NOTE TO USERS

Channelling in non-experimental pharmacoepidemiologic research: Its role in understanding confounding by indication

Lise M. Bjerre, MSc, MD

Department of Epidemiology and Biostatistics

McGill University, Montreal

August 2003

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Your file Votre référence
ISBN: 0-612-98209-2
Our file Notre référence
ISBN: 0-612-98209-2

"Statistical methods are adopted or,
if necessary,
invented to achieve
the epidemiologic objectives"

Kenneth J. Rothman

Table of contents

Dedication

To my unborn child,

whatever your path in Life

May God bless and keep you always,
May your wishes all come true,
May you always do for others
And let others do for you.
May you build a ladder to the stars
And climb on every rung.
...

May you grow up to be righteous,
May you grow up to be true,
May you always know the truth
And see the lights surrounding you.
May you always be courageous,
Stand upright and be strong.
...

May your hands always be busy,
May your feet always be swift,
May you have a strong foundation
When the winds of changes shift.
May your heart always be joyful,
May your song always be sung,
May you stay forever young.

Freely excerpted from "Forever Young"
Words and music by Bob Dylan

Acknowledgements

As the light at the end of the tunnel rapidly approaches, I reflect upon the time gone by and the events that have taken place since I embarked on this PhD degree adventure, almost eight years ago... The first thing that strikes me is that I had no idea then that I would be where I am now! So how did I get here? For those readers who know me well, you may wish to skip the "Prelude" and move to the acknowledgements proper.

Prelude

February 1995. In the throes of completing my Master's degree in epidemiology, I received a letter. Jubilation!: "Dear Ms. Bjerre, it is a pleasure for us to inform you that you have been accepted into McGill's combined MD-PhD program...". May 1995. Handsome and charming German pharmacist attends McGill University's Department of Epidemiology Summer School. Coup de foudre at the welcome reception. Life is great but things are getting complicated... 1995-1999. I attend / go through / survive / enjoy / despise being a medical student at McGill, while passing the PhD comprehensive exams (June '97) and sustaining a transatlantic relationship. Having decided that the timetable of the combined MD-PhD program is no longer compatible with my personal plans, I withdraw from the combined program, continue in the MD stream but also remain registered as a "half-time" PhD student, the net effect being that completion of the PhD is to take place after instead of in the middle of medical school. A wise decision if I ever made one. May 1999. Jubilation! Dr. Lise M. Bjerre, MSc, and now MD, graduates from McGill University Medical School. After much soul-searching and gut-wrenching during the last year of medical school, the situation has become clear: we (Christoph, the above-mentioned handsome and still charming teuton, and I) are moving to Germany. But before that, we take a detour through the autumn-coloured Laurentian mountains

for a wonderful weekend with family and friends and our wedding. Then it's off to the old continent - by boat, to dampen the shock.

To facilitate my integration, I decide to jump right in and take on a job as a resident and research fellow at the Department of Family Medicine of Göttingen University, where part of my time is to be 'protected' for PhD-related endeavours. Little did I know that my McGill Epidemiology training and my "native" English language skills would make me a highly desirable target for the most diverse assignments and requests. Add to that a full clinical workload as an intern, and very little time is left for PhD-related endeavours.

Nonetheless, the thesis committee meets, the project takes shape (several times, different shapes), a data-base is ordered, ethical approval obtained, a computer is bought and imported, via Barcelona... December 2001. All elements are finally in place for The Great Ascent of Thesis Mountain. Only one is missing. Time. Agonizingly, I decide to put clinical work aside temporarily to focus (almost) exclusively on The Ascent. In concrete terms, this means transformation from a beeper-carrying, stethoscope-wielding, sleep-deprived, adrenaline-driven internal medicine resident in a small hospital where being on call (6 to 8 times a month) means doing everything from running the Emergency Room and the Intensive Care Unit to caring for patients on the regular wards while also riding the ambulance through the surrounding countryside (overall rating: never easy, sometimes scary, sometimes lonely, never boring) to being a home-bound keyboard-leashed data junkie whose best friend and worst enemy is a ten-pound laptop computer with a fickle operating system and a hard-drive that's never quite large enough to digest the database it's being fed (overall rating: never easy, never scary, often lonely, sometimes boring). Not an easy transition, to say the least.

After six months of trying this out at home, I decide to protect my sanity and move to a "real" workplace, namely a disaffected office above my husband's pharmacy. This provides me with acceptable working conditions and freedom from extraneous distractions, until the carpenters moved in to transform the place into a "real" doctor's office. So I move three floors up to my mother-in-law's, more specifically to my husband's boyhood room and desk, when I am now putting these thoughts to paper.

Acknowledgements proper

Apart from the unpredictability of life, the other thing that strikes me when I think back over the past eight years is how important *other people* have been in my arriving where I am today by both shaping the circumstances of my life and helping me navigate them. It seems somewhat unfair that only I (assuming all goes well) will be awarded a doctorate, because I really never would have made it alone. Therefore, I have decided to award a few honorary doctorates of my own, as small tokens of recognition and gratitude. Here they are!

For their energy, support, creativity, patience and motivation, I award an honorary Doctorate of Outstanding Supervision to my co-supervisors, Dr. Jacques LeLorier and Dr. Michal Abrahamowicz, who, despite very active lives of their own, gave unreservedly of their time, attention and expertise. Without them this thesis could not have come into being.

Also to Dr. Jacques LeLorier, for his generous and continuous support (both intellectual and material) throughout, for insightful guidance, inspiring leadership and for almost seven years of productive and pleasant scientific collaboration, I would like to award a Doctorate of Exceptional Mentoring. I cannot think of anyone more deserving.

For ever cheerfully and patiently explaining the intricacies of SAS (Statistical Analysis Software) as well as for invaluable and prompt help with programming and data extraction issues, I would like to award an honorary Doctorate of Excellence in Statistical and Computer Consulting to Marc Dorais, research assistant in Dr. LeLorier's team.

For outstanding organizational, word-processing and interpersonal skills, for her patience and cheerfulness, I award a Doctorate of Superior Office (and personnel!) Management to Anita Massicotte, Dr. LeLorier's peerless secretary.

Many thanks also go to other members of his team for their company, stimulating discussions, cheerful support and for providing a pleasant work environment whenever I was in Montreal: Odile Sheehy, the late Jean-Marc Boucher, Lucie Blais and Elham Rahme. The latter two also became members of my thesis committee. For their availability, guidance, and insightful comments I am most grateful.

My gratitude also extends to the fifth and final member of my thesis committee. For supporting me in this – from a German perspective – somewhat exotic endeavour, for providing me with some time-out for PhD-related trips to Canada and for enabling me to rapidly integrate the German medical system, I would like to thank Prof. Michael Kochen, director of the Department of Family Medicine / Primary Care at the Georg-August University in Göttingen.

Last but not least, I turn to the people who, by their presence, encouragement, self-sacrifice, good humour and support, have sustained, pushed, nurtured and encouraged me throughout, and without whom all this would not have been possible. These are the people with whom I have the good fortune of sharing the so-called "private" sphere of my life.

First of all, I wish to thank my husband, Christoph Läer, for his undying support and belief in my ability to make it. He never gave up, even when I was ready to. For this, and for being himself a model of perseverance in the face of very difficult circumstances, I would like to award him a Doctorate of Perseverance and Constancy; both our home and his work are ample proof of his merit for this.

To my parents, for having had the emotional (and financial) generosity to support me through many years of university and for always being positive and optimistic, yet realistic about my projects, I thank you from the bottom of my heart and would like to award you a Doctorate of Selfless Parenting.

To my brother, for immense courage in the face of adversity, and for showing me what life is really worth, I thank you infinitely.

To my friends in Canada, with whom I have continued to be close despite the distance – Gillian Bartlett, Hélène Decaluwe, Alain Gagnon, Matthew Hodge, Andrew Katz, Nathalie Morissette – thank you. My special thanks to Sonia Tremblay for our life-long friendship.

My gratitude also goes out to my friends, relatives and acquaintances in Germany – in particular, Annette Becker, Bernhard Erdlenbruch, Christine Heine, Alice Trojok - who have helped smooth my transition into a new culture, and to my mother-in-law, Leonore Läer, for providing me with a welcoming environment, a second family, and now, a quiet workplace.

A very special thanks goes to my musical friends and band co-members, Stefan Könsgen and Norbert Remmert. Not only for enabling me to discover music in a new light and introducing me to a new musical instrument, but for their unwavering belief in us, the

band, or rather, the "hot-docs", as we are now called, for putting up with my periodic stints in Canada, for inspiring company and magical moments, for the chance to "tune out" and think about something completely different from a PhD thesis, and finally, for leading me to realize that yes, a surgeon, a dentist and a physician can actually play music on the street and make money! For all this, I thank you both *vom ganzen Herzen* – and I would like to award you each an honorary Doctorate of Musical Achievement.

Finally, I wish to thank my friend Matthew Hodge for editing this text and for reminding me that the light at the end of the tunnel is NOT – as it sometimes may seem – a train approaching head-on at high speed, but rather life after the PhD, which he tells me is even better than it seems from a distance. I look forward to it!

Lise M. Bjerre

June 2003
Hildesheim (Germany)

Financial acknowledgements

This project was funded in part by a grant from the "Réseau FRSQ sur l`utilisation des médicaments" of the "Fonds de recherché en santé du Québec" (FRSQ) of the government of the province of Québec. Financial support for several transatlantic trips, conference registration fees, statistical and secretarial support was generously provided by Dr Jacques LeLorier through discretionary funds. Additional discretionary funds were generously provided by my husband and my parents.

Abstract

Background

When a new drug is introduced onto the market, patients treated with it often tend to be sicker than those who are not. Once a drug has been released onto the market, its safety and efficacy can usually only be assessed using non-experimental studies. However, these methods cannot account for the differential assignment of sicker patients to new drugs, which can make a new drug seem detrimental or even dangerous, if these systematic differences are not accounted for. This is a prototypical example of confounding by indication, which is often viewed as an insurmountable problem in pharmacoepidemiologic research. Confounding by indication consists of two components, namely channelling and the risk factor effect. The risk factor effect, i.e. the association between potential confounders and the outcome, is often known *a priori* and can be quantified from previous studies. Typically, this is not the case with channelling, which is the non-random assignment of patients to one or the other treatment alternatives depending on both documented and undocumented characteristics.

Objectives

To develop methods for the assessment of channelling in non-experimental research, with a view to identifying, and possibly correcting for, confounding by indication in data-base studies for post-marketing drug surveillance.

Methods

In this thesis, the index of apparent channelling (IAC) is introduced as a novel tool for the measurement of the degree of channelling. The IAC makes use of propensity scores to quantify the proportion of the variance in treatment assignment that can be

accounted for by documented patient characteristics. However, the IAC can only account for channelling due to documented factors. Thus, it is always possible that there be residual channelling due to undocumented factors. Such residual channelling is of concern mostly to the extent that it leads to confounding of the treatment effect. Consequently, the index of residual channelling (IRC) is developed to estimate residual channelling and a novel approach is proposed to assess the strength of the resulting confounding bias. This model-based approach is based on estimating the interaction between treatment effect and the expected strength of residual channeling on treatment assignments of individual patients, measured by the discrepancy between their predicted treatment and the treatment they actually received.

Results

Using simulated as well as empirical examples, the IAC was shown to discriminate well between various degrees of channelling due to documented factors. The IRC was found to be responsive to the presence of residual channelling, but unable to differentiate between systematic residual channelling and random processes, presumably reflecting individual physicians'/patients' subjective preferences for a given treatment. On the other hand, the empirical model-based approach to assess the magnitude of confounding bias due to residual channelling was shown to be practicable as well as able to differentiate between residual channelling and quasi-random treatment assignment.

Conclusion

The combination of the index of apparent channelling and the model-based approach to residual channelling provides a practical approach to the problem of assessing the impact of confounding by indication in non-experimental studies in the post-marketing evaluation of the safety and efficacy of new drugs. The application of these methods

may enhance the validity of the conclusions drawn in such studies.

Résumé

Introduction

Quand un nouveau médicament est mis en marché, les patients qui le reçoivent comme traitement ont tendance à être plus malades que ceux à qui on prescrit des médicaments déjà existants. Suite à l'introduction d'un nouveau médicament, son efficacité et sa sûreté ne peuvent habituellement être mesurés qu'au moyen d'études non-expérimentales. Cependant, ces méthodes ne tiennent pas compte du fait que les patients plus malades ont une plus forte probabilité de se faire prescrire des médicaments nouveaux, donnant l'impression que ces derniers sont nocifs ou mêmes dangereux, si on ne tient pas compte des différences systématiques entre les deux groupes de patients. Cette situation constitue un exemple typique de biais d'indication, et est généralement perçue comme un obstacle insurmontable en recherche pharmacoépidémiologique. Le biais d'indication est composé de deux éléments, soit la canalisation et l'effet du facteur de risque. Ce dernier, étant l'association entre des facteurs confondants et l'issue finale, est souvent connu a priori et peut être quantifié à l'aide d'études existantes. Cependant, ceci n'est pas possible avec la canalisation, qui consiste en une attribution non-aléatoire des patients à des traitements différents dépendant des caractéristiques documentées et non documentées.

Objectifs

Développer des méthodes d'évaluation de la canalisation en recherche non-expérimentale, dans le but d'identifier et possiblement de corriger le biais d'indication présent dans des études post-marketing de surveillance de médicaments faisant usage de bases de données.

Méthodes

Dans la présente dissertation, l'indice de canalisation apparente (ICA) est présenté en tant qu'outil nouveau permettant la mesure de l'ampleur de la canalisation présente. L'ICA fait usage de scores de propensité afin de quantifier la portion de la variance de l'assignation des traitements qui est causée par des caractéristiques chez les patients qui sont documentées. Cependant, il est toujours possible d'être en présence d'une canalisation résiduelle causée par des facteurs non-documentés. Cette canalisation résiduelle est pertinente dans la mesure où elle induit une incertitude quand à l'effet des traitements. En conséquence, un indice de canalisation résiduelle (ICR) a été développé afin d'estimer la canalisation résiduelle, et une nouvelle approche modélisatrice est proposée pour évaluer l'ampleur du biais d'indication en découlant. Cette approche est basée sur une estimation de l'interaction entre l'effet du traitement et l'ampleur de la canalisation résiduelle telle que mesurée par la différence entre le traitememt prévu et le traitement reçu.

Résultats

Faisant usage d'exemples simulés et empiriques, l'ICA permet de bien discriminer entre divers degrés de canalisation causée par des facteurs documentés. L'ICR s'est avéré sensible à la présence de canalisation résiduelle, mais incapable de différencier entre une canalisation résiduelle systématique et des processus aléatoires, reflétant probablement des préférences subjectives individuelles entre médecins et patients. Cependant, l'approche modélisatrice empirique visant à évaluer l'amplitude du biais d'indication causée par la canalisation résiduelle s'est avérée appliquable ainsi que capable de différencier entre de la canalisation résiduelle et une assignation quasi-

aléatoire des traitements.

Conclusion

La combinaison de l'indice de canalisation apparente et de l'approche modélisatrice visant à évaluer la canalisation résiduelle procure une approche pratique au problème d'évaluation du biais d'indication présent dans les études non-expérimentales de sûreté et d'efficacité de nouveaux médicaments. L'usage de ces méthodes pourrait rehausser la validité des conclusions tirées de telles études.

Traduit de l'anglais par Nikolas Bjerre.

PREFACE

Notes on manuscript-based thesis

This thesis was written as a collection of manuscripts to be submitted for publication. The following section is quoted from the Faculty of Graduate Studies and Research's 'Guidelines for Submitting a Doctoral Thesis' (revised June 2000):

"As an alternative to the traditional thesis format, the dissertation can consist of a collection of papers that have a cohesive, unitary character making them a report of a single program of research. The structure for the manuscript-based thesis must conform to the following:

Candidates have the option of including, as part of the thesis, the text of one or more papers submitted, or to be submitted, for publication, or the clearly duplicated text (not reprints) of one or more published papers. Theses texts must conform to the 'Guidelines for Thesis Preparation' with respect to font size, line spacing and margin sizes and must be bound together as an integral part of the thesis.

The thesis must be more than a collection of manuscripts. All components must be integrated into a cohesive unit with logical progression from one chapter to the next. In order to ensure that the thesis has continuity, connecting texts that provide logical bridges between the different papers are mandatory.

The thesis must conform to all other requirements of the 'Guidelines for Thesis Preparation' in addition to the manuscripts. The thesis must include the following: (a) table of contents; (b) an abstract in English and French; (c) an introduction which clearly states the rationale and objectives of the research; (d) a comprehensive review

of the literature (in addition to that covered in the introduction to each paper); (e) a final conclusion and summary.

As manuscripts for publication are frequently very concise documents, where appropriate, additional material must be provided (e.g., in appendices) in sufficient detail to allow a clear and precise judgement to be made of the importance and originality of the research reported in the thesis.

In general, when co-authored papers are included in a thesis the candidate must have made a substantial contribution to all papers included in the thesis. In addition, the candidate is required to make an explicit statement in the thesis as to who contributed to such work and to what extent. This statement should appear in a single section entitled 'Contribution of Authors' as a preface to the thesis. The supervisor must attest to the accuracy of this statement at the doctoral oral defense."

In case of conflict between journal-specific formatting and thesis guidelines, I have followed the McGill conventions for thesis preparation. Manuscripts included in this thesis have been or will be modified to conform to journal criteria before submission for publication.

Contribution of authors

As PhD candidate and first author of all the manuscripts, I was primarily responsible for conceptualizing, designing and carrying out the research described in this thesis. The overall scope of this research was determined together by myself, Dr. Jacques LeLorier and Dr. Michal Abrahamowicz, based in large part on earlier reflections and ideas of Dr. Jacques LeLorier. Further conceptualization and operationalization of the original concepts were carried out in consultation with members of the thesis committee, in

particular Dr. Lucie Blais and Dr. Elham Rahme.

Based on the literature review that I conducted and wrote and on numerous discussions with Drs LeLorier and Abrahamowicz, I developed the specific study objectives for the three manuscripts, conceived of the formulae for the indices of apparent and residual channelling and constructed hypothetical examples to illustrate the behaviour of these indices. The study cohorts used in the empirical examples were extracted from the raw Régie de l'assurance maladie du Québec (RAMQ) data by Marc Dorais, MSc. I validated the data, performed the statistical analyses, and wrote the manuscripts as well as all other sections of the thesis. Invaluable help with SAS programming was provided by Marc Dorais at various stages along the way and I am very grateful for it.

Members of the thesis committee who are listed as co-authors provided specific feedback on their respective areas of expertise. Dr. LeLorier provided conceptual guidance and inspiration as well as clinical expertise. Dr. Abrahamowicz provided methodological and statistical expertise.

As PhD candidate, I take full responsibility for the scientific quality of the research, the originality of the ideas and the accuracy of the data contained in this thesis.

Statement of originality

The research in this thesis constitutes original scholarship and advances knowledge in the domain of non-experimental applied medical research in general and of data-based pharmacoepidemiologic research in particular.

In the first manuscript, based on some previous work by Dr. LeLorier, a conceptual framework for channelling was delineated and the index of apparent channelling (IAC)

was proposed as a new tool to identify and quantify apparent channelling, an important component of confounding by indication in non-experimental research that is due to documented factors. The index of apparent channelling provides a readily applicable tool for investigators to check for channelling in their own databases. The behaviour of the index of apparent channelling was illustrated in the first manuscript using hypothetical examples, which demonstrated that the index behaved as expected.

In the second manuscript, I focused on particular empirical situations, where different degrees of channelling were expected, in order to illustrate the application and behaviour of the index of channelling to these situation. The IAC was shown to be responsive to the presence of apparent channelling also when applied to these empirical examples

In the third manuscript, I focused on the issue of residual channelling, which is channelling due to undocumented factors. An empirical approach to the identification of residual channelling was developed, partly based on an original idea by Dr. Abrahamowicz. This approach was applied to empirical examples and shown to be practicable.

Overall, the present body of work provides an original step toward addressing the issue of confounding by indication, one of the fundamental problems of non-experimental research on drug utilisation. It provides researchers with an easily applicable tool for the assessment of different aspects of channelling in data-based research. In the present thesis, application of these novel methods yielded some new insights. For example, in Manuscript 3, we demonstrated how residual confounding could lead to an important under-estimation of the protective effect of COX-2 inhibitors. Furthermore, I

expect that the methods and conceptual contributions of this thesis might stimulate further research on new, more refined models to deal with the analytical challenges of non-experimental studies of medications. It is my hope that future applications of these methods in different areas of pharmacoepidemiology will contribute in a small way to making future research results more valid, therefore ultimately contributing to more informed decisions in clinical management, and better patient care.

1 Introduction

1.1 Background

The efficacy and safety of drugs is an issue of prime importance to patients, physicians, health-care institutions, third-party payers, pharmaceutical manufacturers and regulatory agencies. Although the requirements for drug licensing have become more stringent around the world over the past half-century (1), these requirements cannot guarantee that drugs are completely safe, nor that they are efficacious under all circumstances (2). Trials conducted by pharmaceutical manufacturers for purposes of regulatory approval are often relatively small in size, typically up to 5000 patients, which seriously limits the statistical power to assess rare outcomes such as adverse events. They also tend to be of short duration, usually 1-2 years, and they focus on highly selected, usually homogenous groups of patients, often excluding women, children, the elderly, patients with co-morbid conditions, etc. (2;3). Consequently, there is a knowledge gap that needs to be bridged. To do so, it is necessary to conduct long-term post-marketing surveillance to assess both the safety and the efficacy of drugs after they have been released onto the market (2;3).

Post-marketing surveillance studies are, by definition, non-experimental studies, where the choice of therapy is dictated by clinicians' judgement, partly based on patient characteristics, rather than for the purpose of gaining scientific information (3;4) A growing trend toward linking large administrative health databases offers new exciting opportunnites for such studies. These studies present a particular challenge in terms of validity, because factors associated with the choice of treatment (such as the presence or severity of a condition) may also be associated with the outcome of interest. Such

factors – or potential confounders – may artefactually create a seemingly causal association or, on the contrary, mask a real one, a phenomenon called confounding by indication (4). The challenge in assuring the validity of non-experimental study results is to separate the effects due to confounding from the causal association (5). Some approaches have been proposed for dealing with confounding by indication, however, not all of them are applicable to large drug claims databases and none has so far provided a satisfactory way of dealing with unknown or undocumented confounders. Some researchers view this as an impossible task, the Achilles' heel of pharmacoepidemiologic research (2;6;7).

1.2 Objectives

The ultimate goal of this thesis is to tackle some methodological issues related to the general problem of confounding by indication in non-experimental pharmacoepidemiologic research. To investigate this, a conceptual framework will be proposed whereby confounding by indication is broken down into its component elements. Currently available solutions to the problem of confounding by indication will be examined, with a view to identifying problematic areas in need of solutions. In a second phase, possible extensions of currently available methods will be developed, using the proposed conceptual framework. Their application to hypothetical and empirical examples will be presented. Further conceptual work will be presented concerning the specific problem of confounding due to "residual channelling", i.e. to undocumented factors. A method to estimate residual channelling, and its use to infer about the likely extent of resulting confounding will be proposed and illustrated using empirical examples from a large administrative database.

1.3 References

1. Skegg DC. Evaluating the safety of medicines, with particular reference to contraception. Stat Med 2001;20:3557-69.

2. Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. Ann Intern Med 1979;90:257-61.

3. Tamblyn R, Abrahamowicz M. Drug utilization patterns. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: John Wiley & Sons, 1998:1235-47.

4. Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? Am J Med 1984;77:703-8.

5. Miettinen OS. Etiologic research: needed revisions of concepts and principles. Scand J Work Environ Health 1999;25:484-90.

6. Walker AM. Confounding by indication. Epidemiology 1996;7:335-6.

7. Shapiro S. Confounding by indication? Epidemiology 1997;8:110-1.

2 Review of the literature

2.1 Role of non-experimental research

Pharmacoepidemiology has been described as "what you do when you can't randomize" (1). Although this statement may at first glance seem somewhat lapidary and disparaging of the complexity of the field of pharmacoepidemiology and drug safety research, upon reflection, one can only agree: if it were possible to ethically randomize sufficiently large numbers of patients to alternative interventions - be they medical, surgical, psychological or other – and observe them for a sufficiently long time period to document the outcomes of interest, then all clinically relevant questions could - and would - be answered by means of randomized clinical trials (RCTs). No more lengthy discussions about possible sources of uncontrolled biases and how to deal with them. No more "hierarchy of evidence", as there would no longer be any need for case-control or non-experimental (observational) cohort studies. All epidemiologists would become "controlled trialists". The medical literature would consist only of randomized controlled trial reports, and meta-analyses would be updated using the most recent trial results, each one contributing incremental knowledge to the whole. The only operational challenge for the "controlled trialists" would be to ensure patient adherence to treatment and adequate documentation of outcomes, and to minimize losses to follow-up.

However, we are – and most likely always will be – far from this utopian, nearly Orwellian world that might seem like paradise to (some) proponents of evidence-based medicine. Reality is quite different. Randomized controlled trials, though central to the knowledge-base of medicine, cannot answer all clinical questions of interest. Ethical considerations may preclude the initiation – or continuation – of a study, as soon as clinical equipoise is lost. Long latency between exposure and outcome make lengthy

observation periods necessary, which require the use of historical, retrospective data (2). Some outcomes, such as rare adverse reactions to medication, occur so rarely that it would be necessary to carry out RCTs with huge numbers of patients in order to adequately estimate the magnitude of a hypothesized association, the cost of such trials being absolutely prohibitive for any company or government body (2).

There is, therefore, a need for alternative approaches to answer clinical questions in situations where RCTs reach their limits. These methods collectively form what is often referred to as "observational research", emphasizing the role of the epidemiologist as observer rather than intervener (3). Another term used is "non-experimental" research (3). In my opinion, this term is preferable - at least for the purposes of the present discussion - insofar as it emphasizes not the role of the researcher, but rather the contrast between non-experimental ("epidemiologic", "observational") and experimental ("clinical trials") methodology. It is useful to emphasize this contrast, because it highlights the core issues and problems inherent to non-experimental research. Consequently, the term "non-experimental research" will be used instead of "observational research" throughout the text of this thesis.

2.2 Role of post-marketing surveillance

Due to the inherent limitations of randomized controlled trials and to the discrepancies between regulatory requirements and practice requirements (i.e. the knowledge needed to use drugs in practice), there is a need to study the use of drugs continuously after they have been released onto the market (2). Such activities are regrouped under the heading of "post-marketing surveillance", and include both surveillance for safety (i.e. focusing on unintended and undesirable effects of drugs) and surveillance for efficacy

(i.e. study of intended effects) (2). I will examine both in turn in the following sections.

2.2.1 Study of unintended effects (surveillance for safety)

The reasons why post-marketing surveillance for safety is necessary have to do with the limits of the evidence base that is required to permit the marketing of a drug. The main ones are as follows (2;4): 1) The pre-marketing phase 3 studies – usually double blind, randomized clinical trials conducted in human subjects for the purpose of demonstrating the safety and efficacy of a drug for a particular indication - are of necessity limited in size (usually 500 to 5000 participants, maximum). Consequently, any adverse drug reaction that is rare, say that occurs with a frequency of less than 1 per 1000 is unlikely to be detected, and even then, the power of the study to establish statistical significance would be insufficient. 2) The typical study patient is usually quite different from the patients who will be taking the drug in the real world – real patients tend to be older, are often females, and frequently have multiple co-morbid conditions and co-medications, etc. It is therefore questionable whether study results can safely be extrapolated to these patients. 3) Finally, the duration of phase 3 studies is quite short, usually up to a maximum of one or two years, which makes them unsuited to establishing the long-term safety of drugs.

The objectives of post-marketing surveillance for safety are threefold (2): 1) The discovery of side-effects unknown at the time of marketing; 2) The confirmation of side-effects reported from other sources (spontaneous reporting systems, for example); and finally 3) the quantitative evaluation of risks of known adverse effects, whether discovered before or after marketing. The quantitative evaluation of these risks is subject to difficult methodologic problems, in particular confounding by indication, which I will explore in detail in a subsequent section, together with some of the approaches

that have been proposed as solutions to this problem.

2.2.2 Study of intended effects (surveillance for efficacy)

The limitations of Phase 3 studies and other licensing requirements described in the previous section also have consequences for the evaluation of the efficacy of drugs (2). Pre-marketing studies namely do not document long-term efficacy, efficacy in everyday practice or efficacy for new indications. Some of the clinical knowledge that is today undisputed was established in post-marketing surveillance studies of efficacy (5). Such knowledge would not have otherwise become available. It is also not uncommon for a new drug indication to emerge after a drug has been put onto the market. For example, propranolol, a beta-receptor blocking agent, was first marketed as an anti-arrhythmic. The post-marketing discovery that it caused hypotension as a side-effect led to it being widely used as an antihypertensive agent (6).

2.3 Problems inherent to non-experimental research

2.3.1 Documentation

Documentation in non-experimental research is a particularly problematic issue because, unlike clinical trials, the (research) purpose for which the data will be used in the future is often not clear at the time when the data to be collected are defined. For example, a drug claims data base may be set up for purposes of billing reimbursement, but later used for the quantification of risks of newly discovered side-effects. Furthermore, it is usually not possible to record all reasons for prescribing a drug (many of them being perhaps not objectively identifiable, even to the prescribing physician), and many characteristics that are of relevance to prescribing (such as smoking, alcohol use, etc.) are simply not documented in many databases.

2.3.2 Confounding by indication

A central problem that poses a severe threat to the validity – and credibility – of non-experimental research lies not with incomplete documentation, but rather with actual differences in the way drugs are prescribed to patients in everyday practice (7). Patients are generally chosen for treatment with a given drug because it may alter their expected clinical course. It follows that patients assigned to one drug rather than to another may differ in the reasons (the indication - its presence/absence, or severity) that lead to a particular treatment choice. This indication may in turn be related to the outcome – whether beneficial or detrimental – that a patient will experience, thus distorting any estimate of the association between the chosen drug and the outcome of interest that fails to adjust for such systematic differences. This distortion of the estimate of association between drug exposure and outcome by the reason for prescribing the drug is called confounding by indication.(2;8;9)

Therefore, an estimate of the strength of association between an exposure and an outcome can either reflect a true causal association, be confounded by factors that were undocumented or unaccounted for in the analysis, or result from a combination of both causality and confounding (10). It therefore follows that the challenge for applied medical researchers is to eliminate, or at least minimize, the possibility that an estimate of association is confounded. This is a central problem of non-experimental research, which some researchers view as intractable (2;11;12).

2.4 Proposed solutions to the problem of confounding by indication

As researchers have come to better understand the problem of confounding by indication, they have developed tools for the design and analysis of non-experimental

studies with the purpose of reducing, and ideally eliminating, confounding by indication. In the following sections, I will examine these approaches each in turn - restriction, matching, stratification, multivariate analysis, ecological analyses, confounder score, instrumental variables and propensity scores - focusing on their definition, underlying assumptions and limitation, and their ability to control for confounding by indication.

2.4.1 Restriction

To be a confounder, a variable has to be associated with both the treatment choice and the outcome. Yet, a variable that has the same value in all subjects of a given study cannot possibly act as a confounder (13). Restricting admissibility of study subjects to homogeneous group(s) selected on the basis of clearly defined criteria is therefore a very efficient and usually inexpensive means of preventing confounding (13). In fact, restriction has been termed the most effective approach for preventing confounding by a known risk factor (14). However, it should be kept in mind that this does not prevent an estimate of association from being confounded by factors *other* than those used to restrict admissibility. Moreover, restriction can severely shrink the pool of available subjects, leading to less precise estimates of exposure-outcome association (13). Furthermore, it is possible that a homogenous study group resulting from restriction would provide a poor base for "generalization" of results to the population at large. On the other hand, Rothman and Greenland argue that a valid estimate from a homogeneous group may be more precise and useful than imprecise and even conflicting estimates derived from a collection of heterogeneous subgroups (13). Finally, the unequivocal demonstration of an association within a clearly defined subgroup can provide a firm rationale to further investigate this association in other subgroups. Because it is necessary to assess (and ideally document) the characteristics

defining admissibility, restriction is a method that depends on documented – or at least, documentable – factors.

2.4.2 Matching

Matching is the process by which referent subjects (unexposed subjects in a cohort study, or "controls" in a case-control study) are selected in such a way that they are as similar as possible to an index (exposed, "case") subject on one or more potentially confounding factors (13). For example, patients may be matched on gender and age within a one-year bracket. There are different sub-types of matching, depending on whether it is performed for confounder values observed for individual subjects (*individual matching*) or for the distribution of these values in the entire groups (exposed vs non-exposed) of subjects (*frequency matching*). However, although these approaches are different with respect to the design of the study, there are no substantial differences in the way in which they are analyzed (13).

The purpose of matching was originally thought to be validity (15). However, it was later posited that the main purpose of matching is (statistical) efficiency, and not validity (16). This view continues to be held and presented in most current textbooks of epidemiology (13), however Miettinen has recently argued that matching is never justified by the pursuit of either validity or efficiency (17).

Although it is possible to match subjects in both cohort and case-control studies, these are actually two different undertakings, with different consequences (13). Matching in cohort studies is done by matching exposed and unexposed subject before the outcome has occured, and as such it prevents any association between the exposure and the matching factor, and therefore has the ability to prevent confounding by the matched

factors. Despite this advantage, matching in field cohort studies is rarely used, mostly because it entails substantial costs.

Matching in case-control studies, on the other hand, is widely used, and will be the focus of the remainder of this section. If one adheres to the mainstream view that matching can improve statistical efficiency in case-control studies, there are nonetheless a number of drawbacks that have to be kept in mind. First of all, matching may – paradoxically – introduce selection bias into a study (13). This selection bias must be accounted for by controlling for the matching factors in the analysis, even if the matching factors are not themselves risk factors for the disease, lest the validity of the results be jeopardized (13;18). Furthermore, this selection bias can occur whether or not there is confounding in the source population from which the cases were selected (13). Finally, although statistical efficiency may be improved by matching in field case-control studies, this may be at the expense of cost efficiency, as it may be necessary to screen and reject a large number of potential controls before finding one suitable for matching to a case – this problem is particularly pronounced the more matching factors there are. Matching also introduces limitations on the scope of a study: once cases and controls have been matched on a factor, it is no longer possible to estimate the effect of that factor on the exposure-outcome association of interest (13).

Finally, an important issue when considering the use of matching is the potential for overmatching. Overmatching is the (inappropriate) use of matching in such a way that there is a negative effect on either statistical efficiency, validity of results, or cost efficiency (13). Overmatching causing a loss of statistical efficiency could occur if one were to match on a non-confounder associated with exposure but not with disease.

This calls for control by stratification in the analysis, which would have been unnecessary had there been no matching on the said factor (13). Moreover, matching on a factor that lies on the causal pathway between exposure and outcome can cause intractable bias whereby both the crude and stratified estimates of effect would be distorted (13).

Despite all these drawbacks, matching in case-control studies can be beneficial in some very clearly defined situations. If the information to be obtained from cases and controls comes at a high cost – as in the case of expensive blood tests – then it would be desirable to ensure that control of confounding in the analysis will not lead to the loss of information that was expensive to obtain. Matching would, in this a situation, prevent such losses (13).

In summary, the design issues involved in matching are quite complex, more than one would expect from a method with so much intuitive appeal. Although there are clear-cut situations in which matching should or should not be used, there are a host of intermediate situations where it is unclear whether matching is the most efficient approach, both from a statistical as well as from a practical point of view. Finally, matching can only control for factors that are documented – or documentable – or at most, for undocumented factors that are tightly associated with the matching variable (13).

2.4.3 Stratification

Stratification is an analytic method whereby groups of study subjects are formed according to particular values of a potential confounder or combinations thereof. These groups (strata) are, therefore, internally homogenous with respect to the stratification

variable, and, consequently, stratum-specific effect estimates of association should be free of confounding by the stratification variable. There are a number of methods for combining stratum-specific estimates of effect, the discussion of which goes beyond the scope of this literature review, however suffice it to say that the main assumption underlying the computation of an overall estimate of effect is that of a uniform effect across strata, i.e. that each stratum provides an estimate of the same quantity (13).

Stratification plays a central role in the analysis of non-experimental studies. One of the main reasons for performing stratified analyses – although there are others - is the evaluation and control of confounding (13). Even in studies where more sophisticated methods of analysis are ultimately employed, stratification is an important intermediate step that allows investigators to familiarize themselves with important features of their data, such as the distribution of key variables (13). Furthermore, stratification enables the investigator to detect the presence of effect modification (13). Stratification also has other features that make it attractive: for one, it is intuitively appealing. It is also cheap and relatively simple to carry out.

One of the main drawbacks of stratification is the fact that it is inherently difficult to stratify the data simultaneously on several factors, because each new factor adds a new dimension to the matrix over which the data is spread (13). For example, stratification on five binary factors implies $2^5 = 32$ strata! This can rapidly lead to a situation where several individual strata only contain one subject, be it index or referent, and thereby contribute no information about the association of interest. Finally, like the other approaches to confounding presented earlier, stratification can only be performed using documented factors.

