression applied a linear model.

Of all empirical studies, 146 (66%) reported testing for autocorrelation (77% of ARIMA papers, and 73% of segmented regression papers), 68 (31%) reported adjusting for seasonality (52% of ARIMA, and 29% of segmented regression), and 32 (15%) reported testing for non-stationarity (65% of ARIMA, 9% of segmented regression). One-third (35%) of all empirical studies reported the use of a comparison group, 70% reported absolute and/or relative impacts with confidence intervals or standard errors, and 28% reported including lag periods in their models. Most articles (85%) clearly reported the intervention time point(s) of interest, and 84% of studies included a graph, yet only 39% reported the number of pre- and post-intervention data points.
included in their analysis (range: 3 to 72 data points). One-fifth (21%) of applications conducted a sensitivity analysis.

2.6 Discussion

We examined the application and reporting of interrupted time series analysis methods in drug utilization research. Use of interrupted time series analysis has increased since the year 2000, a finding noted in other recent reviews of innovative methods in pharmacoepidemiology [71, 72]. The most common interrupted time series methods were segmented regression (67%), ARIMA models (16%), and linear regression (11%). When executing time series models, several methodological aspects are important and may impact the validity of the model. We identified that the majority of eligible articles employing ARIMA models addressed autocorrelation (77%), non-stationarity (65%), seasonality (52%). Since ARIMA models inherently account for autocorrelation, non-stationarity, and seasonality [73]; it is possible that authors may have chosen not to report these considerations. In contrast, segmented regression models do not intrinsically account for autocorrelation, non-stationarity, and seasonality; and thus it is imperative to consider each [61]. However, only 73% of segmented regression studies reported testing for autocorrelation, 29% reported adjustment for seasonality, and only 9% reported consideration of non-stationarity.

Over 80% of interrupted time series analyses cited a clearly defined intervention time point, and used figures to graphically display results; however, other methodological issues were poorly reported. Explicit reporting of all methodological considerations may improve awareness of their importance as well as the interpretation of interrupted time series studies. Therefore, based on prior suggestions [56, 57, 60, 61, 67, 74], we recommend the following be reported in all
interrupted time series applications: 1) autocorrelation, non-stationarity, and seasonality considerations; 2) intervention time point(s) and lag periods; 3) the number of data points pre-, post- and between-intervention(s); 4) specific statistical regression methods and the appropriateness of a linear model when applied; and 5) changes from baseline (intervention impact) with confidence intervals. We also recommend that all interrupted time series studies: 1) use graphical display with clearly defined time point(s) to present results; 2) comment on: the minimum number of observations per data point, data variability, ceiling or floor effects; and 3) consider the use of a control or comparison group. Authors are also encouraged to discuss possible data or co-intervention confounding issues and provide a rationale if no control or comparison group was considered.

Our systematic review is subject to some limitations related to our literature search strategy and inclusion criteria. First, we recognize that the lack of Medical Subject Headings (MeSH terms) and standardized terminology to describe the interrupted time series design may have resulted in some missed applications. Although segmented regression was first introduced to healthcare research in 1981 [62], and a seminal method paper was published in 2002 [61], we found that many papers did not use the term “segmented regression” to describe their analysis. Therefore particular attention to each study’s statistical analysis was required during data abstraction to determine the type of interrupted time series method used. Second, our search was limited in ability to identify applications that are not indexed in either of the databases used (MEDLINE®, Web of Science®). Indeed, during the peer-review of our manuscript, a blind reviewer identified one eligible article (Appendix B.109) that is not indexed in the databases used and therefore was not identified in our original search.
Third, by restricting inclusion to studies that examined prescription drug utilization defined by the number or proportion of prescription drugs dispensed or patients dispensed a drug; we will have missed interrupted time series analyses with different drug outcomes, such as illicit drug use, drug sales, or drug market share [75, 76]. Fourth, we acknowledge that studies examining single institution interventions (n=59, Appendix C) were excluded so we could focus on population-based interventions that may be more generalizable. Despite potentially missing some applications, we feel that our results that identify an increase in the number of applications in recent years, and conclusions of the general trends of methods and underreporting of statistical considerations, would still hold. Indeed, 66% of the 59 single institution studies used segmented regression analysis, similar to our finding that 67% of studies included in our review employed segmented regression analysis.

Finally, we acknowledge that our review is limited by what authors have reported or presented in their studies, which may not reflect the true methodological rigour of each study. Therefore, the large variation in reporting that we identified may not indicate inappropriate use of interrupted time series methods, but rather a need for reporting standards to facilitate quality reporting, application, and interpretation of interrupted time series results.

A major strength of our systematic review is the use of multiple search strategies to identify articles. Our keyword and citation search yielded 190 eligible articles (86% overall), with only 31% identified in both. We attribute this small overlap to a lack of MeSH terms for time series analysis. The additional reference list search of eligible articles identified another 30 (14% overall) eligible applications not captured in our prior searches. This observation corroborates the importance of using multiple search strategies as identified in prior reviews of new statistical
methods [71, 72]. We encourage future systematic reviews to use a similar proportional Venn diagram to clarify search strategy yield.

In summary, we identified an increase in the number of applications of interrupted time series analysis to examine interventions in drug utilization, particularly in recent years. When properly executed, interrupted time series analysis is a valuable method to evaluate the success, failure, or unintended consequences of healthcare interventions on drug utilization [77]. However, there is large variation in the reporting of interrupted time series methods. Developing methodological and reporting standards for interrupted time series analysis is important to improve its application in drug utilization research.
Table 2.1 Characteristics of empirical applications of time series studies, n=220

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Intervention</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drug policy changes</td>
<td>112</td>
<td>50.9</td>
</tr>
<tr>
<td>Co-payments or cost-sharing</td>
<td>33</td>
<td>15.0</td>
</tr>
<tr>
<td>New drug or drug withdrawal</td>
<td>15</td>
<td>6.8</td>
</tr>
<tr>
<td>Prior authorization</td>
<td>21</td>
<td>9.5</td>
</tr>
<tr>
<td>Reimbursement changes</td>
<td>43</td>
<td>19.5</td>
</tr>
<tr>
<td>New evidence</td>
<td>48</td>
<td>21.8</td>
</tr>
<tr>
<td>Guideline changes</td>
<td>15</td>
<td>6.8</td>
</tr>
<tr>
<td>Major publications</td>
<td>33</td>
<td>15.0</td>
</tr>
<tr>
<td>Quality improvement interventions</td>
<td>15</td>
<td>6.8</td>
</tr>
<tr>
<td>Educational</td>
<td>18</td>
<td>8.2</td>
</tr>
<tr>
<td>Other quality improvement</td>
<td>17</td>
<td>7.7</td>
</tr>
<tr>
<td>Safety advisories</td>
<td>39</td>
<td>17.7</td>
</tr>
<tr>
<td><strong>Data source</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Administrative data</td>
<td>203</td>
<td>92.3</td>
</tr>
<tr>
<td>Medical charts</td>
<td>10</td>
<td>4.5</td>
</tr>
<tr>
<td>Survey data</td>
<td>7</td>
<td>3.2</td>
</tr>
<tr>
<td><strong>Time point intervals</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Monthly</td>
<td>166</td>
<td>75.5</td>
</tr>
<tr>
<td>Quarterly</td>
<td>31</td>
<td>14.1</td>
</tr>
<tr>
<td>Other (e.g., annually, bi-annually, bi-weekly, weekly)</td>
<td>23</td>
<td>10.5</td>
</tr>
<tr>
<td><strong>Primary outcome measure</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Patient</td>
<td>64</td>
<td>29.1</td>
</tr>
<tr>
<td>Number</td>
<td>15</td>
<td>6.8</td>
</tr>
<tr>
<td>Proportion</td>
<td>49</td>
<td>22.3</td>
</tr>
<tr>
<td>Prescriptions</td>
<td>157</td>
<td>71.4</td>
</tr>
<tr>
<td>Number</td>
<td>49</td>
<td>22.3</td>
</tr>
<tr>
<td>Proportion</td>
<td>77</td>
<td>35.0</td>
</tr>
<tr>
<td>Standardized dose (e.g., daily defined dose)</td>
<td>31</td>
<td>14.1</td>
</tr>
<tr>
<td><strong>Interrupted time series method reported</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARIMA models</td>
<td>31</td>
<td>15.5</td>
</tr>
<tr>
<td>Linear regression</td>
<td>22</td>
<td>11.0</td>
</tr>
<tr>
<td>Segmented regression</td>
<td>134</td>
<td>67.0</td>
</tr>
<tr>
<td>Linear models</td>
<td>67</td>
<td>30.0</td>
</tr>
<tr>
<td>Other models (e.g., GEE, logistic, nonlinear)</td>
<td>9</td>
<td>6.7</td>
</tr>
<tr>
<td>Not specified</td>
<td>58</td>
<td>43.3</td>
</tr>
<tr>
<td><strong>Other regression (e.g., GEE, logistic, nonlinear, Poisson)</strong></td>
<td>14</td>
<td>7.0</td>
</tr>
<tr>
<td><strong>Statistical considerations reported</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>146</td>
<td>66.4</td>
</tr>
<tr>
<td>Comparison group</td>
<td>77</td>
<td>35.0</td>
</tr>
<tr>
<td>Confidence intervals or standard errors reported with estimates</td>
<td>153</td>
<td>69.5</td>
</tr>
<tr>
<td>Forecasting using pre-intervention trends</td>
<td>64</td>
<td>29.1</td>
</tr>
<tr>
<td>Graphical figures to display results</td>
<td>184</td>
<td>83.6</td>
</tr>
<tr>
<td>Lag periods</td>
<td>61</td>
<td>27.7</td>
</tr>
<tr>
<td>Non-stationarity</td>
<td>32</td>
<td>14.5</td>
</tr>
<tr>
<td>Number of pre- and post-intervention data points</td>
<td>86</td>
<td>39.1</td>
</tr>
<tr>
<td>Outliers</td>
<td>17</td>
<td>7.7</td>
</tr>
<tr>
<td>Seasonality</td>
<td>68</td>
<td>30.9</td>
</tr>
<tr>
<td>Sensitivity analysis</td>
<td>45</td>
<td>20.5</td>
</tr>
<tr>
<td>Time point clearly defined</td>
<td>186</td>
<td>84.5</td>
</tr>
</tbody>
</table>

ARIMA: Autoregressive integrated moving average; GEE: Generalized estimating equation

*a Some characteristics are not mutually exclusive, thus proportions add to greater than 100%

*b For papers reporting detailed time series methods only, n=200

*c For papers conducting segmented regression only, n=134
Figure 2.1 Flow diagram of systematic search results. MEDLINE® was used for the keyword search using search terms (Appendix A), and Web of Science® was used for the citation search [60-62, 70].
Figure 2.2 Proportional Venn diagram of search result yield of empirical applications by search strategy, n=220. The size of each circle is proportional to the relative number of articles identified. MEDLINE® keyword search terms are listed in Appendix A, and four papers were used in the Web of Science® citation search [60-62, 70]. The reference search included all eligible empirical applications (n=190), methods (n=10), and reviews (n=4) identified by the keyword and citation searches. One article not indexed in MEDLINE® or Web of Science® databases was identified by a reviewer during the peer-review process.
Figure 2.3 Number of interrupted time series empirical applications in drug utilization research, by publication year, n=220
Chapter 3. Methods

3.1 Chapter Overview

This chapter summarizes the methods used to investigate the research objectives of the second study of this thesis. The relevant data sources are listed, potential urban and rural region classifications are reviewed, and the outcome measures are described. Interrupted time series methods and the data analysis plan are also outlined. The two main objectives of this second thesis project were as follows:

1) To assess urban-rural differences in the dispensing of oral bisphosphonate in Ontario following formulary changes using interrupted time series analysis.

**Hypothesis:** Rural physicians will respond more slowly to drug formulary changes that made alendronate and risedronate more accessible to the public, as rural physicians may be less aware of this new information.

2) To examine urban-rural differences in the dispensing of new oral bisphosphonate formulations upon entry into the market compared to existing oral bisphosphonate formulations.

**Hypothesis:** Rural regions will initially have a slower uptake of new oral bisphosphonate formulations, as pharmaceutical marketing is often concentration in urban regions and thus rural physicians may be less aware of these new products.

There were three main components examined in this analysis: 1) regional differences, which refers to differences in the mean proportion of claims dispensed between urban, nonmajor urban, and rural regions; 2) temporal differences, which represent changes in the level or trend (defined in Section 3.5.2) of dispensing following formulary changes or the introduction of new formulations; and 3) geo-temporal, or slope, differences. Slope differences were tested with
interaction terms between region and temporal estimates (i.e., level, trend) to determine any regional impacts on the level or trend of dispensing.

3.2 Data Sources

A comprehensive collection of healthcare administrative databases from the Institute for Clinical Evaluative Sciences (ICES) was used to examine oral bisphosphonate utilization among community-dwelling seniors in Ontario as well as their demographic information (age, sex, region) and the demographic information of prescribing physicians (age, sex, location, speciality). These databases include the ICES Physician Database (IPDB), Ontario Drug Benefit (ODB), and Ontario Registered Persons Database (RPDB). Each database has been well validated and described in the literature, and has previously been utilized to examine dispensing trends in Ontario (Appendix D) [24, 78-82]. It is important to note that pharmacy claims from the ODB database represent drugs that have been dispensed, and may not necessarily reflect what physicians initially prescribed to patients.

3.3 Defining Urban-Rural Regions

There are several methods for defining regions within Ontario for healthcare analysis. Most are derived from Statistics Canada’s Postal Code Conversion File, which recodes postal codes to define regional classifications [81]. The most commonly used coding systems include Statistics Canada’s Standard Geographical Classification, the Ministry of Health and Long-Term Care’s residence coding system, and Local Health Integration Networks, which are based on geographic environments and are further described in Table 3.1. Since the focus of this thesis is on the variation between urban and rural regions, definitions based on geography alone are insufficient and more detailed measures of rurality are needed.
Despite growing interest in the examination of health inequities between regional areas, there is no universally accepted definition of rurality. Depending on the definition used, Canada’s rural population can range from 22% to 38% of the country’s total population [2]. The most commonly used urban-rural classification systems are based on factors other than, or in addition to, geographic boundaries, and include: 1) Forward Sortation Areas, 2) Urban Area Rural Area codes, 3) Statistical Area Classification codes, and 4) Rurality Index for Ontario [81]. Each classification is summarized in Table 3.2 and detailed below.

### 3.3.1 Forward Sortation Areas

Forward Sortation Areas (FSA) comprise the first three characters of the six character Canadian postal code. Rural areas are defined as areas with postal codes containing a zero (0) in the second character of the FSA, and urban areas have a number from 1-9 [81]. Although using the FSA is a quick and simple way to identify rurality, many rural postal codes cross the boundaries of standard geographic areas and thus it can be difficult to identify precise locations. Furthermore, this dichotomous urban-rural indicator is less comprehensive, and assumes two homogeneous categories, which may not be the case [81]. Rurality indicators with more detail are more desirable as they account for the heterogeneity present in rural communities [1]. Therefore the FSA is an inadequate rurality indicator and was not considered further for this research.

### 3.3.2 Urban Area Rural Area Codes

Urban Area Rural Area codes define urban areas based on population size. Urban areas are defined as having at least 1,000 persons and a population density of at least 400 persons per square kilometre [81]. By default, all remaining regions are considered rural. Since the Urban Area Rural Area classification codes are based on population size and density alone, it is a less comprehensive definition of rurality and was therefore deemed insufficient for this research.
3.3.3 Statistical Classification Codes

Statistical Area Classification codes are also based on population size, yet further divide rural regions based on Metropolitan Influenced Zones [1, 2]. The Metropolitan Influenced Zones are ratios which consider the percentage of residents in rural areas who travel to work daily into an urban area, capturing the degree to which larger urban regions have an influence outside of their physical boundaries [81]. Although these “commuter flows” can be used as a proxy for the population’s access to healthcare facilities [1, 83], the Rurality Index of Ontario (described below) encompasses additional variables that better reflect differential access to health services between regions and thus a more appropriate rurality definition for this thesis project.

3.3.4 Rurality Index of Ontario

The Rurality Index of Ontario (RIO) incorporates ten variables of “remoteness” or “isolation,” in addition to population and geography to define rurality, which makes it the best choice of the available rurality definitions. The RIO scores communities from 0 to 100 based on the following ten variables (further described in Table 3.3): travel time to the nearest basic and advanced referral centres, community population, number of active general practitioners, population to general practitioner ratio, presence of a hospital, ambulatory services, social indicators, weather conditions, and selected services. Communities with higher scores are considered to be more rural [81, 84]. Regions in Ontario are considered to be rural if their RIO score is greater than 45 [81, 85]; the remaining urban Ontario communities can also further be divided into major urban areas (RIO score 0 to 9) and nonmajor urban areas (RIO score 10 to 44) [85]. Using the RIO is advantageous as the estimates of travel time to the nearest referral centres can act as an alternate way to measure access to healthcare services [84]. RIO accounts for variables such as the number of physicians, the population to practitioner ratio, and the presence of hospitals, capturing the professional and social isolation of physicians and their patients [5]. Therefore the
RIO is the most appropriate rurality measure for this project. Regions were defined by physician RIO since the thesis hypotheses are based on the lack of awareness of new drug information a physician may have. A sensitivity analysis examining outcomes by patient RIO was also completed. Regions were categorized into one of three groups for this project: major urban, referred to as urban (RIO scores ≤ 9), nonmajor urban (RIO scores 10-44), or rural regions (RIO scores ≥ 45) [85].

3.4 Outcome Measure

All claims (new and refill) for oral bisphosphonates (alendronate, etidronate, and risedronate) indicated for osteoporosis dispensed from January 1, 2000 (the year alendronate was first listed on the formulary) to March 31, 2014 (most recent data available) were identified from the ODB for all persons aged 65 or more years. Oral bisphosphonates dispensed to long-term care patients were excluded, as were prescriptions with missing physician and/or patient postal code data.

The primary objective of this thesis was to assess the effect of ODB formulary listing status changes on the use of oral bisphosphonates. This was done by examining the outcome measure as the percentage of alendronate and/or risedronate prescriptions of the total number of oral bisphosphonate claims. Alendronate and risedronate prescriptions were combined into a single outcome measure as the formulary changes of interest were identical and therefore expected to impact both drugs equally. The secondary objective was to examine differences in uptake of oral bisphosphonates by new formulations (alendronate + vitamin D3, monthly risedronate, and risedronate delayed-release) entering the market. The outcome measure used for these analyses was the percentage of the formulation of interest of the total number of claims for oral bisphosphonates containing the same bisphosphonate molecule. For example, the new weekly
alendronate + vitamin D₃ formulation was calculated as a percentage of the total number of alendronate prescriptions dispensed during the time period. This strategy allowed comparison of use of the same bisphosphonate molecule, thus focusing on preferential prescribing of new formulations. The outcome measure for both objectives was calculated as a percentage of claims within regions to compare proportional use between urban, nonmajor urban, and rural regions.

