2014

Application of Marginal Structural Models in Pharmacoepidemiologic Studies

Shibing Yang
Virginia Commonwealth University

Application of Marginal Structural Models in Pharmacoepidemiologic Studies

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

Shibing Yang
Bachelor of Medicine in Preventive Medicine, Capital Medical University, 2007
Master of Science in Social Medicine and Healthcare Management, Capital Medical University, 2010

Directors:

Juan Lu, Ph.D.
Assistant Professor
Division of Epidemiology, Department of Family Medicine and Population Health, Virginia Commonwealth University

Kate L. Lapane, Ph.D.
Professor
Department of Quantitative Health Sciences, University of Massachusetts Medical School

Virginia Commonwealth University
Richmond, Virginia
June, 2014

Acknowledgements

The author wishes to thank several people. I would like to thank Dr. Juan Lu and Dr. Kate Lapane for their invaluable guidance, wisdom, and moral support over the years as my advisors and in the completion of my dissertation. I would also like to sincerely thank the members of my committee, Dr. Charles Eaton and Dr. Spencer Harpe for their sage advice and time commitments toward this project. I would like to thank Dr. Victor Heh for providing advice for the simulation study and Dr. Rachel Jawahar for her help with the systematic review. Finally, I would like to thank my wife (Qin Shen) and parents for their unconditional love, support, and understanding.

Table of Contents

List of Tables and Figures

Abstract

APPLICATION OF MARGINAL STRUCTURAL MODELS IN PHARMACOEPIDEMIOLOGIC STUDIES

By Shibing Yang

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2014

Directors:
Juan Lu, Ph.D., Assistant Professor
Division of Epidemiology, Department of Family Medicine and Population Health, Virginia Commonwealth University

Kate L. Lapane, Ph.D., Professor
Department of Quantitative Health Sciences, University of Massachusetts Medical School

Background: Inverse-probability-of-treatment-weighted estimation (IPTW) of marginal structural models was proposed to adjust for time-varying confounders that are influenced by prior treatment use. It is unknown whether pharmacoepidemiologic studies that applied IPTW conformed to the recommendations proposed by methodological studies. In addition, no previous study has compared the performance of different analytic strategies adopted in IPTW analyses.

Objectives: This project aims 1) to review the reporting practice of pharmacoepidemiologic studies that applied IPTW, 2) to compare the validity and precision of several approaches to

constructing weight, 3) to use IPTW to estimate the effectiveness of glucosamine and chondroitin in treating osteoarthritis.

Methods: We systematically retrieved pharmacoepidemiologic studies that were published in 2012 and applied IPTW to estimate the effect of a time-varying treatment. Under a variety of simulated scenarios, we assessed the performance of four analytic approaches what were commonly used in studies conducting IPTW analyses. Finally, using data from Osteoarthritis Initiative, we applied IPTW to estimate the long-term effectiveness of glucosamine and chondroitin on treating knee osteoarthritis.

Results: The practice of reporting use of IPTW in pharmacoepidemiologic studies was suboptimal. The majority of reviewed studies did not report that the positivity assumption was assessed, and several studies used unstablized weights or did not report that the stabilized weights were used. With data simulation, we found that intention-to-treat analyses underestimated the actual treatment effect when there was non-null treatment effect and treatment non-adherence. This underestimation was linearly correlated with adherence levels. As-treated analyses that took into account the complex mechanism of treatment use generated approximately unbiased estimates without sacrificing the estimate precision when the treatment effect was non-null. Finally, after adjustment for potential confounders with marginal structural models, we found no clinically meaningful benefits of glucosamine/chondroitin in relieving knee pain, stiffness and physical function or slowing joint space narrowing.

Conclusions: It may be prudent to develop best practices of reporting the use of IPTW. Studies performing intention-to-treat analyses should report the levels of adherence after treatment initiation, and studies performing as-treated analyses should take into the complex mechanism of treatment use in weight construction.

Chapter 1: Background

This dissertation was motivated by the desire to quantify the effect of glucosamine and chondroitin (Glu/Chon) on relieving knee symptoms and slowing joint structural progression among patients knee osteoarthritis (OA). When analyzing the relation between Glu/Chon and knee OA using data from Osteoarthritis Initiative, we were concerned that the data structure might involve time-varying confounders that were affected by previous use of Glu/Chon. To properly control for the confounding bias, we used inverse-probability-of-treatment-weighted (IPTW) estimation of marginal structural models (MSM) in the analysis stage. In this introduction chapter, we briefly discussed 1) the disease burden of OA and existing evidence regarding the efficacy of glucosamine and chondroitin in treating OA; 2) the causal diagram describing the relations between Glu/Chon use, study outcomes and potential confounders; and 3) application of IPTW to control for confounding and assumptions underlying IPTW estimation. At the end of this chapter, the three specific aims of this dissertation were provided.