2.4.4 Regression analysis

Stratification is limited in its ability to deal with many potential confounders simultaneously (see previous section for a more detailed discussion of this issue). Consequently, a vast array of techniques designed to overcome this problem have been developed. These techniques can be unified under the heading of *regression analysis* (13). The focus here is, once again, on their advantages and drawbacks with respect to their ability to control for confounding.

Briefly, all forms of regression analysis involve (at least) two steps, namely first *model specification*, in which the most appropriate model is chosen for a given substantive question, and secondly, *model fitting*, in which data is used to produce the best estimate possible of the parameter(s) included in the model. The estimated coefficients can then be used to assess the association between the variable of interest and the outcome, as well as to identify the confounders and effect modifiers (13).

The overwhelming advantage of regression analysis techniques is their ability to control for several potential confounding covariates simultaneously, and to estimate the independent effect of each covariate on an outcome of interest while controlling for the effects of other covariates (13). This advantage comes at the expense of having to make stronger assumptions about the relationship between covariates (independent variables) and the outcome of interest (the dependent variable). In regression analysis, these assumptions are explicitly defined by the formal properties of a given regression model - a situation which should be advantageous. However, it is not unusual that regression modelling is applied "automatically" without the intended audience – or even the investigator – having verified and/or even understood the implications of these assumptions very clearly (13). This "black box" phenomenon and the resulting

confusion are important drawbacks of regression analysis.

If the chosen model - and its underlying assumptions – are ill-suited to the particular substantive question at issue, then the estimates of effect may be distorted, potentially leading to erroneous conclusions about the exposure-outcome association of interest, or about confounding.

Finally, to include a factor into a model, it has to have been documented, so like other techniques presented in this chapter, regression analysis can only account for confounding caused by documented factors.

2.4.5 Ecological analyses

Ecological studies are conducted using group-level variables - such as hospitals, cities, regions or countries - to draw inferences about an exposure-association of interest. Such studies tend to be easily confounded by factors others than those under study (a phenomenon described as the "ecological fallacy" (19), and furthermore, the potentially confounding factors in ecological studies tend to be difficult to document and to control for (20). Consequently, ecologic studies have generally been regarded as poor substitutes for non-experimental studies of individual-level data (20).

Recently, Wen and Kramer (20) proposed that ecological studies can be useful in the study of intended treatment effects because they are "relatively immune to confounding by indication" (20). They proposed a theoretical framework to carry out such studies, and called upon other researchers to apply it to empirical data (20). The challenge was taken up by Johnston, who demonstrated that there are indeed discrepancies in the results produced by individual-level and group-level analyses, with the former most likely being biased by indication (21). Johnston also carried out a "two-level" analysis – i.e.

an individual-level modeling analysis including some area-level variables – and suggested that such an approach may combine the advantages of both types of analyses, namely the reduced confounding by indication of ecological studies and the increased power of individual-level analyses (21).

This approach has the advantage that it can also account for potentially confounding factors that are not documented, thereby avoiding a problem inherent to all the other approaches to confounding by indication that I have reviewed so far in this chapter. On the other hand, such two-level analyses are dependent on the availability of area-level data, such as data from multi-center studies or multiple regional registries, which may not be available depending on the substantive question of interest.

Nonetheless, this is a promising approach that opens up new avenues for tackling confounding by indication, particularly that due to undocumented – or undocumentable – factors.

2.4.6 Confounder score

In order to bypass the limitations of both stratification and multivariate regression analysis, Miettinen proposed an approach that combined features of both techniques, namely "stratification by a multivariate confounder score" (22). In this approach, a regression model is constructed, with the dependent variable being the outcome. Potential confounders as well as the exposure variable are included in the model, and the regression coefficients are estimated. This "fitted outcome function" is then transformed into a "scoring function" by setting the value of the (independent) exposure variable to zero, thus eliminating the exposure from the scoring function. The scoring function is then used to calculate a score for each subject, and the scores are used to

create strata, within the score-range of both cases and controls. Unconfounded within strata-analyses can then be performed assuming that the strata are sufficiently homogenous with respect to the confounding scores (22). One of the advantages of this technique is that it can help identify and eliminate false confounders – such as, in Miettinen's example, "yellow finger" in the association between smoking and lung cancer (22) - from the causal pathway.

Miettinen and colleagues had made use of this score in a paper (23) they published before the theoretical paper, however this approach was not used or discussed frequently in the subsequent literature, and its theoretical foundation is not as well established as that of other approaches, such as propensity scores, which are the topic of a later section in this literature review (24).

2.4.7 Instrumental variables

Developed in the field of econometrics in the 1920s and widely used in this domain since then, instrumental variables have only recently started to be used in the health sciences and, to a very limited extent, in epidemiology (25). In order for a factor to be used as an instrumental variable, three assumptions must be fulfilled: the factor must be associated with exposure, be independent of potential confounders (both documented and undocumented) and be independent of outcome, except perhaps through an association with the exposure (25). An example of a situation where these criteria are fulfilled is a randomized controlled trial with non-compliance: the instrumental variable is the treatment assignment, and the exposure is the treatment actually received (25). In such a situation, the instrumental variable approach can be used to correct for confounding introduced by non-compliance. Likewise, instrumental variables can be used to correct for bias introduced by measurement errors (misclassification) (25).

In non-experimental studies, instrumental variable approaches are severely limited by the assumptions requiring that they be independent of the potential confounders and of the outcome (25). In fact, using instrumental variables may even introduce additional bias if these assumptions are not satisfied (25).

2.4.8 Propensity scores

Propensity score analysis was proposed in 1983 by Rosenbaum and Rubin for applications including, among others, control of confounding in non-experimental studies using multivariate adjustment by subclassifaction on the propensity score (26;27). Propensity score analysis was only recently been addressed in literature intended for non-statisticians (28;29), and its use has subsequently become more widespread in applied medical research (30-32).

Conceptually and technically, the propensity score approach is similar to Miettinen's confounder score approach (22). In both methods, first a multivariate regression model is estimated using potential confounders, then the fitted model is used to calculate a score for each subject, and this score is used as a new covariate in subsequent analyses, typically as a stratification variable. The main difference between the two approaches is that the confounder score uses the outcome as the dependent variable and ignores the exposure in the estimation of the scoring function, whereas the propensity score approaches uses the exposure, and the outcome plays no role in the calculation of propensity scores. Consequently, the propensity score is the "conditional probability of exposure to a treatment given observed covariates" (24).

Like the confounder score approach, propensity scores reduce the number of confounders to be controlled, combining the advantages of multivariate analysis and

stratification while avoiding the sparse-data problems associated with stratification. However, like the confounder score approach, it can only account for documented factors and does not meaningfully adjust for undocumented factors, unless they are strongly correlated with documented factors (24). The further aspects of propensity score analyses will be addressed in more detail in the chapter on methods ("Overview of study design, data sources and statistical analyses").

2.5 Outstanding problems due to confounding by indication

The features of the various approaches to control for confounding are summarized in table 1 below. It can be seen that no single method can adequately control for confounding due to documented factors while also accounting for the potential effects of undocumented confounders. This remains a central problem to the analysis of non-experimental studies, because in order to produce valid results, it is imperative that confounding be eliminated so that only causation remains as an explanation for a given association (17). This thesis presents an attempt at tackling some aspects of this important problem.

Table 1. Current approaches to the problem of controlling for confounding in non-experimental research.

Approach	Applied to	Advantages	Disadvantages	Accounts for undocumented factors
Restriction	design	• Efficient • Precise estimates • Inexpensive	• does not prevent confounding by factors other than the restriction factor • questionable generalizability	no
Matching	design	• thought to be efficient (controversial)	• does not prevent confounding by factors other than the matching factor • may introduce bias if no control for matching factor in analysis • risk of overmatching • inefficient if matching on several variables	no
Stratification	analysis	• simple • cheap • intuitive	• controlling for several factors simultaneously is difficult (sparse data problem)	no
Multivariate analyses	analysis	• control for multiple variables simultaneously	• relies on strong assumptions	no
Ecological analyses	analysis	• combines advantages of both ecological and individual-level studies without the disadvantages of either	• requires area level data • theoretical basis not very well established	no
Confounder score	analysis	• combines advantages of stratification and multivariate analysis • avoid sparse-data problem • helps eliminate false confounders	• theoretical basis not very well established	no

Instrumental variables	analysis	• can control for undocumented confounders	• strongly dependent on assumptions	yes
Propensity scores	analysis	• decreases number of confounders needing control	• can only account for the effect of documented factors	no

2.6 Pharmacology of drugs used against pain and inflammation

2.6.1 Introductory remarks

Confounding by indication is particularly likely to occur when a new drug is introduced onto the market. If this new drug is believed to be "safer" or more effective, it has been demonstrated that sicker patients tend to be preferentially prescribed this drug, which can then make the drug look "worse" than it is when outcomes are evaluated. In manuscripts 2 and 3, we investigated such a situation by focusing on the introduction of a new class of anti-inflammatory drugs, the selective cyclooxygenase II (COX-2) inhibitors.

Another situation where confounding by indication can arise is when there are two treatment alternatives for a given indication, and one of the alternatives is thought to cause a particular adverse effect, while the other alternative is thought to be free of this adverse effect. Patients having already experienced this adverse effect (whether or not as a result of taking the implicated drug) are more likely to be assigned to the latter, "risk-free" treatment alternative. If these patients are nonetheless at higher risk of developing this adverse effect, the "adverse-effect-free" alternative may appear to cause more such events, simply due to the underlying higher risk of patients assigned to it. This is another example of confounding by indication. In the second empirical example used in this thesis, we illustrate such a situation by focusing on the use of NSAIDs and acetaminophen as two treatment alternatives that might be subject to confounding by indication in patients with a history of GI bleeding.

Finally, we wanted to investigate a situation where little or no confounding by indication was expected. We chose to study users of naproxen and diclofenac, two well

established NSAIDs with fairly similar effectiveness and adverse-effect profiles (33). These drugs were the two most commonly prescribed NSAIDs that were not otherwise available over the counter in the province of Quebec during the period covered by our example (January 1st 1994 to December 31st 1996). In order to set the stage for these empirical examples, it is necessary to briefly review the pharmacology of these various agents. This is the topic of the following section.

2.6.2 Pharmacology

2.6.2.1 Non-Steroidal Anti-Inflammatory Drugs (NSAIDs)

Non-steroidal anti-inflammatory drugs, NSAIDs for short, is a term used to describe a chemically heterogenous group of drugs that share common anti-inflammatory, analgesic and anti-pyretic properties (34). The term "non-steroidal" is used to distinguish these drugs from glucocorticoid anti-inflammatory drugs, and they have also been called "non-narcotic analgesics", to distinguish them from opiod analgesics (34). The term NSAID is the most commonly employed name, and it is used throughout this thesis.

The NSAIDs are classically organized into three subgroups, namely aspirin, non-aspirin NSAIDs (sometimes abbreviated "NANSAIDs") and, more recently, cyclooxygenase-II inhibitors (COX-2 inhibitors). The pharmacology of these drug groups is the topic of the following sections. Acetaminophen, which, due to its lack of anti-inflammatory properties, is not classified as an NSAID, will also be discussed in this review, since it is used for some of the same indications as NSAIDs, either alone or in combination with an NSAID.

NSAIDs constitute one of the most frequently used drug classes, both as prescription drugs and over-the-counter medication (34). Through their side-effects, mainly gastro-intestinal and renal, they also account for a large proportion of drug-related hospitalisations and deaths (34). The NSAIDs currently available on the Quebec market are shown in Table 2 at the end of this chapter, together with their half-life, recommended anti-inflammatory dose and other particular features.

2.6.2.1.1 Aspirin

Aspirin has been used to combat pain and fever for over a hundred years (33). In the past two decades, its benefits as an anti-platelet agent have become evident in the treatment of cardiovascular disease. Consequently, many patients are now prescribed aspirin for indications other than acute and chronic inflammation, for which the other NSAIDs are chiefly prescribed. Consequently, we did not use dispensation of aspirin as a criterion for selection into our study cohorts, since this is not (necessarily) an alternative to the use of acetaminophen, other NSAIDs or COX-2 inhibitors. Aspirin is known to increase the risk of adverse gastrointestinal (GI) effects (33), however this is thought to be mainly a dose-dependent effect that should be minimal at the low doses used for cardiovascular indications (35). Nonetheless, it may influence the choice of other drugs used to treat concomitant conditions, so it was necessary to control for aspirin use by including it into our regression analyses.

Because aspirin was the mainstay of analgesic and anti-inflammatory therapy for many decades, it is the standard against which other drugs are compared. It is therefore appropriate to also briefly review its pharmacology here.

2.6.2.1.1.1 History

For many years, the bark of the willow was used in folk medicine to treat mild pain and fever. The active ingredient is salicin, which, when hydrolysed, yields salicylic acid (33). Acetyl salicylic acid, a chemically related compound, was first synthesized in 1853, but only in 1899 was it found to be effective against arthritic pain, and marketed under the name "Aspirin" (33). Because it was cheaper and more effective than natural extracts and powders from the bark of the willow, aspirin quickly replaced these natural products and has remained a cornerstone of analgesic and anti-inflammatory therapy for over a hundred years (33).

2.6.2.1.1.2 Pharmacodynamics

2.6.2.1.1.2.1 Mechanism of action

Aspirin acts by inhibiting prostaglandin biosynthesis in the body (33). Prostaglandins are vasoactive fatty acids that play a central role in the mediation of inflammation and pain (33). Aspirin irreversibly blocks the enzyme cyclooxygenase-1 (COX-1, also known as prostaglandin synthase), whereas aspirin's metabolite, salicylate, reversibly inhibits prostaglandin synthesis (33).

2.6.2.1.1.2.2 Effects

Aspirin has anti-inflammatory effects mediated through its inhibitory effects on the inflammatory cells of the immune system at the site of inflammation. Asprin has analgesic (pain-reducing) effects mediated via its local anti-inflammatory effects, and it is also believed to have central pain-inhibiting effects, due to inhibition of pain stimuli at the subcortical level (33). Aspirin is an antipyretic agent, that is, it reduces elevated body temperatures, whereas normal body temperatures are barely affected (33). Finally, aspirin affects hemostasis (the blood's coagulation-bleeding equilibrium). This

occurs because aspirin inhibits thromboxane synthesis (another product of cyclooxygenase activitiy), which in turn leads to an inhibition of platelet aggregation. Since the effect of aspirin on thromboxane synthesis is irreversible, its anti-platelet effect lasts until new platelets are formed, that is, about 8 to 10 days (33).

2.6.2.1.1.3 Clinical uses (indications)

Since aspirin is effective in the treatment of mild to moderate pain, it has been used extensively to treat pain of different causes, such as pain of muscular, vascular or dental origin, post-partum pain, arthritis and bursitis (33). Aspirin is not effective in the treatment of severe visceral pain, such as renal colic, acute abdomen, myocardial infarction or pericarditis, for example (33).

The antipyretic properties of aspirin were instrumental in its discovery, and aspirin is still today considered one of the best agents for reducing fever, which, except in a few rare conditions such as neurosyphillis and chronic brucellosis, is not thought to be a useful physiological defense mechanism (33).

Aspirin has been shown to decrease the incidence of transient ischemic attacks ("mini-strokes" lasting less than 24 hours), unstable angina (chest pain of cardiac origin experienced at rest), myocardial infaction and re-infaction, and it is thought to potentially decrease the incidence of occlusion of coronary artery by-pass grafts (33). In a meta-analysis of more than 50 anti-platelet secondary prevention trials, aspirin was found to reduce vascular death (mainly fatal strokes and heart attacks) by 15% and non-fatal vascular events by approximately 30% in various patient groups (36). In recent years, increasing numbers of patients are being prescribed aspirin as primary or secondary prophylaxis for these conditions. With chronic use, the anti-platelet effects of

aspirin are obtained even with low doses (about 100 mg per day), whereas side-effects, which will be addressed in the next section, appear to be dose-dependent (35).

Other potential indications for which aspirin could be used are the prevention of cataracts, although the evidence in this respect is contradictory, and the prevention of colon cancer, for which there is an increasing body of plausible evidence (33).

2.6.2.1.1.4 Adverse effects

The ingestion of aspirin may cause upper abdominal (epigastric) discomfort, more rarely nausea and vomiting (37). Aspirin can also cause gastric ulceration, bleeding, even perforation and hemorrage. Blood losses due to chronic aspirin use are often asymptomatic and can lead to iron deficiency anemia (37). These undesirable effects can be reduced or avoided if aspirin is taken concomitantly with buffers or prostaglandin analogs (such as misoprostol) in patients with a history of GI disturbancies (33). Nonetheless, aspirin has a less favourable profile with respect to GI side-effects than most non-aspirin NSAIDs, which will be discussed in a later section of this literature review (33).

Other adverse side-effects of aspirin include increases in serum uric acid levels, mild, usually asymptomatic hepatitis, decreases in glomerular filtration rates in patients with underlying renal disease and hypersensitivity reactions (33). Hypersensitivity to aspirin is more commonly manifest in patient with asthma and nasal polyps (33). Because of its mild anti-coagulant effect, aspirin should not be taken by patients suffering from hemophilia (33). Aspirin is generally not to be used in pregnancy, and caution should be used in children, as an increased incidence of Reye's syndrom (a rare, acute and sometimes fatal disease of childhood, characterized by recurrent vomiting, elevated liver

enzymes, histologic liver changes and, in some cases, encephalopathy and cerebral edema (38)) has been reported with the use of aspirin shortly after a viral infection.

2.6.2.1.1.5 Toxicity

Aspirin at high or near-toxic doses can induce tinnitus (ringing in the ears), decreased hearing and vertigo (dizziness), a symptom complex collectively known as "salicylism", since these symptoms are a common feature of all forms of salicylate poisoning (33). These symptoms can usually be reversed by simply reducing the dose. At even higher doses, aspirin causes an increased breathing rate, which is mediated via a direct effect on the brainstem. This can lead to respiratory alkalosis. At even higher doses, the balance is shifted toward metabolic acidosis, which is caused by the accumulation of salicylic acid derivatives and by depression of the respiratory centers (33). In severe cases, patients may require ventilatory assistance from a respirator. Aspirin poisoning also causes hyperthermia. The treatment of aspirin overdose focuses on increasing the elimination rate of aspirin, which is done by maintaining a high urine output volume and can be improved by alkalizing the urine, and on symptomatic support and maintenance of vital functions (33).

2.6.2.1.2 Non-Aspirin NSAIDs

2.6.2.1.2.1 History

Phenylbutazone was the first drug to be classified as an NSAID when it was introduced into clinical practice in 1949 (39). In 1963, the anti-inflammatory effects of indomethacin were discovered as a result of a screening program aimed at identifying new substances with anti-inflammatory effects similar to those of aspirin, but with fewer

side-effects (39). The search for better NSAIDs with fewer side effects continued over the following decades and has led to the proliferation of substances available in this class (33).

2.6.2.1.2.2 Pharmacodynamics

2.6.2.1.2.2.1 Mechanism of action

The anti-inflammatory effects of non-aspirin NSAIDs (hereafter NSAIDs for short) are mediated, like those of aspirin, mainly through the inhibition of prostaglandin synthesis, however, unlike aspirin, this inhibition is reversible (33). Recently, it has been established that the inhibition of prostaglandin synthesis caused by NSAIDs is mediated by one or both of two isoenzymes, namely cyclooxygenase (COX) 1 and 2 (34). The different NSAIDs have variable and incomplete selectivity for COX-1 and COX-2 (33). Because the gastro-intestinal toxicity of NSAIDs is caused mainly by the inhibition of COX-1, selective COX-2 inhibitors were developed to provide the same therapeutic benefits of classical NSAIDs without the associated GI side-effects (40). The selective COX-2 inhibitors will be reviewed in a separate section of this literature review.

2.6.2.1.2.2.2 Effects

The NSAIDs are defined as a class by the effects they share in common: anti-inflammatory, analgesic and anti-pyretic effects. In addition, all exhibit anti-platelet effects (33). NSAIDs suppresses the signs and symptoms of inflammation, but they do not treat the underlying cause nor do they modify the course of the underlying disease (34). Their efficacy as anti-inflammatory agents had been well established in a number of randomized placebo-controlled trials, and the efficacy of various NSAIDs compared to one another and to aspirin has been found to be similar (34).

The analgesic effects of NSAIDs are manifest at doses significantly lower than those required for their anti-inflammatory effects (34). In contrast to the central effects of opiod analgesics, NSAIDs are thought to act peripherally, by modulating pain signals to local pain receptors at the site of inflammation (34).

2.6.2.1.2.3 Clinical uses (indications)

NSAIDs are effective in the treatment of pain associated with osteoarthritis, rheumatoid arthritis, ankylosing spondylitis, dysmenorrhea (menstrual pain) as well as pain associated with musculoskeletal injury and inflammatory conditions (34). Some NSAIDs are used for special indications, such as indomethacin in the treatment of acute gout, a condition for which it is the drug of first choice, and in the treatment of patent ductus arteriosus in premature infants, where its use has significantly reduced the need for surgery (33). Ketorolac is used mainly for its analgesic rather than its anti-inflammatory properties and has been used successfully instead of morphine in the management of mild to moderate post-surgical pain (33). Because they do not modify the course of the disease, NSAIDs should be considered only as therapeutic adjuncts to the disease-modifying drugs (DMARDs) in the treatment of rheumatoid arthritis (34).

2.6.2.1.2.4 Adverse effects

No NSAID is completely safe, and as a group, they have the potential of causing gastrointestinal, renal, hepatic, hematopoietic, cutaneous and neurological side-effects (34).

All NSAIDs are gastric irritants, although they tend, as a group, to cause less severe gastric side-effects than aspirin (33). NSAIDs are thought to contribute to the anemia and small bowel protein loss that is often observed in patients with rheumatoid arthritis

who are chronically taking NSAIDs (34). Risk factors known to be associated with gastro-intestinal tract complications of NSAID use include old age, history of GI bleed, history of cardiovascular disease, concomitant use of corticosteroids, and presence of co-morbid conditions (34). Furthermore, some NSAIDs have intrinsic secretory effects, which may cause diarrhea (34).

Hepatic toxicity has been reported with all NSAIDs, however, as a group, the NSAIDs exhibit less hepatic toxicity than aspirin (34). In most patients, these effects are asymptomatic; elevated liver values usually return to normal when the dose is reduced or the drug discontinued (34).

NSAID use can lead to a reduction in renal blood flow and thus in glomerular filtration, leading to fluid retention, edema and elevation of serum creatinin (34). This is particularly problematic in patients with congestive heart failure. NSAIDs can cause hyperkalemia in diabetic patients as well as in patients receiving beta-blockers, ACE-inhibitors or potassium-sparing diuretics such as spironolactone (34). Idiosyncratic renal reactions to NSAIDs have also been observed, causing massive proteinuria and intersitial nephritis (34).

Patients presenting with the triad of vasomotor rhinitis, nasal polyposis and asthma are particularly subject to hypersensitivity reactions and acute asthmatic attacks as result of NSAID use (34). This is thought to be mediated via the inhibition of bronchodilatory prostaglandins. This effect is common to all NSAIDs as well as to aspirin, therefore these drugs should not be used in patients exhibiting such features (34).

Phenylbutazone, and to a lesser degree, indomethacin, are known to cause agranulocytosis and aplastic anemia; phenylbutazone was withdrawn from the US

market for this reason (34). Indomethacin is safer and has replaced phenylbutazone as the agent of first choice in the treatment of acute gouty arthritis (33).

Severe headaches can occur as a result of NSAID use, in particular with indomethacin, and aseptic meningitis has been reported in patients with systemic lupus erythomatosus (SLE) using ibuprofen, sulindac, tolmetin or naproxen (34).

2.6.2.1.2.5 Toxicity

When taken in overdose, NSAIDs are much less toxic than aspirin or other salicylates (34). Symptoms of overdose include central nervous system depression, seizures, disturbances of vision, tinnitus, abdominal pain, nausea, vomiting, cardiovascular and respiratory depression (bradycardia, hypotension, apnea, cardiac arrest) (34). Treatment consists of gastric lavage, fluid administration and supportive therapy (34).

2.6.2.1.2.6 Drug interactions

Because of their wide-spread use and important pharmacologic activity (in particular due to extensive binding to plasma proteins), NSAIDs have the potential of interacting with a vast number of other drugs as well as with the underlying disease conditions of the patient (34). For examples, NSAIDs can diminish the hypotensive effects of diuretics, beta-blockers and ACE inhibitors, so that it may be necessary to adjust the dose of the hypotensive agent in order to achieve the desired effect (34). Most NSAIDs significantly inhibit warfarin metabolism, leading to higher plasma concentrations of warfarin so that it may be necessary to reduce the dose of warfarin accordingly (34). NSAIDs also interact in various ways with sulfonyureas, lithium, methotrexate, phenytoin, aspirin, digoxin and peripheral vasodilators such as hydralazine, to name but a few (34).

Conversely, NSAID metabolism may be affected by a patient's underlying disease

condition(s), and the NSAID dose may have to be reduced accordingly, as is the case for naproxen used in patients with renal insufficiency, or a drug might have to be avoided all together, as is the case, for example, with ibuprofen in severe liver disease (34).

2.6.2.1.3 Cyclooxygenase II (COX-2) Inhibitors

2.6.2.1.3.1 History

The COX-1 isoenzyme is constitutively expressed in a broad variety of human tissues, whereas the COX-2 isoenzyme is mainly induced in endothelial cells, macrophages, chondrocytes, synovial cell and osteoblasts by inflammatory mediators, growth factors and tumour promoters, suggesting a role in inflammation and cancer (34) (41). COX-2 is, however, constitutively expressed in the kidney (34). Because the adverse effects of aspirin and most classical NSAIDs are thought to result mainly from their inhibition of COX-1, the search was on for a specific inhibitor of COX-2, in the hope that this agent would possess the benefits of NSAIDs without their side-effects (33).

2.6.2.1.3.2 Pharmacodynamics

2.6.2.1.3.2.1 Mechanism of action

The COX-2 inhibitors have a very high affinity for the COX-2 isoenzyme receptor site, which they bind reversibly, albeit forming a tight enzyme inhibitor complex that is only slowly dissociable (34). Recently, the international COX-2 study group proposed a new classification of NSAIDs based not on their chemical structure, which is quite heterogenous, but on their degree of clinically defined selectivity for COX-1 or COX-2 (42). The proposed classification comprises four classes: COX-1 specific, COX nonspecific, COX-2 preferential and COX-2 specific.

2.6.2.1.3.2.1.1Effects

The analgesic and anti-inflammatory effects of COX-2 inhibitors have been found to be comparable to those of classical NSAIDs in a number of randomized clinical trials in patients with rheumatoid arthritis (34). The COX-2 inhibitors appear to have no significant effect on platelet aggregation (34).

2.6.2.1.3.3 Clinical uses (indications)

The COX-2 inhibitors are effective in relieving pain and inflammation associated with rheumatoid arthritis and osteoarthritis, as well as alleviating acute pain of musculo-skeletal origin, although the indication for which they are licensed vary from one jurisdiction to another (41). Like the classical NSAIDs, the COX-2 inhibitors help control the symptoms of inflammatory conditions, but they do not appear to treat the underlying causes or modify the course of the disease. As such, they should be considered adjuncts to disease-modifying drugs (DMARDs) in the treatment of chronic inflammatory conditions like rheumatoid arthritis. Furthermore, recent evidence showing that human colon tumours express high levels of COX-2 whereas normal gut tissue does not points to a potential role of COX-2 in tumour growth and progression, implicating that the use of COX-2 inhibitors may be beneficial in preventing or slowing the progression of colon malignancies (34). Finally, it is thought that COX-2 inhibitors may be useful in preventing COX-2 induced neuronal apoptotic cell death, which is thought to play a role in the development of Alzheimer's disease (34).

2.6.2.1.3.4 Adverse effects

The COX-2 inhibitors have been shown in a number of randomized controlled trials in patients with rheumatoid arthritis to cause significantly fewer gastric ulcers than classical NSAIDs (34). Safety, however, cannot be equated with tolerability: patients do

withdraw from COX-2 inhibitor therapy because of dyspepsia (34). Furthermore, COX-2 inhibitors have been shown to be associated with small but potentially clinically relevant increases in blood pressure, as well as decreases in glomerular filtration rates in some patient groups; these adverse effects occurred with a frequency similar to those observed for classical NSAIDs (41). It would appear, therefore, that the fact that COX-2 is constitutionally expressed in the kidney would explain this lack of benefit of COX-2 inhibitors compared to classical NSAIDs with respect to renal adverse effects (34). Furthermore, COX-2 inhibitors are contraindicated in patients who have had hyper-sensitivity reactions after taking classical NSAIDs, aspirin or sulfonamides (34) (43).

Finally, in the Vioxx Gastrointestinal Outcomes Reasearch (VIGOR) trial, a large randomized clinical trial comparing the COX-2 inhibitor rofecoxib with naproxen (an NSAID), patients taking rofecoxib were found to have a significantly higher rate of major cardiovascular events (mainly due to more myocardial infarctions) compared to patients taking naproxen (41) (44). On the other hand, in the Celebrex Long-Term Arthritis Safety Study (CLASS) trial, where celecoxib was compared with diclofenac and ibuprofen, no difference was found in the incidence of major cardiovascular events (45). A number of hypotheses and supporting evidence, ranging from the biological plausibility that COX-2 inhibitors might increase susceptibility to thrombosis, to methodological flaws in study design, to an anti-platelet effect of naproxen (46) have been advanced to explain the differences in cardiovascular outcomes observed in the VIGOR trial (41). What is certain, though, is that more information is needed on the cardiovascular effects of COX-2 inhibitors and their combination with anti-platelet drugs if clinicians are to be able to treat arthritic patients with cardiovascular disease in the most optimal way possible.

2.6.2.1.3.5 Drug interactions

Unlike the classical NSAIDs, no significant drug-drug interactions have been observed between COX-2 inhibitors and the drugs commonly used in the target patient population, including warfarin and methotrexate (34).

2.6.2.2 Acetaminophen

2.6.2.2.1.1 History

Acetaminophen (or paracetamol, as it it called in Europe), is a chemical derivative of acetanilide, an aromatic compound with anti-pyretic properties introduced in 1886 whose excessive toxicity quickly ended its use in clinical practice and prompted the search for related compounds with less toxicity (37). Acetaminophen itself was first used in a medical setting in 1893, but it was first after it was recognized as the main active metabolite of both acetanilide and phenacetin (a related compound with somewhat less toxicity than acetanilide) in 1949 that its use became widespread (37).

2.6.2.2.1.2 Pharmacodynamics

2.6.2.2.1.2.1 Mechanism of action

The mechanism of action of acetaminophen has not been explained satisfactorily so far. Its anti-pyretic activity appears to reside in its aminobenzine structure (37). Acetaminophen is only a weak inhibitor of COX enzymes, particularly in peroxide-rich environments such inflamed tissues (37).

2.6.2.2.1.2.2 Effects

Acetaminophen has analgesic and anti-pyretic effects similar to those of aspirin, however, unlike aspirin, it has only very weak anti-inflammatory effects and no anti-

platelet effects (37). It has no effects on the cardiovascular and respiratory systems, does not disturb the acid-base balance, does not cause gastric irritation, erosion or bleeding, and has no effect on coagulation or excretion of uric acid (37).

2.6.2.2.1.3 Clinical uses (indications)

Acetaminophen can be substituted to aspirin for the treatment of mild to moderate pain, and fever. Because of it fairly benign adverse effect profile, it is particularly suited for patients with a history of gastrointestinal disorders and for patients in whom aspirin is contraindicated, such as pregnant women or children having recently had a viral infection (37). Since it lacks anti-inflammatory properties, it is not suitable for the treatment of inflammatory conditions such as rheumatoid arthritis, however its analgesic properties make it a useful as an adjunct in the treatment of these conditions (33).

2.6.2.2.1.4 Adverse effects

Acetaminophen is usually well tolerated when taken at therapeutic doses. Skin rashes and allergic reactions have been reported occasionally; in some rare case, acetaminophen has been thought to cause neutropenia, thrombocytopenia and pancytopenia (37).

2.6.2.2.1.5 Toxicity

The main toxic effect of acetaminophen that results from overdose is a dose-dependent hepatic necrosis that can potentially be fatal. Intoxication can be accompanied by dizziness, excitement and disorientation as well as nausea, vomiting, anorexia and abdominal pain, particularly in the first 24 hours after ingestion (33;37). Manifestation of hepatic damage first become obvious within 2 to 4 days of ingestion. Nomograms are available to predict the likelihood of hepatotoxicity as a function of plasma level and

time since ingestion (37). In non-fatal cases, the hepatic lesion are reversible, within a period of weeks or months. Acute renal failure may occur as a result of acetaminophen overdose (33;37). The treatment of acetaminophen overdose consists in gastric lavage, administration of the antidote n-acetylcystein and vigorous supportive therapy (37).

2.6.2.2.1.6 Drug interactions

In chronic alcoholics, the formation of hepatotoxic acetaminophen metabolites is thought to be increased, leading to a higher likelihood of hepatic damage in an already damaged tissue (37). Bile-acid binding resins decrease the gastrointestinal absoption of acetaminophen, however their use in the treatment of dyslipidemias has decreased significantly since the introduction of the HMG-Co-A reducatase inhibitors, also known as statins (33).

2.6.3 Conclusion

At the present time, although the available analgesic and anti-inflammatory drugs are effective in providing symptomatic relief of pain and inflammation in a broad variety of conditions, none of them are completely safe. The COX-2 inhibitors possess a better profile with respect to gastrointestinal effects when compared to aspirin and classical NSAIDs, however their renal and cardiovascular side-effects remain of concern, and there is still insufficient data with respect to their long-term safety. In choosing a drug, careful consideration should be given to the patient's characteristics, his or her co-medication and co-morbidities, the desired therapeutic effect and the characteristics of the drug. Sufficient time – usually at least two weeks – should be allotted for a drug trial before switching to another compound, unless significant side-effects become apparent sooner (34).

Finally, it should be kept in mind that inflammation, though painful and bothersome, is a natural process that evolved as a response to tissue injury and pathogens, and has led to improved survival. Given this, it is possible that "better" anti-inflammatory agents may actually, by blocking these natural defenses, do more harm than good (37). In fact, it might be impossible to block physiologically important mechanisms without also causing some degree of toxicity – in other words, the "safe aspirin" may only be an illusory dream.

Table 2: NSAIDs and other analgesics, their half-life, recommended dosage and particular features (compiled from references (33), (34) and (43))

Drug (INN name)	Half-life (hours)	Recommended anti-inflammatory dose (per dose)	Dosing fre-quency *	Particular features
Acetaminophen	2	500-1000	qid	• No anti-inflammatory or anti-platelet effect
Aspirin	0.25	1200 – 1500 mg	tid	• Do not administer concomitantly with other NSAID • Irreversible COX-1 inhibition
Celecoxib	8 -12	100 – 200 mg	bid	• Selective COX-2 inhibitor • No anti-platelet effect
Diclofenac	1.1	50 – 75 mg	qid	
Diflunisal	13	500 mg	bid	
Etodolac	6.5	200 – 300 mg	qid	
Fenoprophen	2.5	600 mg	qid	
Flurbiprofen	3.8	300 mg	tid	
Ibuprofen	2	600 mg	qid	• Avoid in severe hepatic disease
Indomethacin	4 – 5	50 – 70 mg	tid	• For treatment of patent ductus arteriosus and acute gout • May cause blood dyscrasias
Ketoprofen	1.8	70 mg	tid	
Ketorolac	4 – 10	10 mg	qid	• For treatment of acute surgical pain
Meclofenamate	3	100 mg	qid	
Nabumetone	26	1000 – 2000 mg	qd	• Avoid in severe hepatic disease
Naproxen	14	375 mg	bid	
Oxaprozin	58	1200 – 1800 mg	qd	
Phenylbutazone	68	100 mg	qid	• May cause blood dyscrasias
Piroxicam	57	20 mg	qd	
Rofecoxib	17	25 – 50 mg	qd	• Selective COX-2 inhibitor • No anti-platelet effect
Sulindac	8	200 mg	bid	
Tenoxicam		20 mg	qd	
Tiaprofenic acid	1,5 - 3	200 mg	tid	
Tolmetin	1	400 mg	qid	

* qid = 4 times a day, tid = 3 times a day, bid = twice a day, qd = once a day

2.7 References

1. LeLorier, J. Personal communication. 2002.

2. Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. Ann Intern Med 1979;90:257-61.

3. Last JM. A dictionary of epidemiology. 3 ed. Toronto: Oxford University Press, 1995.

4. Skegg DC. Evaluating the safety of medicines, with particular reference to contraception. Stat Med 2001;20:3557-69.

5. Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of treatment on morbidity in hypertension: results in patients with diastolic blood pressure averaging 90 to 114 mm Hg. JAMA 1970;213:1143-52.

6. Zacharias FJ, Cowen KJ, Presst J, Vickers J, Wall BG. Propranolol in hypertension: a study of long-term therapy, 1964-1970. Am Heart J 1972;83:755-61.

7. Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? Am J Med 1985;78:475-80.