3.5 Interrupted Time Series Analysis

To investigate temporal changes and whether formulary updates have impacted oral bisphosphonate dispensing patterns, interrupted time series analysis was used. As detailed in Chapter 2, interrupted time series analysis is the strongest quasi-experimental approach to assess the impact of an interruption, or intervention, on an outcome of interest [61]. The strength of interrupted time series models lies in their ability to control for a pre-existing baseline trend to identify changes in response to an intervention of interest (Figure 3.1) [61]. Data are analyzed at multiple time points before and after the intervention, with the key assumption that data trends prior to the intervention can be extrapolated to accurately predict post-intervention trends had the intervention not occurred. This strengthens the inference that any deviation from the expected underlying trend was caused by the intervention [59, 86]. In addition, temporal changes can be intuitively detected by visual inspection as seen in Figure 3.1. This allows a general understanding of the data and effects for readers who do not have a time series background [61, 70]. However, the design cannot eliminate potential confounding related to other changes occurring at the same time and thus caution must be used when making a conclusion about the extent of the impact of the intervention on the outcome (Section 2.3) [58].
As outlined in Chapter 2, the two main interrupted time series methods used in drug utilization research are: 1) autoregressive integrated moving average (ARIMA) models [87], and 2) segmented regression analysis [61]. Each method is detailed below and their advantages and disadvantages are briefly described in Table 3.4.

3.5.1 ARIMA Models

There are three structural parameters that define an ARIMA time series model: the autoregressive (p), integrated (d), and moving average (q) components. To execute interrupted time series analysis using ARIMA methods, a hypothetical probability model that best represents the data is generated by estimating the value of at least one of these three parameters [88]. Once the values have been identified, the remaining error values can be estimated and any changes resulting from an intervention (i.e., temporal changes) can be tested for significance using interventional ARIMA models [88, 89]. The model is estimated using pre-intervention observations, which are then applied to the entire dataset. An intervention component is added to the model to assess the impact of the intervention on the outcome of interest. The intervention component can assume several forms depending on the assumptions made about the intervention. Interventions with an abrupt and permanent change are modeled with a step function, abrupt temporary changes are modeled by a pulse function, and gradual, constant changes following an intervention are modelled by a ramp function [87, 88].

A major advantage of ARIMA models is their ability to account for: 1) autocorrelation, 2) non-stationarity, and 3) seasonality [87], further described below:

1) **Autocorrelation** is a prevalent issue in time series data and occurs when consecutive observations in time are dependent on one another, resulting in serial correlations of error terms. If autocorrelation exists, the standard errors will be underestimated and produce
artificially low p-values. Since values are in a chronologic time sequence in time series analysis, each point is likely to be correlated to the next and therefore error terms are also likely to be correlated [61, 87].

2) **Non-stationarity** time series data exhibit one or more secular trends, resulting in the outcome mean and variance to fluctuate for reasons other than the intervention.

3) **Seasonality** is a type of non-stationarity that represents fluctuations in time series data that occur seasonally, such as an increase in drug utilization due to the seasonality of some diseases [61, 87]. There may also be an increased number of prescriptions dispensed prior to long-term travel plans, or seasonality in clinical outcomes requiring treatment, such as increased fracture rates during winter [90].

These issues are prevalent in time series studies and if not accounted for, may mask the true impact of the intervention. ARIMA models inherently account for all three issues and thus provide a better estimation of standard errors and an accurate assessment of the effect of an intervention.

**3.5.2 Segmented Regression Analysis**

Segmented regression analysis is an alternative interrupted time series method that fits a regression model to data in the pre-intervention segment and another regression model to data in the post-intervention segment. Most segmented regression models fit a least squares regression line to each segment, assuming a linear relationship between time and the outcome, although polynomial and nonlinear regression models can be used as well [60, 61]. Within the segmented regression model, two regression coefficients represent temporal changes following the intervention: the level and trend parameters. A change in the level parameter represents a change in the intercept between the pre- and post-intervention segments, and constitutes an abrupt intervention effect. The trend parameter represents the slope of the data in each segment and
represents gradual changes following an intervention [61]. Level and trend values represent relative changes from baseline (pre-intervention) values. A positive temporal change indicates an increase in the value immediately after the intervention compared to the value before the intervention (level), or an increase in the slope (trend) of dispensing in the post-intervention segmented compared to the pre-intervention segment. Likewise, negative values indicate an immediate drop in dispensing (level), or a slower rate of dispensing (trend), following the intervention.

Segmented regression is advantageous as it provides a more transparent means of estimating the magnitude of the impact at different time points using the level and trend components, in comparison to ARIMA models. However, unlike ARIMA models, segmented regression does not inherently account for autocorrelation, non-stationarity, or seasonality in its model. Instead, correction for these can be applied separately if necessary. Segmented regression analysis was used in this thesis as it yields a level and trend coefficient, allowing for direct comparison of temporal changes in oral bisphosphonate utilization between regions following ODB formulary updates.

### 3.6 Statistical Analysis

All statistical analyses were performed using SAS statistical software, version 9.3 (SAS Institute Inc., Cary, North Carolina) [91]. A two-sided p-value of 0.05 was used as the threshold of statistical significance for all analyses.
3.6.1 Descriptive Analysis

3.6.1.1 Characteristics of Patients and Physicians
The age and sex of all unique patients dispensed an oral bisphosphonate for the first time during the study period (January 2000 to March 2014), as well as the age, sex, and specialty of all physicians prescribing an oral bisphosphonate were summarized by region using descriptive statistics. Regional differences in means were compared using one-way analysis of variance (ANOVA) test, and when significant, a Tukey’s test was used to determine which regions were statistically different. Chi-squared tests were used to compare proportions with the Bonferroni-adjusted p-value to determine which regions were significantly different.

3.6.1.2 Oral Bisphosphonate Dispensing Over Time
The monthly number of claims of oral bisphosphonates (total, and by type: alendronate, etidronate, risedronate) dispensed to community-dwelling seniors in Ontario were examined. The number of monthly alendronate claims by formulation (daily, weekly, and weekly + vitamin D₃) and the number of monthly risedronate claims by formulation (daily, weekly, monthly, and weekly delayed-release) dispensed in Ontario were also examined and described separately.

3.6.2 Interrupted Time Series Analysis: Utilization of Oral Bisphosphonates Following Formulary Changes
Segmented regression analysis using a linear model was used to assess the temporal impact of the formulary changes on the proportion of alendronate and risedronate dispensed. The following model was used:

\[ Y_t = \beta_0 + \beta_1 \times \text{time}_t + \beta_2 \times \text{intervention}_t + \beta_3 \times \text{time after intervention}_t + e_t \quad \text{Equation 3.1} \]

Where \( Y_t \) represents the proportion of alendronate/risedronate claims at month \( t \); \( \text{time} \) is a continuous variable which represents the time in months at time \( t \) from the start of the
observational period; **intervention** is a dummy variable for the period occurring before (intervention = 0) or after (intervention = 1) the intervention of interest. In this thesis, the intervention of interest is the formulary change for alendronate and risedronate; **time after intervention** is a continuous variable indicating the time since the intervention, where months prior to the intervention are coded as 0 and the months in the post-intervention period are continuous beginning at 1; and e is an error term at time t which depicts the random variation not explained by the model. In the above equation, $\beta_0$ is the baseline level, or proportion of alendronate/risedronate prescriptions, at time zero; $\beta_1$ estimates the baseline trend before the intervention was implemented; $\beta_2$ depicts the level change after the intervention occurs; and $\beta_3$ is an estimation of the change in slope between the pre- and post-intervention phases [61].

Two separate segmented regression analyses were conducted to examine the impact of each formulary changes on dispensing. The formulary changes (interventions of interest) that were examined included:

1) The updated Limited Use listing of alendronate (April 2003) and risedronate (September 2003), and

2) General Benefit listing of alendronate (January 2007) and risedronate (June 2007).

Interventions issued within less than 24 months of each other were combined and treated as a single event based on the occurrence of the first event [61], thus the two intervention time points of interest were April 2003 and January 2007. In addition, claims for alendronate and risedronate were combined into a single outcome measure as formulary changes were expected to impact both drugs equally. Since claims are combined and the interventions occurred at a specified point in time without any gradual implementation, no transition (lag) periods were included in the model. Both models also included interaction terms to compare geo-temporal slope differences in the level ($\text{region*intervention}_t$) and trend ($\text{region*time after intervention}_t$) coefficients.
between regions. Interaction terms were tested and rejected at an alpha level of 5%. The full model for the first (Equation 3.2, change in Limited Use criteria in 2003) and second intervention (Equation 3.3, change to General Benefit in 2007) regressions are below:

\[
Y_{1t} = \beta_0 + \beta_1 \ast \text{time}_t + \beta_2 \ast \text{intervention1}_t + \beta_3 \ast \text{time after intervention1}_t + \beta_4 \ast \text{region}_t + \beta_5 \ast \text{region} \ast \text{intervention1}_t + \beta_6 \ast \text{region} \ast \text{time after intervention1}_t + \epsilon_t
\]

\[
Y_{2t} = \beta_0 + \beta_1 \ast \text{time}_t + \beta_2 \ast \text{intervention2}_t + \beta_3 \ast \text{time after intervention2}_t + \beta_4 \ast \text{region}_t + \beta_5 \ast \text{region} \ast \text{intervention2}_t + \beta_6 \ast \text{region} \ast \text{time after intervention2}_t + \epsilon_t
\]

Equation 3.2

Equation 3.3

Results were stratified by region and segmented regression was conducted separately for each urban, nonmajor urban, and rural regions. Preliminary graphing of the outcome from January 2000 to March 2014 revealed ceiling and floor effects (Appendix E). Floor effects were observed prior to 2001, as very few prescriptions for alendronate/risedronate were dispensed since second generation bisphosphonates were not listed on the ODB until late 2000 (alendronate) and early 2001 (risedronate). Ceiling effects of the outcome were also identified due to the nature of the outcome measure selected; since the outcome measure is a percentage of alendronate/risedronate claims of total oral bisphosphonates, the outcome reaches a ceiling as the proportion of claims dispensed reaches 100%. Thus the segmented regression analysis examining formulary change impacts was focused on claims dispensed between January 1, 2001 (when claims begin to increase) and December 31, 2008 (two years after the second formulary change and prior to any observed ceiling effects) in Ontario overall, and by region. Data since January 2001 (when claims began to increase) was used to forecast data for two years following the first formulary change. Forecasting was also completed for two years following the second formulary change, using pre-intervention data since January 2005 (two years prior to the second formulary change).
The adequacy of a linear model was examined using the following residual diagnostic checks: the residual normality probability plot and the Shapiro-Wilk test for normality, the standardized residual plots and generalized Durbin-Watson statistic for independence, and the DF Fit statistic for outliers. These diagnostic checks indicated that the linear model was an adequate fit and satisfied distribution assumptions overall. No influential outliers were identified.

The augmented Dickey-Fuller test was performed to test for non-stationarity, and differences examined using the PROC ARIMA procedure. The generalized Durbin-Watson statistic was used to detect autocorrelation. If autocorrelation was present, autoregressive terms were included in the model to control for autocorrelation using the AUTOREG procedure available in SAS [92]. Seasonality was also accounted for using PROC AUTOREG by specifying an NLAG option greater than the order of any potential seasonality. Regression estimates of the linear model were estimated using the maximum-likelihood method, a technique commonly applied in segmented regression time series methods [92]. This type of analysis produced more likely estimates than those derived from ordinary least squares regression. Regression estimates represent relative changes in the proportion of alendronate/risedronate dispensed compared to baseline trends, and were reported with 95% confidence intervals.

3.6.3 Utilization of New Oral Bisphosphonate Formulations

Regional, temporal, and slope differences in dispensing of new oral bisphosphonate formulations were examined, as outlined by the secondary objective of this thesis project. All claims (new and refill) for new oral bisphosphonate formulations were examined as a percentage of the new formulation of the total number of oral bisphosphonates claims containing the same bisphosphonate molecule. The new formulations of interest and their ODB listing status were:
1) Alendronate + vitamin D₃, first available as General Benefit in January 2007,
2) Monthly risedronate, first available as General Benefit in June 2009, and
3) Risedronate delayed-release, first available as General Benefit in February 2012.
Patterns of claims dispensed were plotted from the date the drug was first available on the ODB as General Benefit until end of the study period (March 31, 2014). No distinction was made between brand or generic claims for the main thesis analyses, however brand and generic claims were examined separately in another analysis. The mean proportions of new formulation claims were compared between regions using ANOVA tests to determine regional differences. If ANOVA tests were significant, Tukey’s test was conducted to determine which regions were statistically different. Temporal differences in the trends in uptake of new bisphosphonate formulations dispensed were tested for using regression analysis and 95% confidence intervals were calculated. Interaction terms between region and time were included in the regression to examine slope differences in dispensing over time. Results were stratified by region if interaction terms were significant. The regression model used was as follows:

\[
y_t = \beta_0 + \beta_1 \cdot \text{time}_t + \beta_2 \cdot \text{region} + \beta_3 \cdot \text{region} \cdot \text{time}_t + e_t
\]

Equation 3.4

3.7 Data Access and Ethical Considerations
ICES is a Prescribed Entity under Ontario’s privacy law and is thus legally permitted to receive and use personal health information for health services research. ICES UofT is a satellite site of ICES and maintains policies, practices, and procedures that are approved and regularly audited by the Information and Privacy Commissioner of Ontario (www.ipc.on.ca). Ethics approval for this study was obtained from the University of Toronto Research Ethics Board. Drs. Cadarette, Lévesque, and Mamdani are ICES Scientists and Racquel Jandoc received ethical training by
privacy officer Don DeBoer prior to gaining student access to data housed at ICES UofT. Coded data were provided and data analyses were completed solely by Racquel Jandoc at ICES UofT.
### Table 3.1 Geography classifications commonly used in Ontario

<table>
<thead>
<tr>
<th>Geographic Coding System</th>
<th>Description</th>
</tr>
</thead>
</table>
| Postal Codes              | • Canadian postal codes are a combination of six alpha-numeric characters that describe a region in Ontario  
• The first character denotes a province, territory, or a major sector found within the boundaries of a province |
| Local Health Integration Network (LHIN) | • Established by the Ministry of Health and Long-Term Care in 2006 as a way to plan, fund, and manage health services locally (14 in Ontario)  
• Boundaries of the LHINs reflect patient utilization of healthcare services in their communities  
• Between 59.1% to 97.2% of the populations receives health services locally within their LHIN |
| Ministry of Health and Long-Term Care Residence Coding System | • Postal codes are linked to Ontario census divisions, dividing Ontario into four main divisions:  
  • Single-tier municipality: Predominantly urban; created when a former regional municipality consisted of a single dominant urban center and its suburbs  
  • Regional municipality: Cluster of predominantly urban centers and independent cities with no single dominant center  
  • County: Cities, villages, towns, or townships  
  • District: May contain small cities, towns, and townships; mostly comprised of unincorporated land in Northern Ontario |
| Standard Geographical Classification | • Made up of a three-level hierarchy:  
  • Province/territory  
  • Census Division (CD): A group of neighbouring municipalities that are joined together for the purposes of regional planning and managing common services (49 in Ontario)  
  • Census Subdivision (CSD): General term for municipality (as determined by provincial legislation) or an area treated as municipal equivalent (585 in Ontario)  
• Levels are hierarchically related in that CSDs aggregate into CDs, which aggregate into provinces and territories |

Information summarized from the Ontario Ministry of Health and Long-Term Care [81]
Table 3.2 Rurality classifications commonly used in Ontario

<table>
<thead>
<tr>
<th>Name of classification</th>
<th>Brief description of urban-rural region definitions</th>
</tr>
</thead>
</table>
| Forward Sortation Area (FSA) | • The first three characters of the Canadian postal code comprise the FSA  
• Rural areas have a zero (0) in the second character of the FSA  
• Urban areas have a number from 1-9 in the second character of the FSA                                                                 |
| Urban Area Rural Area        | • Based on population size and density  
• Urban: Population > 1000 and a population density of at least 400 persons/km²  
• Rural: All areas not classified as urban                                                                                   |
| Statistical Area Classification | • Based on population size and given a code of 1-7. Codes > 4 (population < 10,000 persons) are classified as rural  
• Additional use of a commuter ratio to define Metropolitan Influenced Zones. Commuter ratios represent the percentage of the rural labour force who travel to work daily into an urban area            |
| Rurality Index for Ontario (RIO) | • A score from 0 to 100 which weight and sum a set of variables (Table 3.3) in order to provide a degree of “rurality”  
• Higher RIO score reflects a higher degree of rurality                                                                       |

Information summarized from the Ontario Ministry of Health and Long-Term Care [81]
## Table 3.3 Variables considered in the Rurality Index of Ontario

<table>
<thead>
<tr>
<th>Variable</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Travel time to nearest basic referral center</td>
<td>• Provide hospital/specialty services, including medical imaging</td>
</tr>
<tr>
<td></td>
<td>• Highest score given to communities with travel times &gt; 170 min</td>
</tr>
<tr>
<td>Travel time to nearest advanced referral center</td>
<td>• Highest score given to communities with travel times &gt; 190 min</td>
</tr>
<tr>
<td>Population</td>
<td>• Populations &lt; 46,000 receive points in a linear fashion</td>
</tr>
<tr>
<td></td>
<td>• Populations &gt; 46,000 receive no points</td>
</tr>
<tr>
<td>Number of active general practitioners</td>
<td>• Highest score given to communities with none</td>
</tr>
<tr>
<td></td>
<td>• No points given to communities with &gt; 20</td>
</tr>
<tr>
<td>Population to general practitioner ratio</td>
<td>• Populations exceeding provincial average receive added points</td>
</tr>
<tr>
<td>Presence of hospital</td>
<td>• Points are added to communities with a hospital</td>
</tr>
<tr>
<td></td>
<td>• Points are deducted as the number of specialists increases</td>
</tr>
<tr>
<td>Ambulance services</td>
<td>• Points are added to communities with no ambulance services</td>
</tr>
<tr>
<td></td>
<td>• No points are added if ambulance services are available</td>
</tr>
<tr>
<td>Social indicators</td>
<td>• Points are added to communities with no airports, university, or community college</td>
</tr>
<tr>
<td></td>
<td>• No points are given if these facilities are present</td>
</tr>
<tr>
<td>Weather conditions</td>
<td>• Points are added to communities with extreme snowfall (exceeds provincial 75th percentile)</td>
</tr>
<tr>
<td></td>
<td>• Cold temperature (below provincial 25th percentile)</td>
</tr>
<tr>
<td>Selected services</td>
<td>• Points are added to communities where general practitioners provide anaesthetic and obstetrical services</td>
</tr>
</tbody>
</table>

Information summarized from Minore et al [84]
Table 3.4 Advantages and disadvantages of interrupted time series analysis techniques

<table>
<thead>
<tr>
<th></th>
<th>Autoregressive integrated moving average (ARIMA) model</th>
<th>Segmented linear regression</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Advantages</strong></td>
<td>Models take autocorrelation into account to provide a better estimation of standard errors and an actuate estimate of significance [61, 87]</td>
<td>Estimates the size of the intervention impact at different time points, as well as variations in trend over time</td>
</tr>
<tr>
<td></td>
<td>Model can fit both linear and non-linear trends in the data</td>
<td>Requires a minimum of only 9 baseline observations for forecasting [67]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Slope and level regression coefficients are a more transparent estimate of the intervention impact [61]</td>
</tr>
<tr>
<td><strong>Disadvantages</strong></td>
<td>Less effective when used to examine shifts in trends that occur at defined time points, such as an intervention [61]</td>
<td>Must be corrected for autocorrelation, non-stationarity, and seasonality using additional methods [61]</td>
</tr>
</tbody>
</table>
Figure 3.1 Interrupted time series model adapted from Schneeweiss et al [93]. Data are collected at multiple time points before and after the intervention (vertical dotted line), with the key assumption that pre-intervention baseline trends can be extrapolated to accurately predict post-intervention trends (dashed line) had the intervention not occurred. Any deviation from expected baseline trends may therefore be attributed to the impact of the intervention. Level changes occur immediately after the intervention, and slope changes are gradual changes in trends following the intervention.
Chapter 4. Results

4.1 Descriptive Analysis: Characteristics of Patients and Physicians

In total, 23,503,576 oral bisphosphonate claims dispensed to ODB-eligible seniors (aged 65+ years) from January 1, 2000 to March 31, 2014 were identified (Figure 4.1). After excluding non-osteoporotic formulations (Appendix F) and claims dispensed to patients in long-term care facilities, 18,381,942 oral bisphosphonate claims were eligible. Of these claims, 89% had complete physician regional data (primary analysis) and 99% had complete patient regional data (Appendix G). This thesis analysis did not distinguish between brand and generic claims when examining trends as it was not part of the main objectives; however brand and generics were plotted separately in a secondary analysis (Appendix H). After excluding claims with missing physician regional data, there were 16,367,752 eligible claims for primary analysis.