Glucosamine/chondroitin and knee osteoarthritis

OA is the most common form of arthritis and nearly 27 million American adults have physician-diagnosed OA.[1] OA typically affects weight-bearing joints such as hips, knees and spine, but can also occur in non-weight-bearing joints.[2] The most common OA symptoms include joint pain and stiffness and reduced range of joint movement.[2] Radiographic evidence of

OA includes progressive narrowing of joint space, formation of subchondral sclerosis and cysts and development of osteophytes.[3] OA has detrimental effects on individuals' physical function and quality of life.[4,5] Because of its high prevalence and the frequent disability that accompanies disease in major joints such as the knee and hip, OA accounts for more difficulty with climbing stairs and walking than any other disease.[6]

Currently, there are no curative remedies for OA and clinical guidelines recommend both pharmacological and non-pharmacological therapies to relieve symptoms.[7] Glucosamine and chondroitin are two dietary supplements commonly used among OA patients in the United Sates. Lapane et al. found that among patients with radiographic knee OA, 31% and 28% reported frequent use of glucosamine and chondroitin, respectively.[8] These supplements hold promise for treating OA because both are essential components of the proteoglycan in normal cartilage and thus may provide substrate or building blocks for the biosynthesis of proteoglycan.[9, 10]

Despite the biologic plausibility, evidence regarding the efficacy of glucosamine and chondroitin in relieving OA symptoms and modifying structural progression is not established. A recently updated Cochrane review reported a moderate clinical treatment benefit for pain reduction in favor of glucosamine over placebo.[11] However, this superiority of glucosamine was not consistently reported by all studies included in the review. When the analysis was restricted to studies with adequate allocation concealment or studies without connection to private industry, no superiority of glucosamine was found.[11] Regarding the efficacy in slowing joint structural progression, two 3-year clinical trials, which were funded by one pharmaceutical company, reported significant beneficial effect from glucosamine,[12,13] whereas another two 2-year publicly-funded trials found that there were no substantial benefits from glucosamine (with or without chondroitin) in retarding joint space narrowing.[14,15]

Causal diagram and time-varying confounders

We analyzed the effectiveness of Glu/Chon on treating OA using data from Osteoarthritis Initiative (OAI), which is a multi-center observational study aimed to identify risk factors for incidence and progression of knee OA.[16] In our study, we included OAI participants with radiographic knee OA at baseline. Annual follow-up surveys and examinations were conducted to collect information on treatment use and changes in knee symptoms and joint structure. Our analyses used information for the first four years.

Prior to modeling the effect of Glu/Chon on knee OA, we drew a causal diagram to help identify potential confounders as well as methods to control for confounding. Figure 1.1 depicts the hypothesized relationships between Glu/chon, study outcomes (including knee symptoms and structural progression), and potential time-varying confounders. Previously measured study outcomes and time-varying confounders may be simultaneously confounders and intermediate variables. For instance, when studying knee pain as the outcome (i.e., $Outcome_t$ in Figure 1.1), the pain severity measured at the previous visit (i.e., $Outcome_{t-1}$) can be a potential confounder because 1) it correlates with pain score measured at current visit (i.e., $Outcome_t$), and 2) patients with more severe pain are more likely to use Glu/Chon (i.e., $Glu/Chon_{t-1}$).[8] Furthermore, if Glu/Chon is effective in relieving pain (which is the hypothesis tested in our study), the previously measured pain score (i.e., $Outcome_{t-1}$) lies on the causal path from prior treatment use (i.e., $Glu/Chon_{t-2}$) and currently measured pain (i.e., $Outcome_t$).

If the causal structure in Figure 1.1 is true, standard regression models adjusting for previous pain severity will produce a biased estimate of the overall treatment effect.[17] Standard regression models adjust for confounding through conditioning analyses on the potential confounders.[17] Under the causal structure in Figure 1.1, conditioning analyses on $Outcome_{t-1}$ can

eliminate its confounding bias to the relation between $Glu/Chon_{t-1}$ and $Outcome_t$. However, conditioning analyses on $Outcome_{t-1}$ also eliminates the indirect effect of $Glu/Chon_{t-2}$ on $Outcome_t$ that is mediated by $Outcome_{t-1}$, and thus generates a biased estimate of the overall treatment effect of Glu/Chon on $Outcome_t$.

Marginal structural models

To properly control for the bias by time-varying confounders that are affected by previous treatment, Robins et al. proposed the IPTW estimation of MSM.[18,19] As the name indicates, IPTW reduces confounding through assigning a weight to each participant, which is proportional to the inverse of conditional probability of receiving his/her observed treatment given those time-varying confounders.[19] In the resulting weighted pseudo-population, treated participants and untreated participants are balanced over those time-varying confounders.[19] Since the analysis is not conditioned on the confounders, IPTW can properly estimate overall treatment. In this section, we illustrated how IPTW can adjust for confounding with a simplified example and briefly discussed the assumptions underlying this method.