8. Miettinen OS, Cook EF. Confounding: essence and detection. Am J Epidemiol 1981;114:593-603.

9. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol 1999;149:981-3.

10. Miettinen OS. Confounding and effect-modification. Am J Epidemiol 1974;100:350-3.

11. Shapiro S. Confounding by indication? Epidemiology 1997;8:110-1.

12. Walker AM. Confounding by indication. Epidemiology 1996;7:335-6.

13. Rothman KJ, Greenland S. Modern epidemiology. 2 ed. Philadelphia, PA: Lippincott-Raven, 1998.

14. Greenland S, Morgenstern H. Confounding in health research. Ann Rev Pub Health 2001;22:189-212.

15. Miettinen OS. Matching and design efficiency in retrospective studies. Am J Epidemiol 1970;91:111-8.

16. Kupper LL, Karon JM, Kleinbaum DG. Matching in epidemiologic studies: validity and efficiency considerations. Biometrics 1981;37:271-92.

17. Miettinen OS. Etiologic research: needed revisions of concepts and principles. Scand J Work Environ Health 1999;25:484-90.

18. Schlesselmann JJ. Case-Control Studies. Oxford: Oxford University Press, 1982.

19. Piantadosi S, Byard DP, Green SB. The ecological fallacy. Am J Epidemiol 1988;127:893-904.

20. Wen SW, Kramer MS. Uses of ecologic studies in the assessment of intended treatment effects. J Clin Epidemiol 1999;52:7-12.

21. Johnston SC. Combining ecological and individual variables to reduce confounding by indication: case study--subarachnoid hemorrhage treatment. J Clin Epidemiol 2000;53:1236-41.

22. Miettinen OS. Stratification by a multivariate confounder score. Am J Epidemiol 1976;104:609-20.

23. Jick H, Miettinen OS, Neff RK, Shapiro S , Heinonen OP, Slone D. Coffee and myocardial infarction. N Engl J Med 1973;289:63-7.

24. Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. Am J Epidemiol 1999;150:327-33.

25. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 1999;29:722-9.

26. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

27. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. JASA 1984;79:516-24.

28. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757-63.

29. D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265-81.

30. Perkins SM, Tu W, Underhill MG, Zhou XH , Murray MD. The use of propensity scores in pharmacoepidemiologic research. Pharmacopepidemiol Drug Safety 2000;9:93-101.

31. Margolis DJ, Kantor J, Santanna J, Strom BL, Berlin JA. Effectiveness of platelet releasate for the treatment of diabetic neuropathic foot ulcers. Diabetes Care 2001;24:483-8.

32. Rahme E, Pettitt D, LeLorier J. Determinants and sequelae associated with utilization of acetaminophen versus traditional nonsteroidal antiinflammatory drugs in an elderly population. Arthritis Rheum 2002;46:3046-54.

33. Katzung BG. Basic and clinical pharmacology. 6th ed. East Norwalk (CT): Appleton and Lange, 1995.

34. Kelley's Textbook of Rheumatology. 6th ed. Philadelphia: W.B. Saunders, 2001.

35. Patrono C, Coller B, Dalen JE et al. Platelet-active drugs: The relationships among dose, effectiveness, and side effects. Chest 2001;119:39S-63S.

36. Antiplatelet Trialists' Collaboration. Collaborative overview of randomized trials of antiplatelet therapy: I. Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. BMJ 1994;308:81-106.

37. Goodman and Gilman's The pharmacological basis of therapeutics. 9th ed. London: McGraw Hill, 1996.

38. Dorland's illustrated medical dictionary. 28th ed. Philadelphia : W.B. Saunders, 1994.

39. Adams SS, Cobb R. A possible basis for the anti-inflammatory activity of salicylates and other nonhormonal anti-rheumatic drugs. Nature 1958;181:773-.

40. Deeks JJ, Smith LA, Bradley MD. Efficacy, tolerability, and upper gastrointestinal safety of celecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials. BMJ 2002;325:619-26.

41. Fitzgerald GA, Patrono C. The coxibs, selective inhibitors of cyclooxygenase-2. N Engl J Med 2001;345:433-42.

42. Lipsky PE, Abramson SB, Crofford L. The classification of cyclooxygenase inhibitors. J Rheumatol 1998;25:2298-.

43. British Medical Association, Royal Pharmaceutical Society of Great Britain. British National Formulary. London: The British Medical Association and the Royal Pharmaceutical Society of Great Britain, 2000.

44. Bombardier C, Laine L, Reicin A. Comparison of upper gastro-intestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. N Engl J Med 2000;343:1520-8.

45. Silverstein FE, Faich G, Goldstein JL. Gastrointestinal toxicity with clecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS stuy: a randomized controlled trial. JAMA 2000;248:1247-55.

46. Rahme E, Pilote L, LeLorier J. Association between naproxen use and protection against acute myocardial infarction. Arch Intern Med 2002;162:1111-5.

3 Overview of data sources and statistical analyses

In this chapter, a more detailed account of the methodology, variable definitions and statistical analyses is provided, because these topics are necessarily dealt with in an abridged form in the three manuscripts, contained in chapters 4 to 6.

3.1 Data sources

3.1.1 Source population

In order to illustrate the theoretical concepts developed in this thesis with empirical examples, historical study cohorts were formed using an administrative prescription database from the province of Quebec's universal health insurance plan, the Régie de l'assurance maladie du Québec (RAMQ). This is a well-validated prescription claims database containing information on prescription drugs, physician visits and other medical services dispensed in both the ambulatory setting and in hospitals to the vast majority (97%) of the population aged 65 and over living in the province of Quebec (1;2).

A data subset was extracted from the RAMQ administrative database for the specific purposes of this study. This subset formed the raw database, from which further extractions were performed as needed. The extraction of this database and its further use were approved by the Access to Information Board for the province of Quebec (Commission d'accès à l'information du Québec) as well as by the Institutional Review Board of McGill University (see Appendix for Certificate of Ethical Approval).

The source database consisted of all prescriptions for acetaminophen, NSAIDs and COX-2 inhibitors that were dispensed between January 1st 1990 and December 31st 2000 by a random sample of 25% (n=2778) of all general practitioners and by all rheumatologists (n=111) practicing in the province of Québec during that period. For all the patients

(n=878 505) having been dispensed any of these prescriptions, we obtained data on all other medications prescribed as well as all medical services received during this time period, even if the medication was prescribed by a physician other than the one who wrote the original prescription that led to inclusion into the database. We also obtained demographic data on all the patients (gender, age, place of residence) and all the physicians (specialty, age, gender, graduation year and university) included in the database. The data was anonymous, eliminating the possibility of retrospectively identifying any individual.

The hierarchical structure of the database was designed in order to have access to complete information both at the level of the physician and of the patient. Consequently, for a given physician, information was available concerning all prescriptions for non-steroidal anti-inflammatory drugs (NSAIDs), acetaminophen, or cyclooxygenase-2 inhibitors (COX-2) written and dispensed during the above-mentioned 11-year period. For patients, complete information about drugs and medical services dispensed was available for this period, regardless of the identity of the prescribing physician.

From the raw database, we extracted the study population of individuals, aged 65 and older on Jan. 1st 1990, who had received a relevant prescription and who were covered by the RAMQ's drug insurance plan. Patients were excluded from the study if they had an invalid age (n=26084, 3.0%), or if they died before the 1st of January 1990 (n=11). Further extractions using subsets of the study population were made as needed to suit the purposes of the various analyses, as described in the respective sub-sections.

3.1.2 Study base for empirical examples

In order to construct cohorts for the specific purpose of analyses illustrating various methodologic issues addressed in this thesis, further extractions were carried from the source population. Each of these cohorts is defined in the following sub-sections.

3.1.2.1 Example 1: NSAID and COX-2 inhibitor current users' cohort

Starting with all patients in the source population (n=878 505), only patients (n=194 815; 22%) having been dispensed at least one prescription (index prescription) for an NSAID or a COX-2 inhibitor (celecoxib being the only COX-2 inhibitor on the Quebec market at the time) during the 6-month period spanning from October 1st 1999 to March 31st 2000 were retained. The starting date of October 1st 1999 was chosen because it corresponds to the introduction of celecoxib (Celebrex®) on the Quebec market; rofecoxib (Vioxx®) was introduced on April 1st 2000. From these patients, those who were older than 66 at the time of dispensation of the index prescription who also had complete insurance coverage by the RAMQ in the period spanning from one year prior to 6 months after the index date (n=86 369) were retained. This complete documentation of drugs and services used by these patients during the study period of interest was necessary to ensure that exclusion criteria pertaining to drug use prior to the index date could be applied equally to all patients. Because some analyses focused on predictors of gastrointestinal (GI) events, continuous insurance coverage with the RAMQ up to at least 6 months following the index date was required, in order that information concerning GI events be available equally for all patients. Patients who were dispensed a gastro-protective agent (GPA) (n=8 507) or Arthrotec® (a combination of diclofenac and misoprotol, which is a GPA) (n=6 744) at the index date were eliminated, because they constitute a higher risk group, potentially different from both the "NSAID

only" users and the COX-2 inhibitor users. Finally, because the primary intention was to follow a cohort of current users of NSAIDs in order to examine the hypothesis that sicker or high risk patients tend to be switched to new drugs, we excluded patients who were not dispensed any prescription for NSAIDs in the year prior to the index date (n=43 871). The remaining patients (n=27747) formed the cohort of current users of NSAID. The cohort selection process is shown in Manuscript 2, Table 1. This cohort was used in Manuscript 2, Example 1 and in Manuscript 3, Example 1 to illustrate the application of the proposed indices of apparent and residual channelling, respectively.

3.1.2.2 Example 2: New users of NSAID and acetaminophen in the pre-COX-2 era

Starting with all patients in the source population (n=878 505), only those having been dispensed at least one prescription (index prescription) for an NSAID or acetaminophen during the 3-year period from January 1st 1994 to December 31st 1996 (n=435 278) were retained. From these patients, we eliminated those who were younger than 66 at the time of dispensation of the index prescription (n=215 631) as well those with incomplete insurance coverage by the RAMQ (n=7 005) at any time during the whole pre-study (January 1st to December 31st 1993) and study period (January 1st 1994 to December 31st 1996. Patients who were dispensed a gastro-protective agent (GPA) or Arthrotec® (n=12 386) were also eliminated. Finally, new users were selected by eliminating patients (n= 82 364) having received one or more prescriptions for NSAIDs or acetaminophen in the year preceding the index date. The time-window for the selection of the cohort was deliberately chosen before the introduction of COX-2 inhibitors onto the Quebec market, in order to avoid interference from a third therapeutic alternative. Based on the index prescription, the patients in the final cohort (n=97 216) were divided into new users of acetaminophen (n=47 258) and new users of

NSAIDs (n=49 958). The cohort selection process is shown in Manuscript 2, Table 2. Because acetaminophen was expected to be preferentially prescribed to sicker, older patients and patients at higher risk of GI events, the cohort was used (Manuscript 2, Example 2) to illustrate the application of the index of apparent channelling to a situation where relatively strong apparent channelling is expected.

3.1.2.3 Example 3: New users of two established NSAIDs: Diclofenac and Naproxen cohorts

To investigate a situation where little apparent channelling is expected, a cohort of new users of diclofenac or naproxen, the two NSAIDs with the biggest market share that are not available over the counter in the province of Québec, was formed. The time-window was the same as for the NSAID cohort described in the previous section (1.1.1994 – 31.12.1996). Only patients whose index prescription was either diclofenac (n= 13 093) or naproxen (n= 20 783) were retained for these analyses. Like in the previous cohorts, we also eliminated patients aged less than 66 at the index date, those without adequate RAMQ coverage as well as those having received a gastro-protective agent (GPA) at the index date. To ensure that the cohort contained only new users of diclofenac or naproxen and no switchers having first tried out other NSAIDs, we eliminated all patients having received *any* NSAID, or acetaminophen in the year prior to the index date. The reason for this is that switchers are known to have a different risk profile for adverse events, compared to first-time users (3). The cohort selection process is shown in Manuscript 2, Table 3. The application of the index of apparent channelling to this cohort is described in Manuscript 2, Example 3.

3.2 Variable definitions

A vast amount of information relating to demographic characteristics as well as medical

and pharmaceutical service utilization is available from the RAMQ database. For each example, the database was searched to identify demographic, diagnostic, medical and pharmaceutical variables that were considered potential determinants of treatment assignment. The following groups of variables were assessed: Demographic variables: patient's age, gender, location of residence (rural vs urban). Indicators of medical service utilization variables in the year preceding the index date: the number of hospitalizations, number of days in hospital, number of visits to GPs, specialists, rhumatologists, gastroenterologists, number of emergency room visits and number of GI events. Diagnostic variables: Diagnoses of cancer, heart failure and musculoskeletal disease in the year prior to the index date. Pharmaceutical service utilization variables: the season of the index dispensation, the total number of dispensations for the month preceding the index date; for the year preceding the index date: the number of prescriptions for NSAIDs, acetaminophen, aspirin, steroids, anticoagulants, opiates and gastroprotective agents and the chronic disease score, which is based mainly on prescription information (4). Whenever there was overlap between a more specific indicator (for example, use of gastroprotective agents) and a compound indicator (for example, the "prophylaxis" category of GI events) that included the specific indicator, the former variable was not included in the initial model. Table 1 identifies all the variables, with their type and units, that were used in the analyses presented in this thesis. Unless otherwise specified, these variables refer to the year preceding the index date. Throughout this thesis, in regression analyses, odds ratios apply to the units reported in Table 1, unless otherwise specified.

3.2.1 Gastrointestinal (GI) events

To describe gastrointestinal (GI) events, composite variables were created using the

medical and pharmaceutical variables as well as the diagnostic variables described in the previous section. Prior GI events, defined as having occurred in the year preceding the index date, were classified in descending order of severity, based on the classification system used by Rahme et al. (5), namely: 1) GI hospitalization: any hospitalization with a primary or secondary diagnosis of gastroduodenal perforation, ulceration, or bleeding (PUB) *or* any hospitalization during which an endoscopy was performed within the first 2 days; 2) ulcer: any physician visit with a diagnosis of perforation, ulceration, or bleeding (PUB) *or* any upper GI diagnostic test plus a prescription or a gastroprotective agent (GPA) (H2-antagonists, proton pump inhibitor, or misoprostol, but excluding Arthrotec®, which is a combination of misoprostol and diclofenac) *or* a gastroenterologist visit plus a prescription for a GPA; 3) dyspepsia: any physician visit with a diagnosis of dyspepsia, *or* any GI diagnostic test alone, *or* any visit to a gastroenterologist alone; 4) prophylaxis: GPA prescription alone.

Patients having experienced events in two or more categories were classified according to the most severe event. To maximize the number of events, and thus the statistical power and precision of the estimates, in most analyses it was necessary to rely on the inclusive definition of "any GI event", which comprises all four of the above-defined categories.

GI events occurring after the index date we described using the same classification system. GI events having occurred within the 6 months following the index date were defined as outcomes and used for the analyses described in Manuscript 3, to investigate residual channeling and its impact on confounding.

Table 1: Description of the variables derived from the RAMQ prescription database.

Variable	Type	Defintion/Units
Sociodemographic variables		
Sex	binary	male = 1
Age	continuous	year
Rural site of residence	binary	rural = 1
Co-morbidity variables		
Diagnoses of musculoskeletal disease in year prior to index date	binary	yes = 1
Diagnoses of cancer in year prior to index date	binary	yes = 1
Diagnoses of heart failure in year prior to index date	ordinal integers	yes = 1
Chronic disease score	ordinal integers	range 0 - 10
Pharmaceutical service utilisation variables		
Season of index prescription	dummy variable (3 terms)	yes = 1
Total number of drugs dispensed in month prior to index date	ordinal integers	one prescription
Number of dispensations for corticosteroids in year prior to index date	ordinal integers	one prescription
Number of dispensations for gastroprotective agents in year prior to index date	ordinal integers	one prescription
Number of dispensations for aspirin in year prior to index date	ordinal integers	one prescription
Number of dispensations for acetaminophen in year prior to index date	ordinal integers	one prescription
Number of dispensations for non-steroidal anti-inflammatory (NSAID) drugs in year prior to index date	ordinal integers	one prescription
Dispensations for anticoagulants in year prior to index date	binary	yes = 1
Dispensations for opiates in year prior to index date	binary	yes = 1
Medical service utilisation variables		
Number of visits to GPs in year prior to index date	ordinal integers	one visit
Number of visits to specialists in year prior to index date	ordinal integers	one visit
Number of visits to rhumatologist in year prior to index date	ordinal integers	one visit
Number of visits to gastroenterologist in year prior to index date	ordinal integers	one visit

Number of hospitalizations in year prior to index date	ordinal integers	one hospitalization
Average length of stay per hospitalization	continuous	days
Number of visits to the emergency room in year prior to index date	ordinal integers	one visit
Treatment assignment		
COX-2 inhibitor	binary	yes = 1
Acetaminophen	binary	yes = 1
Naproxen	binary	yes = 1
Gastrointestinal (GI) events		
Prior GI events* in year preceding index date:		
• GI hospitalization	binary	yes = 1
• Ulcer	binary	yes = 1
• Dyspepsia	binary	yes = 1
• Prophylaxis	binary	yes = 1
Outcomes* within six months following index date:		
• GI hospitalization	binary	yes = 1
• Ulcer	binary	yes = 1
• Dyspepsia	binary	yes = 1
• Prophylaxis	binary	yes = 1

*: see text for more detailed definition

3.3 Statistical analyses

3.3.1 Propensity score estimation

Propensity scores were calculated using the method put forth by Rosenbaum and Rubin (6). Briefly, propensity scores express the probability that a patient will be assigned to one of two treatment alternatives, based on documented covariates. A critical point is that the outcome variable plays *no role* in the prediction of the treatment assignment (7). Calculation of propensity scores involved the following steps: Potential determinants of treatment assignment were identified and entered into a multiple logistic regression model where the dependent variable was the assignment to one of the two treatment alternatives. The best fitting model was determined using backward elimination procedure. Only statistically significant variables (at the $\alpha = 0.05$ level for

the Wald chi-square test, as determined by backward elimination) were retained in the final model. It should be noted that given the very large size of all three cohorts and the relatively balanced shares of alternative drug treatments, all associations that are not statistically significant have to be very close to the null.

Next, values of selected variables for each patient in the cohort were then multiplied by the corresponding regression coefficients from the final logistic regression model, thus producing predicted logits of treatment assignment for this patient. Finally, logistic transformation was employed to convert logits into propensity scores. The propensity score expresses the estimated probability that a particular patient be assigned to the treatment coded as one. In their classical application, propensity scores are used to replace determinants of treatment choice by a single variable. This variable can in turn be included as a covariate in a regression model, or can be used as a stratification variable. In both case, the propensity score provides a parsimonious method of adjusting for all the covariates that went into its estimation (7), and can thus be used as a tool to reduce confounding due to documented factors. The steps involved in the estimation of propensity scores are summarized in Figure 1.

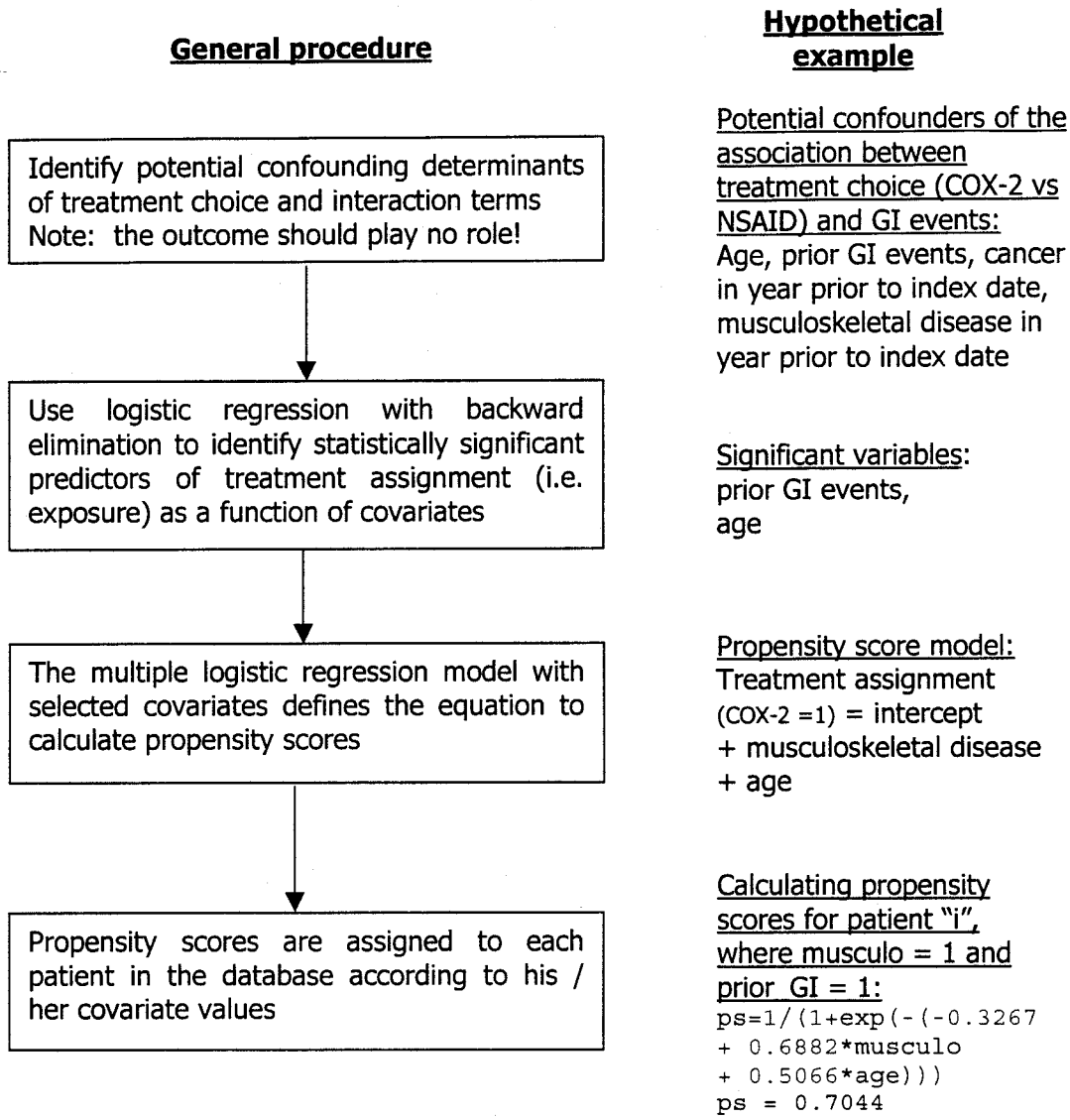

Figure 1. Generating propensity scores

3.3.2 Regression modeling

All regression modelling involved backward elimination with p-value > 0.05 as criterion for exclusion. This is an automated regression technique whereby a full set of *k* variables is included into an initial model and, in a first step, the least significant variable according to Wald's chi-square test (highest p-value) is excluded. In a second step, the model is fit anew, and once again, the least significant of the remaining *k-1* variables is

eliminated from the model. This procedure is repeated until only significant variables (i.e. those with p-values below the pre-specified cut-off point) remain in the model, which then constitutes the final model. This is necessary because the basic idea of propensity score methods is to substitute all confounding variables in a study with a function of these covariates (a single predictor), namely the propensity score (7). This single predictor should combine significant individual predictors of treatment assignment, and yet be "parsimonious" – i.e. avoiding the inclusion of highly inter-correlated variables that would contribute little additional predictive value to the propensity score. The backward elimination procedure accomplishes this by eliminating the least significant variables first and only retaining variables that are significant in the presence of other variables. However, such an automated approach has the drawback of being entirely "data-driven" – in other words, no *a priori* knowledge influences the model selection process, except in the selection of the variables that are entered into the initial model. If power is limited, one runs the risk of eliminating potentially clinically relevant variables simply because they do not reach statistical significance. Moreover, in the smaller datasets, the multiplicity of data-dependent choices results in unstable selections, with the subsets of selected variables varying across repeated samples from the same population, especially when "candidate variables" are inter-correlated. Finally, the fact that variable selection is p-value-driven induces inflation of type I error, because classical statistical inference assumes the regression model has been fully specified *a priori* (8). However, given the very large size of the datasets used in this thesis, these problems are minimized. Any variable that is not statistically significant will have an estimated effect very close to the null value, and can, therefore, be safely eliminated. Similarly, the very large size of the samples ensures stability of selection and, thus,

accuracy of statistical inference (8).

The list of all initial models used in this thesis is provided in Table 2, together with the variables considered in each model. The numbering is consistent throughout the thesis, so that "Model 10" in Manuscript 2 refers to the same initial model as "Model 10" in Manuscript 3, for example.

Table 2. Variables included in the initial step of backward elimination, for the various models used throughout this thesis.

Model No.	Initial variables included into model
1	Full model: sex, age, rural site of residence, season of index prescription (dummy variables: winter, spring, fall), number of dispensations in month prior to index date, and, in year prior to index date: • number of hospitalizations, • average duration of hospitalization • number of visits to GP • number of visits to specialist • number of visits to rheumatologist • number of visits to gastroenterologist • number of visits to emergency room • diagnosis of cancer • diagnosis of heartfailure • diagnosis of musculoskeletal disease • number of dispensations of aspirin • number of dispensations of NSAIDs • number of dispensations of acetaminophen • number of dispensations of steroids • use of anticoagulants • use of opiates • number of GI hospitalizations • ulcer • dyspepsia • prophylaxis
2	Full model including all potential covariates except medical service utilization variables (number of hospitalizations, average duration of hospitalisation, number of visits to GP, number of visits to specialist, number of visits to rheumatologist, number of visits to gastroenterologist, number of visits to emergency room in year prior to index date)
3	Full model including all potential covariates except prior GI events (GI hospitalizations , ulcer, dyspepsia, prophylaxis in year prior to index date)
4	Model including all potential covariates except prior GI events and medical service utilization variables
5	Full model excluding musculoskeletal disease in year prior to index date
6	Full model excluding musculoskeletal disease, GI prophylaxis and sex in year prior to index date
7	Full model excluding GI prophylaxis in year prior to index date
8	Full model excluding GI prophylaxis and number of hospitalizations in year prior to index date
9	Full model excluding rural site of residence

10	Full model excluding musculoskeletal disease, GI prophylaxis, sex and number of NSAID dispensations in year prior to index date
11	Full model excluding number of NSAID dispensations in year prior to index date
12	Full model excluding GI prophylaxis in year prior to index date

Most of the count variables, such as numbers of prescriptions or numbers of visits to health professionals were skewed to the right, some quite severely. While we could have used non-linear functions of generalized linear models to correct for this, we elected not to, because this would have greatly complicated the methods we are proposing, thus making them less accessible to potential users. However, we did screen for severe departures from the assumption of linearity of continuous independent variables by categorizing these variables and running our models using these categorized variables instead. This did not alter our final models or our parameter estimates in any important way, so that we kept the untransformed version of our variables in logistic regression modelling. Nonetheless, it should be kept in mind that the linearity assumption inherent to parametric models such as regression modelling may lead to biased estimates of effects; non-parametric methods me produce less biased estimates under similar conditions (9).

3.3.3 Variable selection

Variables were selected for inclusion into our initial regression models because they were either relevant descriptive variables, such as sex and age, or because they were suspected, on the basis of clinical knowledge or prior studies, to be relevant predictors of treatment assignment to COX-2 inhibitors, acetaminophen or NSAIDs (for example: the number of prescriptions for NSAIDs in the year preceding the index date) or of the risk

of GI events (for example: prior GI events, use of anticoagulants in year prior to index date) (10;11). Backward elimination was then applied, and only variables significant at the 0.05 level for the Wald chi-square test at each step we retained in the model and reassessed at the next elimination step. In other words, no variables were "forced" into the model. Potential drawbacks of backward stepwise regression are of little concern because the very large number of patients in the study bases ensures that results are very stable and that, furthermore, any effect that is statistically non-significant is of no clinical relevance.

3.3.4 Interaction terms

Interaction terms were only considered on an *a priori* knowledge basis. Since there was no reason to suspect any relevant interactions in the clinical situations we considered, no interaction terms were included into our regression models, with the exception of some models in manuscript 3. In this setting, we attempted to assess the extent of residual confounding in the association between treatment assignment and GI events. In doing so, we postulated that the discrepancy between the treatment actually received and the treatment predicted by propensity scores on the basis of documented factors would vary in relation to the magnitude of residual channelling. Furthermore, we postulated that this phenomenon might be different depending on the actual treatment assignment received. In order to test these hypotheses, we performed regression analyses which included a term for the interaction between actual treatment assignment and discrepancy. These issues are addressed in more detail in manuscript 3.

3.3.5 Outliers

In preparing the data for regression analysis, we did not exclude outliers. We deliberately chose to do so because we wanted our data to reflect the full range of

variations encountered in real life. We did, however, screen for the plausibility of the values of outliers. Had obviously unrealistic values been encountered – say 1200 emergency room visits in one year – then this patient would have been excluded from the cohort, however this was not necessary.

3.3.6 Calculating the IAC

To calculate the IAC, we used the following procedure, which we describe here in detail for example 1. The procedure for the other examples was similar, except that the treatment assignment alternatives were different. First, we calculated propensity scores by assessing the importance of potential determinants of treatment assignment. This was done by running a backward logistic regression model with the treatment assignment as the dependent variable (COX-2 (celecoxib) coded as = 1, and all NSAIDS (diclofenac, diflunisal, etololac, fenoprofen, flurbiprofen, ibuprofen, indomethacine, ketoprofen, mefenamate, nabumeton, naproxen, phenylbutazone, piroxicam, salsalate, tenoxicam, tiaprofenate, and tolmetine) coded as = 0). Initially, all the potential determinants of treatment assignment were included in the model as independent variables, and backward elimination was applied.

Propensity scores for individual subjects were then calculated using the regression coefficients estimated in the "final" model. Based on the empirical distribution of propensity scores, the index of apparent channelling (IAC) was then calculated for each patient group (12). The IAC is described in more detail in Manuscript 1 (12).

3.3.7 Sensitivity analyses

For each of the three examples, the responsiveness of the IAC was illustrated by performing sensitivity analyses whereby potential confounders that were *a priori*

considered to be predictors of channelling, namely prior GI events and markers of health services utilization, were eliminated (11). We also performed sensitivity analyses by omitting the variables that were found to be the most significant determinants of treatment assignment, based on their regression coefficients in the full model.

In a first step, we ran a propensity score model including all potential confounders ("full model") and applied the IAC as described previously. We then repeated the procedure without the medical service utilization variables and prior GI events variables, as defined above, individually and together in the multivariable regression model used to estimate propensity scores, and then observed the effect of these omissions on the IAC. It is expected that omitting a known potential confounder should reduce the value of the IAC by shifting the "burden" of channelling, since the part of channelling due to these factors would then fall in the realm of the "undocumented", and thus outside of what can be accounted for by propensity scores.

In additional analyses, we subdivided the patient population into those treated by GPs and those seen by specialists. The propensity scores and the IAC were calculated separately for each sub-cohort and then compared. This was done using the observed proportion of the sub-cohorts as well as propensity scores that had been estimated using the propensity score model obtained from that particular stratum.

3.4 References

1. Tamblyn R, Lavoie G, Petrella L, Monette J. The use of prescription claims databases in pharmacoepidemiological research; the accuracy and comprehensiveness of the prescription claims database in Quebec. J Clin Epidemiol 1995;48:999-1009.

2. Rahme E, Joseph L, Kong XS, Watson DJ, LeLorier J. Gastrointestinal health care resource use and costs associated with nonsteroidal antiinflammatory drugs versus acetaminophen; retrospective cohort study of an elderly population. Arthritis Rheum 2000;43:917-24.

3. Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of beta 2-agonists. Am J Epidemiol 1996;144:1161-9.

4. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992;45:197-203.

5. Rahme E, Pettitt D, LeLorier J. Determinants and sequelae associated with utilization of acetaminophen versus traditional nonsteroidal antiinflammatory drugs in an elderly population. Arthritis Rheum 2002;46:3046-54.

6. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

7. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757-63.

8. Abrahamowicz M, MacKenzie T, Esdaile JM. Time-Dependent Hazard Ratio: Modeling and Hypothesis Testing With Application in Lupus Nephritis. Am J Epidemiol 1996;91:1432.

9. Abrahamowicz M, du BR, Grover SA. Flexible modeling of the effects of serum cholesterol on coronary heart disease mortality. Am J Epidemiol 1997;145:714-29.

10. Rahme E, Marentette MA, Kong XS, LeLorier J. Use of NSAIDs, COX-2 inhibitors, and acetaminophen and associated coprescriptions of gastroprotective agents in an elderly population. Arthritis Rheum 2002;47:595-602.

11. Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.

12. Bjerre, L. M., Abrahamowicz, M. A., and LeLorier, J. Channelling in non-experimental pharmacoepidemiologic research: The estimation of apparent channeling Part I: Theory and hypothetical examples. 2003. Ref Type: Thesis/Dissertation.

4 Preface to Manuscript 1

This is the first in a series of three articles that addresses the phenomenon of channelling in non-experimental research and its impact on confounding by indication. As a first step in the investigation of this broad topic, this manuscript develops a conceptual framework to help better understand the phenomenon of channelling and its relation to confounding by indication. Next, based on this conceptual framework, a novel approach, the Index of Apparent Channelling (IAC), is proposed with a view to quantifying the phenomenon of channelling due to documented factors, or apparent channelling. This IAC is first explored in a theoretical framework, by assessing its behaviour under a variety of clearly defined, often extreme conditions. Then, the IAC is applied to a variety of less extreme, hypothetical examples that make use of fictitious cohorts with pre-specified characteristics. Theses hypothetical examples illustrate situations closer to what could be found in non-experimental database studies. The IAC has been found to be sensitive to the degree of channelling under both extreme and less extreme channelling conditions. Based on the results of these hypothetical examples, the IAC will be applied, in the second article, to empirical data, in order to illustrate its behaviour in real-life analyses of large administrative prescription databases.

Channelling in non-experimental pharmacoepidemiologic research:

Manuscript 1:
The estimation of apparent channeling
Part I: Theory and hypothetical examples

Lise M Bjerre, MSc, MD
PhD candidate
Department of Epidemiology and Biostatistics
McGill University

Michal Abrahamowicz, PhD
James McGill Professor
Department of Epidemiology and Biostatistics
McGill University

Jacques LeLorier MD, PhD
Chief
Pharmacoepidemiology and Pharmacoeconomics Unit
Centre de recherche
Hôtel-Dieu du Centre hospitalier de l'Université de Montréal
Université de Montréal

Corresponding author: Dr. Jacques LeLorier
Centre de recherche, Hôtel-Dieu du Centre hospitalier de l'Université de Montréal
Room 8-223, Pavillon Masson
3850 Saint-Urbain
Montréal, Québec, Canada H2W 1T8
Tel: (514) 843-2918
Fax: (514) 843-2774
E-mail : jacques.le.lorier@UMontreal.Ca

Running head: Measurement and control of apparent channelling

Keywords: Pharmacoepidemiology, bias, non-experimental research, observational research, medications, confounding, confounding by indication, confounding by severity, propensity scores, channelling, apparent channelling.

Text only word count: 3753

4.1 Abstract

PURPOSE: Non-experimental (observational) research is essential to ascertain the safety of drugs once they have been released onto the market. Yet, observational studies are prone to bias. One particularly difficult issue is that of confounding by the presence or absence of an indication or its severity, which results from channelling. It is often impossible to adequately document and measure the variables necessary to control for this phenomenon in the analysis. Therefore, there is a need to develop more adequate approaches to measure, and possibly control for, channelling.

OBJECTIVES: We propose a conceptual framework and an index that allow for the identification and quantification of channeling due to documented factors and have the ability to discriminate between clinically relevant situations.

METHODS: Our approach relies on breaking down the problem of confounding by indication into its component elements, namely (i) channelling and (ii) the risk factor effect. The risk factor effect is often known *a priori* and its magnitude can be approximated from the relative risk estimates of prior studies. To measure channelling, we propose the index of apparent channelling (IAC), which makes use of the observed distribution of propensity scores to assess the strength of the effect of documented factors on treatment assignment.

RESULTS: Hypothetical examples demonstrate the ability of the proposed index to accurately quantify the degree of apparent channelling under various assumptions. Real-life illustrations are provided in the companion paper.

CONCLUSION: The proposed approach and index provide a valid and practicable solution to the measurement of apparent channelling, one of the components of confounding by indication in non-experimental pharmacoepidemiological research.

4.2 Introduction

The safety and effectiveness of drugs released onto the market is of prime concern to consumers, physicians, regulatory drug agencies and the pharmaceutical industry. Although the requirements that need to be met in order to release a drug onto the market have become markedly more stringent over the past decades, they may not be sufficient to ensure safety. Pre-marketing studies required for licensing are limited by their moderate size, short follow-up and often highly selected patient populations (1). Therefore, post-marketing surveillance, which includes adverse drug event monitoring as well as pharmacoepidemiologic studies, is necessary in order to detect adverse drug reactions, as well as to provide further evidence of safety and effectiveness (1). Unlike clinical trials, in the post-marketing context, it is impossible to randomize patients to a given treatment, often for ethical as well as practical reasons. Hypothesized exposure-outcome associations must therefore be estimated from non-experimental studies (2). Among other approaches, computerized databases constitute a powerful and affordable tool for studying associations between drug exposure and adverse drug effects in the post-marketing context. However, they are also subject to the potential biases inherent to non-experimental research.