Overall, 595,597 unique patients were dispensed an eligible oral bisphosphonate for the first time and 19,161 unique physicians dispensed an eligible oral bisphosphonate from January 2000 to March 2014 (Figure 4.2). Demographic characteristics for patients and physicians are summarized in Table 4.1a and Table 4.1b, respectively, based on complete physician RIO data. The majority (92%) of claims in this study had matching patient and physician RIOs. Seventy-two percent of patients resided in urban regions, 21% resided in nonmajor urban regions, and 6% resided in rural regions. The mean age of patients was similar between regions, however nonmajor urban regions had a higher proportion of patients aged 75 or more years. The proportion of female patients was also similar across all regions (approximately 84%). Of 19,161 physicians dispensing an eligible oral bisphosphonate for the first time, 80% practiced in urban regions, 15% practiced in nonmajor urban areas, and 5% practiced in rural regions. The mean age of prescribing physicians was similar between urban (47.1 years) and nonmajor urban (48.8
years), yet was lower for physicians practicing in rural regions (42.8 years). Overall, 32% of prescribers were female. Similar proportions of female physicians were observed between urban (33%) and rural (30%) physicians, however nonmajor urban regions had the lowest proportion of female prescribers (26%). Oral bisphosphonates were prescribed most often by general practitioners in all regions, although there was a higher proportion prescribed by general physicians in rural regions (88%) compared to urban (54%) and nonmajor urban (72%) regions.

4.2 Descriptive Analysis: Oral Bisphosphonate Dispensing Over Time

The total number of oral bisphosphonate claims in Ontario has increased steadily over time from 2000 to 2014, and has been relatively stable since 2009 (Figure 4.3). The number of alendronate and risedronate claims was similar over time, until early 2010; at this time, the number of risedronate claims was slightly higher than claims for alendronate. The gap between risedronate and alendronate claims has persisted and widened since 2013. There were monthly fluctuations observed in all oral bisphosphonate claims. Alendronate and risedronate claims showed similar fluctuations over time, with the highest peaks consistently occurring in October. In addition, the lowest drop in claims was consistently observed in February. Etidronate claims also showed monthly fluctuations, however they were not consistent over the entire study period.

As alendronate and risedronate entered the market in 2000 and became more accessible through the ODB (Limited Use listing in 2000 [alendronate] and 2001 [risedronate], and General Benefit listing in 2007), use of etidronate decreased while claims for alendronate and risedronate increased. The number of alendronate and risedronate claims showed similar patterns over time, until 2009 when alendronate prescriptions began to decrease and risedronate prescriptions continued to increase and eventually plateaud in recent years.
**Figure 4.4** examines the number of alendronate claims by formulation (daily, weekly, and weekly alendronate + vitamin D₃) in Ontario from January 2000 to March 2014. The use of daily alendronate increased after ODB listing in November 2000 and until the weekly formulation entered the market in January 2003. Once weekly alendronate was listed on the ODB in 2003, dispensing of daily alendronate decreased and dispensing of weekly alendronate increased until 2007. Claims for daily formulations have since leveled off, with a slight decline in the final year of observation (2013/2014). Of interest, the number of weekly alendronate + vitamin D₃ claims increased since its General Benefit listing in January 2007. The number of both weekly formulations resulted in an increase in total alendronate dispensing, but both weekly formulations have stabilized since 2012. The fluctuating trends observed in total alendronate and risedronate claims (peaks in October and drops in February) were also observed in this figure, however the patterns were less consistent than those in **Figure 4.3**. Furthermore, these fluctuations were only observed for weekly alendronate and weekly alendronate + vitamin D₃; daily alendronate claims did not seem to fluctuate monthly.

The number of risedronate claims by formulation (daily, weekly, monthly, and weekly delayed-release) from January 2000 to March 2014 is plotted in **Figure 4.5**. Similar to daily alendronate, daily risedronate increased upon formulary availability in March 2001 until weekly risedronate was listed in April 2003. Since then, the number of daily risedronate claims decreased and remained low for the rest of the study period. Dispensing of weekly risedronate increased since market entry in 2003, peaking at 67,184 claims dispensed in June 2009 and then declined. Monthly risedronate was listed on the ODB in June 2009 and dispensing increased until 2012, and has since declined. This decline coincided with the ODB listing of weekly risedronate delayed-release in February 2012. Since being listed on the formulary in 2012, the number of claims for the delayed-release formulation has been increasing. Similar to total risedronate
claims, monthly drops in weekly risedronate claims were observed, however these drops only persist until 2011. Daily and monthly risedronate formulations did not show consistent seasonal patterns. There is not enough data to determine whether weekly risedronate delayed-release follow these secular fluctuations.

4.3 Interrupted Time Series Analysis: Utilization of Oral Bisphosphonates Following Formulary Changes

Results were analyzed in the context of three aspects: 1) regional differences, which refers to differences in the mean proportion of claims dispensed between urban, nonmajor urban, and rural regions; 2) temporal differences, which represent changes in the level or trend of dispensing following formulary changes or the introduction of new formulations. Temporal changes represent relative changes from pre-intervention values; and 3) geo-temporal, or slope, differences. Slope differences were tested with interaction terms between region and temporal estimates (i.e., level, trend) to determine any regional impacts on the level or trend of dispensing.

Figure 4.6 illustrates the proportion of alendronate/risedronate claims of the total number of bisphosphonate claims dispensed to community-dwelling seniors (aged 65+ years) in Ontario from January 2001 to December 2008. The proportion of alendronate/risedronate claims increased since January 2001. Table 4.2a outlines level and trend estimates from segmented regression analysis following the first formulary change, and Table 4.2b depicts estimates following the second formulary change. There were no significant temporal changes following the first formulary change (level estimate = 0.67%, p = 0.33; trend estimate = 0.12%, p = 0.083) in Ontario. There was a significant level change following the second formulary change (level estimate = 1.31%, p = 0.0004), however there was no significant trend change (trend estimate = 0.035%, p = 0.31) in Ontario.
The proportion of alendronate/risedronate dispensing was also examined by region. As seen in Figure 4.7, dispensing patterns of alendronate/risedronate dispensing were similar in urban, nonmajor urban, and rural regions of Ontario, however regional differences in the proportion of alendronate/risedronate dispensed was observed. The mean proportion of alendronate/risedronate claims was higher in urban regions for most of the study period (mean = 68.0% of all oral bisphosphonates dispensed), however it was not significantly different from proportions dispensed in other regions (p = 0.15). The average proportion of alendronate and risedronate claims dispensed was 64.4% in nonmajor urban areas, and 63.3% in rural regions.

Significant interaction terms (Appendix I.1, p = 0.0019) indicated slope differences between the level and trend of the proportion of alendronate/risedronate prescriptions dispensed in each region following the first formulary change thus segmented regression analysis was conducted separately for each region. There was a significant trend change following the first intervention (trend estimate = 0.11%, p = 0.047), and a significant level change following the second intervention (level estimate = 1.25%, p = 0.0005) in urban regions. Nonmajor regions also exhibited a significant change in level (level estimate = 1.57%, p = 0.0005) following the second intervention. Rural areas showed significant temporal changes (level estimate = 3.41%, p < 0.0001; trend estimate = 0.29%, p < 0.0001) following the first intervention, however changes following the second intervention were not significant. These results are depicted graphically in Figures 4.8 (urban), 4.9 (nonmajor urban), and 4.10 (rural).

Examination of 95% confidence intervals in Table 4.2a and Table 4.2b demonstrated a significant difference in the change in level between nonmajor (level estimate = -0.11, 95% CI = -1.65, 1.43) and rural (level estimate = 3.41, 95% CI = 1.76, 5.07) regions, but no other temporal (i.e., level or trend) differences in the response to the formulary changes between regions were
observed, however the relative differences are very small. There were no slope differences between regions following the second intervention.

Overall, all regions exhibited similar patterns in dispensing over time. Urban regions dispensed the highest proportion of claims, but this was not significantly different from other regions. Significant interaction terms indicated slope differences in the impact of the first formulary changes between regions, however no temporal differences were observed for most of the level and trend coefficients. The lack of major temporal differences between the level and trend coefficients indicates that the response to both interventions was similar between regions.

4.4 Utilization of New Oral Bisphosphonate Formulations

4.4.1 Alendronate + Vitamin D₃

Figure 4.11 illustrates the proportion of alendronate + vitamin D₃ claims in each region from January 2007 (initial ODB-listing date) to March 2014. Dispensing trends for all regions exhibited increasing use of brand alendronate + vitamin D₃ over time. Less than 6% of urban patients dispensed alendronate were dispensed a claim for alendronate + vitamin D₃ when it was first available in January 2007. This increased to 55% of all urban alendronate users in December 2013. This increasing trend was observed in nonmajor urban (4% in January 2007 to 38% in December 2013) and rural (1% in January 2007 to 23% in December 2013) regions as well, with the highest proportion of claims dispensed in urban regions. The mean proportion of claims was significantly higher in urban regions (mean = 44%, p < 0.01), followed by nonmajor regions (mean = 31%); dispensing in rural areas (mean = 18%) was significantly lower than both urban and nonmajor urban regions (p < 0.01). These regional differences persist throughout the entire period. The highest proportion dispensed alendronate + vitamin D₃ in urban was 56% of all
alendronate users in October 2013, 39% in nonmajor regions (March 2013), and 24% in rural regions (November 2011). By the end of the study period (March 2014), 24% of oral bisphosphonate users were dispensed alendronate/risedronate in rural regions, 38% dispensed in nonmajor urban regions, and 55% dispensed by urban physicians.

The interaction term (Appendix J.1) between time and region was highly significant (p < 0.0001), indicating slope differences in dispensing between regions and thus separate regressions were conducted for urban, nonmajor urban, and rural regions. Rural regions had the slowest increase (slope = 0.24%), followed by nonmajor (slope = 0.39%), and urban regions (slope = 0.58%) had the fastest uptake (Table 4.3). Examination of 95% confidence intervals indicated significantly different temporal trends between rural (slope = 0.24%, 95% CI = 0.12, 0.36) and urban regions (slope = 0.58%, 95% CI = 0.40, 0.75).

Overall, urban regions had a significantly faster uptake and dispensed a higher proportion of alendronate + vitamin D₃ compared to other regions; these temporal differences persisted for the entire study period.

4.4.2 Monthly Risedronate

The proportion of monthly risedronate claims of the total number of risedronate claims per region was examined from June 2009 (initial ODB-listing date) until March 2014. As illustrated in Figure 4.12, increasing proportions of brand name monthly risedronate dispensing was observed in all regions over time until the end of 2011 (39% dispensed in urban regions, 32% in nonmajor regions, and 26% in rural regions dispensed in December 2011). A subsequent decrease in monthly risedronate dispensing in all regions was then observed. Less than 1% of total risedronate prescriptions were dispensed for monthly risedronate in all regions in June
2009, the first month this formulation was listed on the ODB. This increased to 28% in rural, 33% in nonmajor, and 40% in urban regions in May 2012. Urban, nonmajor urban, and rural regions showed distinct regional differences in monthly risedronate dispensing which were also statistically significant; urban regions had a statistically higher proportion of monthly risedronate claims overall (mean = 29%, p < 0.01) compared to nonmajor urban (mean = 24%) and rural regions (mean = 20%, p < 0.01). The mean proportion dispensed was also significantly different between nonmajor and rural regions (p < 0.05). This regional gradient in dispensing remained throughout the entire study period. The highest proportion of monthly risedronate claims dispensed in urban and nonmajor urban regions was 40% and 33% in January 2012, respectively; the highest proportion of claims in rural regions was 28% in October 2011. By the end of the study period (March 2014), 25% of alendronate/risedronate claims were dispensed to urban patients, 21% to patients in nonmajor regions, and 20% to patients residing in rural areas.

A significant interaction term (Appendix J.2, p = 0.0085) between time and region indicated slope differences in monthly risedronate dispensing, thus separate regressions for each region were conducted. Similar to alendronate + vitamin D₃ trends, rural regions exhibited the slowest increase of monthly risedronate dispensing (slope = 0.33%, 95% CI = 0.05, 0.62). Nonmajor urban areas had an increase of 0.36% (95% CI = 0.09, 0.63) monthly risedronate prescriptions per month, and urban regions had the fastest uptake in dispensing, with an increase of 0.42% (95% CI = 0.12, 0.72) monthly risedronate claims dispensed per month (Table 4.3). As indicated by overlapping 95% confidence intervals, there were no significant temporal differences between regions in the dispensing of monthly risedronate. Despite this, there were significant regional differences in monthly risedronate dispensing, with urban areas dispensing the highest proportion of claims.
4.4.3 Risedronate Delayed-Release

Claims for risedronate delayed-release formulations were examined as a proportion of total risedronate claims in Ontario from February 2012 (initial ODB-listing date) to March 2014 (Figure 4.13). Overall dispensing of risedronate delayed-release has increased steadily since its introduction to the market. There were no prescriptions for risedronate delayed-release dispensed by rural physicians when the drug was first available in February 2012, however the proportion of claims increased to 26% in March 2014. The percentage dispensed was similar in urban and nonmajor urban regions, increasing from less than 1% in February 2012 to 37% in urban regions and 38% in nonmajor urban regions in March 2014. As observed in the other new oral bisphosphonate formulations, risedronate delayed-release claims were significantly lower in rural regions (mean = 14.0%, p < 0.05). However, unlike the alendronate + vitamin D₃ and monthly risedronate where distinct trends between urban, nonmajor urban, and regions were observed, urban (mean = 22%) and nonmajor urban (mean = 22%) dispensing patterns were very similar and were both greater than rural regions. This regional trend was consistent for the entire study period.

The non-significant interaction term (p = 0.37) indicated no slope differences between regions (Appendix J.3), however this may be due to the limited number of data points available for this formulation. Separate regression analyses demonstrate similar temporal trends in the uptake of risedronate delayed-release between urban (slope = 1.46%, 95% CI = 1.26, 1.66) and nonmajor urban regions (slope = 1.51%, 95% CI = 1.28, 1.74); trends were significantly slower in rural regions (slope = 1.11%, 95% CI = 1.05, 1.17).

Overall, significant regional and temporal trends were observed in the proportion of risedronate delayed-release claims dispensed. Dispensing in urban and nonmajor urban regions were very
similar; rural regions had slower uptake of risedronate delayed-release and also dispensed a lower proportion of claims in comparison to urban and nonmajor urban areas. These patterns persisted for the entire observation period.
Table 4.1a Demographic characteristics of patients dispensed an oral bisphosphonate (for initial claims with complete physician regional data)\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Patient Characteristics</th>
<th>Total N=591,744</th>
<th>Urban N=430,626</th>
<th>Nonmajor Urban N=127,971</th>
<th>Rural N=33,147</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>73.15 (7.34)</td>
<td>73.04 (7.31)</td>
<td>73.41 (7.45)</td>
<td>73.18 (7.29)</td>
</tr>
<tr>
<td>65-74</td>
<td>60.74</td>
<td>61.22</td>
<td>58.11</td>
<td>59.95</td>
</tr>
<tr>
<td>75-84</td>
<td>30.82</td>
<td>30.34</td>
<td>32.28</td>
<td>31.37</td>
</tr>
<tr>
<td>≥85</td>
<td>8.71</td>
<td>8.44</td>
<td>9.62</td>
<td>8.69</td>
</tr>
<tr>
<td><strong>Sex (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>83.77</td>
<td>83.53</td>
<td>84.27</td>
<td>84.91</td>
</tr>
</tbody>
</table>

RPDB: Registered Persons Database; SD: Standard deviation
\textsuperscript{a} Regions based on data from the RPDB as of 2013
\textsuperscript{b} Adjusted for missing values (missing patient RIO, n=3,853)

Table 4.1b Demographic characteristics of physicians prescribing an oral bisphosphonate (for initial claims with complete physician regional data)\textsuperscript{a}

<table>
<thead>
<tr>
<th>Physician Characteristics</th>
<th>Total N=19,161</th>
<th>Urban N=15,400</th>
<th>Nonmajor Urban N=2,820</th>
<th>Rural N=941</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age (%)\textsuperscript{b}</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>46.88 (11.87)</td>
<td>47.14 (11.87)</td>
<td>48.84 (11.83)</td>
<td>42.79 (11.36)</td>
</tr>
<tr>
<td>&lt;35</td>
<td>15.80</td>
<td>15.11</td>
<td>15.60</td>
<td>27.93</td>
</tr>
<tr>
<td>35-44</td>
<td>31.24</td>
<td>30.90</td>
<td>32.09</td>
<td>34.28</td>
</tr>
<tr>
<td>45-54</td>
<td>28.04</td>
<td>28.44</td>
<td>28.11</td>
<td>21.03</td>
</tr>
<tr>
<td>55-64</td>
<td>15.68</td>
<td>15.99</td>
<td>15.35</td>
<td>11.61</td>
</tr>
<tr>
<td>65-74</td>
<td>7.31</td>
<td>7.62</td>
<td>6.70</td>
<td>4.16</td>
</tr>
<tr>
<td>≥75</td>
<td>1.93</td>
<td>1.94</td>
<td>2.15</td>
<td>0.99</td>
</tr>
<tr>
<td><strong>Sex (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>31.66</td>
<td>32.85</td>
<td>25.64</td>
<td>30.18</td>
</tr>
<tr>
<td><strong>Specialty (%)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General practice</td>
<td>58.27</td>
<td>53.98</td>
<td>71.91</td>
<td>87.57</td>
</tr>
<tr>
<td>Internal medicine</td>
<td>4.88</td>
<td>5.06</td>
<td>4.57</td>
<td>2.87</td>
</tr>
<tr>
<td>Other specialty</td>
<td>36.85</td>
<td>40.95</td>
<td>23.51</td>
<td>9.56</td>
</tr>
</tbody>
</table>

IPDB: ICES Physician Database; SD: Standard deviation
\textsuperscript{a} Regions based on data from the IPDB as of 2011
\textsuperscript{b} Adjusted for missing values (missing physician age, n=254)
Table 4.2a Results from segmented regression model: relative changes from baseline and 95% confidence intervals of proportion of alendronate/risedronate dispensed following formulary change 1