Broadly, there are three mechanisms in non-experimental research by which estimates of association may be biased: selection bias, information bias (misclassification) and confounding (3). Appropriate study design can reduce or eliminate the first two sources of bias, however confounding poses particular problems requiring more sophisticated solutions (3). Confounding can often be dealt with in the analysis, but this is conditional on adequate documentation of relevant covariates at the time of data collection.

Confounding by indication is a type of confounding bias that is particularly problematic in non-experimental pharmacoepidemiologic studies (4;5). It arises when the presence of a disease or medical condition (the indication), that is prognostically related to the outcome of interest, is distributed unequally across treatment groups. Such an imbalance may lead to a biased estimate of the treatment-outcome association and to the erroneous conclusion that one of the treatment alternatives is associated with poorer or better outcomes (6). Confounding by severity of indication is a special type of confounding by indication whereby the severity of disease – instead of its presence or absence – acts as a confounding factor (7).

Controlling for confounding by indication is problematic in any non-experimental investigation, and particularly in retrospective studies using hospital charts or administrative databases as data sources, where there is often little or no information about potentially confounding patient characteristics and/or disease severity. For this reason, some researchers view confounding by indication as a possibly insurmountable problem, the Achilles' heel of pharmacoepidemiology (6;8;9).

Nonetheless, some approaches have been proposed that attempt to deal with the problem of confounding by indication. Among others, the use of compound indices of severity such as the chronic disease score (10), propensity scores (11) and their use as a stratification variable (12) have been put forth. Although they produce less biased estimates of treatment effect, these methods are limited by their total reliance on documented factors. Furthermore, stratification, usually into quintiles of propensity scores, provides at most a relatively crude adjustment for confounding by indication, raising the issue of residual confounding (13). Therefore, there is a need for a more

refined approach to address this issue in a quantitative way.

In the present study, we propose a conceptual framework whereby confounding by indication is broken down into its component elements, namely channelling and risk factor effect, which describes the association between a potential confounder and the outcome of interest. We then extend the use of propensity scores and propose a new construct, the index of apparent channelling (IAC), to measure channelling due to documented factors in non-experimental studies, a necessary first step toward identification and control of confounding by indication. We illustrate the application of the IAC using hypothetical examples where varying degrees of apparent channelling are expected. Empirical examples will be provided in a companion paper.

4.3 Methods

4.3.1 Conceptual framework and definitions

4.3.1.1 Components of confounding by indication

A factor can act as a confounder for a given exposure-outcome association only if it is related to both the exposure and to the outcome of interest, and does not lie along the causal pathway of exposure to disease (3). By analogy, we suggest to break down confounding by indication into its two components: channelling and risk factor effect, both of which must be present for confounding by indication to occur. Channelling occurs when certain patients are preferentially assigned to one or another treatment, be it for known or unknown reasons (14). The risk factor effect describes the association between a potential confounder and the outcome of interest. The strength of this association can often be estimated from earlier studies. The relationships between these components are illustrated in Figure 1.

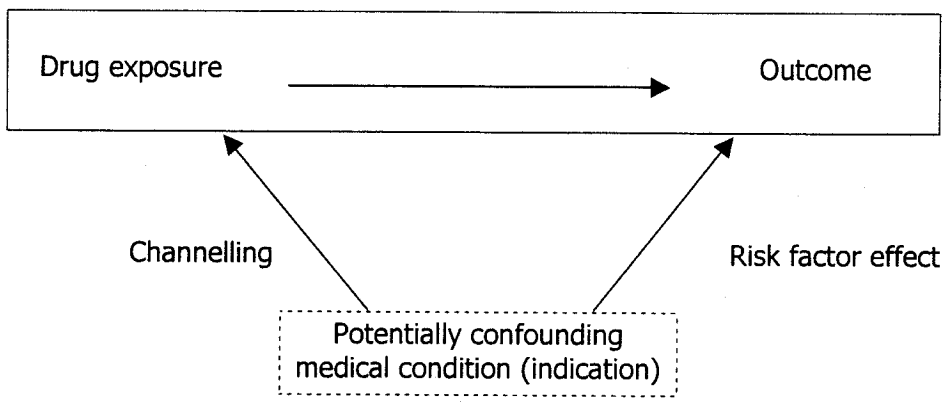

Figure 1. Channelling and risk factor effect in relation to confounding by indication

Although, like the risk factor effect, it is also a *sine qua non* condition for confounding, relatively little attention has been devoted in the literature to the problem of channelling, its identification and quantification. It was first formally defined in the literature by Petri

and Urquhart (14). Later, Blais and colleagues distinguished between two forms of channelling, namely channelling due to differential first-time use by patients with different prognostic characteristics, and channelling resulting from a switch from one product to another within a therapeutic class by patients having failed to improve under the first therapy (15). The authors emphasized the importance of a longitudinal study design in distinguishing between these two forms of channelling and stated that the exposure in case-control studies needs to be assessed over a period of time of sufficient length to cover the complete history of drug use, even including a period of observation before the initiation of therapy. However, information about patients' complete drug histories is often lacking.

Channelling has been shown to occur in a number of non-experimental studies (15-17). Although, as described above, some methods have been proposed to address the overall problem of confounding by indication, to our knowledge, no method has been put forth that specifically focuses on the estimation of channelling. Of the methods proposed to tackle the issue of confounding by indication, propensity score analysis is one that, despite its limitations, has gained in popularity and been applied increasingly in recent years.

4.3.1.2 Propensity scores

Propensity scores were proposed by Rosenbaum and Rubin in 1983 as one approach to control for confounding by indication (11). Propensity scores are obtained by estimating a logistic regression model, with the exposure to the medication of interest being used as the dichotomous dependent variable. Potential predictors of treatment allocation are entered into the model as independent variables and a regression is run to determine

the best fitting model. The estimated logistic regression coefficients are then employed to calculate, for each patient, the expected probability of being assigned to a particular treatment, i.e. the propensity score. Therefore, propensity scores express the probability of a patient being assigned to one or another treatment, based on known covariates that influence treatment allocation (11).

Stratifying patients into categories on the basis of their propensity score allows one to carry out a stratified analysis that has the useful property of adjusting for all the covariates that went into the estimation of the propensity score, no matter how many there are, without having to rely on model-based assumptions and that may be violated, which is the case with standard methods of analysis such as linear or logistic regression (12). For example, Perkins *et al.* (18) and Margolis *et al.* (19) have applied propensity scores to control for confounding by indication in longitudinal non-experimental studies. Using propensity scores as a stratification variable produced a less biased estimate of treatment effect than when stratification was not used (18). Furthermore, Margolis and colleagues assessed the ability of their propensity score model to discriminate between patients assigned to one of two treatment alternatives by measuring the area under the receiver-operating characteristic (ROC) curve, which revealed that the model performed very well (estimated area under ROC of 0.9) (19). However, propensity scores can only control for documented covariates. If data are available on only some of the relevant confounders, the amount of bias will be reduced but residual confounding will still be present (19).

Since propensity scores express the likelihood that a patient will be assigned to one or another treatment based on known covariates, we hypothesized that they should be

distributed differently depending on the presence or absence of strong channelling factors. If a strong, documented channelling factor, such as an absolute contraindication to a treatment, were present, the distribution of propensity scores should be sharply bimodal, with all patients with the contraindication being assigned scores very close to zero. Other patients without the contraindication may have different scores depending on the values of other weaker determinants of the treatment choice. The distribution of propensity scores in such a situation could resemble the graph in Figure 2.

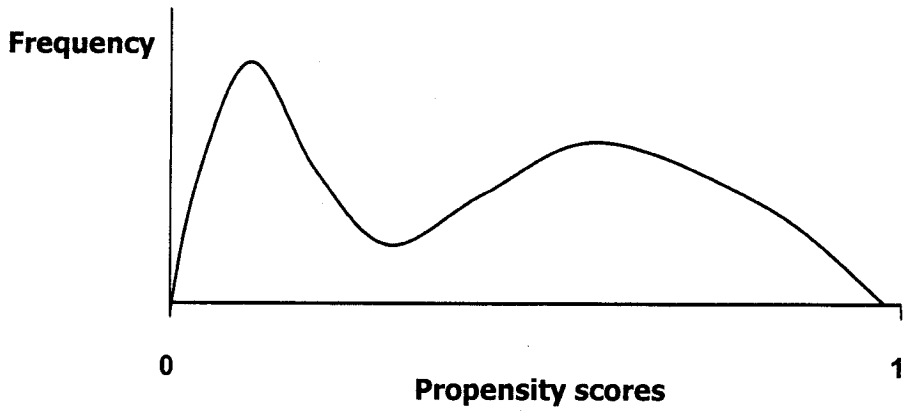

Figure 2. Possible distribution of propensity scores when strong channelling is present

On the other hand, if no documented factor bears strongly on the choice of treatment, the probability of being assigned to one or another treatment as predicted by the propensity score model would be close to the treatment's observed proportion, and the distribution of propensity scores would be unimodal and could look like Figure 3.

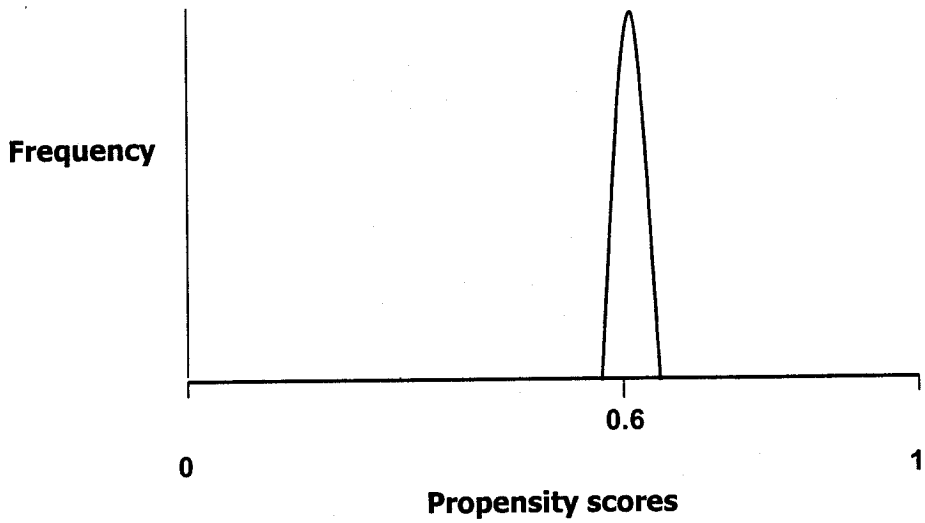

Figure 3. Expected distribution of propensity scores when very little channelling is present (here observed proportion receiving treatment 1 is = 0.6)

Ideally, one would like to be able to measure channelling in order to then be able to control for it in the analysis. Alternatively, if this were not feasible, it would be desirable to identify sub-populations where there is no channelling, and consequently no potential for confounding by indication. Figures 2 and 3 suggest that information conveyed by the distribution of propensity scores can be useful for these purposes.

Although a graph of a propensity score distribution can give a general idea of whether and to what extent channelling is present, it would nonetheless be useful to have a way of expressing the degree of channelling quantitatively. Such a summary score would be particularly useful when assessing and comparing sub-populations with intermediate degrees of channelling.

4.3.2 The Index of Apparent Channelling (IAC)

4.3.2.1 Derivation

In order to measure channelling due to documented factors, which we term "apparent channelling", one could consider the following type of index:

$$Index = \frac{\sum_{i=1}^{n}(p_i - m)^2}{n} \qquad (1)$$

where: p_i = estimated propensity score of individual patient *i*, *i*=1, ...,..., n;

m = observed proportion, i.e. proportion of all patients in the study base who are treated with the drug coded=1, (which is also equal to the average propensity score for all patients in the study base), and

n = total number of patients in the study base

This index quantifies the average spread of propensity scores in a given population and, therefore, expresses the degree of channelling within this population that is due to documented patient characteristics that were included in the multivariate model used to estimate the propensity scores.

The smallest possible value of this index is 0, a situation that would arise if all patients had a propensity score equal to m, the observed proportion – in other words, if there were no channelling at all and patients were randomly assigned to one or the other

treatment, with a probability of *m* and *1-m*, respectively. The theoretical maximum of the index is equal to the maximum variance of the propensity score, which would correspond to a complete polarization of propensity scores at either 0 or 1, i.e., to a situation where characteristics of each individual patient completely determined the choice of treatment. In this case, the variance of propensity scores would be obtained from the binomial distribution, and would equal *m(1-m)*. Thus, the range of possible values of the index in equation (1) depends on the observed proportion.

In order to make this index of channelling comparable across studies of different sizes and with different observed proportions, it is necessary to rescale the index so that it has a fixed range of possible values. Dividing the index in equation (1) by the maximum value it can take on results in the following definition, which we call the "index of apparent channelling" (IAC):

$$IAC = \frac{\sum_{i=1}^{n} (p_i - m)^2}{nm(1-m)} \qquad (2)$$

Definition (2) ensures that the IAC values are restricted to a [0,1] interval. In order to make it easier to interpret, one can express the IAC in percentage points instead of as proportions, by multiplying equation (2) by 100. In the rest of this article, we express the IAC in percentage points. Notice that, for analyses limited to some subsets of the study population, the propensity scores p_i must be obtained from the model estimated from that particular subset of the population, and the observed proportion *m* should also reflect the frequency of a given treatment in the relevant sub-population. Otherwise, the

theoretical range of IAC values may differ from a [0,1] interval.

4.4 Results

4.4.1 Hypothetical examples of applications of the index of apparent channelling

To illustrate the properties of the IAC, it is useful to apply it to simulated data where the distribution of propensity scores can be manipulated. To this end, we simulated a hypothetical cohort of 1000 patients to which we then randomly assigned propensity scores according to pre-specified frequency distributions. These were used to calculate the IAC, using equation (2) for each distribution.

4.4.1.1 Extreme situations

To illustrate the full range of values the IAC can take on, we started by simulating extreme situations. One extreme situation is that of no channelling at all, where all patients have an equal chance of being assigned to one or the other treatment. This should be the case in randomized controlled trials with a 1-to-1 randomization. Such a situation gives rise to propensity scores of 0.5 for each patient, with the IAC therefore being equal to zero as expected. This is illustrated in figure 4.

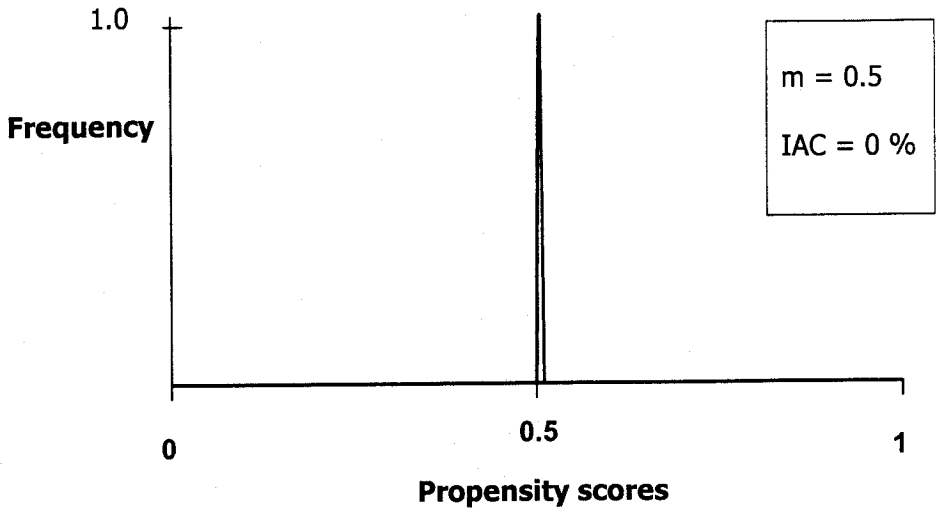

Figure 4. Distribution of propensity scores and IAC for a randomized controlled trial with 1-to-1 randomization

Logically, any randomized controlled trial should be expected to have an IAC of zero, regardless of the randomization ratio (assuming that the randomization has been effective, i.e. assuming a sufficiently large *n* to ensure, on average, an equal distribution of covariates across the two treatment groups). This is indeed the case. If patients are randomized in a 2-to-1 fashion (i.e. 2/3 get treatment A and 1/3 treatment B), then the average propensity score (the "observed proportion") becomes 0.67 (assuming that treatment A is coded as 1), but the IAC is nonetheless equal to zero. This is illustrated in figure 5.

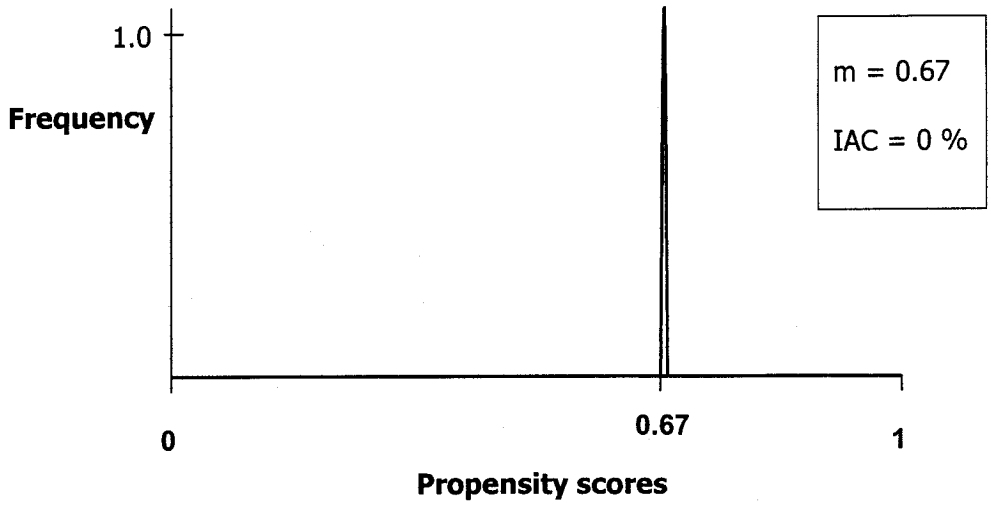

Figure 5. Distribution of propensity scores and IAC for a randomized controlled trial with 2-to-1 randomization

The other extreme is one of absolute channelling, where patients' characteristics completely determine the choice of treatment, so that each individual patient would have a 100% chance of being assigned to a given treatment. In practice this is an unusual situation that could nonetheless arise if there were an absolute contraindication to a treatment that is observed by physicians as well as a restriction (say due to cost) that forbade using this treatment for any patient without the contraindication. In such a case, the propensity scores would be highly segregated, with patients having either a propensity score very close to 1 (no contraindication) or to 0 (contraindication present), with no intermediate values. This could look like the situation depicted in figure 6, assuming a 1:1 ratio of treatment assignment. As expected, the IAC would be equal to 100 %, i.e. it would reach its theoretical maximum.

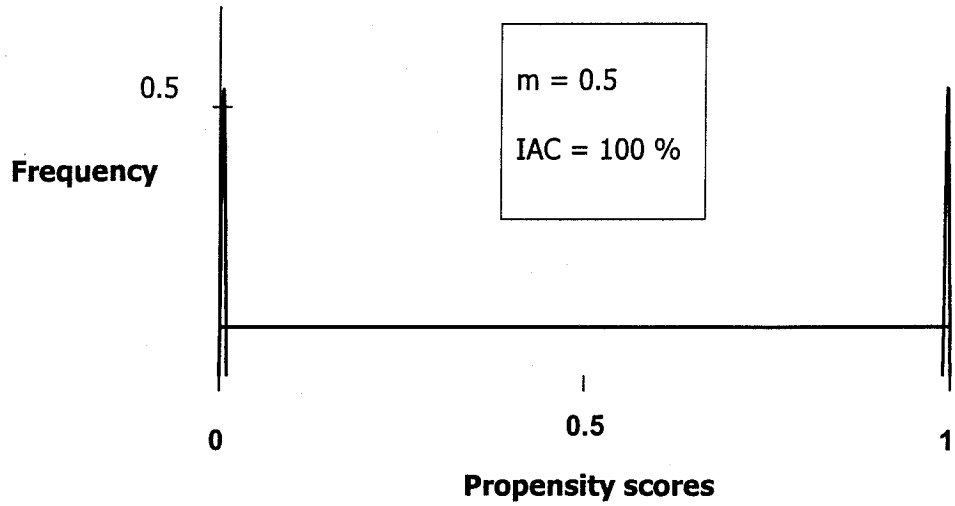

Figure 6. Distribution of propensity scores and IAC when complete channelling is present (equal assignment to treatment groups)

One can also imagine a situation where there is extreme channelling, but where patients are assigned unequally to the two treatment alternatives because the relevant absolute contraindication is very rare. For example, this could be the case if only patients with an allergy to penicillin were given an alternative treatment (for example, erythromycin) for the treatment of an infection. In this situation, the IAC also behaves as expected, taking on its maximum value of 100 %, as illustrated in figure 7.

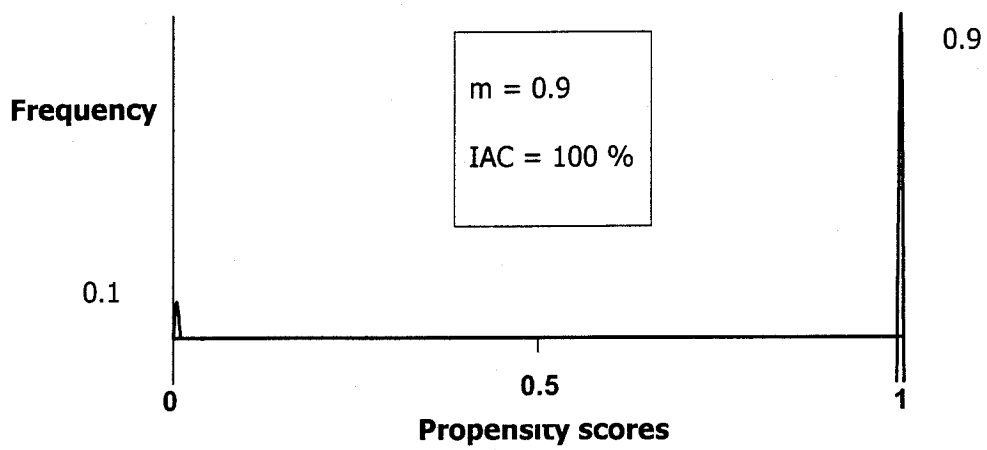

Figure 7. Distribution of propensity scores and IAC when complete channelling is present (unequal assignment to treatment groups)

4.4.1.2 Intermediate situations

To illustrate situations more realistic than the extremes shown in Figures 4 through 7, we constructed hypothetical examples where there is only a mild degree of channelling. Figure 8 illustrates the case where the propensity scores cluster around the "observed proportion" (0.5 in this case) with some degree of spread on both sides. This hypothetical distribution has an IAC of 13 %. This value is close to zero but indicates that there is some degree of channelling. The dotted lines represent the theoretical distribution from which the data was generated.

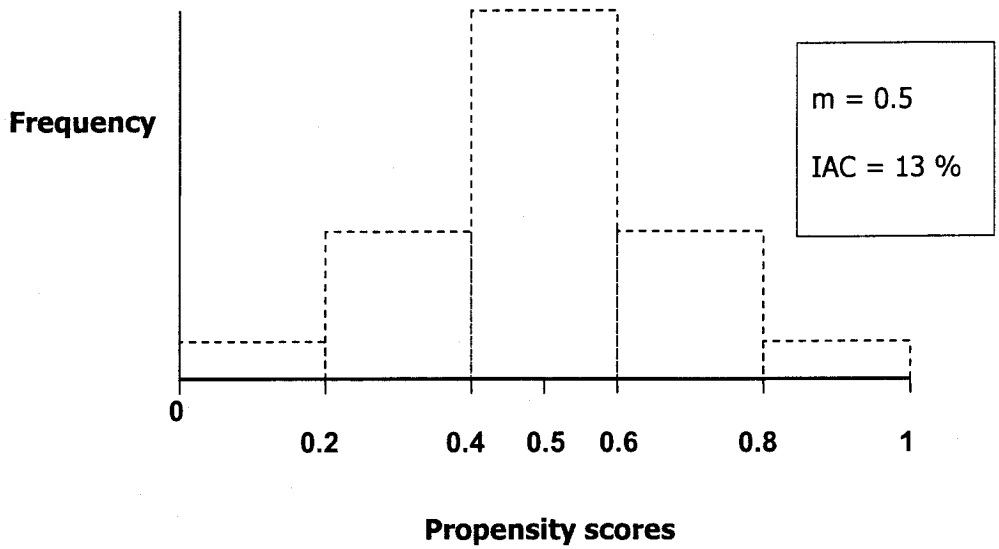

Figure 8. Distribution of propensity scores and IAC for situation with mild degree of channelling

An intermediate situation is depicted in figure 9. Here, there is rather strong apparent channelling. This type of distribution would be expected if there were a strong contraindication to one treatment and subjects who didn't have this contraindication usually received the other treatment. However, in both cases there may be exceptions, so the estimated propensity scores are not concentrated extremely close to either 0 or 1. The corresponding IAC is 55 %, indicating the presence of stronger apparent channelling than in figure 8.

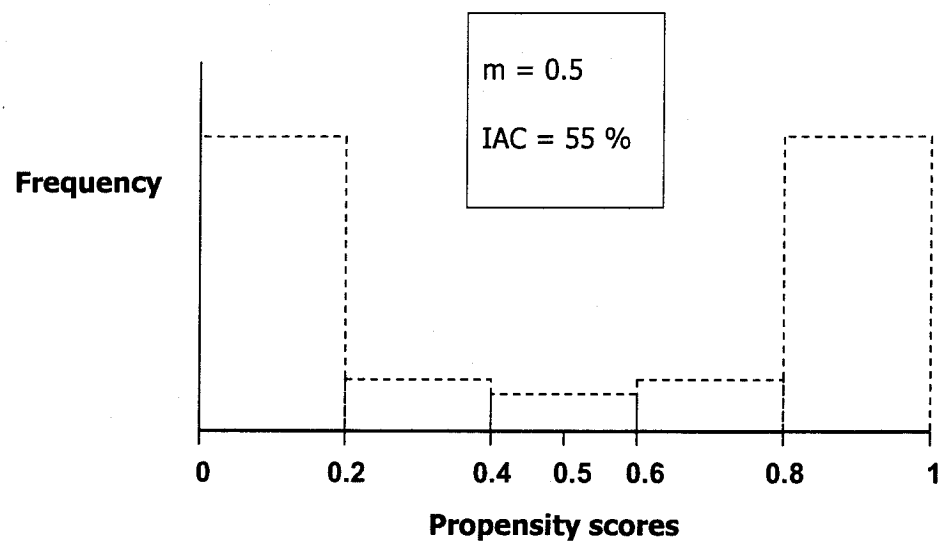

Figure 9. Distribution of propensity scores and IAC for situation with stronger degree of channelling

These two examples illustrate the behaviour of the IAC under different conditions and show that it quantifies well the degree of channelling due to documented factors.

In Figures 8 and 9, it can be seen that the observed proportion corresponds to the average propensity score for the population at issue. Note also that equation (2) ensures that any "asymmetry" in the observed proportion does not strongly influence the IAC. To illustrate this point, Figures 10 and 11 show situations similar to figures 8 and 9 respectively, but with a observed proportion of the drug coded 1 reduced to 30 % (m=0.3). In spite of this asymmetry, the resulting IAC values still discriminate well between the unimodal distribution (less channelling) in figure 10 (IAC = 33 %) and the bimodal distribution (more channelling) in figure 11 (IAC = 49 %).

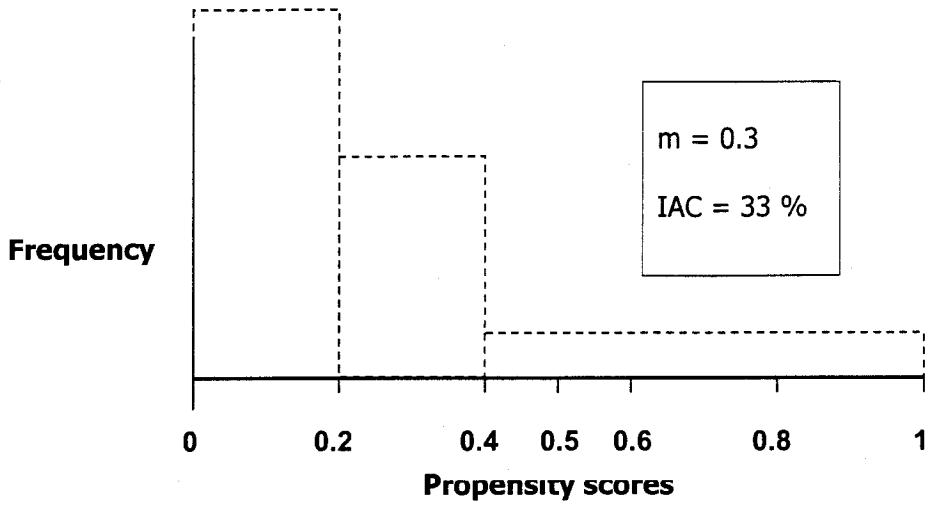

Figure 10. Distribution of propensity scores and IAC for situation with a moderate degree of asymmetric channelling

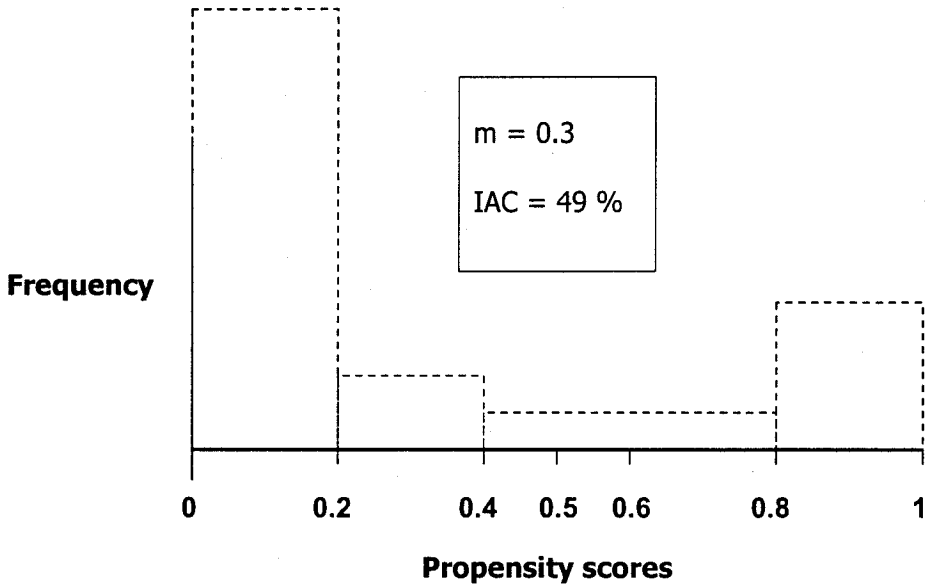

Figure 11. Distribution of propensity scores and IAC for situation with a stronger degree of asymmetric channelling

4.5 Discussion

Confounding by indication is an important source of potential bias in non-experimental drug research. However, one must ask not only "could bias be present?", but also "in what direction does it act, how large is it, and what is its impact on the estimate of association?". Indeed, Rothman and Greenland argue that it is the amount of confounding that is important to evaluate, rather than its mere presence or absence (3). Channelling and risk factor effect are both *sine qua non* conditions for confounding by indication. The quantification of relative risks associated with different factors is at the very core of epidemiological research. In contrast, the quantification of channelling has received relatively little attention so far. In this paper, we propose an explicit conceptual model of confounding by indication, consisting of two necessary components, channelling and risk factor effect. Such a conceptual model of confounding by indication helps to lay the foundations for the development of tools to measure its components, which is a necessary first step in the development of methods to control for confounding by indication.

Propensity scores describe the likelihood of patients being assigned to one or another treatment (11). We took this concept one step further and used the distribution of propensity scores to quantify the overall degree of channelling toward one or the other treatment by summarizing them into a single index of apparent channelling (IAC). We standardized the index so that it ranges from 0 to 1 and its magnitude is independent of the study size and of the observed proportion of the two products of interest, thus allowing for comparisons across studies or patient groups of different sizes and different populations.

In our hypothetical examples, we demonstrated how well the IAC responds to channelling. In extreme situations, the IAC was shown to take on values approaching its theoretical maximum and minimum. In intermediate situations, it was shown to respond to various degrees of spread of propensity scores, regardless of the asymmetry of the propensity score distribution.

The IAC can also be applied to sub-populations within a larger database. It might be tempting, when dealing with a sub-population, to apply the IAC using propensity scores and the observed proportion derived from the entire population in order to measure any channelling that may be present at the level of the sub-population. This is problematic, however. Since the propensity scores and the observed proportion of the two treatment options at issue would be derived from the whole study base, the IAC applied to a sub-population of the study base would express not only the variability of propensity scores within that sub-population, but also the discrepancy between this sub-population's mean propensity score and the mean propensity score of the total population (i.e. with respect to the population observed proportion, which may be different from that of the sub-population). Furthermore, variables that may contribute to channelling at the level of the whole database may not play the same role in a sub-population and vice versa. Therefore, in order to assess channelling in a sub-population, it is necessary to repeat the whole procedure, namely to recalculate the propensity scores and the observed proportion for the patients in the sub-population and to calculate the AIC using these sub-population specific propensity scores and observed proportion values.

A limitation of the IAC is that, like propensity scores (12), it can only account for documented covariates that may impact on treatment assignment. If undocumented

factors play an important role in treatment assignment, this will not be captured by propensity scores and, consequently, will not be expressed by the IAC. Furthermore, the IAC only applies to situations where there is a choice between two drugs or two drug classes (ex.: COX-2 inhibitors vs NSAID). Again, this limitation also applies to the classical application of propensity scores (12). Alternatively, it is possible to artificially "dichotomize" a situation by defining the two treatment alternatives for a given indication, for example, "Drug A" and "all other treatments except drug A". Future research may aim at developing a more general approach to propensity scores based on polytomous regression models for predicting choices between several alternative treatments. A further challenge that needs to be addressed is the measurement of channelling due to undocumented factors.

Nonetheless, the IAC provides a means of measuring the degree of channelling due to documented factors and of expressing it as a single quantitative variable, allowing for comparisons across populations of various sizes and different observed proportions. This cannot be achieved by solely considering the individual propensity scores or even their distribution. As such, the IAC is an important tool that adds to the epidemiologist's armamentum when it comes to tackling the issue of confounding by indication. In the companion paper, we illustrate some practical applications of the IAC in pharmacoepidemiology.

4.6 Acknowledgements

This project was supported by a grant from the "Réseau FRSQ sur l`utilisation des médicaments" of the government of the province of Québec. Michal Abrahamowicz is a scientist of the Canadian Institutes of Health Research (CIHR). Many thanks to Drs Elham Rahme and Lucie Blais for stimulating discussions and insightful comments at various stages in the development of this project.

4.7 References

1. Skegg DC. Evaluating the safety of medicines, with particular reference to contraception. Stat Med 2001;20:3557-69.

2. Strom BL (ed.). Pharmacoepidemiology. 3 ed. Chichester, UK: John Wiley and Sons, 2000.

3. Rothman KJ, Greenland S. Modern epidemiology. 2 ed. Philadelphia, PA: Lippincott-Raven, 1998.

4. Miettinen OS. The need for randomization in the study of intended effects. Stat Med 1983;2:267-71.

5. Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? Am J Med 1985;78:475-80.

6. Walker AM. Confounding by indication. Epidemiology 1996;7:335-6.

7. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol 1999;149:981-3.

8. Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. Ann Intern Med 1979;90:257-61.

9. Shapiro S. Confounding by indication? Epidemiology 1997;8:110-1.

10. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992;45:197-203.

11. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

12. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757-63.

13. Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. Epidemiology 1997;8:429-34.

14. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med 1991;10:577-81.

15. Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of beta 2-agonists. Am J Epidemiol 1996;144:1161-9.

16. Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.

17. Rahme E, Pettitt D, LeLorier J. Determinants and sequelae associated with utilization of acetaminophen versus traditional nonsteroidal antiinflammatory drugs in an elderly population. Arthritis Rheum 2002;46:3046-54.

18. Perkins SM, Tu W, Underhill MG, Zhou XH, Murray MD. The use of propensity scores in pharmacoepidemiologic research. Pharmacopepidemiol Drug Safety 2000;9:93-101.

19. Margolis DJ, Kantor J, Santanna J, Strom BL, Berlin JA. Effectiveness of platelet releasate for the treatment of diabetic neuropathic foot ulcers. Diabetes Care 2001;24:483-8.

5 Preface to Manuscript 2

This is the second in a series of three articles that addresses the phenomenon of channelling in non-experimental research and its impact on confounding by indication. In the first article in this series, a novel approach, the Index of Apparent Channelling (IAC), was proposed with a view to quantifying apparent channelling, the phenomenon of channelling due to documented factors. The behaviour of the IAC was illustrated using hypothetical examples and was shown to be responsive to various degree of apparent channelling in hypothetical patient cohorts.

In the present article, the IAC is applied to empirical data in order to illustrate its behaviour under real-life conditions and with various degrees of apparent channelling. To this end, three cohorts were formed where various degrees of apparent channelling were expected. In example 1, a cohort of current NSAID users was formed and followed to see if there was preferential channelling of patients at higher risk of gastrointestinal events to treatment with COX-2 inhibitors after the introduction of these drugs onto the market. In this case, moderate to strong apparent channelling was expected. In example 2, a cohort of new users of acetaminophen or NSAID was formed, and predictors of treatment assignment were examined to assess whether there is preferential channelling of patients at higher risk of gastrointestinal events toward therapy with acetaminophen. In this case, mild to moderate apparent channelling was expected. Finally, the third example involved a cohort of new users of two similar NSAIDs, naproxen and diclofenac. In this case, no preferential treatment assignment was expected.

The IAC is shown to respond reasonably well to the various degrees of channelling

present in these empirical cohorts, however, as expected, the range of values the IAC takes on is more restricted than in the hypothetical examples in Manuscript 1. Finally, the IAC remains limited by its sole reliance on documented factors. The issue of channelling by undocumented factors is the subject of the third article.

Channelling in non-experimental pharmacoepidemiologic research:

Manuscript 2: The estimation of apparent channelling Part II: Empirical examples

Lise M Bjerre, MSc, MD
PhD candidate
Department of Epidemiology and Biostatistics
McGill University

Michal Abrahamowicz, PhD
James McGill Professor
Department of Epidemiology and Biostatistics
McGill University

Marc Dorais, MSc
Biostatistician
Pharmacoepidemiology and Pharmacoeconomics Unit
Centre de recherche
Hôtel-Dieu du Centre hospitalier de l'Université de Montréal
Université de Montréal

Jacques LeLorier MD, PhD
Chief
Pharmacoepidemiology and Pharmacoeconomics Unit
Centre de recherche
Hôtel-Dieu du Centre hospitalier de l'Université de Montréal
Université de Montréal

Corresponding author: Dr. Jacques LeLorier
Centre de recherche, Hôtel-Dieu du Centre hospitalier de l'Université de Montréal
Room 8-223, Pavillon Masson
3850 Saint-Urbain
Montréal, Québec, Canada H2W 1T8
Tel: (514) 843-2918
Fax: (514) 843-2774
E-mail : jacques.le.lorier@umontreal.ca

Keywords: Bias, non-experimental research, observational research, medication, confounding, confounding by indication, confounding by severity, propensity scores, channelling, apparent channelling, pharmacoepidemiology

Text only word count: 6500

5.1 Abstract

PURPOSE: Confounding by indication is a serious threat to the validity of non-experimental studies in pharmacepidemiology. In the companion article, we introduced the index of apparent channelling (IAC) as a tool for the measurement of channelling due to documented factors, or apparent channelling. In the present article, we applied this index to empirical data and illustrated its behaviour in situations with various degrees of apparent channelling.

METHODS: We used data from the province of Québec's administrative prescription database. First, we applied the index of apparent channelling to a situation where channelling was previously shown to be present in a survey study, namely after the introduction of a new drug class (COX-2 inhibitors vs NSAIDs) onto the market. We also applied the IAC to other situations with differing degrees of expected channelling, to illustrate its performance. These situations concerned choices between: (a) NSAIDs and acetaminophen (moderate to strong channelling expected), and (b) two well-established NSAIDs (naproxen vs diclofenac - little channelling expected). We also carried out sub-group analyses separately for patients for whom the prescribing physician was a GP and a specialist (GPs being expected to channel less). Finally, we illustrated the reponsiveness of the IAC by assessing the impact of omitting known confounders from the propensity score models used to calculate it.

RESULTS: The IAC performed well when applied to real data, as had been predicted from our simulations presented in the companion article. The values for the IAC ranged from 0.3 % in the naproxen-diclofenac example to 17.1 % for the COX-2 inhibitor vs NSAID example. Interestingly, specialists were not found to channel

systematically more than GPs. In fact, this varied considerably from one example to the other.

CONCLUSION: We demonstrated the application of the IAC to real data. The IAC is a sensitive tool to measure channelling due to documented factors that responds reasonably well to the presence or absence of known confounders, and that can be used to compare apparent channelling across different populations.

5.2 Introduction

This is the second of two articles addressing some aspects of confounding by indication in non-experimental pharmacoepidemiologic research, and more specifically one of its components, namely channelling. Confounding by indication can arise when a disease or medical condition (the indication) that is prognostically related to an outcome of interest is distributed unequally across treatment groups. This may erroneously lead to the conclusion that one of the treatment alternatives is associated with poorer outcomes (1). Although some means of dealing with confounding by indication have been developed (2;3), these remain limited by their inability to quantify the degree of confounding. Confounding by indication continues to be considered one of the central weaknesses of non-experimental pharmacoepidemiologic research and especially of case-control studies attempting to assess adverse effects of drugs (4).

Confounding by indication can be broken down into its two components, risk factor effect and channelling, both of which are necessary conditions for confounding by indication. The risk factor effect is simply the association between the potential confounder and the outcome of interest. Channelling is the mechanism by which patients with particular constellations of the potential confounders may be systematically more likely to receive one treatment rather than another (5).

Although channelling has long been recognized – either implicitly or explicitly – as a component of confounding by indication (5), relatively little attention has been devoted to its quantification. Blais *et al.* demonstrated the presence of channelling using documented markers of severity and disease control in a prescription database of asthma drugs (6). However, this approach requires longitudinal data of sufficient

duration, which may often not be available.

To allow for the assessment of channelling in a wider range of studies, we proposed, in the companion paper, a tool for the measurement of channelling due to documented factors, namely the index of apparent channelling (IAC) (7). The IAC makes use of propensity scores, originally put forth by Rosenbaum and Rubin (3) that quantify the likelihood of a patient being assigned to one of two treatment alternatives, based on documented covariates and independently of the outcome. The IAC expands on the propensity score methodology and quantifies the overall tendency for individual patients to be channelled toward one or the other treatment by quantifying the spread of propensity scores with respect to their mean in the same group of patients. Highly segregated propensity scores (bimodal distribution) indicate strong channelling due to documented factors, whereas a distribution of propensity scores tightly centered around the average propensity score of the population (which corresponds to the observed proportion of the drug coded as 1) means there is little or no channelling due to documented factors (7). Hypothetical examples showed that the IAC is responsive to various degrees of channelling (7).

In the present article, we applied the IAC to *empirical* data to illustrate its performance under "real-life" conditions. By applying the IAC to situations where different degrees of apparent channelling are suspected – or known – to be at play, we demonstrated that the IAC is a measure of apparent channelling sufficiently responsive to be of practical use in non-experimental database studies.

5.3 Methods

5.3.1 Overview of empirical examples

In this study, we applied the index of apparent channelling (IAC) to real data, focusing on situations where more or less channelling is expected, based on prior substantive knowledge. First, we investigated a situation for which the presence of channelling after the introduction of a new drug class (COX-2 inhibitors vs NSAIDs) had been demonstrated in a previous non-experimental field study and, thus, where fairly strong channelling was expected (8). To do this, a cohort of current users of NSAIDs was formed and followed over time for six months after the introduction of the first COX-2 inhibitor on the Québec market to assess whether there was channelling of sicker patients to the new drug. Next, we considered two other situations where various degrees of channelling were expected: In the second example, we formed a cohort of new users of NSAIDs or acetaminophen at a point in time before the introduction of COX-2 inhibitors, and we assessed whether sicker patients were channelled preferentially to acetaminophen. Here we expected at least moderate channelling to be present. Finally, in the third example, we formed a cohort of new users of diclofenac and naproxen, two well-established anti-inflammatory drugs with fairly similar effectiveness and adverse effect profiles, and assessed whether there was differential channelling to these drugs. In this case, little or no channelling was expected *a priori*.

For each example, we also carried out analyses for different sub-groups within the cohorts (patients of GPs vs patients of specialists, with GPs expected to channel less). Finally, to assess the responsiveness of the IAC to the presence or absence of known confounders, we performed sensitivity analyses by in turn excluding and including these confounders from the propensity score model, and comparing the two resulting IAC

values.

5.3.2 Definition of cohorts

Specific cohorts were defined for each of the empirical examples, as detailed in the following sections.

5.3.2.1 Example 1: NSAIDs vs COX-2 Inhibitors

Our first example illustrates a situation where channelling occurs as a result of the introduction of a new drug on the market. Recently, Wolfe *et al.* attempted to measure the extent of channelling in a large group of patients seen by US rheumatologists shortly before and shortly after the introduction on the market of the new cyclooxygenase-2 (COX-2) inhibitor class of non-steroidal anti-inflammatory drugs (8). The drugs were marketed as having the same anti-inflammatory and analgesic effects as traditional non-steroidal anti-inflammatory drugs (NSAIDs), but are thought to be associated with significantly fewer gastrointestinal side effects, which is a major consideration in the use of NSAIDs. Six thousand six hundred and thirty-seven (6637) patients being seen by 433 rheumatologists for rheumatoid arthritis or osteoarthritis were asked to fill out two sets of detailed questionnaires concerning the last six months of 1998 (period 1) and the first six months of 1999 (period 2), respectively. The questionnaires were administered at the end of each period. These periods correspond to periods before and after the release of the two COX-2 inhibitors, celecoxib (Celebrex®) and rofecoxib (Vioxx®) on the US drug market. The authors then compared, on the basis of the questionnaire responses provided by the patients, the baseline (pre-COX-2 inhibitor era) characteristics of patients who later switched to a COX-2 inhibitor with those who did not switch. Switchers were found to have a more frequent lifetime history of adverse drug reactions of all kinds but particularly of gastrointestinal drug reactions. They also had more

severe pain scores, greater functional disability, more fatigue, helplessness and higher global severity than patients who did not switch. Finally, they also used health care services, both as in- and outpatients, more frequently than non-switchers (8).

Wolfe's study elegantly demonstrates the presence of channelling of sicker patients to new drugs (8). By designing the study prospectively, the authors ensured that the documentation of potential confounders was not influenced by the fact of switching drugs or by the success or failure of the new therapy. Based on their findings, the authors advocate the use of propensity scores that incorporate the above-mentioned variables as a means of controlling for confounding by indication (8). Unfortunately, in most database studies, the necessary information, particularly data relating to severity of disease, functional disability and quality of life, are not available.

In order to demonstrate the IAC's ability to detect apparent channelling, we applied it to a Québec prescription database that covers the period extending from October 1st 1999 to March 31st 2000. The first COX-2 inhibitor (celecoxib; Celebrex®) was introduced onto the Québec market on October 1st 1999. The other COX-2 inhibitor, rofecoxib (Vioxx®), was introduced on April 1st 2000, exactly six months after Celebrex®, so there is no possibility for patients to be assigned to this treatment alternative during the period we are considering in this example. Based on the findings by Wolfe *et al.* (8), we expected the IAC to reveal at least moderate apparent channelling.

The study cohort was extracted from the overall study population by selecting patients aged 66 or older who were prescribed an NSAID at least once during the period spanning October 1st 1998 to September 30th 1999 (the 12 months preceding the introduction of celecoxib (Celebrex®) on the Québec market) AND who received either

an NSAID or Celebrex® at least once during the period extending from October 1st 1999 to March 31st 2000 (from the introduction of celebrex to 6 months after). The treatment assignment (NSAID or COX-2 (Celebrex®)) was assessed by considering the first prescription dispensed for one of these drugs during the 6 months following the introduction of Celebrex®. Patients having concomitantly received a gastroprotective agent (GPA) at the index date were excluded from the study cohort, because they could represent a high-risk sub-population with a different risk profile. The date of dispensation of this first prescription was defined as the index date. The time-frame of the cohort selection process is illustrated in Figure 1, and the numerical results of this selection are shown in Table 1. The final size of the cohort was 27 247 elderly patients, i.e. 31.0% (27 747 / 87 853) of all elderly NSAID and/or COX-2 inhibitor users.

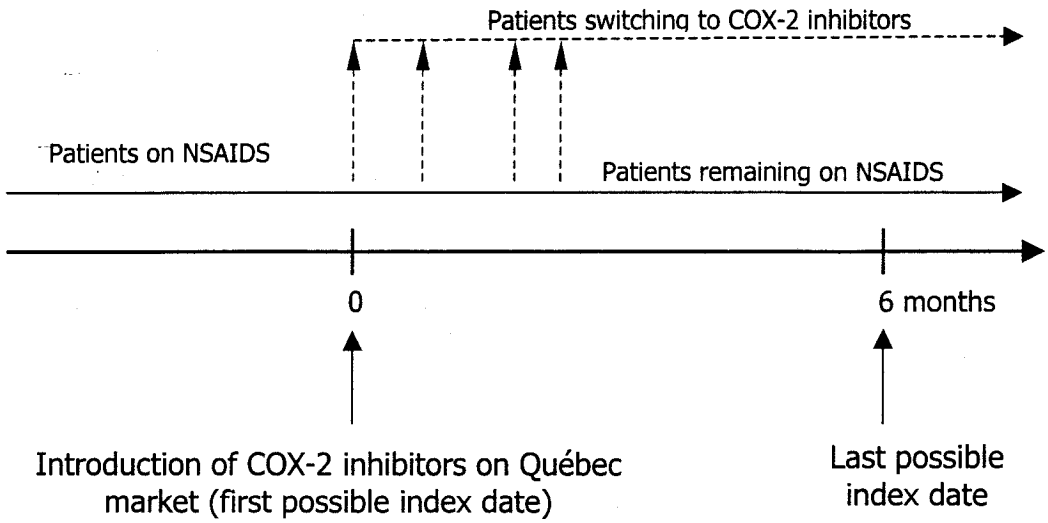

Figure 1. Time-frame of cohort selection for Example 1

Table 1. Cohort selection process for example 1

Total patients in database:	**878 505**	
Exclusion Criteria	**Number excluded**	**Remaining patients**
Patients without dispensation of COX-2 inhibitor or NSAID between October 1st, 1999 and March 31st, 2000	683 690	194 815
Patients aged less than 66 years at the index date	106 962	87 853
Incomplete coverage (including the year prior the index date and the 6 months after)	1 484	86 369
Patients having received a gastro-protective agent at the index date	8 507	77 862
Patients having received Arthrotec (diclofenac + misoprostol) at the index date	6 744	71 118
Patients without NSAID (including Arthrotec) in the year prior to the index date	43 871	27 247
Selected patients:		**27 247**

5.3.2.2 Example 2: NSAIDs vs Acetaminophen

It is expected that patients with rheumatoid arthritis or osteoarthritis who are at higher risk of experiencing gastrointestinal complications would be more likely to be treated for their pain symptoms with acetaminophen rather than with NSAIDs, due to the different side-effect profiles of the drugs (9). Furthermore, current clinical practice guidelines endorsed by the American College of Rheumatology recommend the use of acetaminophen for patients with a history of GI disease or NSAID intolerance (10;11). Channelling is therefore expected to be present in this example.

The study cohort was extracted from the overall study population by selecting patients who were prescribed either an NSAID or acetaminophen at least once over the course of a three-year period extending from January 1st 1994 to December 31st 1996. This first dispensation was considered the index dispensation. We selected first-time users by eliminating patients having received a prescription for any one of these drugs in the year preceding the index date (January 1st to December 31st 1993). Like in example 1, we excluded patients having received a gastro-protective agent at the index date. We deliberately chose a time-window before the introduction of COX-2 inhibitors onto the market to avoid having to deal with channelling to three treatment alternatives (COX-2 inhibitors, NSAIDs and acetaminophen). Of all patients in this fixed cohort (n=97 216), 49 958 (51.4 %) were taking an NSAID and the remainder (n=47 258, 48.6%) were taking acetaminophen. The cohort selection process is illustrated in table 2.

Table 2. Cohort selection process for example 2

Total patients in database:	**878 505**	
Exclusion Criteria	**Number excluded**	**Remaining patients**
Patients without dispensation of NSAID or acetaminophen between January 1st, 1994 and December 31st, 1996	443 227	435 278
Patients aged less than 66 years at the index date	215 631	219 647
Incomplete coverage (year prior the index date and the 6 months after)	7 005	212 642
Patients having received a gastro-protective agent at the index date	20 676	191 966
Patients having received Arthrotec (diclofenac + misoprostol) at the index date	12 386	179 580
Patients with NSAID or acetaminophen in the year prior to the index date	82 364	97 216
Selected patients:		**97 216**

5.3.2.3 Example 3: NSAID (naproxen) vs NSAID (diclofenac)

The third example concerns the choice between naproxen and diclofenac, two NSAIDs used extensively in the treatment of osteoarthritis, rheumatoid arthritis and musculoskeletal pain. Because both are well-established drugs perceived to have fairly similar side-effect and effectiveness profiles, both have been on the market for several years, are widely used in general medicine, rheumatology and surgery, and are comparable in cost, we do not expect any marked channelling in this case (12).

The study cohort was extracted from the overall study population by selecting patients who were prescribed either naproxen or diclofenac at least once over the course of a three-year period extending from January 1st 1994 to December 31st 1996. We selected first-time users by eliminating patients having received a prescription for any one of

these drugs as well as acetaminophen or any other NSAID in the year preceding the index date. We deliberately chose a time-window before the introduction of COX-2 inhibitors onto the market to avoid dealing with channelling to three treatment alternatives (COX-2 inhibitors, naproxen and diclofenac). Of all patients in this fixed cohort (n=33875), 20782 (61.3%) were taking naproxen and the remainder (n=13093, 38.7%) were taking diclofenac. The cohort selection process is illustrated in table 3.

Table 3. Cohort selection process for example 3

Total patients in database:	**878 505**	
Exclusion Criteria	**Number excluded**	**Remaining patients**
Patients without dispensation of naproxen or diclofenac between January 1st, 1994 and December 31st, 1996	659 178	219 327
Patients aged less than 66 years at the index date	111 868	107 459
Incomplete coverage (year prior the index date and the 6 months after)	2 918	104 541
Patients having received a gastro-protective agent at the index date	14 844	89 697
Patients having received Arthrotec (diclofenac + misoprostol) at the index date	46	89 651
Patients with NSAID or acetaminophen in the year prior to the index date	55 776	33 875
Selected patients:		**33 875**

5.3.3 Assessment of potential determinants of treatment assignment

For each example, we searched the respective databases to identify demographic, diagnostic, medical and pharmaceutical variables that were potential determinants of treatment assignment. The following variables were assessed: **Demographic variables:** patient's age, gender, location of residence (rural vs urban). **Medical service utilization variables:** for the year preceding the index date, the number of

hospitalizations, average number of days in hospital, number of visits to GPs, specialists, rhumatologists, gastroenterologists, number of emergency room visits and number of GI events. **Diagnostic variables:** whether diagnoses of cancer, heart failure and / or musculoskeletal disease were made in the year prior to the index date. **Pharmaceutical service utilization variables:** the season of the index dispensation, the total number of prescriptions for the month preceding the index date; for the year preceding the index date, the number of prescriptions for NSAIDs, acetaminophen, aspirin, steroids, anticoagulants, opiates and gastroprotective agents, and the chronic disease score (2). Whenever there was overlap between a single variable (for example, use of gastroprotective agents) and a compound variable (for example, the "prophylaxis" category of GI events) because it included the single variable, the single variable was not included in the initial model.

5.3.4 Gastrointestinal (GI) events

Prior GI events were classified in descending order of severity, based on the classification system used by Rahme *et al.* (9), namely **1) GI hospitalization** (any hospitalization with a primary or secondary diagnosis of gastroduodenal perforation, ulceration, or bleeding (PUB) *or* any hospitalization during which an endoscopy was performed within the first 2 days); **2) ulcer** (any physician visit with a diagnosis of peptic ulcer bleeding (PUB) *or* any upper GI diagnostic test plus a prescription or a gastroprotective agent (GPA) (H2-antagonists, proton pump inhibitor, or misoprostol, but excluding Arthrotec®, which is a combination of misoprostol and diclofenac) *or* a gastroenterologist visit plus a prescription for a GPA; **3) dyspepsia** (any physician visit with a diagnosis of dyspepsia, *or* any GI diagnostic test alone, *or* any visit to a gastroenterologist alone); **4) prophylaxis** (GPA prescription alone).

5.3.5 Statistical analyses

5.3.5.1 Identifying determinants of treatment assignment

To assess the importance of potential determinants of treatment assignment, we ran a backward (stepwise elimination) logistic regression model with the treatment assignment as the dependent variable (COX-2 inhibitor (celecoxib) coded as = 1, and all NSAIDS (diclofenac, diflunisal, etololac, fenoprofen, flurbiprofen, ibuprofen, indomethacine, ketoprofen, mefenamate, nabumeton, naproxen, phenylbutazone, piroxicam, salsalate, tenoxicam, tiaprofenate, and tolmetine) coded as = 0). Initially, all the potential determinants of treatment assignment were included in the model as independent variables. Only statistically significant variables (at the α = 0.05 level for the Wald chi-square test) were retained in the final model. It should be noted that given the very large sizes of all three cohorts and the relatively balanced shares of alternative drug treatments, all associations that are not statistically significant have to be very close to the null.

Propensity scores or individual subjects were then calculated using the regression coefficient estimated in the final model. Based on the empirical distribution of propensity scores, the index of apparent channelling (IAC) was then calculated for each patient group (7).

5.3.5.2 Sensitivity analyses

For each of the examples described above, we illustrated how responsive the IAC is by performing a sensitivity analysis using potential confounders that are thought to be predictors of channelling, namely prior GI events and markers of health services utilization (8). We also performed sensitivity analyses by omitting the variables that were found to be the most significant determinants of treatment assignment, based on

their regression coefficients in the full model.

In a first step, we ran a propensity score model including all potential confounders ("full model") and applied the IAC as described previously. We then repeated the procedure without the medical service utilization variables and prior GI events variables, as defined above, individually and together in the multi-variate regression model used to estimate propensity scores, and then observed the effect of these omissions on the IAC. It is expected that omitting a known potential confounder would reduce the magnitude of the IAC by shifting the "burden" of channelling, since the part of channelling due to these factors would then fall in the realm of the "undocumented", and thus outside of what can be accounted for by propensity scores, and thus by the IAC.

In additional analyses, we subdivided the patient population into those treated by GPs and those seen by specialists. The propensity scores and the IAC were calculated separately for each sub-cohort and then compared. This was done using the observed proportion of the sub-cohorts as well as propensity scores that had been estimated using the propensity score model obtained from that particular sub-cohort.

5.4 Results

5.4.1 Example 1: NSAIDs vs COX-2 inhibitors

5.4.1.1 Patient characteristics

A total of 27 247 patients had filled either a prescription for an NSAID (n=15 186, 55.7%) or for celecoxib (n= 12 061, 42.3%) at the index date. These constituted the two treatment choices within this cohort. The main patient characteristics are shown in Table 4. Patients prescribed a COX-2 inhibitor were more likely to have had a diagnosis of musculoskeletal disease and to have received GI prophylaxis in the year prior to the index date. They had also received fewer dispensation for NSAIDs in the year preceding the index date.

5.4.1.2 Determinants of COX-2 inhibitor use

In addition to patient characteristics, Table 4 also presents the results of the backward logistic regression analysis used to model the determinants of COX-2 use and to produce the propensity score model. Only statistically significant variables (at the 0.05 significance according to Wald's chi-square test) retained in the final model are shown. The odds ratios were calculated for the continuous variables; the categorical frequency distributions are only provided for orientational purposes, to enable quick visual comparisons between the characteristics of the two groups.

The two most important predictors of COX-2 inhibitor dispensation at the index date are a diagnosis of musculoskeletal disease (OR=2.12 (95% CI: 2.00 - 2.25)) and the use of GI prophylaxis in the year prior to the index date (1.67 (95% CI: 1.58 – 1.77)). On the contrary, the number of NSAID dispensations in the year prior to the index date is predictive of assignment to treatment with an NSAID (OR=0.82 (95% CI: 0.81 – 0.83)).

These results are consistent with those of a recent database study on the determinants of COX-2 inhibitor vs NSAID use (13).

Table 4. Example 1: Patient characteristics and determinants of COX-2 use

Patient characteristics	NSAID users, % (n= 15 186)	COX-2 users, % (n= 12 061)	Adjusted odds ratio (95% CI)
Male sex	38.7	29.2	0.71 (0.67 - 0.75)
Musculoskeletal disease diagnosis in year prior to index date	56.0	75.6	2.12 (2.00 - 2.25)
Use of **GI prophylaxis** in year prior to index date	27.7	43.9	1.67 (1.58 – 1.77)
Total number of drugs dispensed in **month** prior to index date: 0	22.6	16.4	1.02 (1.01 - 1.03)
1 - 2	27.4	26.2	
3 - 5	30.4	32.7	
6 – 10	16.2	19.2	
> 10	3.5	5.5	
Number of dispensations for **acetaminophen** in year prior to index date: 0	66.7	55.1	1.04 (1.03 – 1.05)
1 - 2	16.6	20.5	
3 - 5	7.5	10.0	
6 – 10	5.5	8.7	
> 10	3.7	5.7	
Number of dispensations for **NSAIDS** in year prior to index date: 0	0	0	0.82 (0.81 – 0.83)
1 - 2	33.1	57.6	
3 - 5	26.6	24.1	
6 – 10	25.6	14.5	
> 10	14.8	3.9	
Number of dispensations for **aspirin** in year prior to index date: 0	76.3	73.7	1.01 (1.00 – 1.01)
1 - 2	3.0	3.8	
3 - 5	2.9	3.5	
6 – 10	6.6	7.6	
> 10	11.2	11.4	
Dispensations for **anticoagulants** in year prior to index date	2.17	3.22	1.02 (1.00 – 1.03)
Number of visits to **GP** in year prior to index date: 0	9.2	5.8	1.04 (1.04 – 1.05)
1 - 2	18.3	9.9	
3 - 5	32.5	26.7	
6 – 10	27.4	35.3	
> 10	12.7	22.4	

Number of visits to **rheumatologist** in year prior to index date: 0 1 - 2 >= 3	88.2 7.1 4.7	83.1 9.5 7.4	1.10 (1.08 – 1.13)
Number of visits to **gastroenterologist** in year prior to index date: 0 1 - 2 >= 3	95.5 3.6 0.9	92.0 5.9 2.0	1.08 (1.03 – 1.13)
Number of **emergency room** visits in year prior to index date: 0 1 - 2 >= 3	69.7 22.2 8.1	59.6 26.8 13.6	1.04 (1.02 - 1.05)
Rural place of residence	25.0	20.8	0.90 (0.85 - 0.97)
Index prescription dispensed in **winter**	15.2	18.2	0.87 (0.81 - 0.94)

- Only statistically significant determinants are reported.
- OR>1 indicates increased probability of receiving COX-2 inhibitor; OR<1 indicates increased probability of receiving NSAID.
- OR were calculated using continuous/ordinal variables; categorical data is provided only for orientation.

5.4.1.3 Propensity score model and IAC

On the basis the model in Table 4, we calculated propensity score for each patient. The distribution of propensity scores is shown in Figure 4; descriptive statistics for this propensity score distribution are shown in Table 5.

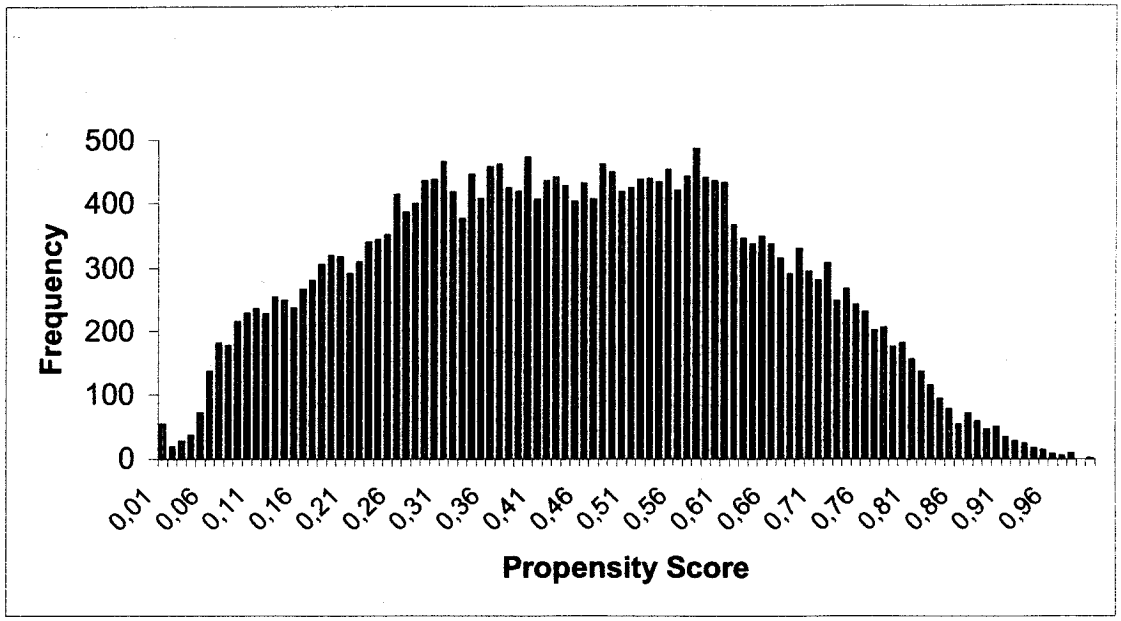

Figure 4. Distribution of propensity scores for patients in Example 1

Table 5. Descriptive statistics of the propensity score distribution for Ex. 1.

Descriptive statistic	Value
Mean	0.4427
Range	0.00003273 – 0.995700
25th percentile	0.2835
50th percentile (Median)	0.4414
75th percentile	0.5971

By looking at the characteristics of patients with extreme propensity scores, it can be seen how the various factors influence the value of the final propensity score. For example, the patient with the lowest propensity score (PS=0.0000746, which virtually guarantees – at least based on documented factors – that this patient should receive an NSAID) had the following characteristics: it was a female patient, she had had a diagnosis of musculoskeletal disease in the year prior to the index date, had received no GI prophylaxis, but had filled 52 prescriptions for NSAIDs in the year prior to the index date. She had also filled a total of 12 prescriptions (regardless of the drug) in the month prior to the index date and made four visits to a GP in the year prior to the index

date. Except for the number of NSAID prescriptions and the lack of GI prophylaxis, all other factors are in fact predictors of COX-2 inhibitor dispensation, however the number of NSAID dispensations was so large that it outweighed the effect of all other factors, leading to a propensity score strongly predictive of an NSAID as index prescription. Indeed, this patient received naproxen at the index date.

At the other extreme, the patient with the propensity score closest to one (PS=0.9977940, i.e. very strongly predictive of assignment to COX-2 inhibitors) was a female patient with a diagnosis of musculoskeletal disease as well as use of GI prophylasis in the year prior to the index date; she had filled a total of 50 prescriptions in the month prior to the index date, and in the year prior to the index date, she had filled 81 prescriptions for acetaminophen, 37 for aspirin and only 4 for NSAIDs, had made 9 visits to a GP, 8 to a specialist (however, none of these were to see a rheumatologist or gastroenterologist), and 11 visits to an emergency room; she lived in a non-rural setting and had received her index prescription in the fall. Except for the fact that this patient did not see either a rheumatologist or a gastroenterologist, all other factors are predictors of COX-2 inhibitor use. Despite all this, this patient received salsalate (an NSAID) as her index prescription, and not a COX-2 inhibitor. This underlines the fact that propensity scores predict treatment assignment on the basis of documented factors but that they say nothing about actual treatment assignment.

The IAC for this overall cohort in example 1 was 17.07 % (Table 12), indicating, as expected, a moderate degree of channelling.

5.4.1.4 Sensitivity analyses

5.4.1.4.1 Effect of known confounders

To illustrate the responsiveness of the IAC, we repeated the calculation of the IAC both with and without *a priori* predictors of COX-2 use postulated by Wolfe *et al.* (8): prior GI events and medical service utilization variables. As expected, removing a known confounder decreases the ability of the propensity score model to predict treatment assignment; consequently, the value of the IAC also decreases when a known confounder is excluded, since treatment assignment cannot be predicted as accurately on the basis of the remaining documented factors. By how much the IAC decreases depends on the strength of association between the known confounders and treatment assignment. The results of this comparative analysis are shown in Table 12 below, together with the results of similar analyses for examples 2 and 3.

5.4.1.4.2 Comparing patients of GPs and specialists

We then subdivided the patient population into those treated by GPs and those seen by specialists. We expected apparent channelling to be more pronounced in the sub-group treated by specialists, because specialists may be more likely to be sensitive to patient characteristics which may make switching to COX-2 inhibitors beneficial, so more channelling may be expected for their patients. On the other hand, one may argue that GPs know their patients better than specialists, so they may better adjust their treatment choice to patient characteristics. Moreover, specialists may see systematically sicker patients (i.e. their patients may have more undocumented characteristics prompting COX-2 inhibitor use, so there would be more residual channelling and less apparent channelling for specialists). Both these arguments suggest that there could be more channelling for GPs.

The logistic model was estimated and propensity scores were calculated separately for each sub-group (patients seen by GPs and specialists, respectively) using the full set of potential confounders in the initial model, and the IAC was calculated for each sub-group and then compared. The characteristics of the GP and specialist sub-cohorts are shown in Table 6 below. The IACs are compared in Table 15, together with the results of the analyses for examples 2 and 3.

Table 6. Descriptive statistics of the GP and specialist sub-cohorts for Ex. 1.

Cohort	Number of patients (n)	Proportion getting a COX-2 inhibitor at index date (m, observed proportion)
GPs	**22 120**	**0.4405**
Specialists	**4 848**	**0.4581**
TOTAL	**27 247**	**0.4427**

It can be seen from table 6 that the two sub-cohorts differ in size, as could be expected since there are much fewer rheumatologists than GPs, but that the proportion of patients having received a COX-2 inhibitor as the index prescription does not differ significantly.

5.4.2 Example 2: NSAIDs vs acetaminophen

5.4.2.1 Patient characteristics

A total of 49 958 patients (51.4%) had filled an NSAID prescription at the index date, the remainder a prescription for acetaminophen (n= 47 258, 48.6%). The main patient characteristics are shown in Table 7. Patients in the acetaminophen group were more likely to be older, to have been dispensed anticoagulants and to have been hospitalized more frequently in the year preceding the index date. They were also slightly more likely to have had an ulcer or received GI prophylaxis in that year.

5.4.2.2 Determinants of acetaminophen use

In addition to patient characteristics, Table 7 also presents the results of the backward logistic regression analysis used to model the determinants of acetaminophen use and to produce the propensity score model. Only statistically significant variables (at the 0.05 significance according to Wald's chi-square test) retained in the final model are shown. The odds ratios were calculated for the continuous (ordinal) variables; the categorical frequency distributions are only provided for orientational purposes, to enable quick visual comparisons between the characteristics of the two groups.

The two most important predictors of acetaminophen dispensation at the index date are the use of GI prophylaxis (OR=1.38 (95% CI: 1.33 – 1.44)) and the number of hospitalizations in the year prior to the index date (OR=1.28 (95% CI: 1.26 – 1.31)). Having had an ulcer as a GI event in the year preceding the index date is also a predictor, albeit not as strong (OR=1.16 (95% CI: 1.07 – 1.26)).

Table 7. Example 2: Patient characteristics and determinants of acetaminophen use

Patient characteristics	NSAID users % (n= 49 958)	Acetaminophen users % (n= 47 258)	Adjusted odds ratio (95% CI)
Age 66-74 yrs	67.1	51.6	
75-84 yrs	28.5	37.5	1.06 (1.06 - 1.06)
≥ 85 yrs	4.4	10.9	
Male sex	40.1	39.0	0.95 (0.93 - 0.98)
Cancer diagnosis in year prior to index date	8.7	13.3	1.19 (1.14 – 1.25)
Musculoskeletal disease diagnosis in year prior to index date	22.4	21.4	0.84 (0.81 - 0.86)
Ulcer in year prior to index date	3.4	6.0	1.16 (1.07 – 1.26)
Use of **GI prophylaxis** in year prior to index date	14.3	21.8	1.38 (1.33 – 1.44)
Total number of drugs dispensed in **month** prior to index date: 0	33.1	28.3	
1 - 2	34.9	31.8	
3 - 5	24.5	27.4	1.02 (1.01 – 1.02)
6 – 10	6.8	10.8	
> 10	0.8	1.8	
Dispensations for **anticoagulants** in year prior to index date	1.4	4.5	1.07 (1.06 – 1.08)
Dispensations for **opiates** in year prior to index date	0.8	1.1	0.96 (0.93 - 0.99)
Number of dispensations for **corticosteroids** in year prior to index date: 0	84.1	80.7	1.01 (1.01 – 1.02)
1 - 2	11.4	12.9	
>=3	4.5	6.4	
Number of visits to **specialists** in year prior to index date: 0	30.5	23.6	
1 - 2	29.0	27.0	
3 - 5	22.0	24.3	1.01 (1.00 – 1.01)
6 – 10	13.1	16.7	
> 10	5.5	8.4	
Number of visits to **gastroenterologists** in year prior to index date: 0	95.8	93.9	
1 - 2	3.2	4.5	1.04 (1.01 – 1.07)
>=3	1.0	1.7	

Number of **emergency room** visits in year prior to index date: 0 1 - 2 >= 3	69.6 24.2 6.2	53.6 34.1 12.3	1.06 (1.05 – 1.07)
Number of **hospitalizations** in year prior to index date: 0 1 - 2 >= 3	81.4 13.0 5.7	62.0 21.9 16.1	1.28 (1.26 – 1.31)
Average **length-of-stay** per hospitalization in year prior to index date (median, in days)	2.0	3.5	1.05 (1.04 – 1.04)
Index prescription dispensed in **spring**	29.4	27.3	0.84 (0.82 - 0.87)
Index prescription dispensed in **summer**	24.8	22.1	0.81 (0.79 - 0.84)

- Only statistically significant determinants are reported.
- OR>1 indicates increased probability of receiving acetaminophen; OR<1 indicates increased probability of receiving NSAID.
- OR were calculated using continuous/ordinal variables; categorical data is provided only for orientation.