<table>
<thead>
<tr>
<th></th>
<th>Ontario</th>
<th>Urban</th>
<th>Nonmajor Urban</th>
<th>Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level change 1</td>
<td>0.67</td>
<td>1.18</td>
<td>-0.11</td>
<td>3.41*</td>
</tr>
<tr>
<td>(intervention\textsubscript{1})</td>
<td>(-0.65, 2.01)</td>
<td>(-0.11, 2.47)</td>
<td>(-1.65, 1.43)</td>
<td>(1.76, 5.07)</td>
</tr>
<tr>
<td>Trend change 1</td>
<td>0.12</td>
<td>0.11*</td>
<td>0.18</td>
<td>0.29*</td>
</tr>
<tr>
<td>(time after intervention\textsubscript{1})</td>
<td>(-0.013, 0.26)</td>
<td>(0.0046, 0.22)</td>
<td>(-0.17, 0.53)</td>
<td>(0.18, 0.39)</td>
</tr>
</tbody>
</table>

\* Formulary change 1: change in Limited Use criteria, April 2003
* Significant at p < 0.05

Table 4.2b Results from segmented regression model: relative changes from baseline and 95% confidence intervals of proportion of alendronate/risedronate dispensed following formulary change 2

<table>
<thead>
<tr>
<th></th>
<th>Ontario</th>
<th>Urban</th>
<th>Nonmajor Urban</th>
<th>Rural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level change 2</td>
<td>1.31*</td>
<td>1.25*</td>
<td>1.57*</td>
<td>0.61</td>
</tr>
<tr>
<td>(intervention\textsubscript{2})</td>
<td>(0.65, 1.97)</td>
<td>(0.60, 1.89)</td>
<td>(0.76, 2.38)</td>
<td>(-0.25, 1.47)</td>
</tr>
<tr>
<td>Trend change 2</td>
<td>0.035</td>
<td>0.021</td>
<td>0.021</td>
<td>0.046</td>
</tr>
<tr>
<td>(time after intervention\textsubscript{2})</td>
<td>(-0.031, 0.10)</td>
<td>(-0.046, 0.087)</td>
<td>(-0.042, 0.085)</td>
<td>(-0.012, 0.10)</td>
</tr>
</tbody>
</table>

\* Formulary change 2: change in ODB formulary listing from Limited Use to General Benefit, January 2007
* Significant at p < 0.05
Table 4.3 Results from regression analysis: slope estimates and 95% confidence intervals of proportion of new oral bisphosphonate formulations dispensed

<table>
<thead>
<tr>
<th></th>
<th>Regression Estimates</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Urban</td>
<td>Nonmajor Urban</td>
<td>Rural</td>
<td></td>
</tr>
<tr>
<td>Alendronate</td>
<td>0.58</td>
<td>(0.40, 0.75)</td>
<td>0.39</td>
<td>(0.25, 0.53)</td>
</tr>
<tr>
<td>+ Vitamin D₃</td>
<td>0.42</td>
<td>(0.12, 0.72)</td>
<td>0.36</td>
<td>(0.09, 0.63)</td>
</tr>
<tr>
<td>Monthly Risedronate</td>
<td>1.46</td>
<td>(1.26, 1.66)</td>
<td>1.51</td>
<td>(1.28, 1.74)</td>
</tr>
</tbody>
</table>
Figure 4.1 Identification of eligible oral bisphosphonate (BP) claims dispensed to community-dwelling seniors in Ontario (Refer to Appendix F for excluded formulations)
Figure 4.2 Identification of unique patients dispensed an eligible oral bisphosphonate (BP) and physicians who dispensed an eligible oral BP claim in Ontario (a Missing patient RIO, N=3,853)
Figure 4.3 Number of oral bisphosphonate claims dispensed to community-dwelling seniors in Ontario, January 2000 – March 2014

Figure 4.4 Number of alendronate (ALN) claims dispensed to community-dwelling seniors in Ontario, by formulation, January 2000 – March 2014
Figure 4.5 Number of risedronate (RSD) claims dispensed to community-dwelling seniors in Ontario, by formulation, January 2000 – March 2014

Figure 4.6 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in Ontario following formulary changes, January 2001 – December 2008 (N=7,579,128). Forecasting was completed for two years following each formulary change.
Figure 4.7 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in Ontario, by region, January 2001 – December 2008 (N=7,579,128). Forecasting was completed for two years following each formulary change.

Figure 4.8 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in urban regions following formulary changes, January 2001 – December 2008 (N=5,848,858). Forecasting was completed for two years following each formulary change.
**Figure 4.9** Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in nonmajor urban regions following formulary changes, January 2001 – December 2008 (N=1,369,623). Forecasting was completed for two years following each formulary change.

**Figure 4.10** Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in rural regions following formulary changes, January 2001 – December 2008 (N=360,647). Forecasting was completed for two years following each formulary change.
Figure 4.11 Proportion of alendronate + vitamin D₃ claims of total alendronate claims dispensed to community-dwelling seniors in Ontario, by region, January 2007 – March 2014 (N=4,961,458)

Figure 4.12 Proportion of monthly risedronate claims of total risedronate claims dispensed to community-dwelling seniors in Ontario, by region, June 2009 – March 2014 (N=4,033,497)
Figure 4.13 Proportion of risedronate delayed-release claims of total risedronate claims dispensed to community-dwelling seniors in Ontario, by region, February 2012 – March 2014 (N=1,870,993)
Chapter 5. Discussion

5.1 Summary of Results

Results demonstrated that oral bisphosphonate dispensing has increased over time and plateaued in recent years, which is consistent with prior studies examining osteoporosis medication utilization [30, 94, 95]. In addition, as second generation oral bisphosphonates became available via ODB formulary changes, there was an observed shift in utilization from etidronate to alendronate and risedronate. There were also monthly fluctuations in the number of oral bisphosphonate claims dispensed over time. Alendronate and risedronate claims showed similar fluctuations, with the highest peaks consistently occurring in October and the lowest drops in February. When examining trends by formulation, only weekly alendronate and weekly alendronate + vitamin D₃ showed similar patterns. Weekly risedronate also showed monthly drops in February, but only until 2011. This coincided with the availability of weekly risedronate delayed-release in 2012, however there was not enough data to determine whether risedronate delayed-release claims will follow the same fluctuations. The occurrences of these peaks and drops in dispensing coincides with seniors leaving Ontario during the coldest winter months for vacation, and thus may be attributed to requests for additional prescriptions prior to traveling. Etidronate claims also showed monthly fluctuations, however they were not consistent over the entire study period. This may be attributed to the 90-day supply interval of etidronate, with 14 tablets of etidronate taken once daily, and 76 daily tablets of calcium, thus multiple claims for etidronate are not required prior to travel.

In contrast to the first hypothesis of this thesis, results showed that rural regions did not respond at a significantly slower rate to the formulary changes. Indeed, temporal trends between regions were similar over time. Despite little observed urban-rural differences in trends of oral
bisphosphonate dispensing, there were significant regional differences. Rural areas always had the lowest proportion of dispensed claims compared to urban regions. This pattern was also observed when examining the proportion of new bisphosphonate formulation (alendronate + vitamin D$_3$, monthly risedronate, and risedronate delayed-release) claims dispensed in each region. In addition to dispensing a lower proportion, rural regions also consistently had a slower trend of dispensing of the new formulations. The regional differences observed in the proportion and trend of new formulation dispensing are indicative of a slower uptake and coincide with the study’s hypothesis that rural physicians may be less aware of new products than their urban counterparts. This regional gradient in dispensing may also indicate that rural physicians are slow adopters and lag in response to the introduction of new drugs to the market compared to urban regions.

Prior studies have identified formularies to be an effective means of controlling drug utilization [32-36], however the level and trend impacts of the formulary changes were only significant for some regions. In addition, all changes were very minor despite their significance, potentially due to the large sample size of claims. One reason for the lack of impact following formulary changes may be the availability of alendronate and risedronate for two years prior to the first formulary change in 2003, thus their use was already increasing prior to the formulary change. Indeed, visual representation of results showed that the proportion of alendronate/risedronate dispensed began to increase prior to the initial formulary change in April 2003. This increase began in late 2002, which also coincided with the release of updated osteoporosis clinical practice guidelines (November 2002) that officially recommended oral bisphosphonates as first-line therapy for osteoporosis management [31]. In addition, weekly formulations of alendronate and risedronate were also first listed on the ODB formulary in February (alendronate) and December (risedronate) 2002 [37]. Thus the release of the 2002 guidelines and the availability of
The weekly formulations may have acted as co-interventions and impacted dispensing trends. In addition, the increase may also be attributed to increased marketing efforts by pharmaceutical sales representatives in anticipation of formulary listing. Indeed, a recent study in Australia showed increased expenditure on ‘free’ doctor samples for risedronate prior to increased formulary access [13].

The lag in uptake and lower proportion of dispensing of new products in rural regions may be attributed to the fact that physicians in these areas have less exposure to pharmaceutical promotion [8], a reportedly large and influential source of new drug information. Following pharmaceutical promotion, academic and professional literature were reported as the next information source for new drugs [10]. However, research papers and journal articles do not convey new information as quickly [14] which may also contribute to the lag in rural dispensing. In addition, the lack of health services in rural regions may impact dispensing, as suggested by one study attributing differences in health services use to a higher concentration of specialist physicians located in urban centers [96]. Indeed, the results of this thesis showed much higher proportions of bisphosphonates dispensed by specialist physicians rather than general physicians in urban regions, yet a higher proportion of general practitioners dispensing in rural regions. This may indicate a higher proportion of specialist physicians in urban regions compared to rural regions, contributing to overall health service disparities between regions.

Alendronate + vitamin D₃ and monthly risedronate dispensing exhibited a regional gradient in dispensing, with three distinct slopes of dispensing: urban regions consistently dispensed the highest proportion, followed by nonmajor urban regions, and rural regions consistently had the lowest proportion dispensed. Dispensing of risedronate delayed-release in rural regions was also lowest overall, however unlike alendronate + vitamin D₃ and monthly risedronate patterns,
nonmajor urban dispensing was almost identical to urban dispensing. This may relate to the earlier availability of another risedronate delayed-release formulation manufactured by the same pharmaceutical company in the US. The US risedronate delayed-release formulation was approved and available in October 2010 [97], before risedronate delayed-release was listed on the formulary in Canada (February 2012). This may have influenced physician prescribing behaviour.

A unique decreasing trend of monthly risedronate claims was also observed over time, which occurred after weekly risedronate delayed-release became available on the ODB formulary. This observation may potentially indicate a preference for specific dosing regimens of risedronate. As described in Chapter 1, monthly dosing may improve adherence over weekly bisphosphonates. This may be due to the convenience and practicality of less frequent dosing intervals [45], however the literature shows conflicting information [98]. Some research has shown women to prefer the monthly dosage over the weekly regimen due to its simplicity while other studies have speculated that infrequent dosing of monthly formulations may actually result in skipped doses [99]. The results of this thesis indicate a possible preference for risedronate delayed-release (a weekly formulation) over monthly risedronate, which may be attributed to the novel combination of risedronate with a chelating compound that eliminates the need for fasting prior to administration. Monthly risedronate tablets must be taken on an empty stomach at least 30 minutes before consuming food, however risedronate delayed-release may be taken before or after a meal. This novel aspect increases its convenience despite a more frequent dosing interval [49]. It is also noted that the first generic formulation of monthly risedronate was listed on the ODB formulary in 2012, which may have also impacted dispensing patterns. Indeed, a recent analysis identified a switch from brand name formulations to the generic equivalent when it was
available on the formulary [100]. Switching between brand and generic formulations is beyond the scope of this thesis, yet this may be of interest for exploration in future analyses.

Although oral bisphosphonates are the most commonly dispensed drug for patients initiating therapy for osteoporosis [30], this study identified a plateau in oral bisphosphonate dispensing in recent years. The plateau may be due to the availability of new agents, such as zoledronic acid, an intravenous bisphosphonate formulation taken once a year [101], and denosumab, a twice-yearly injection [102]. Zoledronic acid was first available on the ODB under the Exceptional Access Program in November 2007, and has since been listed as Limited Use in September 2012 [37]; denosumab has been listed as Limited Use since February 2012 [37]. Physicians may be dispensing these newer agents instead of oral bisphosphonates, thus potentially contributing to the plateau in oral bisphosphonate dispensing. In addition, there has been a significant amount of concern from government and safety media safety advisories associating oral bisphosphonates with adverse events, such as osteonecrosis of the jaw [103] and atypical fractures [104], which may have also impacted the uptake of bisphosphonates.

5.2 Significance of Research

The results of this thesis project highlight differences between major urban, nonmajor urban, and rural dispensing of new oral bisphosphonate drugs, with rural physicians being slow adopters to new formulations. There were temporal differences between regions, with rural regions having a slower uptake of new formulations compared to urban and nonmajor urban regions. In addition, regional differences indicated a consistently lower proportion of new formulations dispensed in rural areas; these temporal and regional differences were persistent over time. It is difficult to conclude whether this is due to inequity in utilization of the drugs themselves, as the observed
regional variation may instead indicate disparities in dissemination of healthcare information between regions. The observed regional variation in dispensing of new therapies is an important finding as these differences may result in clinically significant differences in outcome. It is widely known that adherence to oral bisphosphonates is suboptimal, resulting in reduced drug effectiveness [105, 106]. Alendronate + vitamin D₃, monthly risedronate, and risedronate-delayed release have been developed to improve patient adherence to therapy and therefore patient outcomes. However if rural patients are not being dispensed these new products, adherence may not improve and may be significantly poorer compared to urban patients. This may translate into increased risk of outcomes such as fractures and related morbidity in patients residing in rural areas, and therefore actual health disparities between urban and rural regions. It is important to note that there is currently no evidence for improved fracture outcomes with the new bisphosphonate formulations. Thus more research is required to determine if regional differences in dispensing are result in differences in outcomes. If regional variation in outcomes is found to exist, the results of this thesis may provide relevant evidence-based material to target specific regions of Ontario to improve the standard of care.

Disparities in health services have also been described in osteoporosis management. For example, diagnosis of osteoporosis by BMD testing can positively influence the decision to initiate osteoporosis therapy [19, 94, 107]. However healthcare professionals in rural areas use BMD tests less frequently, are less likely to have local access, and have less confidence in its use than urban physicians [22, 108]. Furthermore, women residing in areas with access to BMD facilities had 2.6 to 2.9-fold higher odds of ever having undergone testing compared to women in areas with limited accessibility [22]. BMD testing machines are mostly located in urban areas [109], consistent with previous literature that document women residing in rural areas being less
likely to have had a BMD test [22, 23]; regional variation in BMD testing may also potentially impact treatment trends.

The results of this thesis research demonstrate that differential utilization of new oral bisphosphonate formulations exists between major urban, nonmajor urban, and rural areas, which may signal a potential need to increase academic detailing and dissemination of healthcare information to rural areas. This disparity may also exist in other aspects of osteoporosis management as outlined above. Thus, these results suggest that health inequities due to rurality should be addressed. Interventions targeted to specific areas of Ontario may improve overall management of osteoporosis.

5.3 Strengths and Limitations

There are several potential limitations with this thesis that must be considered. First, misclassification of location may occur, as patient (RPDB) and physician (IPDB) regional data are based on RIO scores in 2013 and 2011 respectively. Thus the regions for individual patients and physicians identified in this study may not necessarily reflect the patient’s residence or physician’s practice location at the time of dispensing. This thesis assumed that individuals did not move regions over time. Misclassification of physician location may also occur due to locum practice. All physicians are required to register with the Ontario Ministry of Health and Long-Term Care with their primary practice site address to receive a billing number to submit claims for insured services [110]. A locum is a physician who temporarily covers another physician’s clinical duties while they are away, and may practice in any geographic region [111]. Locum physicians must submit claims for services they provide using their own billing number rather than the billing number of the host physician [111]. Consequently, locum physicians may be
prescribing in a region that has a different RIO score than their billing address (often urban physicians practicing in rural areas), resulting in potential misclassification of physician location. Nineteen percent of family physicians across Canada provide locum services [112], however the majority (92%) of claims in this study had physician RIOs that matched the RIO of the patient receiving the prescription, therefore misclassification bias due to locum practice is expected to be minimal.

There are also a number of limitations associated with the databases used for this thesis. The ODB is limited to seniors (aged 65+ years) living in Ontario thus younger patients were not be captured in the analyses. However, since the risk for osteoporosis and osteoporotic fractures increases dramatically with age [18-20], it is expected that the majority of patients who should be treated for osteoporosis were captured and therefore analyzed in this study. In addition, only publically funded drug claims in Ontario are captured by the ODB, and thus drugs dispensed under private insurance or paid for out-of-pocket were not captured [30]. However, private insurance plans are often offered as an employee benefit and since the majority of seniors are likely retired, private insurance plans would likely account for few individuals in this age group [82]. Furthermore, the ODB database represents drug dispensing, which does not necessarily reflect what physicians initially prescribed to patients. Another limitation to using the ODB database is the inability to capture free drug samples provided by pharmaceutical representatives as part of product promotion, resulting in underestimation of the use of new drugs. Most pharmaceutical marketing occurs in urban regions, thus regional differences in the percentage of claims dispensed may be even more significant.

Since this study is ecological and assesses urban-rural differences at the population-level, it is difficult to make assumptions about individuals. For instance, it is difficult to disentangle
whether prescribing changes are due to physician or patient behaviour [13, 20]. For the purposes of this thesis, physicians were assumed to have decided what formulation was dispensed, but patient factors have been shown to influence physician prescribing [10]. Particular caution must also be taken when interpreting regional differences due to the lack of a universally accepted definition of rurality [81]. Depending on the definition used, Canada’s rural population can range from 22% to 38% of the country’s total population [2]. However, this study used the RIO to define urban, nonmajor urban, and rural regions, a classification often used by the Ontario Ministry of Health and Long-Term Care and the Ontario Medical Association to determine financial incentives for physician and specialists that are based on a community’s degree of rurality [84]. Other potential study limitations include the lack of a comparison group to control for potential effects from external interventions occurring at the same time as the studied intervention. Despite this, all interventions examined in this study occur province-wide and thus there are limited options for control groups.

This thesis project included the use of a large, well-validated administrative database to capture all oral bisphosphonate claims dispensed to community-dwelling seniors in Ontario and examined regional trends in dispensing. These results demonstrate that regional variation in bisphosphonate utilization exists, with declining proportions of claims as rurality increased (as measured by RIO scores). This is consistent with other studies examining prescribing differences between regions, for example, in statin use [113], prescribing of risperidone long acting injections [114], and treatment for chronic heart failure [115].

5.4 Future Directions

Oral bisphosphonates were used as a case example to study urban-rural differences in medication utilization. Since a regional gradient in dispensing was observed for newly marketed
formulations in osteoporosis drugs between urban and rural regions, similar trends may occur in other disease areas. Additional research exploring potential regional variation and its impact in other chronic disease areas is warranted.

Although urban-rural differences in utilization of new bisphosphonate formulations are observed and results suggest potential regional differences in adherence, this is beyond the scope of this project and warrants further examination into differences in adherence. Furthermore, there is not enough information to determine whether regional differences in treatment were appropriate, and whether it results in differences in osteoporosis-related outcomes. Additional research to determine the clinical significance is required as well.

A difference in utilization trends of monthly risedronate and risedronate delayed-release was also observed in this study, indicating a potential preference for specific bisphosphonate formulations. As previously mentioned, there is no conclusive evidence that physicians and patients prefer the delayed-release over the monthly formulation, or whether there are differences in their adherence or effectiveness. Future research to explore this is necessary.