5.4.2.3 Propensity score model and IAC

On this basis, we calculated a propensity score for each patient. The distribution of propensity scores is shown in Figure 5; descriptive statistics for this propensity score distribution are shown in Table 8.

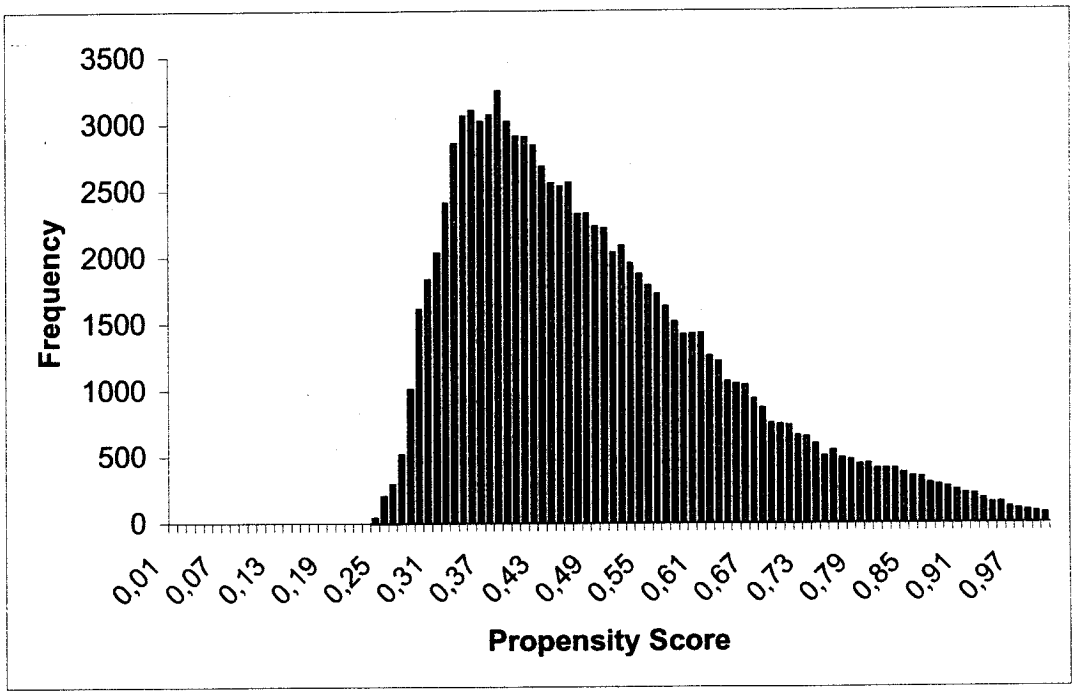

Figure 5. Distribution of propensity scores for patients in Example 2

Table 8. Descriptive statistics of the propensity score distribution for Ex. 2.

Descriptive statistic	Value
Mean	0.4861
Range	0.0509 – 0.9999
25th percentile	0.3677
50th percentile (Median)	0.4539
75th percentile	0.5737

From Table 8 it can be seen that the likelihood of getting acetaminophen or an NSAID is quite similar (0.4861 vs 0.5139, respectively; the average probability of getting the drug coded as one (here acetaminophen) is equal to the mean propensity score, namely 0.4861) and that the range of propensity scores is quite broad, indicating a relatively good predictive ability of the propensity score model. The IAC was 9.29 % (Table 12), indicating a mild degree of channelling due to documented factors.

5.4.2.4 Sensitivity analyses

5.4.2.4.1 Effect of known confounders

To illustrate the responsiveness of the IAC, we repeated the calculation of the IAC both with and without *a priori* predictors of acetaminophen use (prior GI events, and medical service utilization variables, as defined above) in the propensity score model (9). The results of this comparative analysis are shown in Table 12 below. We also performed sensitivity analyses by removing the two strongest predictors of treatment assignment as predicted by the propensity score model. The results of these analyses are shown in Table 13.

5.4.2.4.2 Comparing patients of GPs and specialists

Like for the previous example, we carried out sub-group analyses for patients whose index prescription had been written by a GP or a specialist, respectively. The characteristics of the GP and specialist sub-cohorts are shown in Table 9 below. The IACs are compared in Table 15, together with the results of the analyses for examples 2 and 3.

Table 9. Descriptive statistics of the GP and specialist sub-cohorts for Ex. 2.

Cohort	Number of patients (n)	Proportion getting acetaminophen at index date (m, observed proportion)
GPs	81 081	0.4517
Specialists	15 043	0.6775
TOTAL	97 216	0.4861

It can be seen from Table 9 that the two sub-cohorts differ in size, as could be expected since there are much fewer rheumatologists than GPs, and that the proportion of patients having received acetaminophen as the index prescription is higher among patients who got their index prescription from a specialist (0.6775, vs 0.4517 for GPs).

This lends support to the hypothesis that specialists tend to see a patient population that is generally sicker and in particular, more likely to have suffered from adverse GI events in the past.

5.4.3 Example 3: NSAID vs NSAID

5.4.3.1 Patient characteristics

A total of 20 782 (61.3%) patients had filled a prescription for naproxen at the index date, the remainder a prescription for diclofenac (n= 13 093, 38.7%). These two groups formed the naproxen and diclofenac users' groups respectively. The main patient characteristics are shown in Table 10.

5.4.3.2 Determinants of naproxen use

In addition to patient characteristics, Table 10 also presents the results of the backward logistic regression analysis used to model the determinants of naproxen use and to produce the propensity score model. Only statistically significant variables retained in the final model are shown.

The two most important predictors of naproxen use are a rural site of residence (OR=1.20 (95% CI: 1.14 – 1.26)) and a diagnosis of cancer in the year prior to the index date (OR=1.15 (95% CI: 1.07 – 1.25)). The different predilection of rural vs urban physicians for naproxen cannot be explained on substantive grounds but may rather have to do with regional practice variations. As for a diagnosis of cancer favouring the use of naproxen, this could be explained by the longer half-life of naproxen, which needs to be taken only twice a day, as opposed to four times a day for diclofenac.

Table 10. Example 3: Patient characteristics and determinants of naproxen use

Patient characteristics	Diclofenac users % (n= 13093)	Naproxen users % (n= 20782)	Adjusted odds ratio (95% CI)
Male sex	41.5	40.1	0.92 (0.88 - 0.96)
Cancer diagnosis in year prior to index date	8.9	9.9	1.15 (1.07 – 1.25)
Number of visits to **GP** in year prior to index date: 0	7.0	8.7	0.98 (0.98 - 0.99)
1 - 2	23.7	23.2	
3 - 5	33.9	34.3	
6 – 10	25.6	24.7	
> 10	9.9	9.1	
Number of visits to **specialists** in year prior to index date: 0	29.4	29.9	0.99 (0.98 - 0.99)
1 - 2	28.6	28.8	
3 - 5	21.9	22.4	
6 – 10	13.9	13.1	
> 10	6.2	5.9	
Number of **emergency room** visits in year prior to index date: 0	72.7	63.1	1.12 (1.09 – 1.14)
1 - 2	21.8	29.2	
>= 3	5.5	7.7	
Number of **hospitalizations** in year prior to index date: 0	82.5	79.5	1.05 (1.02 – 1.08)
1 - 2	12.1	14.0	
>= 3	5.5	6.6	
Total number of drugs dispensed in **month** prior to index date: 0	34.4	32.6	1.01 (1.00 – 1.02)
1 - 2	35.2	35.4	
3 - 5	23.3	24.4	
6 – 10	6.4	6.8	
> 10	0.7	0.8	
Rural place of residence	23.3	27.1	1.20 (1.14 – 1.26)
Index prescription dispensed in **winter**	23.6	22.4	0.94 (0.89 – 0.99)

- Only statistically significant determinants are reported.
- OR>1 indicates increased probability of receiving naproxen; OR<1 indicates increased probability of receiving diclofenac.
- OR were calculated using continuous/ordinal variables; categorical data is provided only for orientation.

5.4.3.3 Propensity score model and IAC

On the basis of the model shown in Table 10, we calculated propensity score for each patient. The distribution of propensity scores is shown in figure 6; descriptive statistics for this propensity score distribution are presented in Table 11.

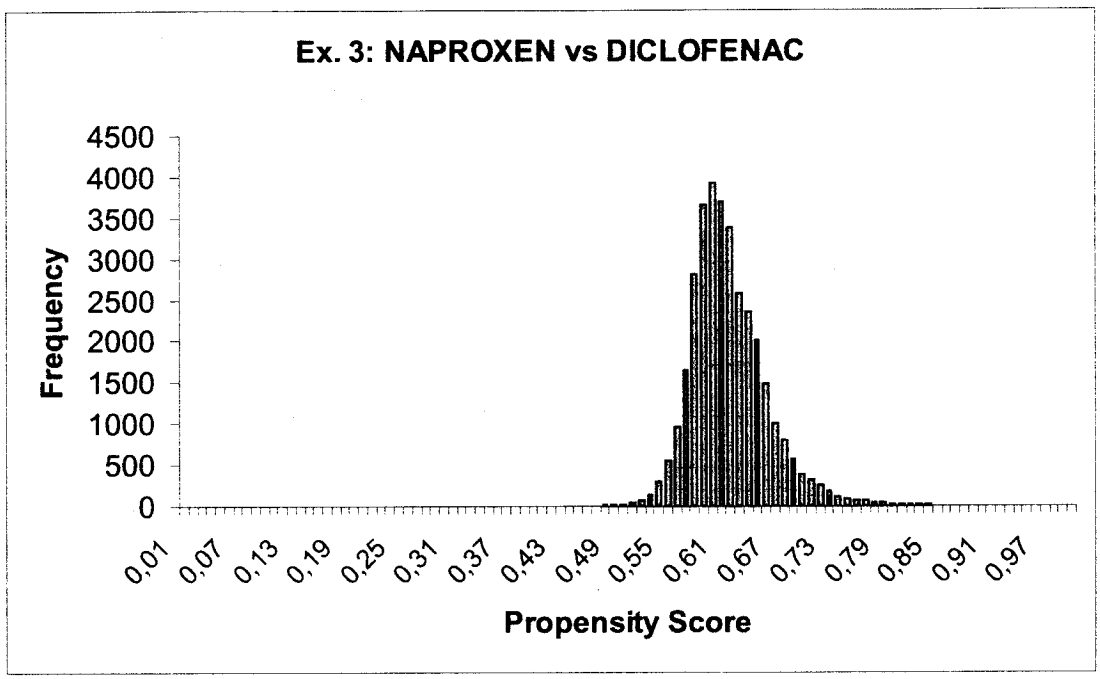

Figure 6. Distribution of propensity scores for patients in Example 3

Table 11. Descriptive statistics of the propensity score distribution for Ex. 3.

Descriptive statistic	Value
Mean	0.613491
Range	0.1958 – 0.9844
25th percentile	0.585597
50th percentile (Median)	0.607497
75th percentile	0.636545

In this example, the probability of getting the two treatment alternatives is not nearly equal, as was the case in the other examples; here, patients have, on average a 60% chance of receiving naproxen, which corresponds to the drug's observed proportion in this cohort. The range of the propensity score distribution is somewhat narrower than in

the other examples, and more importantly, the 25th and 75th percentile are very close together, indicating an important clustering of the propensity scores around the observed proportion of patients receiving acetaminophen. This is consistent with the propensity score model having a limited ability to predict treatment outcome, a situation that arises when there is little or no apparent channelling (7). Indeed, the IAC was 0.85%, indicating virtual absence of apparent channelling (Table 12).

5.4.3.4 Sensitivity analyses

5.4.3.4.1 Effect of known confounders

To illustrate the responsiveness of the IAC, we repeated the calculation of the IAC both with and without prior GI events and medical service utilization variables, as defined earlier, in the propensity score model. However, since these factors are not known to preferentially affect treatment assignment toward diclofenac or naproxen, they were not expected to significantly affect the IAC. The results of this comparative analysis are shown in Table 12 below, together with the results of the analyses for examples 1 and 2. Indeed, since the IAC already indicated a near absence of apparent channelling, it only changed minimally when weak predictors of treatment assignment were removed. We also performed a sensitivity analysis by removing the strongest predictor of treatment assignment (rural vs non-rural site of residence) from the initial propensity score model. The results of this analysis is shown in Table 13. As expected, this only changed the IAC minimally.

Table 12. Index of apparent channelling with and without including known confounder in the propensity score model

No.	Model	IAC (in %)		
		Ex.1: COX-2 vs NSAID	**Ex.2:** Acetaminophen vs NSAID	**Ex.3:** Naproxen vs Diclofenac
1	Full model including all potential covariates	17.07	9.29	0.85
2	Model including all potential covariates except **medical service utilization** variables	16.10	6.27	0.27
3	Model including all potential covariates except **prior GI events**	16.07	8.93	0.85
4	Model including all potential covariates except **prior GI events** and **medical service utilization** variables	14.69	5.52	0.27

Table 13. Index of apparent channelling with and without strongest predictors of treatment in the full model

No.	Model	IAC (in %)		
		Ex.1: COX-2 vs NSAID	**Ex.2:** Acetaminophen vs NSAID	**Ex.3:** Naproxen vs Diclofenac
1	Full model including all potential covariates	17.07	9.29	0.85
12	Full model (Ex. 1), excluding GI prophylaxis in year prior to index date	16.13	----	----
5	Full model (Ex. 1), excluding musculoskeletal disease in year prior to index date	14.97	----	-----
6	Full model (Ex. 1), excluding musculoskeletal disease, GI prophylaxis and sex in year prior to index date (three strongest determinants of treatment)	13.28	----	----
11	Full model (Ex. 1), excluding number of NSAID dispensations in year prior to index date	9.16	----	----
10	Full model (Ex. 1), excluding musculoskeletal disease, GI prophylaxis, sex and number of NSAID dispensations in year prior to index date (four strongest determinants of treatment)	5.40	----	----
7	Full model (Ex. 2), excluding GI prophylaxis in year prior to index date	-----	9.04	----
8	Full model (Ex. 2), excluding GI prophylaxis and number of hospitalizations in year prior to index date (two strongest determinants of treatment)	----	8.24	----
9	Full model (Ex. 3), rural site of residence (strongest determinants of treatment)	----	----	0.71

5.4.3.4.2 Comparing patients of GPs and specialists

Similar to the previous example, we performed sub-group analyses for patients whose index prescription had been written by GPs and specialists, respectively. The

characteristics of the GP and specialist sub-cohorts are shown in Table 14 below. The IACs are compared in Table 15, together with the results of the analyses for examples 2 and 3.

Table 14. Descriptive statistics of the GP and specialist sub-cohorts for example 3.

Cohort	Number of patients (n)	Proportion getting naproxen at index date (m, observed proportion)
GPs	30 705	0.6180
Specialists	3 127	0.5683
TOTAL	33 875	0.6135

It can be seen from Table 14 that the two sub-cohorts differ in size, as could be expected since there are much fewer rheumatologists than GPs, and that the proportion of patients having received naproxenen as the index prescription is slightly higher among patients who got their index prescription from a GP. This could be explained by the fact that GPs are more likely than specialists to be seeing cancer patients, the strongest predictor of assignment to acetaminophen in this example.

Table 15. Index of apparent channelling for patients seen by GPs and specialists, respectively.

	IAC (in %)		
	Ex.1: COX-2 vs NSAID	**Ex.2:** Acetaminophen vs NSAID	**Ex.3:** Naproxen vs Diclofenac
Full cohort:	17.07	9.29	0.85
Sub-cohorts:			
GP	17.72	9.71	0.89
Specialist	15.76	9.90	1.20

Table 15 compares the IAC values for the GP and specialists sub-cohorts across the three examples. In example 1, GPs have a higher IAC value than specialists, indicating a higher degree of apparent channelling in this sub-cohort. On the contrary, in examples 2 and 3, specialists have slightly higher IAC values, which is consistent with the converse situation. Furthermore, in examples 2 and 3, both GPs and specialists have higher IAC values than the overall IAC for the whole cohort. This would indicate that some factors, which favour one particular treatment assignment in one sub-cohort would in fact favour the other treatment assignment in the other cohort. In the overall cohort, these effects would cancel each other out, thus rendering the propensity score model less able to accurately predict treatment assignment – i.e. "shifting" propensity score toward their mean (the observed proportion), and thus leading to a lower value of the IAC for the overall cohort.

5.5 Discussion

Since channelling is a necessary condition for confounding by indication, it would be useful, as a first step toward controlling for confounding by indication, to be able to assess the magnitude of channelling. In the first article of this series, we proposed the index of apparent channelling (IAC) as a means to estimate the importance of channelling due to documented factors (7). We then illustrated the behaviour of the IAC by applying it to hypothetical examples. The present study was designed to investigate the behaviour of the IAC when applied to real-life situations with differing expected degrees of apparent channelling. To attain this goal, three examples were used. The first example (NSAID vs COX-2 inhibitor) was chosen because it was possible to predict at least moderate channelling based on results from a previous prospective non-experimental field study (8). The second example (NSAID vs acetaminophen) was chosen because a moderate degree of channelling was expected on clinical grounds and because prior studies have demonstrated that acetaminophen tends to be prescribed preferentially to sicker patients (13). The third example (naproxen vs diclofenac) was chosen because no channelling was expected on clinical grounds, since both drugs are well established in clinical practice, available as generics at similar cost and perceived to be relatively similar in their efficacy and adverse effects.

In example 1, the IAC was equal to 17.07%, which corresponds to a moderate amount of apparent channelling, and is considerably higher than the IAC in the two other examples. This result is in accordance with the results of the non-experimental field study by Wolfe *et al.* (8). In the sensitivity analyses (Table 12), we used the determinants of treatment that had been found by Wolfe *et al.* (8) and Rahme et al.

(9;13) to be significant predictors of treatment with COX-2 and acetaminophen, namely medical service utilization variables and prior GI events. Surprisingly, removing these variable groups each in turn or simultaneously reduced the value of the IAC by only a small amount. However, this can be readily explained by the fact that these variables did not include musculoskeletal disease in the year prior to the index date, which is the strongest predictor of treatment assignment in our database. Wolfe *et al.* had not found this variable to be significant because, by virtue of their cohort definition, all their patients suffered from musculoskeletal disease (8). The fact that they restricted their cohort in a way that eliminated any variability in this variable prevented them from drawing any conclusions about its importance.

Consequently, we expanded the sensitivity analyses by excluding from the propensity score model factors found to be the strongest predictors of treatment assignment in our database. Surprisingly, removing musculoskeletal diagnosis from the full model did not alter the value of the IAC in an important way (see Table 8). However, removing the two other most important predictors (use of GI prophylaxis and number of NSAID prescriptions in year prior to index date) in addition to musculoskeletal disease led to an important reduction in the value of the IAC (Table 13). The lack of effect on the IAC of the musculoskeletal disease variable alone can be explained in large part by some degree of overlap with the two other variables.

In example 2, the IAC took on a value of 9.29 %, indicating weaker channelling. The predictors of treatment assignment revealed by the modelling process (Table 7) are in accordance with results of a previous study of NSAID and acetaminophen that made use of propensity scores in the classical way to control for confounding by indication (9).

Removing the most significant predictors of treatment assignment in the sensitivity analyses (GI prophylaxis and number of hospitalizations) (Tables 13) led to a reduction in the value of the IAC down to 8.24%, a relative reductions in the value of 11%, thereby demonstrating moderate responsiveness of the IAC to the presence or absence of such predictors in this example.

Finally, example 3 illustrated a situation where little or no channelling is expected. Accordingly, the IAC was very close to zero (0.85%), indicating only minimal channelling and leaving little manoeuvring room for fluctuations in the sensitivity analyses. Nonetheless, the IAC was mildly responsive to the omission of the most important predictor of treatment assignment (rural site of residence), dropping to a mere 0.71% when this variable was omitted (Table 13). Conversely, removing a non-significant determinant of treatment assignment (in this case prior GI events) leaves the IAC unchanged, as would be expected (Table 12).

The interpretation of the analyses focusing on the comparison between GPs and specialists is not as straightforward. In examples 2 and 3, GPs have slightly lower IAC values than specialists, whereas the contrary is true for example 1. There could be many reasons for this phenomenon, one of them being that GPs and specialists could be influenced by different factors in their treatment choices. Another reason could be the unequal distribution of treatment determinants across the two sub-cohorts (GPs vs specialists). For instance, in example 1, the rural site of residence of a patient increases the likelihood that this patient will be prescribed an NSAID rather than a COX-2. On the other hand, GPs have 25.7 % of their patients living in rural areas whereas specialists only have 12.0%. More balanced distribution of the two categories of a binary variable

(in this case, rural site of residence) leads to larger variance (0.1910 for GPs, vs 0.1056 for specialists). Therefore, channelling due to rural/urban residence will induce larger variance in propensity scores among patients of GPs than among those seen by specialists. Furthermore, the large discrepancy in numbers, and therefore in statistical power, between the two sub-cohorts might lead to non-significant results, which could result in a variable being excluded from the final regression model.

Therefore, this study allowed us to demonstrate that the IAC behaves as expected from prior hypothetical examples (7). Furthermore, it was shown that the IAC can be used to compare the degree of apparent channelling across different databases, or across sub-sets of a same database, regardless of cohort size and observed proportion of the different treatments. It can thereby potentially identify cohorts that are free or nearly free of apparent channelling, as illustrated by our example 3. Populations – or sub-populations – that are free of apparent channelling may be of interest to eliminate concerns about confounding by indication, but this requires strong assumptions about lack of channeling due to undocumented factors, a complex issue that will be addressed in the third and last article of this series.

The IAC also enables, by means of sensitivity analyses, the quantification of the channelling effect of individual treatment determinants or groups thereof. This channelling effect of a given determinant can be expressed concisely by taking the ratio of the difference between the IAC of the full model to that of the reduced model omitting this determinant, and dividing this quantity by the IAC for the full model. We call this the "relative channelling effect" (RCE). For instance, in example 1, the RCE for the determinant "number of NSAID dispensations in the year prior to index date" is

equal to (17.07 - 9.16) / 17.07 = 0.463 or 46.3 % (Table 13), indicating that this particular determinant explains about 46% of the apparent channeling accounted for by the full model. This is a feature that the classical propensity score approach does not possess, an advantage unique to the IAC.

The IAC also has its limitations. The most important one is undoubtedly the fact that it can only account for documented factors. However, this is a feature it shares with most of the other approaches to the control of confounding. Consequently, the possibility remains that there is residual confounding due to undocumented factors. Another feature of the IAC is that with empirical data, its value does not vary as widely as in the hypothetical examples provided in the first article (7). One reason for these more modest fluctuations is that real-life situations are just not as extreme as the examples used in our hypothetical examples; another reason is possibly the lack of documented information on treatment determinants in the database. However, the limited range of real-life parameter estimates is not an unknown phenomenon: the odds ratio, for example, can – theoretically – take on values ranging from near-zero to infinity, however, in practice we are usually confronted with estimates in the range of 0.5 to 2.5. Nonetheless, it is our hope that as researchers familiarize themselves with the IAC, the interpretation of its value will become more meaningful with time. We invite other researchers to apply the IAC to cohorts where various degrees of channelling are expected and to report the results of their findings.

Undoubtedly, the biggest limitation of the IAC is the fact that it cannot account for undocumented factors. If one is to eliminate the possibility that confounding by indication is a plausible explanation for a particular association of interest, the possibility

that undocumented factors are confounding this association must be eliminated. This may at first glance seem like an impossible task – how can one control for something that is, by definition, not documented? – however, we will turn to this challenge in the third and final paper of this series and attempt to develop some practical solutions to this problem (14).

5.6 Acknowledgements

This project was supported by a grant from the "Réseau FRSQ sur l`utilisation des médicaments" of the government of the province of Québec and by discretionary funds from Dr Jacques LeLorier. Dr Michal Abrahamowicz is a scientist of the Canadian Institutes of Health Research (CIHR).

Many thanks to Drs Elham Rahme and Lucie Blais for stimulating discussions and insightful comments at various stages in the development of this project.

5.7 References

1. Walker AM. Confounding by indication. Epidemiology 1996;7:335-6.

2. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992;45:197-203.

3. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

4. Miettinen OS. The need for randomization in the study of intended effects. Stat Med 1983;2:267-71.

5. Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med 1991;10:577-81.

6. Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of beta 2-agonists. Am J Epidemiol 1996;144:1161-9.

7. Bjerre, L. M., Abrahamowicz, M. A., and LeLorier, J. Channelling in non-experimental pharmacoepidemiologic research: The estimation of apparent channeling Part I: Theory and hypothetical examples. 2003. Ref Type: Thesis/Dissertation

8. Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.

9. Rahme E, Pettitt D, LeLorier J. Determinants and sequelae associated with utilization of acetaminophen versus traditional nonsteroidal antiinflammatory drugs in an elderly population. Arthritis Rheum 2002;46:3046-54.

10. Hochberg MC, Altman RD, Brandt KD et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum 1995;38:1541-6.

11. American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. Guidelines for the management of rheumatoid arthritis. Arthritis Rheum 1996;39:713-22.

12. Katzung BG. Basic and clinical pharmacology. 6th ed. East Norwalk (CT): Appleton and Lange, 1995.

13. Rahme E, Marentette MA, Kong XS, LeLorier J. Use of NSAIDs, COX-2 inhibitors, and acetaminophen and associated coprescriptions of gastroprotective agents in an elderly population. Arthritis Rheum 2002;47:595-602.

14. Bjerre, L. M., LeLorier, J., Dorais, M., and Abrahamowicz, M. A. Channelling in non-experimental pharmacoepidemiologic research: Assessing the impact of residual channelling on confounding by indication. 2003. Ref Type: Thesis/Dissertation.

6 Preface to Manuscript 3

This is the third in a series of three articles that addresses the phenomenon of channelling in non-experimental research and its impact on confounding by indication. In the first article in this series, a novel approach, the Index of Apparent Channelling (IAC), was proposed with a view to quantifying apparent channelling, the phenomenon of channelling due to documented factors. The behaviour of the IAC was illustrated using hypothetical examples, and it was shown to be responsive to various experimentally controlled degrees of channelling.

In the second article of the series, the IAC was applied to empirical data in order to illustrate its behaviour under real-life conditions and with various expected degrees of channelling. The IAC was shown to perform consistently with *a priori* expectations. However, because it is based on propensity scores, which in turn rely on documented factors to predict treatment assignment, the IAC remains limited by its sole reliance on documented factors. Consequently, it cannot account for the effect of undocumented factors on treatment assignment. The issue of channelling by undocumented factors, or residual channelling, is the subject of this third article.

In this article, various methods for assessing residual channelling are explored. First, the index of residual channelling (IRC) is proposed as a tool to summarize the overall degree of discrepancy between the actual treatment received and the treatment predicted from available covariates. This discrepancy is postulated to be proportional to residual channelling. A propensity score-based method to assess to what extent residual channelling can bias the treatment effect was then developed. This method is based on modelling the interaction between the treatment received and the measure of residual channelling. The proposed methods allow for the assessment of the impact of residual

channelling on confounding by indication, thereby providing a practicable tool for reducing bias in the estimation of treatment effects obtained from non-experimental studies of medications.

Channelling in non-experimental pharmacoepidemiologic research:

Manuscript 3: Assessing the impact of residual channelling on confounding by indication

Lise M Bjerre, MSc, MD
PhD candidate
Department of Epidemiology and Biostatistics
McGill University

Jacques LeLorier MD, PhD
Chief
Pharmacoepidemiology and Pharmacoeconomics Unit
Centre de recherche
Hôtel-Dieu du Centre hospitalier de l'Université de Montréal
Université de Montréal

Marc Dorais, MSc
Biostatistician
Pharmacoepidemiology and Pharmacoeconomics Unit
Centre de recherche
Hôtel-Dieu du Centre hospitalier de l'Université de Montréal
Université de Montréal

Michal Abrahamowicz, PhD
James McGill Professor
Department of Epidemiology and Biostatistics
McGill University

Corresponding author: Michal Abrahamowicz, PhD
James McGill Professor
Department of Epidemiology and Biostatistics
McGill University
Division of Clinical Epidemiology
Livingston Hall, Montreal General Hospital
1650 Cedar Ave
Montréal, Québec, Canada H2W 1T8
Tel: (514) 934-1934 ext. 44712
Fax: (514) 934-8293
E-mail : michal@michal.ri.mgh.mcgill.ca

Keywords: Bias, non-experimental research, observational research, confounding, confounding by indication, confounding by severity, propensity scores, channelling, residual channelling
Text only word count: 5848

6.1 Abstract

BACKGROUND: Confounding by indication is a serious threat to the validity of non-experimental pharmacoepidemiologic studies. Channelling is a necessary condition for the existence of confounding by indication. In the first article of this series, the index of apparent channelling (IAC) was introduced as a tool for the quantification of apparent channelling, which is due to documented factors. The IAC is limited by its inability to account for undocumented factors.

METHODS: In the present paper, various approaches to estimate channelling due to undocumented factors (residual channelling), and its impact on confounding by indication were explored. First, we postulated that the discrepancy between actual and predicted treatment should be a function of residual channelling. Then, the index of residual channelling (IRC) was proposed as a tool to summarize the overall degree of discrepancy between the treatment actually received and the treatment predicted from documented covariates. A propensity score-based method to assess to what extent residual channelling can bias the treatment effect was then developed. The method is based on modelling the interaction between the treatment and discrepancy, the measure of residual channelling.

RESULTS: We applied the proposed propensity-score based method to empirical examples, and indeed, the method was able to detect and control for residual channelling: for instance, in example 1 (risk of GI events in COX-2 inhibitor vs NSAID users), the estimate of effect obtained by logistic regression without accounting for residual channelling was OR=0.83 (0.60-1.16) for all COX-2 inhibitor users, whereas with correction for residual channelling (i.e. in non-discrepant COX-2 inhibitor users), the

OR was 0.55 (0.47-0.64), a risk estimate consistent with those obtained from randomized controlled trials. Moreover, these results suggested that less biased estimates of treatment effect may be obtained by assessing the expected effect of treatment in the hypothetical situation where residual channelling is absent. Finally, when residual channelling was purposefully introduced by deliberately excluding known predictors of treatment assignment from the model (i.e. running a "reduced model"), the method was still able to control for residual channelling to a large extent: In example 1, the risk estimate without controlling for residual channelling was OR=1.13 (1.07 – 1.20) for the reduced model, with the interaction term controlling for residual channelling, it was 0.66 (0.49 – 0.87), a value still quite consistent with the results of randomized controlled trials.

CONCLUSION: The proposed method allows for the assessment of the impact of residual channelling on confounding by indication, thereby providing a practicable tool for reducing the bias in the treatment effects estimated in non-experimental studies of medications.

6.2 Introduction

Confounding by indication is a serious threat to the validity of non-experimental studies in pharmacepidemiology (1-3). Confounding by indication can be broken down into its two components, channelling and risk factor effect. In the companion article, we proposed the index of apparent channelling (IAC) as a tool to estimate channelling due to documented factors (4). However, the IAC does not account for undocumented factors. Yet, it is clear from observational studies that factors not usually documented in drug claims databases may often act as important channelling factors. For example, Wolfe et al. (5), in a prospective cohort study, demonstrated that rheumatoid and osteoarthritis patients who were switched from NSAIDs to COX-2 inhibitors after the introduction of the latter onto the US market, tended to report more pain, less functionality and more helplessness at cohort inception (i.e. before being switched to COX-2 inhibitors) than patients who remained on NSAIDs. These are all factors that are not routinely documented in administrative claims drug databases. Consequently, the estimation of a treatment effect may be biased by the investigators' inability to control for these variables, even if the treatment is adjusted for all available covariates or their aggregates such as propensity scores (6;7). Furthermore, this inability to control for unknown or undocumented confounders is at the root of the common perception that observational studies can only provide limited evidence about treatment effects.

The methods proposed so far to control for confounding by indication – among others, propensity scores (6-8), confounder score (9) and instrumental variables (10) – all suffer from the same limitation, namely reliance on documented factors only.

To reduce the risk of biased estimation of treatment effect, it is necessary to control for

confounding due to both documented and undocumented factors. A necessary step in order to achieve the latter is the assessment of the extent of residual channelling, i.e. treatment choices that are based on patient characteristics not available in a given database.

This may at first glance seem like an impossible task. How can one assess something that is not even documented? The approach proposed in this article is based on the observation that what really matters is the extent to which channeling based on undocumented factors affects the estimated treatment effect. This, perhaps, can be assessed by comparing how the treatment effect estimate changes across situations with different expected magnitudes of residual confounding. Thus, we propose an indirect method for assessing the impact of residual channeling on confounding by indication and explore two approaches for applying this method.

The overall objective of this study was to develop empirical methods for the assessment of both the extent of potential residual channeling and the degree to which residual channelling may bias the treatment effect. The proposed methods make use of propensity scores to measure the discrepancy between the treatment actually received and the treatment predicted on the basis of documented factors. We first suggest the index of residual channelling (IRC) as a tool to summarize the overall degree of such discrepancies. We then develop a propensity score-based method to assess to what extent residual channelling can bias the treatment effect. The method is based on modelling the interaction between the treatment and the discrepancy described above, which measures the expected extent of residual channelling for individual subjects.

6.3 Methods

In our recent work, we defined apparent channelling as channelling due to documented factors (4). Conversely, we now define residual channelling as channelling due to undocumented factors, i.e due to patient characteristics that are not available in a given database. These factors can be subdivided into variables that are unavailable because they are simply never documented in the database – for example, functional limitations or pain scores – and variables that are documented but that were ignored in the case of a particular patient because they were present only in the past, before the "beginning" of the database - such as for example prior use of NSAIDs in the more distant past. This is a common problem, as databases are often "left truncated", i.e. information about patient characteristics is often only available for a limited time period before the index date (point in time at which the patient enters the study base). In our empirical examples, for instance, this pre-index date period was limited to one year (11).

The expected treatment is determined solely by the documented factors included in the propensity score model (7). The treatment actually received, on the other hand, is determined by both documented and undocumented factors, representing, respectively, apparent and residual channeling, as well as by random processes not accounted for by either type of channeling, such as individual physicians' and / or patients' preferences or haphazard decisions (12). It then follows that the observed discrepancy between the expected and the actual treatment assignment reflects the joint impact of i) residual channeling, and ii) random processes. The relationships between these entities are illustrated in figure 1.

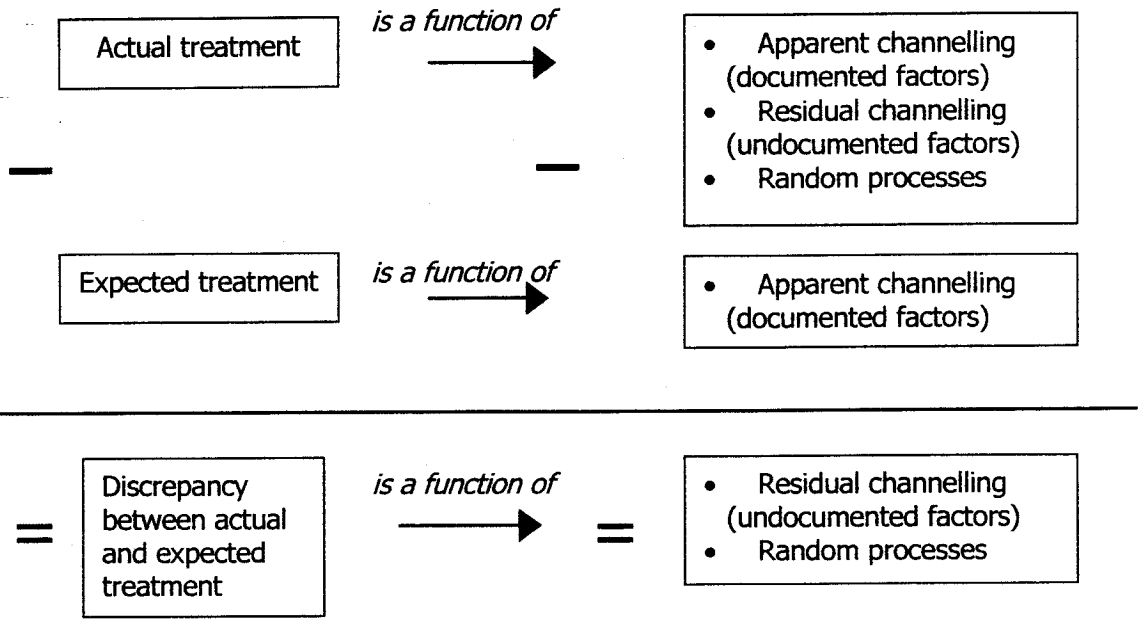

Figure 1. Relationships between expected and actual treatment, apparent and residual channelling.

We, therefore, postulate that the potential extent of residual channelling should be approximated by the discrepancy between actual and expected treatment. However, the observed discrepancy may also be affected by some degree of "random" processes leading to the choice of a given treatment (Figure 1). Whether residual channelling and random variation can be teased apart will be addressed later in this article. Residual channeling may lead to biased estimates of treatment effect, to the extent that it results in a differential assignment of patients with differing prognoses to two treatment alternatives. In contrast, random processes should have little impact on bias, since they are likely to influence treatment assignment in a non-differential way, acting independently of a patient's characteristics, i.e. of his / her prognosis.

6.3.1 Estimation of actual-expected discrepancy in treatment assignment

The discrepancy between actual and expected treatment can be estimated using either a categorical or a continuous approach. The categorical approach would be limited to estimating the frequencies of discrepancies in both directions (for example: discrepant, received drug A when expected to get drug B). The "natural choice" for the categorical estimation of discrepancies between expected and actual treatment would be to use discrepancy = 0.5 as a cut-off point. Patients with discrepancies below this level are considered "concordant", those above, "discordant".

The continuous approach would be somewhat more refined in that it would attempt to estimate the degree of discrepancy, the assumption being that the larger the discrepancy between actual and expected treatment, the more room there is for residual channelling to affect the treatment effect, even after adjusting for documented factors. Specifically, we propose to quantify the discrepancy as the difference between observed treatment (coded as 0 or 1) and propensity score, i.e. the expected probability of receiving the treatment coded as 1 (7). For example, a patient whose propensity score predicted only a 0.2 probability of being assigned to the drug coded as 1, but who nonetheless was prescribed this drug (actual treatment = 1) would have a discrepancy of 0.8 (actual – expected = 1 – 0.2 = +0.8). Discrepancy can range from – 1 to +1. Notice that patients prescribed drug "1" will always have positive discrepancies and those prescribed drug "0", negative discrepancies.

In order to summarize the overall magnitude of treatment discrepancies for a given group of patients, we propose the Index of Residual Channelling (IRC), which takes on

the following form:

$$IRC = \frac{\sum_{i=1}^{n} |o_i - p_i|}{n} \qquad (1)$$

where IRC = Index of residual channelling

o_i = actual (observed) treatment (o_i = 0 or 1) for patient i, i=1,..., n;

p_i = estimated propensity score, i.e. expected probability of receiving treatment "1" for individual patient i, i=1, ...,..., n;

n = total number of patients in the group

Because we defined "discrepancy" as the difference between actual and expected treatment, the IRC is also equal to the average absolute magnitude of the discrepancies. Since maximum likelihood estimation ensures that, on average, predicted treatment is closer to the actual treatment than to the alternative one, the average discrepancy cannot exceed 0.5. Thus, the IRC has a theoretical range that goes from 0 to 0.5 – or 0% to 50%. Values of discrepancy close to zero indicate that, for these populations, almost all variation in observed treatment has been explained by documented factors, so that there is only little room for residual confounding. Notice, however, that IRC values close to 0 would be obtained only if documented factors included patient characteristics that almost completely determined treatment choice so that most propensity scores would be very close to either 0 or 1.

6.3.2 Assessing the impact of residual channeling on estimated treatment effect

Since patients with values of discrepancies close to zero would be subject to very little if

any residual channelling, estimates of treatment effect after adjustment for documented covariates obtained from such patient groups should be nearly free of confounding by indication. At first glance, it would therefore appear that one approach to tackling the issue of residual channelling might be to use discrepancy as a stratification variable. The empirical data in a given database would be used to form subgroups of patients by classifying them according to their degree and direction of discrepancy between actual and expected treatment. Estimates from the stratum with the least discrepancy would be least subject to bias, and if estimates from other strata of discrepancy were not significantly different from it, then the overall estimate of effect obtained from the whole cohort could be considered as being reasonably free of confounding by indication. Since the range of possible values for discrepancy is fixed, it would be desirable to use limits selected *a priori* to define such strata, rather than data-dependent limits, such as empirical quintiles, for example, that depend on the particular distribution of discrepancy within a given database. Another advantage of *a priori* limits would be that they would allow for comparisons across different databases. For example, observed discrepancies could be divided into categories with the following boundaries: -1 to –0.6, -0.6 to –0.2, -0.2 to 0.2, 0.2 to 0.6 and 0.6 to 1.0.

However, a major problem with such a stratification is that the sign of a discrepancy is always determined by actual treatment: patients treated with drug "1" always have positive discrepancies, and those receiving drug "0", negative discrepancies. Thus, one cannot estimate the treatment effects in any stratum other than the middle, or "neutral", stratum (-0.2, 0.2), because all other strata will only contain patients having exclusively received one of the two treatment alternatives. Therefore, a treatment effect could be estimated only within the stratum with the least discrepancy (the neutral

stratum).

Furthermore, it is possible that this middle stratum – i.e. the stratum of primary interest – has too few data points. This can be remedied by gradually increasing the bounds of the stratum until it contains enough data to produce a stable estimate of treatment effect. However, doing so amounts to including more discordant patients into the stratum and, thus, increasing the risk of residual channelling within this stratum. At any rate, if the middle stratum contained too little data to produce a stable estimate of effect, this should be heeded as a warning sign that there may be residual channelling at play, despite stratification by discrepancy. In the extreme, it is possible that data in the middle stratum are so sparse that it is impossible to obtain a stable yet relatively unconfounded estimate of treatment effect using this stratified approach.

To avoid the limitations of the stratification by discrepancy, we propose an alternative approach. This approach is based on using regression modelling to assess how the estimated treatment effect changes depending on the magnitude of discrepancy. Like in the stratified approach, it is expected that patient sub-groups that are free of discrepancy should also be free of residual confounding.

One alternative to stratification is to model the dependence of the treatment effect on the magnitude of the discrepancy through interaction modelling. If the effect of treatment A vs treatment B depends systematically on the magnitude of the discrepancy between actual and expected treatment, then a statistically significant treatment-by-discrepancy interaction should be found. A conventional approach to interaction modelling would require including the main effects of both treatment and discrepancy, as well as their interaction, in addition to all the covariates. However, in our specific

context such a model may be affected by a near-multi-collinearity problem. The reason is that the discrepancy is in fact a function of (i) the actual treatment and (ii) the patient's characteristics used to estimate the propensity scores, most of which are also included in the same model as separate covariates. Thus, the model with both main effects and interaction, as well as individual covariates, would in fact represent the effects of most covariates twice, leading to a near-multi-collinearity, which could affect both the numerical stability and the interpretability of the regression coefficients.

An alternative to this approach is to exclude the main term for discrepancy from the above-mentioned model but to keep the interaction term. This takes care of the multi-collinearity problem, but, the resulting model is somewhat "asymmetrical" insofar as the reference category for the treatment effect is composed of all patients receiving the drug coded "0" regardless of discrepancy, and the interaction (the effect of discrepancy) is "active" only for patients having received drug "1". Thus, such a model only enables one to estimate the effect of discrepancy in the treatment group coded "1". This problem can, however, be solved easily, by simply running a similar regression model using "reverse coding" for the drugs: the drug previously coded "1" becomes "0", and vice versa. The absolute value of the discrepancy is not affected by such a change and thus does not need to be recalculated.

The disadvantage of this "asymmetric" interaction model is that it does not allow for a direct comparison between the two treatment groups regardless of discrepancy. However, this estimate is easily obtained by running the full model without interaction terms. The advantage of the "asymmetric" interaction model applied separately with both coding strategies is that it enables one to estimate both the treatment effect in the

absence of discrepancy as well as the effect of discrepancy in each treatment group, while avoiding the multi-collinearity problem.

6.3.3 Application of the interaction method to assess the effects of treatment on GI events

The binary outcome variable for all our regression models was any GI event (coded as "1") vs none in the six months following the index date. This classification was described in detail in our previous study (11). A large number of patients experienced GI events in the 6 months following the index date in all three examples. The distribution of GI events according to actual treatment assignment is shown in Table 1.

Table 1. Proportion of patients having experienced GI events in the 6 months following their index date, according to actual treatment assignment.

Example	Actual Treatment	Patients with any GI event, % (n)	Total number of patients, % (n)
1 COX-2 vs NSAID (switchers) (n = 27 247)	COX-2	34.7% (4181)	12 061
	NSAID	26.6% (4046)	15 186
2 Acetaminophen vs NSAID (new users) (n = 97 216)	Acetam.	22.0% (10392)	47 258
	NSAID	17.1% (8564)	49 958
3 Naproxen vs Diclofenac (new users) (n = 33 875)	Naproxen	17.1% (3560)	20 782
	Diclofenac	17.4% (2278)	13 093

Next, we estimated the effect of treatment assignment (COX-2 inhibitor vs NSAID) on GI events up to six months after the index date, while controlling for all other available potential factors that might have an effect on GI events. These variables were the same as those entered into the full propensity score model: patient's age, gender, location of residence (rural vs urban); for the year preceding the index date: the number of hospitalizations, number of days in hospital, number of visits to GPs, specialists, rhumatologists, gastroenterologists, number of emergency room visits and number of GI events, number of diagnoses of cancer, heart failure, musculoskeletal disease; the season of the index dispensation, the total number of prescriptions for the month

preceding the index date; for the year preceding the index date, the number of prescriptions for NSAIDs, acetaminophen, aspirin, steroids, anticoagulants, opiates and gastroprotective agents in particular, and the chronic disease score (13)). To estimate the adjusted effect of treatment, we used backward logistic regression (stepwise elimination), with the presence of any GI events as the binary dependent variable and the treatment indicators (coded as COX-2 inhibitor = 1 and NSAID = 0) as independent variable of primary interest ("exposure"). Variables not significant at the 0.05 level according to Wald's chi-square test were sequentially eliminated using backward elimination, and only statistically significant variables were retained in the final model. We also estimated the same multivariable model using only the patients in the neutral stratum of discrepancy (range: -0.2, 0.2). Finally, for comparison purposes, we also estimated the crude estimate of treatment effect from a simple logistic regression analysis where the model contained the treatment assignment as the only independent variable.

In order to examine the performance of the proposed method in situations with comparatively more residual channelling, we applied them while purposefully excluding the most significant predictors of treatment assignment for each of the three examples (11). This implied, on average, larger discrepancies between actual and expected treatment. As some of the excluded variables were also important risk factors for the clinical outcome (GI events), we expected that the magnitude of the resulting discrepancies would correlate with the strength of residual confounding.

6.4 Results

Here we illustrate the proposed methods using three empirical study bases used in our previous study to illustrate situations with various degrees of apparent channeling (11).

6.4.1 Example 1: COX-2 and NSAID users

The cohort used in example 1 contains current NSAID users who either remained on NSAID therapy or switched to COX-2 inhibitors after their introduction onto the market. On the basis of prior non-experimental studies (5), we suspected that there should be considerable apparent channeling and likely residual confounding in this cohort. First, sicker patients with rheumatoid arthritis or osteoarthritis who did not improve on standard NSAID therapy would be more likely to have been switched to COX-2 inhibitors when they were introduced onto the market. Moreover, as some of the patient characteristics determining the treatment choice are not available in our database, the resulting variation in treatment assignment would fall into the domain of residual channeling. Finally, some of these undocumented predictors of treatment assignment would also be risk factors for the clinical outcome (GI events), thus producing confounding by indication.

The categorical estimation of discrepancies revealed that there were 8 602 patients with absolute discrepancies greater than 0.5, representing 31.6% of a total of 27 247 cohort members. Of these discrepant patients, 4 782 (55.6%) got a COX-2 inhibitor when they were more likely to get an NSAID, the converse being true for 3820 patients (44.4 %), who got an NSAID when their propensity score predicted that they were more likely to receive a COX-2 inhibitor. The results of categorical discrepancy estimations are shown in Table 2 and the distribution of discordant pairs according to treatment assignment is

shown in Table 3 for all three examples. Although the IRC is quite similar in the three examples (Table 2), we expect that the relative shares of residual channelling vs random processes are quite different.

Table 2. Categorical assessment of discrepancy, and corresponding IRC

Example	Model (no.)*	% discrepant patients (n)	IRC (%)
1 COX-2 vs NSAID (switchers) (n = 27 247)	Full (1)	31.6 % (8602)	40.8
	Reduced (10)	39.0 % (10624)	45.8
2 Acetaminophen vs NSAID (new users) (n = 97 216)	Full (1)	36.5 % (35524)	45.2
	Reduced (8)	37.2 % (36187)	45.7
3 Naproxen vs Diclofenac (new users) (n = 33 875)	Full (1)	38.6 % (13089)	47.0
	Reduced (9)	38.6 % (13087)	47.1

* Model numbers are used consistently throughout this thesis. See Chapter 3, Table 2 for detailed definition of models.

Table 3. Distribution of discordant pairs according to treatment assignment.

Example	Model (no.)*	Actual Treatment	Expected treatment	Patients, % (n)
1 COX-2 vs NSAID (switchers) (n = 27 247)	Full (1)	COX-2	NSAID	55.6% (4782)
		NSAID	COX-2	44.4% (3820)
	Reduced (10)	COX-2	NSAID	74.6% (7925)
		NSAID	COX-2	25.4% (2699)
2 Acetaminophen vs NSAID (new users) (n = 97 216)	Full (1)	Acetam.	NSAID	63.1% (22420)
		NSAID	Acetam.	36.9% (13104)
	Reduced (8)	Acetam.	NSAID	63.2% (22871)
		NSAID	Acetam.	36.8% (13316)
3 Naproxen vs Diclofenac (new users) (n = 33 875)	Full (1)	Naproxen	Diclofenac	0.42% (55)
		Diclofenac	Naproxen	99.58% (13034)
	Reduced (9)	Naproxen	Diclofenac	0.36% (47)
		Diclofenac	Naproxen	99.64% (13040)

* Model numbers are used consistently throughout this thesis. See Chapter 3, Table 2 for detailed definition of models.

The continuous estimation of discrepancy for example 1, which uses the actual value of discrepancy for each patient yielded a more differentiated picture, which is illustrated in

Figure 2:

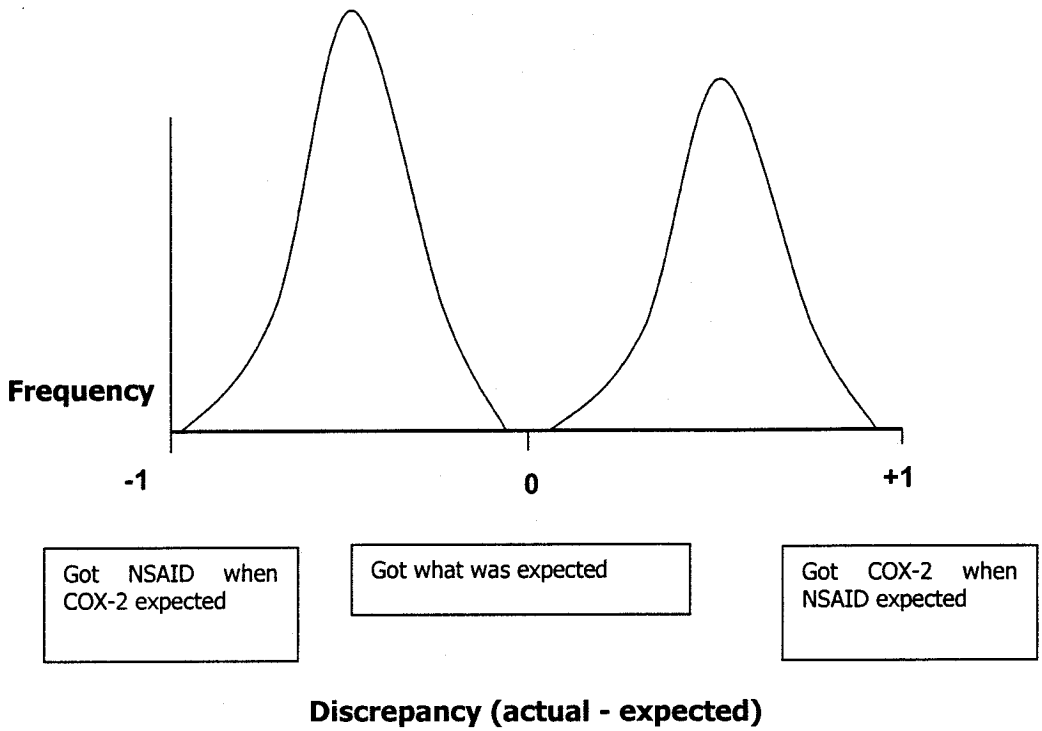

Figure 2. **Continuous estimation of discrepancy between expected and actual treatment for COX-2 and NSAID users (Example 1, full model (Model 1)).**

6.4.1.1 Index of residual channelling

The index of residual channelling (IRC) (Table 2) varied little between examples, but increased, as expected, when residual channelling was artificially introduced by running a reduced model. In example 3, the IRC was higher than in all other examples. This illustrates the fact that the IRC – and thus the quantitative estimation of average discrepancy – cannot distinguish between true residual channelling and quasi-random treatment assignment.

6.4.1.2 Assessing the impact of residual channelling on treatment effect through interaction modelling

Table 4 shows the estimated effect of treatment with COX-2 inhibitors, relative to NSAIDs, on GI events obtained from five different logistic regression models. The crude estimate of effect (OR=1.46 ((1.39 – 1.54)) would appear to indicate that users of COX-2 inhibitors have an almost 50% higher risk of all GI events during the 6 months after the index date compared to NSAID users. However, this estimate is subject to bias, which is corrected at least in part when the full model is run, thereby adjusting for those documented factors that were identified as statistically significant risk factors for GI events. Such adjustment yields an OR of 0.82 (0.77 - 0.82) for users of COX-2 inhibitors compared to NSAID users (2nd row of Table 4), indicating, contrary to the crude estimate, a protective effect of COX-2 inhibitors against GI events, as would be expected based on prior knowledge (14;15). However, it is likely that this estimate is still affected by confounding due to channelling by some undocumented risk factors. To explore this issue further, in the remaining models in Table, 4, we consider the effect of discrepancy between actual and expected treatment. The full model applied to the "neutral" stratum with only minimal discrepancies yields essentially the same point estimate (0.83 (0.60 – 1.16), indicating that either residual channelling has little impact on the treatment effect estimate, or the proposed stratification method is not sufficiently sensitive to detect residual channelling that may be present in this specific situation. Note that the 95 % confidence interval for the OR in the analysis restricted to non-discrepant subjects includes 1, which reflects reduced power and precision of such analyses, indicating another limitation of the stratified approach.

Table 4. Effect of adding interaction term for treatment assignment and absolute value of discrepancy (Example 1, full model (Model 1) and reduced (Model 10)).

Model (all models were applied to whole cohort, n = 27 247)	**Actual exposure**	**Odds ratio (OR) of all GI events at 6 months (95% CI)**
Crude model (only treatment assignment as independent variable)	NSAID COX-2	OR = 1 OR = 1.46 (1.39 – 1.54)
Full model	NSAID COX-2	OR = 1 OR = 0.82 (0.77 – 0.88)
Full model, Stratum 3 (neutral) Discrepancy -0.2, 0.2 (n = 4121)	NSAID COX-2	OR = 1 OR = 0.83* (0.60 – 1.16)
Full model with interaction term for treatment assignment and absolute value of discrepancy	NSAID COX-2 (Coding A)	OR = 1 OR (tx) = 0.55 (0.47 – 0.64) OR (tx * adiscr)= 2.29 (1.71 – 3.07)
	COX-2 NSAID (Coding B)	OR = 1 OR (tx) = 0.94* (0.80 – 1.10) OR (tx * adiscr)= 1.68* (1.23 – 2.29)
Reduced model (Model 10)	NSAID COX-2	OR = 1 OR = 1.13 (1.07 – 1.20)
Reduced model (Model 10) with interaction term for treatment assignment and absolute value of discrepancy	NSAID COX-2 (Coding A)	OR = 1 OR (tx) = 0.66 (0.49 – 0.87) OR (tx * adiscr)= 2.71 (1.64 – 4.49)
	COX-2 NSAID (Coding B)	OR = 1 OR (tx) = 0.93* (0.72 – 1.21) OR (tx * adiscr)= 1.83* (1.05 – 3.19)

*: estimate obtained from a logistic regression where all variables found to be significant in a first stepwise backward logistic regression were forced into the model, and the non-significant variables (either treatment assignment variable and/or the interaction variable) were also included but not forced.

adiscr = absolute value of discrepancy (actual – expected treatment)
tx = actual treatment

Coding A: NSAID = 0, COX-2 inhibitor =1
Coding B: NSAID = 1, COX-2 inhibitor =0

To avoid such limitations of stratified analysis, we then added a term for the interaction between treatment effect and absolute discrepancy to the full model and estimated this model in the entire cohort. The results of these analyses are shown in Table 4. Using coding A (with interaction applied to COX-2 inhibitor users only), the interaction term is statistically highly significant ($p<0.001$), which indicates that the estimated relative risks for COX-2 inhibitor users vary systematically depending on the discrepancy between actual and predicted treatment, suggesting the impact of residual channelling on the estimated treatment effect. The reference category in this model is comprised of all NSAID users regardless of their individual discrepancies. In the presence of interaction, the effect of treatment in the absence of discrepancy can be estimated from the term for treatment alone (16). Thus, the full model with interaction suggests that among COX-2 inhibitor users with no discrepancy (i.e. patients not likely to be affected by confounding due to residual channelling, and thus, residual confounding), COX-2 inhibitors appear to have an even stronger protective effect against GI events when compared to all NSAIDs users (OR=0.55 (0.47 – 0.64)) than was predicted from the model without interaction (OR=0.82 (0.77 – 0.88)). It is interesting to note that this estimate of effect for COX-2 inhibitors is in agreement with results of a recent systematic review of 8 randomized controlled trials (RCTs) containing a total of 5425 patients, where the relative risk of any GI adverse event was estimated at 0.54 (0.42 – 0.71) (15).

According to the model with interaction, for COX-2 users with an extreme discrepancy of 1 (i.e. patients who received a COX-2 inhibitor, in spite of being strongly expected, based on documented factors, to get an NSAID), the risk of all GI events during the six months following the index date is slightly increased, compared to all NSAID users (OR

= 1.26 (= 0.55 * 2.29)). In other words, patients who, on the basis of documented characteristics, were likely to get an NSAID but nonetheless got a COX-2 inhbitor, were at slightly higher risk of developing a GI event compared to all NSAID users. This finding may be explained by the more general interpretation of the model with interaction presented in the next paragraph.

In general, the fact that the OR for the interaction is significantly higher than 1 (OR = 2.29 (1.71 – 3.07)) indicates that, among patients who received a COX-2 inhibitor, the more they were likely to receive an NSAID as predicted by their propensity scores, the higher their risk of a GI event within the first six months after the index date. This result may seem paradoxical at first glance, but it can be interpreted logically if one keeps in mind that all the odds ratios shown above are adjusted for the documented but not for the undocumented risk factors. More specifically, they are adjusted for those documented patient characteristics that are predictors of GI events, and of treatment with COX-2 inhibitors. Patients who got a COX-2 inhibitor and were expected to get one based on documented factors probably did not have too many factors *other* than the documented ones that might increase their risk of experiencing a GI event. Thus, for non-discrepant COX-2 inhibitor users, the residual confounding due to undocumented factors is less likely to substantially affect the adjusted estimate of treatment effect. On the other hand, the more a patient who actually got a COX-2 inhibitor seems likely to receive an NSAID based on documented factors only, the more likely it is that she / he has some *undocumented* reasons motivating this treatment assignment. Because of the perception by physicians of COX-2 inhibitors as a safer treatment, these undocumented factors are likely to be risk factors for GI events. Because these factors cannot be adjusted for in the analysis, they create a spurious increase in the risk associated with

COX-2 inhibitors for more discrepant patients. Had these reasons been documented, they would have been adjusted for in the analysis and the interaction term would likely have been non-significant. Therefore, the fact that the interaction term between treatment assignment and discrepancy is statistically significant indicates that there is residual confounding at play – in other words, there are some undocumented factors associated with both treatment assignment and outcome that bias the treatment-outcome relationship, even after having adjusted for documented factors.

Next, in order to examine a situation where comparatively more residual channelling would be expected, we applied the methods described above using a model based on modified propensity scores, calculated without taking into consideration the most significant predictors of treatment assignment (11). Specifically, the reduced model used to estimated modified propensity scores excluded musculoskeletal disease, GI prophylaxis, sex and number of NSAID dispensations in the year prior to the index date, which were the four strongest predictors for treatment assignment in example 1 (11).

Because the reduced model is subject to more residual channelling, we expected that the estimated adjusted effect of COX-2 inhibitors from the model without interaction would be more confounded. Indeed, the "reduced" model approach yielded an adjusted OR=1.13 (1.07 – 1.20) that suggested that COX-2 inhibitors are associated with an increase in the risk of GI events, in contrast to the protective adjusted estimates in the "full" model (Table 4). For the same reason, it was also expected that the estimate of effect for the interaction (which represents the effect of discrepancy among COX-2 inhibitor users) would be stronger in the reduced model, compared to that for the full model. This was indeed the case, the OR for the interaction term being 2.71 (1.64 –

4.49), as opposed to 2.29 (1.71 – 3.07) for the full model. These results corroborate our expectation that the strength of the interaction between treatment and discrepancy may be used as an indicator of the strength of residual confounding. Interestingly, the estimate of treatment effect for COX-2 users without discrepancy (OR=0.66 (0.49 – 0.87)) yielded by the "reduced " model with interaction was quite similar to that for the full model with interaction term (OR=0.55 (0.47 – 0.64)), even if the corresponding estimates from the models without interaction were quite discrepant (adjusted OR of 0.82 vs 1.13). This is a potentially important finding, suggesting that the interaction model approach may have the ability of producing relatively unbiased estimates of main treatment effect that remain relatively unaffected by the degree of residual channelling present. Thus, the proposed approach appears to have the double advantage of quantifying, at least roughly, the magnitude of confounding due to residual channelling at play, and of giving a less biased estimate of treatment effect, by assessing this effect in patients without discrepancy.

To assess the effect of discrepancy on the estimated effect of NSAIDS, we simply reversed the original coding of the drugs and estimated the model with interaction using this new reverse coding ("Coding B": COX-2 inhibitor=0, NSAID=1). The interaction term now expressed the effect of discrepancy for NSAID users, with the reference category composed of all COX-2 users, regardless of their discrepancy. With this reverse coding, NSAID users without discrepancy are not at a higher risk of GI events than the "average" COX-2 user (Table 4, Coding "B": OR=0.94 (0.80 – 1.10)). This is plausible, since patients for whom NSAIDs was a strongly expected treatment were probably at relatively low risk for GI events and had been able to tolerate NSAIDs well in the past (it should be noted that, in this example, the NSAID users are those who

remained on NSAID therapy, whereas the COX-2 inhibitor users switched from NSAIDs to COX-2 inhibitors). A statistically significant interaction term ($p<0.05$) indicates that, with increasing discrepancy, the risk of GI events increases for NSAID users. In the case of an extreme discrepancy of 1, an NSAID user who was strongly expected to get a COX-2 inhibitor is estimated to have about 60% higher risk of subsequent GI events than a COX-2 inhibitor user with the same "risk profile" (Table 4: OR = 1.58 (=0.94*1.68)). This is also plausible since discrepant NSAID users are those whose characteristics – both documented *and* undocumented – would have made them likely to receive a COX-2 inhibitor. The increased risk of GI events in these patients is thus not surprising. Notice that OR = 1.58 ≅ 1/0.55 = inverse of COX-2 / NSAID for non-discrepant COX-2 users (Table 4, coding "A"), i.e for patients highly expected to get COX-2 inhibitors. Thus, both interaction models indicate that for patients whose documented factors strongly predispose them to receive a COX-2 inhibitor, use of NSAIDs is associated with about 60% increased risk of GI events.

In the reduced model (Table 4, Model 10, coding "B"), similar phenomenona are observed, with non-discordant NSAID users having the same risk as COX-2 inhibitor users (OR=0.93 (0.72 - 1.21), and with the risk for NSAID users increasing with increasing discrepancy. As expected, the OR for the interaction term in the reduced model (OR=1.83 (1.05 – 3.19) is slightly higher than that of the full model (OR=1.68 (1.23 – 2.29), reflecting, as expected, more residual confounding in the reduced model. However, the difference between the magnitude of the two interaction terms for NSAIDs (1.83 vs 1.68) is smaller than for COX-2 inhibitors (2.71 vs 2.29), which may reflect the fact that exclusion of major risk factors for GI events induced more confounding due to residual channelling among users of COX-2 inhibitors than users of NSAIDs.

Overall, the models with interactions suggest that relative risks of COX-2 inhibitors vs NSAIDs depend on the "risk profile". For subjects with no evident GI risk factors, NSAIDs appear to be as safe as COX-2 inhibitors (OR for NSAIDs= 0.94 (0.80 – 1.10)). For subjects highly expected to get COX-2 inhibitors, their use reduces the risk of GI events by about 40% to 50% (OR for COX-2 inhibitors = 0.55 (0.47 – 0.64)).

6.4.1.3 Example 2: Acetaminophen vs NSAIDs

Example 2 consists of a cohort of new users of either NSAID or acetaminophen before the introduction of COX-2 inhibitors onto the market (11). It is suspected that at least some residual channelling will be present, since sicker patients and patients at higher risk of GI events tend to be preferentially prescribed acetaminophen (17;18). Like for example 1, we performed univariate analyses with full as well as reduced models. The full model was also applied to the discrepancy-neutral stratum. The results of these analyses are shown in Table 5.

The crude estimate of effect (OR= 1.36 (1.32 –1.41)) would appear to indicate that users of acetaminophen have a higher risk of GI events during the 6 months following the index date compared to NSAID users. However, adjustment for all covariates included in the full model suggests that the risk of GI events for acetaminophen users is not significantly different than that of NSAID users (OR = 1.04 (0.99 – 1.08)). The full model applied to the neutral stratum (discrepancy –0.25 to 0.25) yields OR= 0.77 (0.48 – 1.22), suggesting a "protective" effect of acetaminophen, but the result is statistically non-significant and less precise than the estimate from the full model. It should be noted that the neutral stratum had to be widened, lest it be uninformative; this in itself raises the suspicion that residual channelling may be at play, since there are few highly concordant pairs.

Table 5. Effect of adding interaction term for treatment assignment and absolute value of discrepancy (Example 2, full model (Model 1) and reduced (Model 8)).

Model (all models were applied to whole cohort, n =97 216)	**Actual exposure**	**Odds ratio (OR) of <u>all</u> GI events at 6 months (95% CI)**
Crude model (only treatment assignment as independent variable)	NSAID Acetaminophen	OR = 1 OR = 1.36 (1.32 –1.41)
Full model	NSAID Acetaminophen	OR = 1 OR =1.04* (0.99 – 1.08)
Stratum 3: neutral Discrepancy -0.25, 0.25 (n =5819) note: neutral stratum had to be widened to include patients having received a COX-2 inhibitor, otherwise it was uninformative	NSAID Acetaminophen	OR = 1 OR = 0.77* (0.48 – 1.22)
Full model <u>with interaction</u> term for treatment assignment and absolute value of discrepancy	NSAID Acetaminophen <u>(Coding A)</u>	OR = 1 OR (tx) = 0.96* (0.87 – 1.04) OR (tx * adiscr)= 1.17* (0.98 – 1.40)
	Acetaminophen NSAID <u>(Coding B)</u>	OR = 1 OR (tx) = 0.82 (0.72 – 0.90) OR (tx * adiscr)= 1.47 (1.17 – 1.84)
Reduced model (Model 8)	NSAID Acetaminophen	OR = 1 OR = 1.17 (1.13 – 1.21)
Reduced model (Model 8) <u>with interaction term</u> for treatment assignment and absolute value of discrepancy	NSAID Acetaminophen <u>(Coding A)</u>	OR = 1 OR (tx) = 0.99* (0.90 – 1.08) OR (tx * adiscr)=1.41 (1.18 – 1.68)
	Acetaminophen NSAID <u>(Coding B)</u>	OR = 1 OR (tx) = 0.76 (0.67 – 0.85) OR (tx * adiscr)=1.68 (1.32 – 2.12)

*: estimate obtained from a logistic regression where all variables found to be significant in a first stepwise backward logistic regression were forced into the model, and the non-significant variables (either treatment assignment variable and/or the interaction variable) were also included but not forced.

Next, we applied to the full cohort a logistic regression model containing a term for the interaction between treatment effect and absolute discrepancy, as well as all other potential predictors of GI events. The reference category in this model is comprised of all NSAID users regardless of their individual discrepancies. The interaction term is statistically marginally non-significant (OR=1.17 (0.98 – 1.40)), indicating that the estimated relative risks for acetaminophen users do not vary systematically with the discrepancy between actual and predicted treatment. Indeed, acetaminophen users with a discrepancy = 1 would have an OR for GI events at 6 months equal to 1.12 = (0.96*1.17) (most likely non-significant), indicating that, in this group of patients too, discrepancy does not appear to predispose to GI events in an important way. This is to be expected, since patients likely to get NSAIDs who nonetheless get acetaminophen could hardly be put at higher risk of GI events, given the excellent safety profile of this drug (19). This is indicative of a lack of residual channelling in this situation. In this model, the main treatment effect in the absence of discrepancy is also non-significant (0.96 (0.87 – 1.04)), suggesting that patients with no discrepancy (i.e. therefore not subject to residual channelling) are also not at a higher – or lower – risk of GI events than all NSAID users combined. This result is also in accordance with that of the full model without interaction term (OR = 1.04 (0.99 – 1.08)).

The reverse coding approach showed that concordant NSAID users (i.e. not likely to be subject to residual channelling) are actually protected from GI events when compared to all acetaminophen users (OR=0.82 (0.72-0.90)). This phenomenon could be explained if the acetaminophen users were overall at higher risk of GI event - which could be the case if patients receiving NSAIDs were overall less severely ill than patients receiving

acetaminophen - due to both documented and undocumented factors, which would lead to residual confounding. Discordance is a stronger predictor of GI events among NSAID users than among acetaminophen users, as evidenced by the significant odds ratio (1.47 (1.17 – 1.84)) for the interaction term. The overall odds ratio for NSAID users with a discordance = 1 is indicative of a moderately elevated risk of GI events (OR=1.21 (=0.82*1.47), as could be expected of patients who, based on their documented characteristics, would have been more likely to get acetaminophen, but were nonetheless prescribed an NSAID.

The reduced model excluded the most significant predictors of treatment assignment (use of GI prophylaxis and number of hospitalizations in year prior to index date). Because this model is subject to more residual channelling, it was expected that the estimate of effect from the reduced model without interaction (OR=1.17 (1.13 – 1.21)) would be more confounded, and thus stronger than the estimate of effect from the full model without interaction (OR =1.04 (0.99 – 1.08)), which indeed was the case, the reduced model suggesting an increased risk of GI events for acetaminophen users. For the same reason, it was also expected that the estimate of effect for the interaction would be stronger in the reduced model, compared to that for the full model. This was indeed the case, the OR for the interaction term for the reduced model being 1.41 (1.18 – 1.68) for coding "A", as opposed to 1.17* (0.98 – 1.40) for the full model, and likewise for coding "B" (Reduced OR= 1.68 (1.32 – 2.12) vs Full OR= 1.47 (1.17 – 1.84)). It is noteworthy that, like in example 1, the OR for the interaction term in the reduced model was higher than in the full model, regardless of the coding strategy used. This is consistent with our hypothesis that discrepancy is a proxy for residual channelling. Finally, as in example 1, the OR for the "zero discrepancy" groups were

very similar for the "full" and "reduced" models, both for acetaminophen (0.96 vs 0.99) and NSAIDs (0.82 vs 0.76). Thus, even eliminating the most significant predictors of treatment assignment does not affect the estimates for the non-discrepant patients, who are least affected by residual channelling.

6.4.1.4 Example 3: Naproxen vs Diclofenac

The cohort used in example 3 consists of new users of naproxen or diclofenac, before the introduction of COX-2 inhibitors on the market, who had not been dispensed any NSAID prescription in the year prior to the index date. This example was chosen to illustrate a situation where little channelling – both apparent and residual – was expected. This is so because naproxen and diclofenac are two well-established NSAIDs that are perceived to be relatively similar in their therapeutic properties and side effects and are available as generics at comparable cost.

Like for examples 1 and 2, we performed similar analyses with full and reduced models. The results of these analyses are shown in Table 6. The crude, full model and stratum-specific model all yielded non-significant estimates of treatment effect, as would be expected from this example. The reference category is composed of all patients using diclofenac.