5.5 Conclusions

This is the first study to examine urban-rural differences in utilization of oral bisphosphonates. Despite little differences in temporal trends between urban, nonmajor urban, and rural regions over time following formulary changes, regional differences were observed, with urban regions dispensing a higher proportion of claims compared to rural areas. Significant regional and temporal differences were observed when claims were examined by new formulations, with rural areas consistently dispensing a lower proportion and having a slower uptake of new
formulations. The results of this study indicate possible disparities in information dissemination to rural physicians, which may result in rural physicians being less aware of the availability of new drugs. Strategies to enhance academic detailing and dissemination of healthcare information to rural areas may be necessary. Further research examining the implications of urban-rural differences in drug utilization and its impact on adherence to treatment and clinical outcomes is warranted.
References


[22] Cadarette SM, Gignac MA, Jaglal SB, Beaton DE, Hawker GA. Access to osteoporosis treatment is critically linked to access to dual-energy x-ray absorptiometry testing. Med Care 2007;45:896-901.


[63] Lagarde M. How to do (or not to do) ... Assessing the impact of a policy change with routine longitudinal data. Health Policy Plan 2012;27:76-83.


Appendices

Appendix A. List of keyword terms used for MEDLINE® search

1. Time series.tw.
2. Time trend$.tw.
3. Trend analys$.tw.
4. Time series analys$.tw.
5. Forecast model$.tw.
6. Intervention analys$.tw.
7. Drug.mp.
8. Medicat$.mp.
11. Pharmacoepidemiolog$.mp.
12. Dispens$.mp.
13. Drug utili#ation.mp.
14. Or/1-6
15. Or/7-13
16. 14 and 15
17. Limit 16 to English language

Search terms 8 and 9 were adopted from Green et al [116]
Appendix B. List of identified references (methodological, reviews, and empirical papers)


B.11 Lagarde M. How to do (or not to do) ... Assessing the impact of a policy change with routine longitudinal data. Health Policy Plan 2012;27:76-83.


B.21 Austin PC, Mamdani MM. Impact of the pravastatin or atorvastatin evaluation and infection therapy-thrombolysis in myocardial infarction 22/Reversal of Atherosclerosis with Aggressive Lipid Lowering trials on trends in intensive versus moderate statin therapy in Ontario, Canada. Circulation 2005;112:1296-300.


and results of an education intervention to ensure appropriate use. Int Psychogeriatr 2005;17:631-52.


B.189 Vegter S, Kolling P, Toben M, Visser ST, de Jong-van den Berg LTW. Replacing hormone therapy—is the decline in prescribing sustained, and are nonhormonal drugs substituted? Menopause 2009;16:329-35.


Appendix C. List of references examining single institutional interventions


## Appendix D. Database information table

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
<th>Key Variables</th>
<th>Validity</th>
</tr>
</thead>
</table>
| ICES Physician Database (IPDB)  | - Includes detailed information about physicians in active practice in Ontario since 1992  
   - Comprises information from the Corporate Provider Database, The Ontario Physician Human Resource Data Centre Database, and the OHIP database of physician billings | - Physician demographics, specialty training and certification  
   - Physician practice location (by postal code) | - Data from the Corporate Provider Database encompasses approximately 80,000 individual providers in Ontario and approximately 8,000 organizations [81]  
   - Accuracy for physician address may be an issue, as physicians may work from one office and bill for services from another location (e.g., locum physicians) [81] |
| Ontario Drug Benefit (ODB)      | - Includes all drugs dispensed in community pharmacies and long-term care/nursing facilities  
   - Covers all seniors in Ontario (aged 65+ years) and those on social assistance for all prescriptions listed in the provincial formulary [24] | - Drug Identification Number of dispensed drug  
   - Date of prescription fill | - Second largest prescription database in Canada, representing approximately 40% of prescription drug spending [80]  
   - At least 95% of seniors filled one prescription in ODB over a five year period. However, 15-20% filled a prescription from a private insurer [82]  
   - High coding reliability overall error rate of 0.7% (95% CI: 0.5%-0.9%) [79]  
   - Drugs dispensed during hospitalizations are not captured |
| Registered Persons Database (RPDB) | - RPDB contains personal and demographic data for all current and previous Ontario residents who are registered with OHIP  
   - Linked by health number [78] | - Residential information (postal code)  
   - Health number  
   - Sex  
   - Date of birth  
   - Provider encrypted location  
   - Provider fiscal specialty | - Continuously updated  
   - May not contain up-to-date information on death status  
   - Approximately 11,000 residents have a missing or invalid postal code [78] |

Information summarized from the Ontario Ministry of Health and Long-Term Care [81] unless otherwise cited
Appendix E. Results for total oral bisphosphonate dispensing in Ontario, January 2000 to March 2014

Figure F.1 Proportion of alendronate and risedronate claims of total oral bisphosphonate claims dispensed to community-dwelling seniors in Ontario, January 2000 – March 2014 (N=16,367,752)
Appendix F. Non-osteoporosis oral bisphosphonate formulations excluded from analysis

<table>
<thead>
<tr>
<th>DIN</th>
<th>Drug Name</th>
<th>Product Name</th>
<th>Strength</th>
</tr>
</thead>
<tbody>
<tr>
<td>2258102</td>
<td>Alendronate</td>
<td>Co Alendronate</td>
<td>40 mg</td>
</tr>
<tr>
<td>2201038</td>
<td>Alendronate</td>
<td>Fosamax</td>
<td>40 mg</td>
</tr>
<tr>
<td>2239146</td>
<td>Risedronate Sodium</td>
<td>Actonel</td>
<td>30 mg</td>
</tr>
<tr>
<td>2298384</td>
<td>Risedronate Sodium</td>
<td>Novo-Risedronate</td>
<td>30 mg</td>
</tr>
<tr>
<td>2248686</td>
<td>Etidronate Disodium</td>
<td>Co Etidronate</td>
<td>200 mg</td>
</tr>
<tr>
<td>1997629</td>
<td>Etidronate Disodium</td>
<td>Didronel</td>
<td>200 mg</td>
</tr>
<tr>
<td>2245330</td>
<td>Etidronate Disodium</td>
<td>Mylan-Etidronate</td>
<td>200 mg</td>
</tr>
</tbody>
</table>

DIN: Drug Identification Number
Appendix G. Results using patient region data

**Figure G.1** Proportion of alendronate/risedronate claims of total oral bisphosphonate claims, by region, January 2000 – March 2008 (N=8,576,244)

**Figure G.2** Proportion of alendronate + vitamin D₃ claims of total alendronate claims, by region, January 2007 – March 2014 (N=5,490,696)
Figure G.3 Proportion of monthly risedronate claims of total risedronate claims, by region, June 2009 – March 2014 (N=4,431,544)

Figure G.4 Proportion of risedronate delayed-release claims of total risedronate claims, by region, February 2012 – March 2014 (N=2,050,211)
Appendix H. Results using physician region data for new oral bisphosphonate formulations, separated by brand and generic

**Figure H.1** Proportion of alendronate + vitamin D₃ claims of total alendronate claims dispensed to community-dwelling seniors in Ontario, by region, January 2007 – March 2014 (N=4,961,458)

**Figure H.2** Proportion of monthly risedronate claims of total risedronate claims dispensed to community-dwelling seniors in Ontario, by region, June 2009 – March 2014 (N=4,033,497)
Appendix I. Full segmented model regression estimates for objective 1

Table I.1 Results from full segmented regression model: relative changes in proportion of alendronate/risedronate dispensed following intervention 1\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>9.74</td>
<td>6.03, 13.45</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Time</td>
<td>1.022</td>
<td>0.89, 1.15</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Level change 1\textsuperscript{a} (intervention1\textsubscript{1})</td>
<td>2.43</td>
<td>0.60, 4.25</td>
<td>0.010</td>
</tr>
<tr>
<td>Trend change 1\textsuperscript{a} (time after intervention1\textsubscript{1})</td>
<td>0.16</td>
<td>-0.090, 0.41</td>
<td>0.21</td>
</tr>
<tr>
<td>Region</td>
<td>-1.78</td>
<td>-3.76, 0.21</td>
<td>0.081</td>
</tr>
<tr>
<td>Slope change 1\textsuperscript{a} (region*intervention1\textsubscript{1})</td>
<td>-2.13</td>
<td>-3.44, -0.81</td>
<td>0.0019</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Intervention 1: Change in Limited Use criteria, April 2003
\textsuperscript{b} Two interaction terms (listed below) were tested separately and rejected at an alpha level of 5%. Once one interaction term was significant, indicating geo-temporal slope differences, results were stratified by region. The estimates and 95% confidence intervals presented are for the first significant interaction term.

- Region*intervention1\textsubscript{1} (p=0.0019)
- Region*time after intervention1\textsubscript{1} (p=0.0001)

Table I.2 Results from full segmented regression model: relative changes in proportion of alendronate/risedronate dispensed following intervention 2\textsuperscript{a,b}

<table>
<thead>
<tr>
<th>Variable</th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>64.72</td>
<td>63.56, 65.88</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Time</td>
<td>0.54</td>
<td>0.47, 0.61</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Level change 2\textsuperscript{a} (intervention2\textsubscript{2})</td>
<td>1.17</td>
<td>0.43, 1.91</td>
<td>0.0023</td>
</tr>
<tr>
<td>Trend change 2\textsuperscript{a} (time after intervention2\textsubscript{2})</td>
<td>-0.0010</td>
<td>-0.13, 0.13</td>
<td>0.99</td>
</tr>
<tr>
<td>Region</td>
<td>3.34</td>
<td>2.82, 3.87</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Slope change 1\textsuperscript{a} (region*intervention2\textsubscript{2})</td>
<td>-0.021</td>
<td>-0.55, 0.50</td>
<td>0.94</td>
</tr>
</tbody>
</table>

\textsuperscript{a} Intervention 2: Change in ODB formulary listing from Limited Use to General Benefit, January 2007
\textsuperscript{b} Two interaction terms (listed below) were tested separately and rejected at an alpha level of 5%. Although both interaction terms were not significant, results were still stratified by region. The estimates and 95% confidence intervals presented are for the first interaction term.

- Region*intervention2\textsubscript{2} (p=0.94)
- Region*time after intervention2\textsubscript{2} (p=0.37)
Appendix J. Full regression estimates for objective 2

Table J.1 Results from regression analysis: changes in proportion of alendronate + vitamin D₃ dispensed

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>11.34</td>
<td>11.27, 11.40</td>
<td>0.0614</td>
</tr>
<tr>
<td>Time</td>
<td>0.58</td>
<td>0.51, 0.68</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Region</td>
<td>-1.69</td>
<td>-7.42, 4.031</td>
<td>0.5627</td>
</tr>
<tr>
<td>Slope change</td>
<td>-0.18</td>
<td>-0.21, -0.15</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

Table J.2 Results from regression analysis: changes in proportion of monthly risedronate dispensed

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>14.83</td>
<td>10.49, 19.17</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Time</td>
<td>0.41</td>
<td>0.34, 0.48</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Region</td>
<td>-1.68</td>
<td>-4.44, 1.082</td>
<td>0.2348</td>
</tr>
<tr>
<td>Slope change</td>
<td>-0.065</td>
<td>-0.11, -0.17</td>
<td>0.0085</td>
</tr>
</tbody>
</table>

Table J.3 Results from regression analysis: changes in proportion of risedronate delayed-release dispensed

<table>
<thead>
<tr>
<th></th>
<th>Estimate</th>
<th>95% Confidence Interval</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.88</td>
<td>-3.24, 9.00</td>
<td>0.3595</td>
</tr>
<tr>
<td>Time</td>
<td>1.32</td>
<td>1.78, 1.47</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Region</td>
<td>-3.61</td>
<td>-7.11, -0.10</td>
<td>0.0473</td>
</tr>
<tr>
<td>Slope change</td>
<td>0.028</td>
<td>-0.033, 0.089</td>
<td>0.3734</td>
</tr>
</tbody>
</table>
Appendix K. SAS code

libname home '/home/andrew/level3/racquel';
**************************************************************
/* CREATING VARIABLES FOR ENTIRE STUDY */
**************************************************************
data thesisall;
  length druggrp $ 9 pregion $ 9 mdregion $ 9;
  set home.cohort;
  Year=year(servdate);
  Month=month(servdate);
  Frequency=1;
/* creating druggrp variables */
  if din in ('02248728','02288087','02270129','02384701','02381486','02388545','02394863','02381478','
    02248727','02384698','02248251') then druggrp='adg';
  else if din in ('02201011','02233055') then druggrp='adb';
  else if din in('02275279','02288109','02248730','02258110','02261715','02273179','02284006','02286335','02299712
    ','02352966','02381494','02384728','02385031','02394871') then druggrp='awg';
  else if din in ('02248625','02245329') then druggrp='awb';
  else if din in ('02276429','02314940') then druggrp='awb_vitd3';
  else if din in ('02403641','02403633') then druggrp='awg_vitd3';
  else if din in ('02298376') then druggrp='rdg';
  else if din in ('02242518') then druggrp='rdb';
  else if din in ('02316838') then druggrp='eg';
  else druggrp='na';
/* creating patient region variables */
  if RIO_pt ge 45 then pregion='rural';
  else if RIO_pt in (10:44) then pregion='nonmajor';
  else if RIO_pt in (0:9) then pregion='urban';
  else pregion='missing';
/* creating physician region variables */
  if phys_RIO ge 45 then mdregion='rural';
  else if phys_RIO in (10:44) then mdregion='nonmajor';
  else if phys_RIO in (0:9) then mdregion='urban';
  else mdregion='missing';
run;
**************************************************************
/* INFO FOR STUDY FLOW DIAGRAM AND PT/MD CHARACTERISTICS*/
**************************************************************
/* all oral bisphosphonates 1996-2014 (Exclude: Non-oral bisphosphonates) */
data all_obp;
   set thesisall;
   where subclnam='BISPHOSPHONATES';
run;

data obp_post2000;
   set all_obp;
   where Year ge 2000;
run;

/* all oral bisphosphonates indicated for osteoporosis (excluded: non-OP indicated) */
data op_obp;
   set obp_post2000;
   where druggrp ne 'na';
run;

/* all community oral bisphosphonates (excluded: ltc claims) */
data obp_community;
   set op_obp;
   where ltc='0';
run;

/* all included prescriptions with complete md region (excluded: missing md region) */
data obp_mdregion;
   set obp_community;
   where mdregion ne 'missing';
run;

/* all included prescriptions with complete pt region (excluded: missing pt region) */
data obp_ptregion;
   set obp_community;
   where ptregion ne 'missing';
run;

/* number of claims with rural, nonmajor, urban md */
proc freq data=obp_mdregion;
   tables ptregion*mdregion / nopercent nocum norow nocol;
run;

/* number of claims with rural, nonmajor, and urban pt */
proc freq data=obp_ptregion;
   tables ptregion*mdregion / nopercent nocum norow nocol;
run;

/* mdregion = ptregion after applying above exclusion criteria */
data mdptmatch_incl;
   set obp_mdregion;
   if mdregion=ptregion;
   keep mdregion ptregion;
run;

/* ptregion=rural and mdregion ne rural after applying above exclusion criteria */
data ptrural_mdother;
   set obp_mdregion;
   if ptregion='rural' and mdregion ne 'rural';
run;
/* PHYSICIAN */

/* Examining missing physnum data */
data missphysnum;
  set obp_community;
  where physnum=' '; 
run;

/* if all physnum have mdregion = missing, then delete */
data eligibleclaims;
  set obp_community;
  if physnum=' ' then delete;
run;

/* Table 1: individual MD */
proc sql;
  create table table1_md as
    select Year, Month, physnum, mdregion, mainspecialty, phys_sex, phys_age, ptregion,
    sum(frequency) as SumFreq
    from eligibleclaims
    group by Year, Month;
quit;

/* population now includes all claims of eligible BP, in community, w/o missing physnum */
/* now need to remove duplicates to get individual md */

/* 1. Sort by date */
proc sort data=table1_md;
  by physnum;
run;

/* 2. Remove duplicate physnum */
proc sort data=table1_md nodupkey;
  by physnum;
run

/* 2. Examine mdregion by ptregion */
proc freq data=table1_md;
  tables ptregion*mdregion / nopercent nocum norow nocol;
run;

/* for table 1: renaming variables before proc freq for simplicity */
data table1_mdchar (rename=(phys_sex=mdsex));
  length mdspec $ 9;
  set table1_md;
  if mdregion='missing' then delete;

/* physician age */
  if phys_age in (35:44) then mdage='35-44';
  else if phys_age in (45:54) then mdage='45-54';
  else if phys_age in (55:64) then mdage='55-64';
  else if phys_age in (65:74) then mdage='65-74';
else if phys_age ge 75 then mdage='75+';
else if phys_age=. then mdage='miss';
else mdage='<35';
/* physician specialty */
if mainspecialty='GP/FP' then mdspec='general';
else if mainspecialty='GERIATRIC MEDICINE' then mdspec='geriatric';
else if mainspecialty='INTERNAL MEDICINE' then mdspec='internal';
else mdspec='other';
run;
***********************************************************************
*****************************************;
/* Frequencies for MD table 1 */

title 'Table 1: Physician age groups by MD region';
proc freq data=table1_mdchar;
   tables mdage*mdregion / nopercent nocum norow nocol;
run;

title;

title 'Table 1: Physician sex by MD region';
proc freq data=table1_mdchar;
   tables mdsex*mdregion / nopercent nocum norow nocol;
run;

title;

title 'Table 1: Physician specialty by MD region';
proc freq data=table1_mdchar;
   tables mdspec*mdregion / nopercent nocum norow nocol;
run;

title;

***********************************************************************
***************************
*************

/* PATIENT* /

/* examining missing ikn data */
data missikn;
   set obp_community;
   where ikn='';
run; /* result: no missing ikn therefore no need to delete any ikn */

/* Table 1: individual PT */
proc sql;
   create table table1_pt as
      select Year, Month, ikn, ptregion, age, sex, mdregion, sum(frequency) as SumFreq
      from obp_community
         group by Year,
                Month;
quit;

/* population now includes all claims of eligible BP, in community */
/* now need to remove duplicates to get individual pt */

/* 1. Sort by date */
proc sort data=table1_pt;
   by ikn;
run;

/* 2. Remove duplicate ikn */
proc sort data=table1_pt nodupkey;
by ikn;
run;

/* 3. Examine mdregion by ptregion */
proc freq data=table1_pt;
   tables ptregion*mdregion / nopercent nocum norow nocol;
run;

*********************************************************************
**
****************************************
*/
/* Renaming some variables before proc freq for simplicity */

data table1_ptchar (rename=(sex=ptsex));
   set table1_pt;
/* patient age */
   if age in (65:69) then ptage='65-69';
   else if age in (70:74) then ptage='71-74';
   else if age in (75:79) then ptage='75-79';
   else if age in (80:84) then ptage='80-84';
   else if age ge 85 then ptage='85+';
   else if age=. then ptage='miss';
   else ptage='<65';
/* exclusions */
   if mdregion='missing' then delete;
run;

***********************************************************************
****************************************
**
Frequencies for Table 1 */

title 'Table 1: Patient age groups by PT region';
proc freq data=table1_ptchar;
   tables ptage*ptregion / nopercent nocum norow nocol;
run;
title;

begin

title 'Table 1: Patient sex by PT region';
proc freq data=table1_ptchar;
   tables ptsex*ptregion / nopercent nocum norow nocol;
run;
title;

eND

/* Finding mean ages */

/* Physicians */

proc means data=table1_mdchar;
   var phys_age;
run;

proc means data=table1_mdchar;
   var phys_age;
   where mdregion='rural';
run;

proc means data=table1_mdchar;
   var phys_age;
   where mdregion='nonmajor';
run;
proc means data=table1_mdchar;
    var phys_age;
    where mdregion='urban';
run;

proc anova data=table1_mdchar;
    class mdregion;
    model phys_age = mdregion;
    means mdregion / tukey cldiff;
run;

******************************************************************************

Patients

********************************************************************************

proc means data=table1_ptchar;
    var age;
run;

proc means data=table1_ptchar;
    var age;
    where ptregion='rural';
run;

proc means data=table1_ptchar;
    var age;
    where ptregion='nonmajor';
run;

proc means data=table1_ptchar;
    var age;
    where ptregion='urban';
run;

proc anova data=table1_ptchar;
    class ptregion;
    model age = ptregion;
    means ptregion / tukey cldiff;
run;

******************************************************************************

Objective 1

******************************************************************************

data obj1;
    length bisphos2 $ 10;
    set thesisall;
    if druggrp in ('adg','adb','awg','awb','awb_vitd3','awg_vitd3') then bisphos='aln';
    else if druggrp in ('rdg','rdb','rwg','rwb','rwb_DR','rmg','rmb') then bisphos='rsd';
    else if druggrp in ('eb','eg') then bisphos='etd';
    else bisphos='none';
    if bisphos in ('aln','rsd') then bisphos2='firstline';
    else if bisphos = 'etd' then bisphos2='secondline';
    else delete;
    keep quantity bisphos2 Month Year servdate ltc mdregion ptregion frequency druggrp;
run;
/* Creating Time Series Tables */

/* RURAL MD COMMUNITY */

/* 1. Alendronate/Risedronate PROC SQL */
proc sql;
    create table obj1_sqlfirstrural as
        select Year, Month, sum(frequency) as SumFirst
        from obj1
        where bisphos2='firstline' and ltc='0' and mdregion='rural'
        group by Year, Month;
quit;

/* 2. Etidronate PROC SQL */
proc sql;
    create table obj1_sqlsecondrural as
        select Year, Month, sum(frequency) as SumSecond
        from obj1
        where bisphos2='secondline' and ltc='0' and mdregion='rural'
        group by Year, Month;
quit;

/* 3. Merging PROC SQL 1 + 2 */
data obj1rural;
    merge obj1_sqlfirstrural obj1_sqlsecondrural;
    by Year Month;
    if SumFirst=. then SumFirst=0;
    if SumSecond=. then SumSecond=0;
    SumAll=SumFirst+SumSecond;
    RateFirst=(SumFirst/SumAll)*100;
run;

/* Creating tables for ITSA */

/* Entire time period with both interventions included */
data obj1rural_time;
    set obj1rural;
    /* first intervention */
    if year lt 2001 then delete;
    if year le 2003 then intvn1=0;
    else if year gt 2003 then intvn1=1;
    if year eq 2003 and month gt 4 then intvn1=1;
    if intvn1=1 then TimeAfter1+1;
    else TimeAfter1=0;
    /* second intervention */
    if year le 2007 then intvn2=0;
    else if year gt 2007 then intvn2=1;
    if year eq 2007 and month gt 1 then intvn2=1;
    if intvn2=1 then TimeAfter2+1;
    else TimeAfter2=0;
    /* delete everything after 2009 */
    if year ge 2009 then delete;
    /* Time variable */
Time+1;
/* Variables for interaction terms */
mdregion=2;
/* for stationarity */
d_RateFirst=dif(RateFirst);
run;
/* First intervention only */
data obj1rural_time1;
  set obj1rural;
  /* first intervention */
  if year < 2001 then delete;
  if year >= 2006 then delete;
  if month >= 4 and year=2005 then delete;
  if year le 2003 then intvn1=0;
  else if year gt 2003 then intvn1=1;
  if year eq 2003 and month ge 4 then intvn1=1;
  if intvn1=1 then TimeAfter1+1;
  else TimeAfter1=0;
  /* Time variable */
  Time1+1;
  /* Variables for interaction terms */
  mdregion=2;
  /* for stationarity */
  d_RateFirst=dif(RateFirst);
run;
/* Second intervention only */
data obj1rural_time2;
  set obj1rural;
  /* delete first intervention */
  if year < 2005 then delete;
  /* second intervention */
  if year lt 2007 then intvn2=0;
  else if year ge 2007 then intvn2=1;
  if intvn2=1 then TimeAfter2+1;
  else TimeAfter2=0;
  /* delete everything after 2009 */
  if year ge 2009 then delete;
  /* Time variable */
  Time2+1;
  /* Variables for interaction terms */
  mdregion=2;
  /* for stationarity */
  d_RateFirst=dif(RateFirst);
run;
***********************************************************************
************************************************************
/* NONMAJOR MD COMMUNITY */
/* 1. Alendronate/Risedronate PROC SQL */
proc sql;
  create table obj1_sqlfirstnonmajor as
  select Year, Month, sum(frequency) as SumFirst
  from obj1
  where bisphos2='firstline' and ltc='0' and mdregion='nonmajor'
  group by Year, Month;
quit;
/* 2. Etidronate PROC SQL */
proc sql;
  create table obj1_sqlsecondnonmajor as
    select Year, Month, sum(frequency) as SumSecond
    from obj1
    where bisphos2='secondline' and ltc='0' and mdregion='nonmajor'
    group by Year, Month;
quit;
/* 3. Merging PROC SQL 1 + 2 */
data obj1nonmajor;
  merge obj1_sqlfirstnonmajor obj1_sqlsecondnonmajor;
  by Year Month;
  if SumFirst=. then SumFirst=0;
  if SumSecond=. then SumSecond=0;
  SumAll=SumFirst+SumSecond;
  RateFirst=(SumFirst/SumAll)*100;
run;
***********************************************************************
***********************
**********
*********************************************************
	/* Creating tables for ITSA */
***********************************************************************
****************************************
	/* Entire time period with both interventions included */
data obj1nonmajor_time;
  set obj1nonmajor;
  /* first intervention */
    if year lt 2001 then delete;
    if year le 2003 then intvn1=0;
    else if year gt 2003 then intvn1=1;
    if year eq 2003 and month gt 4 then intvn1=1;
    if intvn1=1 then TimeAfter1+1;
    else TimeAfter1=0;
  /* second intervention */
    if year le 2007 then intvn2=0;
    else if year gt 2007 then intvn2=1;
    if year eq 2007 and month gt 1 then intvn2=1;
    if intvn2=1 then TimeAfter2+1;
    else TimeAfter2=0;
  /* delete everything after 2009 */
    if year ge 2009 then delete;
  /* Time variable */
    Time+1;
  /* Variables for interaction terms */
    mdregion=1;
  /* for stationarity */
    d_RateFirst=dif(RateFirst);
run;
/* First intervention only */
data obj1nonmajor_time1;
  set obj1nonmajor;
  /* first intervention */
    if year lt 2001 then delete;
    if year ge 2006 then delete;
    if month ge 4 and year=2005 then delete;
    if year le 2003 then intvn1=0;
    else if year gt 2003 then intvn1=1;
    if year eq 2003 and month ge 4 then intvn1=1;
if intvn1=1 then TimeAfter1+1;
else TimeAfter1=0;

/* Time variable */
Time1+1;

/* Variables for interaction terms */
mdregion=1;

/* for stationarity */
d_RateFirst=dif(RateFirst);
run;

/* Second intervention only */
data obj1nonmajor_time2;
set obj1nonmajor;

/* delete first intervention */
if year lt 2005 then delete;

/* second intervention */
if year lt 2007 then intvn2=0;
else if year ge 2007 then intvn2=1;
if intvn2=1 then TimeAfter2+1;
else TimeAfter2=0;

/* delete everything after 2009 */
if year ge 2009 then delete;

/* Time variable */
Time2+1;

/* Variables for interaction terms */
mdregion=1;

/* for stationarity */
d_RateFirst=dif(RateFirst);
run;

***********************************************************************
****************************************
***************** 
****************************************************
***********************************************************************

/* URBAN MD COMMUNITY */

/* 1. Alendronate/Risedronate PROC SQL */
proc sql;
create table obj1_sqlfirsturban as
select Year, Month, sum(frequency) as SumFirst
from obj1
where bisphos2='firstline' and ltc='0' and mdregion='urban'
group by Year, Month;
quit;

/* 2. Etidronate PROC SQL */
proc sql;
create table obj1_sqlsecondurban as
select Year, Month, sum(frequency) as SumSecond
from obj1
where bisphos2='secondline' and ltc='0' and mdregion='urban'
group by Year, Month;
quit;

/* 3. Merging PROC SQL 1 + 2 */
data obj1urban;
merge obj1_sqlfirsturban obj1_sqlsecondurban;
by Year Month;
if SumFirst=. then SumFirst=0;
if SumSecond=. then SumSecond=0;
SumAll=SumFirst+SumSecond;
RateFirst=(SumFirst/SumAll)*100;
run;

title 'Rate of ALN/RSD prescribing in Urban, Community, MD';
proc print data=obj1urban;
run;
title;
 **********************************************************************
****** Creating tables for ITSA ******
**********************************************************************;
/* Entire time period with both interventions included */
data obj1urban_time;
  set obj1urban;
/* first intervention */
  if year lt 2001 then delete;
  if year le 2003 then intvn1=0;
  else if year gt 2003 then intvn1=1;
  if year eq 2003 and month gt 4 then intvn1=1;
  if intvn1=1 then TimeAfter1+1;
  else TimeAfter1=0;
/* second intervention */
  if year le 2007 then intvn2=0;
  else if year gt 2007 then intvn2=1;
  if year eq 2007 and month gt 1 then intvn2=1;
  if intvn2=1 then TimeAfter2+1;
  else TimeAfter2=0;
/* delete everything after 2009 */
  if year ge 2009 then delete;
/* Time variable */
  Time+1;
/* Variables for interaction terms */
  mdregion=0;
/* for stationarity */
  d_RateFirst=dif(RateFirst);
run;
/* First intervention only */
data obj1urban_time1;
  set obj1urban;
/* first intervention */
  if year lt 2001 then delete;
  if year ge 2006 then delete;
  if month ge 4 and year=2005 then delete;
  if year le 2003 then intvn1=0;
  else if year gt 2003 then intvn1=1;
  if year eq 2003 and month ge 4 then intvn1=1;
  if intvn1=1 then TimeAfter1+1;
  else TimeAfter1=0;
/* Time variable */
  Time1+1;
/* Variables for interaction terms */
  mdregion=0;
/* for stationarity */
  d_RateFirst=dif(RateFirst);
run;
/* Second intervention only */
data obj1urban_time2;
  set obj1urban;
  /* delete first intervention */
  if year lt 2005 then delete;
  /* second intervention */
  if year lt 2007 then intvn2=0;
  else if year ge 2007 then intvn2=1;
  if intvn2=1 then TimeAfter2+1;
  else TimeAfter2=0;
  /* delete everything after 2009 */
  if year ge 2009 then delete;
  /* Time variable */
  Time2+1;
  /* Variables for interaction terms */
  mdregion=0;
  /* for stationarity */
  d_RateFirst=d(RateFirst);
run;

***********************************************************************
****************************************;
/* Merge all regions together into one table */
***********************************************************************

/* Entire time period with both interventions included */
data obj1all_time;
  set obj1urban_time obj1nonmajor_time obj1rural_time;
  /* Variables for interaction terms */
  mdregion_intvn1=mdregion*intvn1;
  mdregion_TimeAfter1=mdregion*TimeAfter1;
  mdregion_intvn2=mdregion*intvn2;
  mdregion_TimeAfter2=mdregion*TimeAfter2;
run;

/* First intervention only */
data obj1all_time1;
  set obj1urban_time1 obj1nonmajor_time1 obj1rural_time1;
  /* Variables for interaction terms */
  mdregion_intvn1=mdregion*intvn1;
  mdregion_TimeAfter1=mdregion*TimeAfter1;
run;

/* Second intervention only */
data obj1all_time2;
  set obj1urban_time2 obj1nonmajor_time2 obj1rural_time2;
  /* Variables for interaction terms */
  mdregion_intvn2=mdregion*intvn2;
  mdregion_TimeAfter2=mdregion*TimeAfter2;
run;

***********************************************************************
***********************************************************************;
/* TESTING LINEAR REGRESSION ASSUMPTIONS */
***********************************************************************

/* RURAL intervention 1 */

/* 1a. Normality of residuals */
title "Rural 1: Normality of residuals";
proc reg data=obj1rural_time1;
  model RateFirst=Time1 intvn1 TimeAfter1 / clb;
  output out=residdata1 r=residual1;
run;

proc univariate data=residdata1 normal;
  var residual1;
  histogram;
 qqplot;
run;
title;

/* 1b. No influential outliers */

title 'Rural 1: No influential outliers';
proc reg data=obj1rural_time1;
  model RateFirst = Time1 intvn1 TimeAfter1 / influence;
  id Year Month;
  output out=diag1 df
        fits=cd_dffit1 rstudent=cd_rstudent1;
run;

proc print data=diag1;
  var Year Month cd_dffit1 cd_rstudent1;
  where cd_dffit1>2 or cd_rstudent1>2;
run;
title;

/* 2. Straight line relationship and 3. Homoscedasticity */

title 'Rural 1: Straight line relationship & Homoscedasticity';
proc reg data=obj1rural_time1;
  model RateFirst = Time1 intvn1 TimeAfter1 / partial;
  plot student.*predicted.;
run;
title;

/* 4. Independent error terms -- used DW test */
***********************************************************************
****************************************

/* RURAL intervention 2 */

/* 1a. Normality of residuals */

title 'Rural 2: Normality of residuals';
proc reg data=obj1rural_time2;
  model RateFirst=Time2 intvn2 TimeAfter2 / clb;
  output out=residdata2 r=residual2;
run;

proc univariate data=residdata2 normal;
  var residual2;
  histogram;
  qqplot;
run;
title;

/* 1b. No influential outliers */
title 'Rural 2: No influential outliers';
proc reg data=obj1rural_time2;
   model RateFirst = Time2 intvn2 TimeAfter2 / influence;
   id Year Month;
   output out=diag2 dffits=cd_dffit2 rstudent=cd_rstudent2;
run;
proc print data=diag2;
   var Year Month cd_dffit2 cd_rstudent2;
   where cd_dffit2>2 or cd_rstudent2>2;
run;
title;

/* 2. Straight line relationship and 3. Homoscedasticity */

/* nonmajor intervention 1 */

/* 1a. Normality of residuals */

/* 1b. No influential outliers */

/* 2. Straight line relationship and 3. Homoscedasticity */
title 'nonmajor 1: Straight line relationship & Homoscedasticity';
proc reg data=obj1nonmajor_time1;
    model RateFirst = Time1 intvn1 TimeAfter1 / partial;
    plot student.*predicted.;
run;
title;
/* 4. Independent error terms -- used DW test */
***********************************************************************
****************************************

/* nonmajor intervention 2 */
/* 1a. Normality of residuals */
title 'nonmajor 2: Normality of residuals';
proc reg data=obj1nonmajor_time2;
    model RateFirst=Time2 intvn2 TimeAfter2/ clb;
    output out=residdata2 r=residual2;
run;
proc univariate data=residdata2 normal;
    var residual2;
    histogram;
    qqplot;
run;
title;
/* 1a. Normality of residuals -- retest since found non-normality */
title 'Nonmajor 2: Normality of residuals (retest)';
proc autoreg data=obj1nonmajor_time2;
    model RateFirst=Time2 intvn2 TimeAfter2/ method=ml nlag=13 backstep dwprob;
    output out=residdata2 r=residual2;
run;
proc univariate data=residdata2 normal;
    var residual2;
    histogram;
    qqplot;
run;
title;
/* 1b. No influential outliers */
title 'nonmajor 2: No influential outliers';
proc reg data=obj1nonmajor_time2;
    model RateFirst = Time2 intvn2 TimeAfter2 / influence;
    id Year Month;
    output out=diag2 dfits=cd_dfhit2 rstudent=cd_rstudent2;
run;
proc print data=diag2;
    var Year Month cd_dfhit2 cd_rstudent2;
    where cd_dfhit2>2 or cd_rstudent2>2;
run;
title;
/* 2. Straight line relationship and 3. Homoscedasticity */
title 'nonmajor 2: Straight line relationship & Homoscedasticity ';
proc reg data=obj1nonmajor_time2;
   model RateFirst = Time2 intvn2 TimeAfter2 / partial;
   plot student.*predicted.;
run;
title;

/* 4. Independent error terms -- used DW test */

******************************************************************************
******************************************************************************

/* urban intervention 1 */

/* 1a. Normality of residuals */

title 'urban 1: Normality of residuals';
proc reg data=obj1urban_time1;
   model RateFirst=Time1 intvn1 TimeAfter1 / clb;
   output out=residdata1 r=residual1;
run;

proc univariate data=residdata1 normal;
   var residual1;
   histogram;
   qqplot;
run;
title;

/* 1a. Normality of residuals -- retest since found non-normality */

title 'Urban 1: Normality of residuals (retest)';
proc autoreg data=obj1urban_time1;
   model RateFirst=Time1 intvn1 TimeAfter1 / method=ml nlag=13 backstep dwprob;
   output out=residdata1 r=residual1;
run;

proc univariate data=residdata1 normal;
   var residual1;
   histogram;
   qqplot;
run;
title;

/* 1b. No influential outliers */

title 'urban 1: No influential outliers';
proc reg data=obj1urban_time1;
   model RateFirst = Time1 intvn1 TimeAfter1 / influence;
   id Year Month;
   output out=diag1 dffits=cd_dffit1 rstudent=cd_rstudent1;
run;

proc print data=diag1;
   var Year Month cd_dffit1 cd_rstudent1;
   where cd_dffit1>2 or cd_rstudent1>2;
run;
title;
/ 2. Straight line relationship and 3. Homoscedasticity /

title 'urban 1: Straight line relationship & Homoscedasticity';
proc reg data=obj1urban_time1;
   model RateFirst = Time1 intvn1 TimeAfter1 / partial;
   plot student.*predicted.;
run;
title;

/* 4. Independent error terms -- used DW test */

*******************************************************************************
****************************************
*/

/* urban intervention 2 */

/* 1a. Normality of residuals */

title 'urban 2: Normality of residuals';
proc reg data=obj1urban_time2;
   model RateFirst=Time2 intvn2 TimeAfter2/ clb;
   output out=residdata2 r=residual2;
run;

proc univariate data=residdata2 normal;
   var residual2;
   histogram;
   qqplot;
run;
title;

/* 1b. No influential outliers */

title 'urban 2: No influential outliers';
proc reg data=obj1urban_time2;
   model RateFirst = Time2 intvn2 TimeAfter2 / influence;
   id Year Month;
   output out=diag2 dffits=cd_dffit2 rstudent=cd_rstudent2;
run;

proc print data=diag2;
   var Year Month cd_dffit2 cd_rstudent2;
   where cd_dffit2>2 or cd_rstudent2>2;
run;
title;

/* 2. Straight line relationship and 3. Homoscedasticity */

proc reg data=obj1urban_time2;
   model RateFirst = Time2 intvn2 TimeAfter2 / partial;
   plot student.*predicted.;
run;
title;

/* 4. Independent error terms -- used DW test */

*******************************************************************************
*******************************************************************************

/* INTERRUPTED TIME SERIES MODELS WITH INTERACTION TERMS */
Title 'Obj1: All interaction terms';
proc autoreg data=obj1all_time outest=obj1all_time_out;
  model RateFirst=Time intvn1 TimeAfter1 intvn2 TimeAfter2 mdregion mdregion_intvn1 mdregion_TimeAfter1 mdregion_intvn2 mdregion_TimeAfter2/ method=ml nlag=13 backstep dwprob;
  output out=pred_mdall p=pvar r=rvar; /* forecasting */
run;
title;