The interaction models yielded non-significant estimates of both treatment effect and interaction for both coding strategies. This is consistent with a situation where there is little or no channelling due to undocumented factors, as one would expect here. Fifty percent of patients in this cohort have a propensity score between 0.58 and 0.63 – thus discrepancies between 0.37 and 0.42. This highly centered distribution of propensity scores means that estimates for extremes of discrepancy (0 or 1) are essentially

extrapolations, and thus also very unstable, as can be seen from their confidence intervals. Therefore, this interaction model approach should probably not be applied to a situation where propensity scores, and thus discrepancies, are highly centered.

It is noteworthy that the point estimates for the interaction terms, though non-significant, are nonetheless higher than in any of the other three examples. It could be tempting to take this as meaning that important residual channelling is present, however the fact that they are not significant – unlike the interaction terms in examples 1 and 2 - casts doubt on this.

The reduced model excluded the most significant predictor of treatment assignment (rural site of residence) and yielded results essentially similar to those of the full model, except that, regardless of the coding strategy used, the odds ratio for the interaction term was slightly *stronger* (OR= 0.65 (0.25 – 1.27) and OR=2.64 (0.76 – 9.25)) than that of the full model (OR=0.77 (0.32 – 1.86) and OR=3.01 (0.94 – 9.63), for coding strategies A and B, respectively). Since all estimates were non-significant and had wide confidence intervals, such observations should be interpreted with caution, as they are likely only the product of random variations, and thus not reproducible.

Table 6. Effect of adding interaction term for treatment assignment and absolute value of discrepancy (Example 3, full model (Model 1) and reduced (Model 9)).

Model **(all models were applied to whole cohort,** **n = 27 247)**	**Actual exposure**	**Odds ratio (OR) of <u>all</u> GI events at 6 months (95% CI)**
Crude model (only treatment assignment as independent variable)	Diclofenac Naproxen	OR = 1 OR = 0.98* (0.93 – 1.04)
Full model (Model 1)	Diclofenac Naproxen	OR = 1 OR = 0.97* (0.91 – 1.04)
Stratum 3: neutral Discrepancy -0.45, 0.45 (n =20 140) note: neutral stratum had to be widened to include patients having received naproxen, otherwise it was uninformative	Diclofenac Naproxen	OR = 1 OR = 1.12* (0.31 – 4.14)
Full model <u>with interaction</u> term for treatment assignment and absolute value of discrepancy	Diclofenac Naproxen <u>(Coding A)</u>	OR = 1 OR (tx) = 1.08* (0.76 – 1.52) OR (tx * adiscr) = 0.77* (0.32 – 1.86)
	Naproxen Diclofenac <u>(Coding B)</u>	OR = 1 OR (tx) = 0.52* (0.26 – 1.07) OR (tx * adiscr)= 3.01* (0.94 – 9.63)
Reduced model (Model 9)	Diclofenac Naproxen	OR = 1 OR = 0.97 (0.91- 1.04)
Reduced model (Model 9) **<u>with interaction term</u>** for treatment assignment and absolute value of discrepancy	Diclofenac Naproxen <u>(Coding A)</u>	OR = 1 OR (tx) = 1.15* (0.79 – 1.66) OR (tx * adiscr)= 0.65* (0.25 – 1.67)
	Naproxen Diclofenac <u>(Coding B)</u>	OR = 1 OR (tx) = 0.57* (0.26 – 1.22) OR (tx * adiscr)= 2.64* (0.76 – 9.25)

*: estimate obtained from a logistic regression where all variables found to be significant in a first stepwise backward logistic regression were forced into the model, and the non-significant variables (either treatment assignment variable and/or the interaction variable) were also included but not forced.

6.5 Discussion

Confounding by indication is a central problem in non-experimental pharmacoepidemiologic research. Confounding by indication can be broken down conceptually into its two components, the risk factor effect and channelling. Should one of these two components be completely "controllable", then there would be no room for confounding by indication. Since the risk factor effect is an inherent, substantive association between some potential confounders and the outcome of interest, it cannot be eliminated if the potential confounders are not documented, even if one suspects its presence based on other studies. Channelling, on the other hand, is a phenomenon that is particular to a given patient population. Thus, we attempted to estimate the presence of channelling with a view to controlling it in the analysis (4;11). However, the index of apparent channelling, which we proposed in our recent work, provides only a partial solution to the problem of channelling because propensity scores can only account for covariates that were documented in the database and considered in the propensity score model (4). Therefore, in the present article, we developed new methods to deal with the problem of channelling due to undocumented factors, which we termed residual channelling.

6.5.1 Estimating discrepancy between expected and actual treatment

First, we postulated that residual channelling should be quantifiable by considering the discrepancy between the actual treatment received by a patient and the treatment predicted on the basis of that patient's propensity score. This discrepancy is due to i) residual channelling caused by undocumented factors, or ii) the effects of any random processes at play in the database, or a combination of these two phenomena. These

relationships were illustrated graphically in Figure 1. We proposed two ways of measuring discrepancy. The simplest approach, based on classifying as "discrepant" those patients who have absolute discrepancies greater than 0.5 and therefore received one treatment when they were more likely to get the other, is a fairly crude categorical classification that only allows for the identification of the presence and directionality of such discrepancies. Nonetheless, this gives some preliminary idea of the overall potential extent of residual channelling that may be present. For example, if only five percent of subjects in a database are discrepant, it is less likely that there is residual channelling than if thirty percent are discrepant. The problem is that this approach does not differentiate between patients with a very small discrepancy, say 0.05, and a large discrepancy, say 0.45, that is nonetheless smaller than 0.5. Finally, it should always be kept in mind that discrepancies may in part – or, in extreme cases – in totality, be due to random processes in a database.

The second, quantitative, approach to estimate discrepancy goes one step further and estimates the degree of discrepancy present by taking the difference between the actual treatment received, coded as 0 or 1, and the expected treatment, expressed as the propensity score. The overall, average discrepancy for a cohort can be calculated using the proposed index of residual channelling (IRC), which, on average, is equal to the mean difference between the actual and predicted treatments. It should be noted, however, that situations may arise whereby all patients within a sub-group (for example, the patients of a given MD) get the same treatment. Although this is strictly speaking a case of "absolute channelling" (all propensity score would necessarily be equal to zero, and the IRC would also equal zero), there should be no residual confounding possible if it could be determined with certainty that the determinant of treatment (here, the

physician) is completely unrelated to the treatment outcome.

6.5.2 Stratified approach

In the stratified approach, we formed a stratum based on discrepancy that contained the least discrepant patients, namely those with absolute discrepancies below 0.2. This was done because these patients are thought to be less subject to residual channelling, because their treatment assignment is well explained by documented factors. We then applied the regression model containing all potential documented predictors of GI events in order to estimate the effect of actual treatment on the risk of all GI events at 6 months. In all three examples, this yielded results that were similar to those produced when the same model was applied to the whole cohort. It is not possible, on the basis of these results, to say whether this is the case because there is no residual channelling at play, or because this method is simply too insensitive to detect it. However, the fact that, in examples 2 and 3, the bounds of the neutral stratum had to be widened to include enough patients to obtain a stable regression estimate should be seen as a warning signal that there were potentially too few non-discrepant patients in these databases, which may be indicative of the presence of residual channelling. Thus, the stratum-specific estimate obtained from strata with the "widened" boundaries should be considered with caution.

6.5.3 The Index of residual channelling

The index of residual channelling (IRC) was calculated for each of the models (full and reduced) for each of the three examples, and the results of these analyses were shown in Table 1. It can be seen that the IRC increases in each of the three examples as one goes from full to reduced model. This is to be expected, since excluding strong predictors of treatment assignment from the propensity score model pushes their

channelling effect into the realm of the undocumented, thus contributing to residual channelling. This is in fact the converse of what was observed with the IAC, which decreased when these factors were removed (11). At first glance, it may seem that the IRC would allow one to compare the degree of residual channelling across different databases. However, in doing so, it should be kept in mind that the IRC, by definition (since it corresponds to the average discrepancy), necessarily accounts not only for residual channelling but also for random variations. This is clearly illustrated by the fact that example 3 has the highest IRC, yet this is the example in which the least residual channelling is expected, based on prior knowledge. This seeming contradiction can be reconciled if one considers that the worse a propensity model is at predicting treatment assignment, the more individual propensity scores will be close to 0.5, and thus the larger the discrepancy (which, as mentioned earlier, is the difference between actual treatment (coded 0 or 1) and propensity score). An extreme case would be a randomized controlled trial, where all patients necessarily have a propensity score of 0.5. Here the IRC would be equal to its theoretical maximum, namely 50%. Example 3 illustrates the futility of applying the IRC to a database where little residual channelling is expected, both on substantive grounds and as evidenced by the lack of significance of all interaction terms in this model (Table 6).

6.5.4 Regression modelling approach

We then explored the possibility of estimating the effect of discrepancies – and thus of residual channelling – by using regression modelling instead of stratification. We applied a model that included, in addition to all suspected predictors of GI outcomes, a term for actual treatment and a term for the interaction between actual treatment and discrepancy. The reference category for the odds ratio of treatment effect in such a

model are all patients having received the drug coded "0", regardless of discrepancy. This model enables the calculation of how the odds ratio changes with increasing level of discrepancy.

Classical biostatistical dogma has it that an interaction term should be entered into a model only if both interaction variables are also entered independently as main terms. This would amount to running a model with three separate terms for treatment effect, absolute discrepancy, and the interaction between the two, so that the reference category would be non-discrepant patients using drug "0". However, because the actual treatment enters into all three terms, this leads to problems of multi-colinearity and instability of the estimates. To avoid the multi-colinearity problem, we used only the terms for treatment assignment and interaction between treatment assignment and absolute discrepancy. However, such a model does not enable us to assess the effect of discrepancy independently for patients taking drug "0". This can be overcome by simply reversing the coding of the drugs and re-estimating the model again. It was then possible to assess the effect of discrepancy in users of the "other" drug, and also to assess the risk of GI bleed in non-discrepant patients using this drug.

In the first example (Table 4), we were able to estimate the effect of treatment with COX-2 inhibitors in non-discrepant patients, revealing a protective effect (OR= 0.55 (0.47- 0.64)) that was stronger than that suggested by the conventional analyses based on the full model without interaction term (0.82 (0.77 – 0.88)). Furthermore, this result is in accordance with the results of prior randomized controlled trials that provided evidence for a protective effect of COX-2 inhibitors compared to NSAIDs (14;15). Benefits of treatment with COX-2 inhibitors had probably been underestimated because

sicker patients were being assigned to them. In this case, the COX-2 inhibitors nonetheless have such a beneficial side-effect profile that even results of the overall analyses (without interaction) that are subject to residual channelling were able to demonstrate their superiority. However, this may not always be the case, and the margin of error that makes the difference between beneficial and neutral, or worse, between neutral and detrimental, may be small enough that it could be masked by residual confounding, should it not be adequately taken into account and controlled for. Finally, the significant OR for the interaction term (2.29 (1.71 – 3.07)) shows that, as discrepancy increases, the protective effect of COX-2 inhibitors decreases. In fact, at maximal discrepancy – i.e. maximum residual channelling – the risk of GI events is higher (OR = 0.55*2.29 = 1.26) in COX-2 inhibitor users compared to NSAID users. In other words, the (undocumented) risk profile of these patients is such that they experience more GI events than NSAID users, despite the more benign side-effect profile of COX-2 inhibitors.

In the second example (Table 5), the main treatment effect estimated using the interaction model was not significantly different from that of the model without interaction term for non-discrepant patients taking acetaminophen, however for non-discrepant patients taking NSAIDs (Coding B), the interaction model suggested a protective effect against all GI events. This could be plausible if patients receiving NSAIDs were overall less severely ill than patients receiving acetaminophen.

On the contrary, in the third example (Table 6), neither the treatment effect term nor the interaction term were significant. The fact that the interaction terms were not significant in the third model is consistent with the expectation that there should be very

little residual channelling if any in this example.

An important limitation of our proposed model-based approach is the difficulty involved in interpreting the results it yields, as well as the "asymmetry" of the contrasts it can produce, which calls for the use of two different coding strategies. On the other hand, this complexity of the proposed methods forces the user to establish clarity about the contrasts of interests for which estimates of treatment effect are to be obtained. Another potential limitation of our proposed approach becomes an issue in situations where there is virtually no apparent channelling (IAC $\approx$ 0). In such a situation, most patients would necessarily have a discrepancy close to 0.5, and it would therefore become very difficult to obtain a stable estimate of the coefficient for the interaction term between discrepancy and actual treatment. In other words, even if strong residual channelling were present, it would probably be difficult to detect it.

6.5.5 Conclusion

Our proposed interaction modelling approach enables investigators to obtain both an estimate of the importance of residual channelling and a less biased estimate of treatment effect in non-discrepant patients, which is proportional to the magnitude of the coefficient for the interaction term. By applying this method, we were able to use the proposed approach in a non-experimental study to demonstrate that COX-2 inhibitors, when compared to NSAIDs, probably have an even stronger protective effect against GI events than would have been suspected on the basis of the overall analyses without interaction terms.

Given the unique and central role of non-experimental research in establishing the post-marketing safety and efficacy of new drugs (3), it is our hope that our proposed

approach will lead not only to more valid conclusions being drawn from non-experimental research, but also to a greater appreciation and regard for non-experimental research in general and for database research in particular as powerful and valid means to answer relevant clinical questions and ultimately, improve patient care.

6.6 Acknowledgements

This project was supported by a grant from the "Réseau FRSQ sur l`utilisation des médicaments" of the government of the province of Québec. Michal Abrahamowicz is a scientist of the Canadian Institutes of Health Research (CIHR).

Many thanks to Drs Elham Rahme and Lucie Blais for stimulating discussions and insightful comments at various stages in the development of this project.

6.7 References

1. Strom BL, Melmon KL , Miettinen OS. Post-marketing studies of drug efficacy: why? Am J Med 1985; 78:475-80.

2. Strom BL (ed.). Pharmacoepidemiology. 3 ed. Chichester, UK: John Wiley and Sons, 2000.

3. Skegg DC. Evaluating the safety of medicines, with particular reference to contraception. Stat Med 2001;20:3557-69.

4. Bjerre, L. M., Abrahamowicz, M. A., and LeLorier, J. Channelling in non-experimental pharmacoepidemiologic research: The estimation of apparent channeling Part I: Theory and hypothetical examples. 2003. Ref Type: Thesis/Dissertation.

5. Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.

6. Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. JASA 1984;79:516-24.

7. Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757-63.

8. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

9. Miettinen OS. Stratification by a multivariate confounder score. Am J Epidemiol 1976;104:609-20.

10. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 1999;29:722-9.

11. Bjerre, L. M., LeLorier, J., and Abrahamowicz, M. A. Channelling in non-experimental pharmacoepidemiologic research: The estimation of apparent channeling Part II: Empirical examples. 2003. Ref Type: Thesis/Dissertation.

12. Tamblyn R, Abrahamowicz M. Drug utilization patterns. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: John Wiley & Sons, 1998:1235-47.

13. Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992;45:197-203.

14. Silverstein FE, Faich G, Goldstein JL. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS study: a randomized controlled trial. JAMA 2000;248:1247-55.

15. Deeks JJ, Smith LA , Bradley MD. Efficacy, tolerability, and upper gastrointestinal safety of celecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials. BMJ 2002;325:619-26.

16. Kleinbaum DG, Kupper LL, Morgenstern H. Multivariate regression analysis. 3rd ed. 1998.

17. Rahme E, Marentette MA, Kong XS, LeLorier J. Use of NSAIDs, COX-2 inhibitors, and acetaminophen and associated coprescriptions of gastroprotective agents in an elderly population. Arthritis Rheum 2002;47:595-602.

18. Rahme E, Pettitt D , LeLorier J. Determinants and sequelae associated with utilization of acetaminophen versus traditional nonsteroidal antiinflammatory drugs in an elderly population. Arthritis Rheum 2002;46:3046-54.

19. Katzung BG. Basic and clinical pharmacology. 6th ed. East Norwalk (CT): Appleton and Lange, 1995.

7 Summary and conclusion

7.1 Summary

This thesis addressed the phenomenon of channelling in non-experimental research and its impact on confounding by indication, which is a major threat to the validity of non-experimental studies of medications.

First, I proposed a conceptual framework to better understand the phenomenon of channelling, which is a necessary condition for confounding by indication to occur. I distinguished between channelling due to documented factors, or apparent channelling, and channelling due to undocumented factors, or residual channelling. I then proposed a novel approach, the Index of Apparent Channelling (IAC), to quantify the extent of apparent channelling. This method makes use of propensity scores, which express the estimated probability of a patient being assigned to one of two treatment alternatives, based on documented covariates. I illustrated the IAC's behaviour using both hypothetical and empirical examples, and demonstrated that it responds well to the degree of apparent channelling present in a patient cohort. Nonetheless, the IAC is limited by its sole reliance on documented factors. Undocumented factors can also lead to channelling, or residual channelling, and thus to confounding by indication. Therefore, I explored various methods to assess the presence and extent of residual channelling, and its impact on confounding by indication.

First, I postulated that channelling due to undocumented factors should vary in relation to the discrepancy between actual treatment received and predicted treatment, as predicted by documented factors. The less observed treatment assignment could be explained by documented factors, the greater the potential for residual channelling. I

proposed the Index of Residual Channelling (IRC) as a means to summarize the overall degree of discrepancies for patients in a cohort, and thus the potential extent of residual channelling in this cohort. Although the IRC is relatively sensitive to the presence of residual channelling, it cannot differentiate between large discrepancies due to residual channelling based on undocumented factors, and large discrepancies due to random processes, such a for example subjective treatment preferences. Therefore, it was necessary to explore other methods to assess the presence and impact of residual channelling.

Whereas, in a given application, it may be difficult to separate residual channelling and "random" subjective treatment preferences, what matters the most is the extent of bias induced by residual channelling. Therefore, I proposed a propensity score-based method to assess to what extent residual channelling can bias the treatment effect. The method is based on modelling the interaction between the actual treatment and the measure of residual channelling, operationally defined as the discrepancy between actual and predicted treatment. Empirical examples suggest that this method can provide less biased estimates of treatment effect and that it can help assess the magnitude of residual channelling present in a cohort. The proposed approach thereby provides a practicable tool for reducing the bias in the estimates of treatment effect obtained from non-experimental studies of medications. Figure 1 illustrates the steps to be considered in the application of the methods proposed in this thesis to the analysis of an administrative drugs claims database. This graphical representation is necessarily simplified, but is nonetheless an attempt to convey the essence of the process.

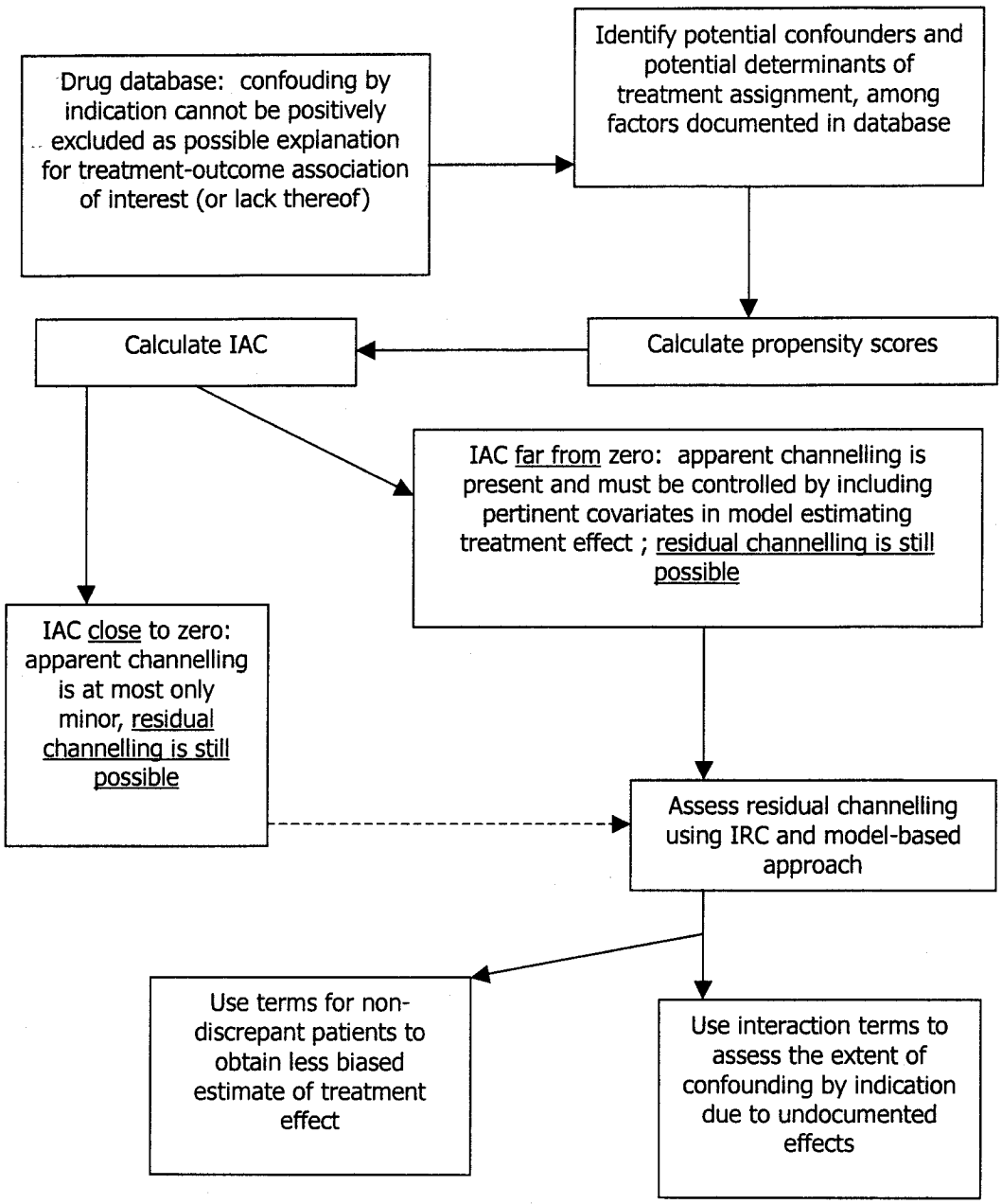

Figure 1. Applying the IAC, the IRC and the model-based approach for the quantification of apparent and residual channelling.

7.2 Strengths and Limitations

When compared with earlier approaches to the problem of confounding by indication such as ecological analyses, confounder scores and instrumental variables (1-3), the methods proposed in this thesis stand out in a number of ways. First of all, unlike ecological analyses (1) and confounder scores (2), the proposed methods enable a quantification of channelling due to undocumented factors. However, unlike instrumental variables (3), with whom they share the property of accounting for undocumented variables, they are not strongly dependent on assumptions that are difficult to verify, let alone satisfy. They do, however, rely on the assumptions inherent to multivariate analyses. Furthermore, to my knowledge, our approach is unique, in that it breaks down the problem of confounding by indication into its component elements, and specifically focuses on the issue of channelling and its quantification.

A specific limitation of our approach to residual channelling is certainly its complexity. The design of the analyses involves many steps and the interpretation of their results is more complex than is usually the case with logistic regression analyses. However, this forces the user to clearly define the contrasts of interest, as the interpretation of the parameter estimates obtained from the modelling approach depends on a clear definition of these contrasts. A further limitation of the proposed approach is that it may be misleading, or even inappropriate to apply it under some circumstances. In particular, if there is no evidence of apparent channelling (IAC $\approx$ 0), then it becomes very difficult to estimate the coefficient of the interaction term between treatment and absolute discrepancy, since most patients would have a discrepancy close to 0.5. In such a situation, our proposed method would probably fail to detect residual channelling even if it were present. This caveat is illustrated in Figure 1 using a dashed arrow at

this step in the process.

Our empirical examples could not avoid some of the common limitations associated with database studies. One of the study limitations was the lack of information about prescriptions dispensed in hospital. It is quite possible, for example, that therapy with an NSAID, COX-2 inhibitor or acetaminophen could have been initiated while a patient was in hospital. Had this occurred toward the end of the index date window, or had the patient not renewed his / her prescription after being discharged from hospital, this patient would have not have been recruited into the relevant cohort, even though all other inclusion and exclusion criteria may have been satisfied. However, since this loss to recruitment would most likely be non-differential between the various treatment alternatives, it is unlikely that it would have biased the results of our analyses.

Another limitation of administrative drug claims databases is that they do not enable the verification of actual consumption. The index date is the date at which the index prescription was dispensed, however there is no way of knowing whether the patient actually took the drug. Had patients not taken their drugs, this would lead to a weakening of any association between drug exposure and, say, GI events. However, the estimates of treatment effect we obtained from our empirical examples in Manuscript 3 were in agreement with the results of prior randomized controlled trials, so we are reasonably confident that non-adherence to treatment did not significantly bias our results.

However, since these limitations are inherent to most administrative drug claims databases, we felt that it was in fact appropriate that our methods be subject to the same limitations – indeed, this even seemed desirable, for what would be the use of

developing a novel method that works marvellously well under ideal conditions, but breaks down when confronted with the limitations and imperfections of the "real world"? To confirm the validity of our proposed methods, whenever possible, we compared our results with those of either randomized controlled trials or descriptive survey studies (4). These comparisons supported the validity of our proposed approaches.

7.3 Areas for future research

As a first step in consolidating and validating our proposed methods, it is essential that they be applied to other databases where various degrees of channelling are suspected, either on the basis of prior knowledge, or, ideally, because this knowledge is supported by data from randomized controlled trials, or at least descriptive survey studies. Second, it is important that researchers develop a feel for the range of values taken on under real-life conditions by our indices (IAC, IRC) and by the interaction term in the model to assess the presence of residual confounding.

An important area for future work concerns the situation where there is little or no apparent channelling, but substantial residual channelling. As detailed above and in Manuscript 3, the proposed methods may have difficulty in detecting the presence of residual channelling in such a situation. The importance of this problem needs to be established using both hypothetical as well as empirical examples.

Another area for investigation would be a more detailed analysis of the role of individual physicians and of physicians' specialty on the degree of channelling present in a database. There is most likely a number of different 'types' of prescribers, and examining their behaviour may help us identify sub-cohorts that are free of channelling, and thus free of confounding by indication. Since the actual treatment received by a

patient is intimately linked to the prescribing physician, this area might warrant high priority in future investigations.

7.4 Final conclusion

The methods developed in this thesis provide new insights into the role of channelling – both apparent and residual – and its impact on confounding by indication. These methods may have important implications in the analysis of administrative drug claims databases. Although they do not provide a definitive solution to the problem of confounding by indication in non-experimental research, it is our hope that this reflection and the methods that resulted from it will be put to use, and that they may contribute in a small way to enhancing the validity of study results in pharmacoepidemiologic research.

7.5 References

1. Wen SW, Kramer MS. Uses of ecologic studies in the assessment of intended treatment effects. J Clin Epidemiol 1999;52:7-12.

2. Miettinen OS. Stratification by a multivariate confounder score. Am J Epidemiol 1976;104:609-20.

3. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 1999;29:722-9.

4. Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.

Thesis Reference List

Abrahamowicz M, du Berger R, Grover SA. Flexible modeling of the effects of serum cholesterol on coronary heart disease mortality. Am J Epidemiol 1997;145:714-29.

Abrahamowicz M, MacKenzie T, Esdaile JM. Time-Dependent Hazard Ratio: Modeling and Hypothesis Testing With Application in Lupus Nephritis. Am J Epidemiol 1996;91: 1432.

Adams SS, Cobb R. A possible basis for the anti-inflammatory activity of salicylates and other nonhormonal anti-rheumatic drugs. Nature 1958;181:773-.

American College of Rheumatology Ad Hoc Committee on Clinical Guidelines. Guidelines for the management of rheumatoid arthritis. Arthritis Rheum 1996;39:713-22.

Antiplatelet Trialists' Collaboration. Collaborative overview of randomized trials of antiplatelet therapy: I. Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. BMJ 1994;308:81-106.

Bjerre, L. M., Abrahamowicz, M. A., and LeLorier, J. Channelling in non-experimental pharmacoepidemiologic research: The estimation of apparent channeling Part I: Theory and hypothetical examples. 2003. Ref Type: Thesis/Dissertation.

Bjerre, L. M., LeLorier, J., and Abrahamowicz, M. A. Channelling in non-experimental pharmacoepidemiologic research: The estimation of apparent channeling Part II: Empirical examples. 2003. Ref Type: Thesis/Dissertation.

Bjerre, L. M., LeLorier, J., Dorais, M., and Abrahamowicz, M. A. Channelling in non-experimental pharmacoepidemiologic research: Assessing the impact of residual channelling on confounding by indication. 2003. Ref Type: Thesis/Dissertation.

Blais L, Ernst P, Suissa S. Confounding by indication and channeling over time: the risks of beta 2-agonists. Am J Epidemiol 1996;144:1161-9.

Bombardier C, Laine L, Reicin A. Comparison of upper gastro-intestinal toxicity of rofecoxib and naproxen in patients with rheumatoid arthritis. N Engl J Med 2000;343:1520-8.

Brenner H, Blettner M. Controlling for continuous confounders in epidemiologic research. Epidemiology 1997;8:429-34.

British Medical Association, Royal Pharmaceutical Society of Great Britain. British National Formulary. London: The British Medical Association and the Royal Pharmaceutical Society of Great Britain, 2000.

D'Agostino RB, Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med 1998;17:2265-81.

Deeks JJ, Smith LA, Bradley MD. Efficacy, tolerability, and upper gastrointestinal safety of celecoxib for treatment of osteoarthritis and rheumatoid arthritis: systematic review of randomised controlled trials. BMJ 2002;325:619-26.

Dorland's illustrated medical dictionary. 28th ed. Philadelphia: W.B. Saunders, 1994.

Fitzgerald GA, Patrono C. The coxibs, selective inhibitors of cyclooxygenase-2. N Engl J Med 2001;345:433-42.

Goodman and Gilman's The pharmacological basis of therapeutics. 9th ed. London: McGraw Hill, 1996.

Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 1999;29:722-9.

Greenland S, Morgenstern H. Confounding in health research. Ann Rev Pub Health 2001;22:189-212.

Hochberg MC, Altman RD, Brandt KD et al. Guidelines for the medical management of osteoarthritis. Part II. Osteoarthritis of the knee. Arthritis Rheum 1995;38:1541-6.

Jick H, Miettinen OS, Neff RK, Shapiro S, Heinonen OP, Slone D. Coffee and myocardial infarction. N Engl J Med 1973;289:63-7.

Joffe MM, Rosenbaum PR. Invited commentary: propensity scores. Am J Epidemiol 1999;150:327-33.

Johnston SC. Combining ecological and individual variables to reduce confounding by indication: case study--subarachnoid hemorrhage treatment. J Clin Epidemiol 2000;53:1236-41.

Katzung BG. Basic and clinical pharmacology. 6th ed. East Norwalk (CT): Appleton and Lange, 1995.

Kelley's Textbook of Rheumatology. 6th ed. Philadelphia: W.B. Saunders, 2001.

Kleinbaum DG, Kupper LL, Morgenstern H. Multivariate regression analysis. 3rd ed. 1998.

Kupper LL, Karon JM, Kleinbaum DG. Matching in epidemiologic studies: validity and efficiency considerations. Biometrics 1981;37:271-92.

Last JM. A dictionary of epidemiology. 3 ed. Toronto: Oxford University Press, 1995.

LeLorier, J. Personal communication. 2002.

Lipsky PE, Abramson SB, Crofford L. The classification of cyclooxygenase inhibitors. J Rheumatol 1998;25:2298-.

Margolis DJ, Kantor J, Santanna J, Strom BL, Berlin JA. Effectiveness of platelet releasate for the treatment of diabetic neuropathic foot ulcers. Diabetes Care 2001;24:483-8.

Miettinen OS. Matching and design efficiency in retrospective studies. Am J Epidemiol 1970;91:111-8.

Miettinen OS. Confounding and effect-modification. Am J Epidemiol 1974;100:350-3.

Miettinen OS. Stratification by a multivariate confounder score. Am J Epidemiol 1976;104:609-20.

Miettinen OS. The need for randomization in the study of intended effects. Stat Med 1983;2:267-71.

Miettinen OS. Etiologic research: needed revisions of concepts and principles. Scand J Work Environ Health 1999;25:484-90.

Miettinen OS. Etiologic research: needed revisions of concepts and principles. Scand J Work Environ Health 1999;25:484-90.

Miettinen OS, Cook EF. Confounding: essence and detection. Am J Epidemiol 1981;114:593-603.

Patrono C, Coller B, Dalen JE et al. Platelet-active drugs: The relationships among dose, effectiveness, and side effects. Chest 2001;119:39S-63S.

Perkins SM, Tu W, Underhill MG, Zhou XH, Murray MD. The use of propensity scores in pharmacoepidemiologic research. Pharmacopepidemiol Drug Safety 2000;9:93-101.

Petri H, Urquhart J. Channeling bias in the interpretation of drug effects. Stat Med 1991;10:577-81.

Piantadosi S, Byard DP, Green SB. The ecological fallacy. Am J Epidemiol 1988;127:893-904.

Rahme E, Joseph L, Kong XS, Watson DJ, LeLorier J. Gastrointestinal health care resource use and costs associated with nonsteroidal antiinflammatory drugs versus acetaminophen; retrospective cohort study of an elderly population. Arthritis Rheum 2000;43:917-24.

Rahme E, Marentette MA, Kong XS, LeLorier J. Use of NSAIDs, COX-2 inhibitors, and acetaminophen and associated coprescriptions of gastroprotective agents in an elderly population. Arthritis Rheum 2002;47:595-602.

Rahme E, Pettitt D, LeLorier J. Determinants and sequelae associated with utilization of acetaminophen versus traditional nonsteroidal antiinflammatory drugs in an elderly population. Arthritis Rheum 2002;46:3046-54.

Rahme E, Pilote L, LeLorier J. Association between naproxen use and protection against acute myocardial infarction. Arch Intern Med 2002;162:1111-5.

Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41-55.

Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. JASA 1984;79:516-24.

Rothman KJ, Greenland S. Modern epidemiology. 2 ed. Philadelphia, PA: Lippincott-Raven, 1998.

Rubin DB. Estimating causal effects from large data sets using propensity scores. Ann Intern Med 1997;127:757-63.

Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. Am J Epidemiol 1999;149:981-3.

Schlesselmann JJ. Case-Control Studies. Oxford: Oxford University Press, 1982.

Shapiro S. Confounding by indication? Epidemiology 1997;8:110-1.

Silverstein FE, Faich G, Goldstein JL. Gastrointestinal toxicity with celecoxib vs nonsteroidal anti-inflammatory drugs for osteoarthritis and rheumatoid arthritis: the CLASS stuy: a randomized controlled trial. JAMA 2000;248:1247-55.

Skegg DC. Evaluating the safety of medicines, with particular reference to contraception. Stat Med 2001;20:3557-69.

Slone D, Shapiro S, Miettinen OS, Finkle WD, Stolley PD. Drug evaluation after marketing. Ann Intern Med 1979;90:257-61.

Strom BL, Melmon KL, Miettinen OS. Post-marketing studies of drug efficacy: why? Am J Med 1985;78:475-80.

Strom BL, Miettinen OS, Melmon KL. Post-marketing studies of drug efficacy: how? Am J Med 1984;77:703-8.

Strom BL (ed.). Pharmacoepidemiology. 3 ed. Chichester, UK: John Wiley and Sons, 2000.

Tamblyn R, Abrahamowicz M. Drug utilization patterns. In: Armitage P, Colton T, eds. Encyclopedia of biostatistics. New York: John Wiley & Sons, 1998:1235-47.

Tamblyn R, Lavoie G, Petrella L, Monette J. The use of prescription claims databases in pharmacoepidemiological research; the accuracy and comprehensiveness of the prescription claims database in Quebec. J Clin Epidemiol 1995;48:999-1009.

Veterans Administration Cooperative Study Group on Antihypertensive Agents. Effects of treatment on morbidity in hypertension: results in patients with diastolic blood pressure averaging 90 to 114 mm Hg. JAMA 1970;213:1143-52.

Von Korff M, Wagner EH, Saunders K. A chronic disease score from automated pharmacy data. J Clin Epidemiol 1992;45:197-203.

Walker AM. Confounding by indication. Epidemiology 1996;7:335-6.

Wen SW, Kramer MS. Uses of ecologic studies in the assessment of intended treatment effects. J Clin Epidemiol 1999;52:7-12.

Wolfe F, Flowers N, Burke TA, Arguelles LM, Pettitt D. Increase in lifetime adverse drug reactions, service utilization, and disease severity among patients who will start COX-2 specific inhibitors: quantitative assessment of channeling bias and confounding by indication in 6689 patients with rheumatoid arthritis and osteoarthritis. J Rheumatol 2002;29:1015-22.

Zacharias FJ, Cowen KJ, Presst J, Vickers J, Wall BG. Propranolol in hypertension: a study of long-term therapy, 1964-1970. Am Heart J 1972;83:755-61.