/* Intervention 1 */

Title 'Interaction term added: mdregion_intvn1 mdregion_TimeAfter1';
proc autoreg data=obj1all_time outest=obj1all_time1_out;
  model RateFirst=Time1 intvn1 TimeAfter1 mdregion mdregion_intvn1 mdregion_TimeAfter1 / method=ml nlag=13 backstep dwprob;
  output out=pred_mdall p=pvar r=rvar; /* forecasting */
run;
title;

/* One interaction term at a time: mdregion_intvn1 */
Title 'Interaction term added: mdregion_intvn1';
proc autoreg data=obj1all_time outest=obj1all_time1a_out;
  model RateFirst=Time1 intvn1 TimeAfter1 mdregion mdregion_intvn1 / method=ml nlag=13 backstep dwprob;
  output out=pred_mdall p=pvar r=rvar; /* forecasting */
run;
title;

/* One interaction term at a time: mdregion_TimeAfter1 */
Title 'Interaction term added: mdregion_TimeAfter1';
proc autoreg data=obj1all_time outest=obj1all_time1b_out;
  model RateFirst=Time1 intvn1 TimeAfter1 mdregion mdregion_TimeAfter1 / method=ml nlag=13 backstep dwprob;
  output out=pred_mdall p=pvar r=rvar; /* forecasting */
run;
title;

/* Intervention 2 */

/* Both interaction terms for Intervention 2: mdregion_intvn2, mdregion_TimeAfter2 */
Title 'Interaction term added: mdregion_intvn2 mdregion_TimeAfter2';
proc autoreg data=obj1all_time outest=obj1all_time2_out;
  model RateFirst=Time2 intvn2 TimeAfter2 mdregion mdregion_intvn2 mdregion_TimeAfter2 / method=ml nlag=13 backstep dwprob;
  output out=pred_mdall p=pvar r=rvar; /* forecasting */
run;
title;

/* One interaction term at a time: mdregion_intvn2 */
Title 'Interaction term added: mdregion_intvn2';
proc autoreg data=obj1all_time outest=obj1all_time2a_out;
model RateFirst=Time2 intvn2 TimeAfter2 mdregion mdregion_intvn2 / method=ml nlag=13
   backstep dwprob;
output out=pred_mdall p=pvar r=rvar; /* forecasting */
run;
title;

/*/ One interaction term at a time: mdregion_TimeAfter2 */
title 'Interaction term added: mdregion_TimeAfter2';
proc autoreg data=obj1all_time2 outest=obj1all_time2b_out;
model RateFirst=Time2 intvn2 TimeAfter2 mdregion mdregion_TimeAfter2 / method=ml nlag=13
   backstep dwprob;
output out=pred_mdall p=pvar r=rvar; /* forecasting */
run;
title;

**********************************************************************
*                                                                 *
**********************************************************************

/* INTERRUPTED TIME SERIES – STRATIFIED BY REGION */
**********************************************************************

/* Rural */
title 'ITSA: Rural 1, Community, MD';
proc autoreg data=obj1rural_time1 outest=obj1rural_time1_out;
   model RateFirst=Time1 intvn1 TimeAfter1/ method=ml nlag=13 backstep dwprob;
   output out=pred_mdrural1 p=pvar r=rvar; /* forecasting */
run;
title;

title 'ITSA: Rural 2, Community, MD';
proc autoreg data=obj1rural_time2 outest=obj1rural_time2_out;
   model RateFirst=Time2 intvn2 TimeAfter2/ method=ml nlag=13 backstep dwprob;
   output out=pred_mdrural2 p=pvar r=rvar; /* forecasting */
run;
title;

**********************************************************************

/* Nonmajor */
title 'ITSA: Nonmajor Urban 1, Community, MD';
proc autoreg data=obj1nonmajor_time1 outest=obj1nonmajor_time1_out;
   model RateFirst=Time1 intvn1 TimeAfter1/ method=ml nlag=13 backstep dwprob;
   output out=pred_mdnonmajor1 p=pvar r=rvar; /* forecasting */
run;
title;

title 'ITSA: Nonmajor Urban 2, Community, MD';
proc autoreg data=obj1nonmajor_time2 outest=obj1nonmajor_time2_out;
   model RateFirst=Time2 intvn2 TimeAfter2/ method=ml nlag=13 backstep dwprob;
   output out=pred_mdnonmajor2 p=pvar r=rvar; /* forecasting */
run;
title;

**********************************************************************

/* Urban*/
title 'ITSA: Urban 1, Community, MD';
proc autoreg data=obj1urban_time1 outest=obj1urban_time1_out;
model RateFirst=Time1 intvn1 TimeAfter1 / method=ml nlag=13 backstep dwprob;
output out=pred_mdurban1 p=pvar r=rvar; /* forecasting */
run;
title;

title 'ITSA: Urban 2, Community, MD';
proc autoreg data=obj1urban_time2 outest=obj1urban_time2_out;
model RateFirst=Time2 intvn2 TimeAfter2/ method=ml nlag=13 backstep dwprob;
output out=pred_mdurban2 p=pvar r=rvar; /* forecasting */
run;
title;

*************************************************************************
*************************************************************************
*****************************************
**********************************************************************
*************************************************;
/* ONTARIO MD COMMUNITY */

/* 1. Alendronate/Risedronate PROC SQL */
proc sql;
  create table obj1_sqlfirstontario as
    select Year, Month, sum(frequency) as SumFirst
    from obj1
    where bisphos2='firstline' and ltc='0' and mdregion ne 'missing'
    group by Year, Month;
quit;

/* 2. Etidronate PROC SQL */
proc sql;
  create table obj1_sqlsecondontario as
    select Year, Month, sum(frequency) as SumSecond
    from obj1
    where bisphos2='secondline' and ltc='0' and mdregion ne 'missing'
    group by Year, Month;
quit;

/* 3. Merging PROC SQL 1 + 2 */
data obj1ontario;
  merge obj1_sqlfirstontario obj1_sqlsecondontario;
  by Year Month;
  if SumFirst=. then SumFirst=0;
  if SumSecond=. then SumSecond=0;
  SumAll=SumFirst+SumSecond;
  RateFirst=(SumFirst/SumAll)*100;
run;

*************************************************************************
*************************************************************************
*****************************************
**********************************************************************

/* Creating tables for ITSA */
*************************************************************************

/* Entire time period with both interventions included */
data obj1ontario_time;
  set obj1ontario;
/* first intervention */
  if year lt 2001 then delete;
  if year le 2003 then intvn1=0;
  else if year gt 2003 then intvn1=1;
  if year eq 2003 and month gt 4 then intvn1=1;
  if intvn1=1 then TimeAfter1+1;
  else TimeAfter1=0;
/* second intervention */
if year le 2007 then intvn2=0;
else if year gt 2007 then intvn2=1;
if year eq 2007 and month gt 1 then intvn2=1;
if intvn2=1 then TimeAfter2+1;
else TimeAfter2=0;
/* delete everything after 2009 */
if year ge 2009 then delete;
/* Time variable */
   Time+1;
run;

/* First intervention only */
data obj1ontario_time1;
   set obj1ontario;
*/ first intervention */
   if year lt 2001 then delete;
   if year ge 2006 then delete;
   if month ge 4 and year=2005 then delete;
   if year le 2003 then intvn1=0;
else if year gt 2003 then intvn1=1;
if year eq 2003 and month ge 4 then intvn1=1;
if intvn1=1 then TimeAfter1+1;
else TimeAfter1=0;
/* Time variable */
   Time1+1;
/* for stationarity */
   d_RateFirst=dif(RateFirst);
run;

/* Second intervention only */
data obj1ontario_time2;
   set obj1ontario;
*/ delete first intervention */
   if year lt 2005 then delete;
/* second intervention */
   if year lt 2007 then intvn2=0;
else if year ge 2007 then intvn2=1;
if intvn2=1 then TimeAfter2+1;
else TimeAfter2=0;
/* delete everything after 2009 */
if year ge 2009 then delete;
/* Time variable */
   Time2+1;
/* for stationarity */
   d_RateFirst=dif(RateFirst);
run;

==============================================*
/* TESTING LINEAR REGRESSION ASSUMPTIONS -- ONTARIO */

/* ontario intervention 1 */

/* 1a. Normality of residuals */

title 'ontario 1: Normality of residuals';
proc reg data=obj1ontario_time1;
   model RateFirst=Time1 intvn1 TimeAfter1/ clb;
   output out=residdata1 r=residual1;
run;

proc univariate data=residdata1 normal;
   var residual1;
   histogram;
   qqplot;
run;
title;

/* 1a. Normality of residuals -- retest since found non-normality */

title 'Ontario 1: Normality of residuals (retest)';
proc autoreg data=obj1ontario_time1;
   model RateFirst=Time1 intvn1 TimeAfter1 / method=ml nlag=13 backstep dwprob;
   output out=residdata1 r=residual1;
run;

proc univariate data=residdata1 normal;
   var residual1;
   histogram;
   qqplot;
run;
title;

/* 1b. No influential outliers */

title 'Ontario 1: No influential outliers';
proc reg data=obj1ontario_time1;
   model RateFirst = Time1 intvn1 TimeAfter1 / influence;
   id Year Month;
   output out=diag1 dffits=cd_dffit1 rstudent=cd_rstudent1;
run;

proc print data=diag1;
   var Year Month cd_dffit1 cd_rstudent1;
   where cd_dffit1>2 or cd_rstudent1>2;
run;
title;

/* 2. Straight line relationship and 3. Homoscedasticity */

title 'Ontario 1: Straight line relationship & Homoscedasticity';
proc reg data=obj1ontario_time1;
   model RateFirst = Time1 intvn1 TimeAfter1 / partial;
   plot student.*predicted.;
run;
title;

/* 4. Independent error terms -- used DW test */

***********************************************************************
************************************;

/* ontario intervention 2 */

/* 1a. Normality of residuals */

title 'Ontario 2: Normality of residuals';
proc reg data=obj1ontario_time2;
   model RateFirst=Time2 intvn2 TimeAfter2 / clb;
output out=residdata2 r=residual2;
run;

proc univariate data=residdata2 normal;
  var residual2;
  histogram;
  qqplot;
run;
title;

/* 1a. Normality of residuals -- retest since found non-normality */
title 'Ontario 2: Normality of residuals (retest)';
proc autoreg data=obj1ontario_time2;
  model RateFirst=Time2 intvn2 TimeAfter2/ method=ml nlag=13 backstep dwprob;
  output out=residdata2 r=residual2;
run;

proc univariate data=residdata2 normal;
  var residual2;
  histogram;
  qqplot;
run;
title;

/* 1b. No influential outliers */
title 'ontario 2: No influential outliers';
proc reg data=obj1ontario_time2;
  model RateFirst = Time2 intvn2 TimeAfter2 / influence;
  id Year Month;
  output out=diag2 dffits=cd_dffit2 rstudent=cd_rstudent2;
run;

proc print data=diag2;
  var Year Month cd_dffit2 cd_rstudent2;
  where cd_dffit2>2 or cd_rstudent2>2;
run;
title;

/* 2. Straight line relationship and 3. Homoscedasticity */
title 'ontario 2: Straight line relationship & Homoscedasticity ';
proc reg data=obj1ontario_time2;
  model RateFirst = Time2 intvn2 TimeAfter2 / partial;
  plot student.*predicted.;
run;
title;

/* 4. Independent error terms -- used DW test */

***********************************************************************
****************************************
*/
/* ITSA */
***********************************************************************
****************************************

/* Both interventions in one model */
title 'ITSA: Ontario, Community, MD';
proc autoreg data=obj1ontario_time outest=obj1ontario_time_out;
  model RateFirst=Time intvn1 TimeAfter1 intvn2 TimeAfter2/ method=ml nlag=13 backstep dwprob;
/* Intervention 1 */
title 'ITSA: Ontario 1, Community, MD';
proc autoreg data=obj1ontario_time1 outest=obj1ontario_time1_out;
    model RateFirst=Time1 intvn1 TimeAfter1 / method=ml nlag=13 backstep dwprob;
    output out=pred_mdontario p=pvar r=rvar; /* forecasting */
run;
title;

/* Intervention 2 */
title 'ITSA: Ontario 2, Community, MD';
proc autoreg data=obj1ontario_time2 outest=obj1ontario_time2_out;
    model RateFirst=Time2 intvn2 TimeAfter2 / method=ml nlag=13 backstep dwprob;
    output out=pred_mdontario p=pvar r=rvar; /* forecasting */
run;
title;

***********************************************************************
****************************************
***********************************************************************
/* FORECASTING using proc forecast procedure */
***********************************************************************
***********************************************************************

/* Ontario - intervention 1 */
data ontario_preddata1;
    set obj1ontario_time1;
    if intvn1=1 then delete;
run;

proc forecast data=ontario_preddata1 interval=month lead=24 out=ontario_pred1 outfull;
    var RateFirst;
run;

data ontariopred1_forecast ontariopred1_l95 ontariopred1_u95;
    set ontario_pred1;
    if _TYPE_='FORECAST' then output ontariopred1_forecast;
    else if _TYPE_='L95' then output ontariopred1_l95;
    else if _TYPE_='U95' then output ontariopred1_u95;
    where _LEAD_ ne 0;
run;

title 'Ontario Predicted 1 - Lower 95% CI Values';
proc print data=ontariopred1_l95;
run;
title;

title 'Ontario Predicted 1 - Forecasted Values';
proc print data=ontariopred1_forecast;
run;
title;

title 'Ontario Predicted 1 - Upper 95% CI Values';
proc print data=ontariopred1_u95;
run;
title;
/* Ontario - intervention 2 */

data ontario_preddata2;
  set obj1ontario_time2;
  if intvn2=1 then delete;
run;

proc forecast data=ontario_preddata2 interval=month lead=24 out=ontario_pred2 outfull;
  var RateFirst;
run;

data ontariopred2_forecast ontariopred2_l95 ontariopred2_u95;
  set ontario_pred2;
  if _TYPE_='FORECAST' then output ontariopred2_forecast;
  else if _TYPE_='L95' then output ontariopred2_l95;
  else if _TYPE_='U95' then output ontariopred2_u95;
  where _LEAD_ ne 0;
run;

title 'ontario Predicted 2 - Lower 95% CI Values';
proc print data=ontariopred2_l95;
run;
title;

title 'ontario Predicted 2 - Forecasted Values';
proc print data=ontariopred2_forecast;
run;
title;

title 'ontario Predicted 2 - Upper 95% CI Values';
proc print data=ontariopred2_u95;
run;
title;

/* Rural - intervention 1 */

data rural_preddata1;
  set obj1rural_time1;
  if intvn1=1 then delete;
run;

proc forecast data=rural_preddata1 interval=month lead=24 out=rural_pred1 outfull;
  var RateFirst;
run;

data ruralpred1_forecast ruralpred1_l95 ruralpred1_u95;
  set rural_pred1;
  if _TYPE_='FORECAST' then output ruralpred1_forecast;
  else if _TYPE_='L95' then output ruralpred1_l95;
  else if _TYPE_='U95' then output ruralpred1_u95;
  where _LEAD_ ne 0;
run;

title 'rural Predicted 1 - Lower 95% CI Values';
proc print data=ruralpred1_l95;
run;
title;
title 'rural Predicted 1 - Forecasted Values';
proc print data=ruralpred1_fore;
run;
title;
title 'rural Predicted 1 - Upper 95% CI Values';
proc print data=ruralpred1_u95;
run;
title;
***********************************************************************
****************************************
*/ Rural - intervention 2 */
data rural_preddata2;
   set obj1rural_time2;
      if intvn2=1 then delete;
run;
proc forecast data=rural_preddata2 interval=month lead=24 out=rural_pred2 outfull;
   var RateFirst;
run;
data ruralpred2_forecast ruralpred2_l95 ruralpred2_u95;
   set rural_pred2;
      if _TYPE_='FORECAST' then output ruralpred2_forecast;
      else if _TYPE_='L95' then output ruralpred2_l95;
      else if _TYPE_='U95' then output ruralpred2_u95;
      where _LEAD_ ne 0;
run;
title 'rural Predicted 2 - Lower 95% CI Values';
proc print data=ruralpred2_l95;
run;
title;
title 'rural Predicted 2 - Forecasted Values';
proc print data=ruralpred2_forecast;
run;
title;
title 'rural Predicted 2 - Upper 95% CI Values';
proc print data=ruralpred2_u95;
run;
title;
***********************************************************************
****************************************
***********************************************************************
usted}
data nonmajorpred1_forecast nonmajorpred1_l95 nonmajorpred1_u95;
  set nonmajor_pred1;
  if _TYPE_='FORECAST' then output nonmajorpred1_forecast;
  else if _TYPE_='L95' then output nonmajorpred1_l95;
  else if _TYPE_='U95' then output nonmajorpred1_u95;
  where _LEAD_ ne 0;
run;

title 'nonmajor Predicted 1 - Lower 95% CI Values';
proc print data=nonmajorpred1_l95;
run;
title;

title 'nonmajor Predicted 1 - Forecasted Values';
proc print data=nonmajorpred1_forecast;
run;
title;

title 'nonmajor Predicted 1 - Upper 95% CI Values';
proc print data=nonmajorpred1_u95;
run;
title;

******************************************************************************
************************************************************************************

/* Nonmajor - intervention 2 */

data nonmajor_preddata2;
  set obj1nonmajor_time2;
  if intv2=1 then delete;
run;

proc forecast data=nonmajor_preddata2 interval=month lead=24 out=nonmajor_pred2 outfull;
  var RateFirst;
run;

data nonmajorpred2_forecast nonmajorpred2_l95 nonmajorpred2_u95;
  set nonmajor_pred2;
  if _TYPE_='FORECAST' then output nonmajorpred2_forecast;
  else if _TYPE_='L95' then output nonmajorpred2_l95;
  else if _TYPE_='U95' then output nonmajorpred2_u95;
  where _LEAD_ ne 0;
run;

title 'nonmajor Predicted 2 - Lower 95% CI Values';
proc print data=nonmajorpred2_l95;
run;
title;

title 'nonmajor Predicted 2 - Forecasted Values';
proc print data=nonmajorpred2_forecast;
run;
title;

title 'nonmajor Predicted 2 - Upper 95% CI Values';
proc print data=nonmajorpred2_u95;
run;
title;

******************************************************************************
************************************************************************************
/* Urban - intervention 1 */

data urban_preddata1;
  set obj1urban_time1;
  if intvn1=1 then delete;
run;

proc forecast data=urban_preddata1 interval=month lead=24 out=urban_pred1 outfull;
  var RateFirst;
run;

data urbanpred1_forecast urbanpred1_l95 urbanpred1_u95;
  set urban_pred1;
  if _TYPE_='FORECAST' then output urbanpred1_forecast;
  else if _TYPE_='L95' then output urbanpred1_l95;
  else if _TYPE_='U95' then output urbanpred1_u95;
  where _LEAD_ ne 0;
run;

title 'urban Predicted 1 - Lower 95% CI Values';
proc print data=urbanpred1_l95;
run;

title 'urban Predicted 1 - Forecasted Values';
proc print data=urbanpred1_forecast;
run;

title 'urban Predicted 1 - Upper 95% CI Values';
proc print data=urbanpred1_u95;
run;

*******************************************************************************;

/* Urban - intervention 2 */

data urban_preddata2;
  set obj1urban_time2;
  if intvn2=1 then delete;
run;

proc forecast data=urban_preddata2 interval=month lead=24 out=urban_pred2 outfull;
  var RateFirst;
run;

data urbanpred2_forecast urbanpred2_l95 urbanpred2_u95;
  set urban_pred2;
  if _TYPE_='FORECAST' then output urbanpred2_forecast;
  else if _TYPE_='L95' then output urbanpred2_l95;
  else if _TYPE_='U95' then output urbanpred2_u95;
  where _LEAD_ ne 0;
run;

title 'urban Predicted 2 - Lower 95% CI Values';
proc print data=urbanpred2_l95;
run;

title;
title 'urban Predicted 2 - Forecasted Values';
proc print data=urbanpred2_forecast;
run;
title;

title 'urban Predicted 2 - Upper 95% CI Values';
proc print data=urbanpred2_u95;
run;
title;

*******************************************************************************
**************************************************
***********************************************************************
**************** OBJECTIVE 1 ANOVA TESTING ****************
***********************************************************************
**************************************************
*******************************************************************************

/* OBJECTIVE 1 ANOVA TESTING */
*******************************************************************************

/* Using ITSA tables created above */

data obj1anova;
  length mdregion $ 9;
  set obj1rural obj1nonmajor obj1urban;
  /* removing time points prior to availability (January 2000) and after plateau (December 2008) */
  if year lt 2001 then delete;
  if year gt 2008 then delete;
run;

proc anova data=obj1anova;
  class mdregion;
  model RateFirst = mdregion;
  means mdregion / tukey cldiff;
run;

*******************************************************************************
**************************************************
***********************************************************************
***********************************************************************

/* OBJECTIVE 2 */
*******************************************************************************

/* Objective 2 tables */
*******************************************************************************

data obj2;
  length bpform $ 10;
  set thesisall;
  if druggrp in ('adg','adb','awg','awb','awb_vitd3','awg_vitd3') then bisphos='aln';
  else if druggrp in ('rdg','rdb','rwg','rwb','rwb_DR','rmg','rmb') then bisphos='rsd';
  else if druggrp in ('eb','eg') then bisphos='etd';
  else bisphos='none';
  if druggrp in ('awb_vitd3','awg_vitd3') then bpform='aln_vitd3';
  else if druggrp in ('rmg','rmb') then bpform='rsd_month';
  else if druggrp='rwb_DR' then bpform='rsd_dr';
  else bpform='na';
  keep quantity bisphos bpform Month Year servdate ltc mdregion ptregion frequency druggrp;
run;

*******************************************************************************
***********************************************************************
***********************************************************************

/* Objectives 2 tables */
*******************************************************************************

152
/* ALN+VITD3 rural */

/* 1. Alendronate+VitD3 PROC SQL */

proc sql;
   create table obj2_sqlalnvitd3_rural as
      select Year, Month, sum(frequency) as SumALNVITD3
      from obj2
      where bpform='aln_vitd3' and ltc='0' and mdregion='rural'
      group by Year, Month;
   quit;

/* 2. Total alendronate PROC SQL */
proc sql;
   create table obj2_sqlaln_rural as
      select Year, Month, sum(frequency) as SumALN
      from obj2
      where bisphos='aln' and ltc='0' and mdregion='rural'
      group by Year, Month;
   quit;

/* 3. Merging PROC SQL 1 + 2 */
data obj2_alnvitd3_rural;
   merge obj2_sqlalnvitd3_rural obj2_sqlaln_rural;
      by Year Month;
   if SumALNVITD3=. then SumALNVITD3=0;
   if SumALN=. then SumALN=0;
   RateALNVITD3=(SumALNVITD3/SumALN)*100;
   mdregion='rural'; if Year lt 2007 then delete;
   mdregion2=2;
   Time+1;
run;

title 'Rate of ALN+VITD3 prescribing in Rural, Community, MD';
proc print data=obj2_alnvitd3_rural;
run;
title;
***********************************************************************
***************************
*************

/* ALN+VITD3 nonmajor */

/* 1. Alendronate+VitD3 PROC SQL */
proc sql;
   create table obj2_sqlalnvitd3_nonmajor as
      select Year, Month, sum(frequency) as SumALNVITD3
      from obj2
      where bpform='aln_vitd3' and ltc='0' and mdregion='nonmajor'
      group by Year, Month;
   quit;

/* 2. Total alendronate PROC SQL */
proc sql;
   create table obj2_sqlaln_nonmajor as
      select Year, Month, sum(frequency) as SumALN
      from obj2
      where bisphos='aln' and ltc='0' and mdregion='nonmajor'
      group by Year, Month;
   quit;

proc print data=obj2_sqlaln_nonmajor;
run;

/* 3. Merging PROC SQL 1 + 2 */
data obj2_alnvitd3_nonmajor;
  merge obj2_sqlalnvitd3_nonmajor obj2_sqlaln_nonmajor;
  by Year Month;
  if SumALNVITD3=. then SumALNVITD3=0;
  if SumALN=. then SumALN=0;
  RateALNVITD3=(SumALNVITD3/SumALN)*100;
  mdregion='nonmajor'; if Year lt 2007 then delete;
  mdregion2=1;
  Time+1;
run;

title 'Rate of ALN+VITD3 prescribing in nonmajor, Community, MD';
proc print data=obj2_alnvitd3_nonmajor;
run;
title;

***********************************************************************
*************************************************
/* ALN+VITD3 urban */

/* 1. Alendronate+VitD3 PROC SQL */
proc sql;
  create table obj2_sqlalnvitd3_urban as
  select Year, Month, sum(frequency) as SumALNVITD3
  from obj2
  where bpform='aln_vitd3' and ltc='0' and mdregion='urban'
  group by Year, Month;
quit;

/* 2. Total alendronate PROC SQL */
proc sql;
  create table obj2_sqlaln_urban as
  select Year, Month, sum(frequency) as SumALN
  from obj2
  where bisphos='aln' and ltc='0' and mdregion='urban'
  group by Year, Month;
quit;

/* 3. Merging PROC SQL 1 + 2 */
data obj2_alnvitd3_urban;
  merge obj2_sqlalnvitd3_urban obj2_sqlaln_urban;
  by Year Month;
  if SumALNVITD3=. then SumALNVITD3=0;
  if SumALN=. then SumALN=0;
  RateALNVITD3=(SumALNVITD3/SumALN)*100;
  mdregion='urban'; if Year lt 2007 then delete;
  mdregion2=0;
  Time+1;
run;

***********************************************************************
*************************************************
/* Merge */

data obj2_alnvitd3_all;
  set obj2_alnvitd3_urban obj2_alnvitd3_nonmajor obj2_alnvitd3_rural;
/* Variables for interaction terms */
/* Interaction terms */

```sas
mdregion_time=mdregion2*Time;
run;
```

```
/* Interaction terms */

title 'ALN+VitD3: Interaction term mdregion*Time';
proc autoreg data=obj2_alnvitd3_all outest=obj2_alnvitd3_all_out;
   model RateALNVITD3=Time mdregion2 mdregion_time / method=ml nlag=20 backstep dwprob;
run;
title;
```

```
/* Interaction term significant, therefore analyze separately */

title 'ALN+VitD3: Rural regression';
proc autoreg data=obj2_alnvitd3_rural outest=obj2_alnvitd3_rural_out;
   model RateALNVITD3=Time / method=ml nlag=13 backstep dwprob;
run;
title;
```

```
title 'ALN+VitD3: Nonmajor urban regression';
proc autoreg data=obj2_alnvitd3_nonmajor outest=obj2_alnvitd3_nonmajor_out;
   model RateALNVITD3=Time / method=ml nlag=13 backstep dwprob;
run;
title;
```

```
title 'ALN+VitD3: Urban regression';
proc autoreg data=obj2_alnvitd3_urban outest=obj2_alnvitd3_urban_out;
   model RateALNVITD3=Time / method=ml nlag=13 backstep dwprob;
run;
title;
```

***********************************************************************
***********************************************************************
***********************************************************************
```
/* RSD MONTH rural */
```
```
/* 1. RSD MONTH PROC SQL */
proc sql;
   create table obj2_sqlrsdmonth_rural as
      select Year, Month, sum(frequency) as SumRSDMONTH
      from obj2
         where bpform='rsd_month' and ltc='0' and mdregion='rural'
      group by Year, Month;
quit;
```
```
/* 2. Total RSD PROC SQL */
proc sql;
   create table obj2_sqlrsd_rural as
      select Year, Month, sum(frequency) as SumRSD
      from obj2
         where bisphos='rsd' and ltc='0' and mdregion='rural'
      group by Year, Month;
quit;
```
```
/* 3. Merging PROC SQL 1 + 2 */
data obj2_rsdmonth_rural;
   merge obj2_sqlrsdmonth_rural obj2_sqlrsd_rural;
      by Year, Month;
   if SumRSDMONTH=. then SumRSDMONTH=0;
```
```
if SumRSD=. then SumRSD=0;
RateRSDMONTH=(SumRSDMONTH/SumRSD)*100;
mdregion='rural';
mdregion2=2;
if Year lt 2009 then delete; if Year = 2009 and Month in (1:5) then delete;
Time+1;
run;
***********************************************************************
*****************************************;
/* RSD MONTH nonmajor */

/* 1. RSD MONTH PROC SQL */
proc sql;
create table obj2_sqlrsdmonth_nonmajor as
select Year, Month, sum(frequency) as SumRSDMONTH
from obj2
where bpform='rsd_month' and ltc='0' and mdregion='nonmajor'
group by Year, Month;
quit;

/* 2. Total RSD PROC SQL */
proc sql;
create table obj2_sqlrsd_nonmajor as
select Year, Month, sum(frequency) as SumRSD
from obj2
where bisphos='rsd' and ltc='0' and mdregion='nonmajor'
group by Year, Month;
quit;

/* 3. Merging PROC SQL 1 + 2 */
data obj2_rsdmonth_nonmajor;
merge obj2_sqlrsdmonth_nonmajor obj2_sqlrsd_nonmajor;
by Year Month;
if SumRSDMONTH=. then SumRSDMONTH=0;
if SumRSD=. then SumRSD=0;
RateRSDMONTH=(SumRSDMONTH/SumRSD)*100;
mdregion='nonmajor';
mdregion2=1;
if Year lt 2009 then delete; if Year = 2009 and Month in (1:5) then delete;
Time+1;
run;
***********************************************************************
*****************************************;
/* RSD MONTH urban */

/* 1. RSD MONTH PROC SQL */
proc sql;
create table obj2_sqlrsdmonth_urban as
select Year, Month, sum(frequency) as SumRSDMONTH
from obj2
where bpform='rsd_month' and ltc='0' and mdregion='urban'
group by Year, Month;
quit;
proc print data=obj2_sqlrsdmonth_urban;
run;

/* 2. Total RSD PROC SQL */
proc sql;
create table obj2_sqlrsd_urban as
select Year, Month, sum(frequency) as SumRSD
from obj2
where bisphos='rsd' and ltc='0' and mdregion='urban'
group by Year, Month;
quit;
/* 3. Merging PROC SQL 1 + 2 */
data obj2_rsdmonth_urban;
merge obj2_sqlrsdmonth_urban obj2_sqlrsd_urban;
by Year Month;
if SumRSDMONTH=. then SumRSDMONTH=0;
if SumRSD=. then SumRSD=0;
RateRSDMONTH=(SumRSDMONTH/SumRSD)*100;
mdregion='urban';
mdregion2=0;
if Year lt 2009 then delete; if Year = 2009 and Month in (1:5) then delete;
Time+1;
run;
***********************************************************************
****
************************************
***********************************************************************
***;
/* Merge */
data obj2_rsdmonth_all;
set obj2_rsdmonth_urban obj2_rsdmonth_nonmajor obj2_rsdmonth_rural;
/* Variables for interaction terms */
mdregion_time=mdregion2*Time;
run;
/* Interaction term */
title 'RSD Monthly: Interaction term mdregion*Time';
proc autoreg data=obj2_rsdmonth_all outest=obj2_rsdmonth_all_out;
   model RateRSDMONTH=Time mdregion2 mdregion_Time / method=ml nlag=13 backstep dwprob;
run;
title;
/* Interaction term significant, therefore analyze separately */
title 'RSD Monthly: Rural regression';
proc autoreg data=obj2_rsdmonth_rural outest=obj2_rsdmonth_rural_out;
   model RateRSDMONTH=Time / method=ml nlag=13 backstep dwprob;
run;
title;
title 'RSD Monthly: Nonmajor urban regression';
proc autoreg data=obj2_rsdmonth_nonmajor outest=obj2_rsdmonth_nonmajor_out;
   model RateRSDMONTH=Time / method=ml nlag=13 backstep dwprob;
run;
title;
title 'RSD Monthly: Urban regression';
proc autoreg data=obj2_rsdmonth_urban outest=obj2_rsdmonth_urban_out;
   model RateRSDMONTH=Time / method=ml nlag=13 backstep dwprob;
run;
title;
/* RSD DR rural */

/* 1. RSD DR PROC SQL */
proc sql;
    create table obj2_sqlrsddr_rural as
    select Year, Month, sum(frequency) as SumRSDDR
    from obj2
    where bpform='rsd_dr' and ltc='0' and mdregion='rural'
    group by Year, Month;
quit;

/* 2. Total RSD PROC SQL */
proc sql;
    create table obj2_sqlrsd_rural as
    select Year, Month, sum(frequency) as SumRSD
    from obj2
    where bisphos='rsd' and ltc='0' and mdregion='rural'
    group by Year, Month;
quit;

/* 3. Merging PROC SQL 1 + 2 */
data obj2_rsddr_rural;
    merge obj2_sqlrsddr_rural obj2_sqlrsd_rural;
    by Year Month;
    if SumRSDDR=. then SumRSDDR=0;
    if SumRSD=. then SumRSD=0;
    RateRSDDR=(SumRSDDR/SumRSD)*100;
    mdregion='rural';
    if Year lt 2012 then delete; if Year = 2012 and month = 1 then delete;
    Time+1;
    mdregion2=2;
run;

/* RSD DR nonmajor */

/* 1. RSD DR PROC SQL */
proc sql;
    create table obj2_sqlrsddr_nonmajor as
    select Year, Month, sum(frequency) as SumRSDDR
    from obj2
    where bpform='rsd_dr' and ltc='0' and mdregion='nonmajor'
    group by Year, Month;
quit;

/* 2. Total RSD PROC SQL */
proc sql;
    create table obj2_sqlrsd_nonmajor as
    select Year, Month, sum(frequency) as SumRSD
    from obj2
    where bisphos='rsd' and ltc='0' and mdregion='nonmajor'
    group by Year, Month;
quit;

/* 3. Merging PROC SQL 1 + 2 */
data obj2_rsddr_nonmajor;
    merge obj2_sqlrsddr_nonmajor obj2_sqlrsd_nonmajor;
    by Year Month;
    if SumRSDDR=. then SumRSDDR=0;
    if SumRSD=. then SumRSD=0;
    RateRSDDR=(SumRSDDR/SumRSD)*100;
    mdregion='nonmajor';
    if Year lt 2012 then delete; if Year = 2012 and month = 1 then delete;
    Time+1;
    mdregion2=2;
run;
by Year Month;
if SumRSDDR=. then SumRSDDR=0;
if SumRSD=. then SumRSD=0;
RateRSDDR=(SumRSDDR/SumRSD)*100;
mdregion='nonmajor';
if Year lt 2012 then delete; if Year = 2012 and month = 1 then delete;
Time+1;
mdregion2=1;
run;
***********************************************************************
****************************************
/* RSD DR urban */
/* 1. RSD DR PROC SQL */
proc sql;
create table obj2_sqlrsddr_urban as
    select Year, Month, sum(frequency) as SumRSDDR
    from obj2
    where bpform='rsd_dr' and ltc='0' and mdregion='urban'
group by Year, Month;
quit;
/* 2. Total RSD PROC SQL */
proc sql;
create table obj2_sqlrsd_urban as
    select Year, Month, sum(frequency) as SumRSD
    from obj2
    where bisphos='rsd' and ltc='0' and mdregion='urban'
group by Year, Month;
quit;
/* 3. Merging PROC SQL 1 + 2 */
data obj2_rsddr_urban;
    merge obj2_sqlrsddr_urban obj2_sqlrsd_urban;
    by Year Month;
    if SumRSDDR=. then SumRSDDR=0;
    if SumRSD=. then SumRSD=0;
    RateRSDDR=(SumRSDDR/SumRSD)*100;
    mdregion='urban';
    if Year lt 2012 then delete; if Year = 2012 and month = 1 then delete;
    Time+1;
    mdregion2=0;
run;
***********************************************************************
****************************************
/*/ * Merge */
data obj2_rsddr_all;
    set obj2_rsddr_urban obj2_rsddr_nonmajor obj2_rsddr_rural;
/* Variables for interaction terms */
    mdregion_time=mdregion2*Time;
run;
/* Interaction term */
title 'RSD Delayed-release: Interaction term mdregion*Time';
proc autoreg data=obj2_rsddr_all outest=obj2_rsddr_all_out;
    model RateRSDDR=Time mdregion2 mdregion_Time / method=ml nlag=13 backstep dwprob;
run;
/* Interaction term non-significant, but analyzing separately anyways */

title 'RSD Delayed-release: Rural regression';
proc autoreg data=obj2_rsddr_rural outest=obj2_rsddr_rural_out;
   model RateRSDDR=Time / method=ml nlag=13 backstep dwprob;
run;
title;

/* Nonmajor urban regression */
proc autoreg data=obj2_rsddr_nonmajor outest=obj2_rsddr_nonmajor_out;
   model RateRSDDR=Time / method=ml nlag=13 backstep dwprob;
run;
title;

/* Urban regression */
proc autoreg data=obj2_rsddr_urban outest=obj2_rsddr_urban_out;
   model RateRSDDR=Time / method=ml nlag=13 backstep dwprob;
run;
title;

***********************************************************************
************************************************************************
***********************************************************************
/* OBJECTIVE 2 - ANOVA TESTING (proportions) */
************************************************************************
***********************************************************************

/* ALENDRONATE + VITAMIN D3 */
data obj2_alnvitd3_anova;
   length mdregion $ 9;
   set obj2_alnvitd3_rural obj2_alnvitd3_nonmajor obj2_alnvitd3_urban;
   if year lt 2007 then delete;
run;

proc anova data=obj2_alnvitd3_anova;
   class mdregion;
   model RateALNVITD3 = mdregion;
   means mdregion / tukey cldiff;
run;

***********************************************************************
************************************************************************
***********************************************************************
/* RSD MONTHLY */
***********************************************************************

/* removing time points prior to availability (June 2009) */
if year lt 2009 then delete;
if year = 2009 and month in (1:5) then delete;
run;

proc anova data=obj2_rsdmonth_anova;
   class mdregion;
   model RateRSDMONTH = mdregion;
   means mdregion / tukey cldiff;
run;
run;

/******************************************************************************/
data obj2_rsddr_anova;
  length mdregion $ 9;
  set obj2_rsddr_rural obj2_rsddr_nonmajor obj2_rsddr_urban;
  /* removing time points prior to availability (FEBRUARY 2012) */
  if year lt 2012 then delete;
  if year = 2012 and month = 1 then delete;
run;

proc anova data=obj2_rsddr_anova;
  class mdregion;
  model RateRSDDR = mdregion;
  means mdregion / tukey cldiff;
run;

/******************************************************************************/

/******************************************************************************/
/* END :) */
/******************************************************************************/
/******************************************************************************/