A simplified example with one-time-point treatment

For illustrative purposes, we focused only on the relation between Glu/Chon and knee pain at one time point (shown in Figure 1.2). For simplicity, we assumed that there was only one confounder, i.e., baseline pain severity. It's likely that patients with more severe pain (i.e., $Pain_0$ in Figure 1.2) were more likely to use Glu/Chon ($Glu/Chon_0$) and also more likely to report severe pain one year later ($Pain_1$). For simplicity, we assumed that $Pain_0$, $Glu/Chon_0$ and $Pain_1$ were all binary variables, with value 1 indicating severe baseline pain, using Glu/Chon at baseline and reporting severe pain at Year 1, respectively.

We generated a hypothetical dataset including 100 OA participants. Distributions of the three variables are listed in Table 1.1. The first column describes the strata formed by levels of $Pain_0$ and $Glu/Chon_0$. The second column shows number of patients in each stratum. $Pr(Pain_1=1)$ is the probability of reporting severe pain at Year 1 and $Pr(Glu/Chon_0|Pain_0)$ represents the conditional probability of receiving observed treatment (either 0 or 1) given $Pain_0$.

In this hypothetical sample, 60% of the patients had severe pain at baseline. 66.7% and 50% took Glu/Chon among patients with severe and mild baseline pain, respectively. The probability of having severe pain at Year 1 was 40% among patients with severe baseline pain, and 20% among patients with mild baseline pain. Since there was no difference in the probability of having severe pain at Year 1 between those treated and untreated with Glu/Chon within the stratum of baseline pain severity, there was actually no treatment effect in this hypothetical example. However, if we calculated the crude association between Glu/Chon and $Pain_1$, crude relative risk (RR) $=\frac{(40\times0.4+20\times0.2)/60}{(20\times0.4+20\times0.2)/40}$=1.11, which was apparently biased.

IPTW estimation can control for the confounding by $Pain_0$ through assigning weights to patients. The first step of IPTW is to construct weight. For the hypothetical sample, weight was calculated as the inverse of conditional probability of receiving observed treatment given $Pain_0$. For instance, among patients with $Pain_0$=1, the possibility of receiving treatment is 66.7%, and possibility of not receiving treatment is 1-66.7%=33.3%. Thus the weight assigned to patients who had severe baseline pain and received treatment was 1/0.667=1.5, and weight for those not receiving treatment was 1/0.333=3. Accordingly, we calculated weights for patients with mild pain at baseline. We listed the weights for each stratum of patients in the fifth column in Table 1.1.

The relative risk in the weighted population was subsequently calculated as follows:

Weighted RR=$\frac{(60\times0.4+40\times0.2)/100}{(60\times0.4+40\times0.2)/100}$=1.0, which was unbiased.

Time-varying treatment and confounders

IPTW estimation for the simplified situation depicted in Figure 1.2 can be generalized to situations involving time-varying treatment and confounders, for example, the causal structure in Figure 1.1. For studies with a time-varying treatment and confounders, weights are first constructed at each assessment (time) point, which can be calculated as the unconditional probability of receiving observed treatment divided by conditional probability of receiving observed treatment given potential confounders.[19] The final weight for each participant is the product of his/her weights constructed at all available time points. The inclusion of the numerator, i.e., unconditional probability of receiving observed treatment, is to improve the precision of the final estimate.[19]

Identifiability assumptions

There are three conditions or assumptions, under which consistent causal effects can be identified from non-experimental data: exchangeability, positivity and consistency.[20, 21]

Exchangeability assumption is also known as the assumption of no unmeasured confounders.[20,22] For longitudinal studies with time-varying exposures, exchangeability assumption holds when there are no unmeasured confounders for treatment use at each follow-up assessment, given the history of measured confounders and previous treatment. Exchangeability assumption is untestable with observed data[20] and is violated to some degree in almost all epidemiologic studies. Notwithstanding, if the investigators have adequate substantive knowledge with respect to the direction and magnitude of unmeasured confounding, sensitivity analysis can be conducted to test the robustness of IPTW estimates to unmeasured confounding.[23]

Positivity assumption states that each possible treatment level occurs with some positive probability at every level of observed confounders in the study population.[24,25] IPTW estimation is more sensitive to violations of positivity assumption than standard regression models.[20,25] If the conditional probability of receiving a certain level of treatment is zero, the weight is undefined. Furthermore, when positivity is nearly-violated, e.g., a very small proportion of the study sample within one or more covariate strata are treated, the weights for these few participants become very large. The disproportionate reliance of the effect estimate on the experience of a few unusual individuals can result in substantial bias.[25]

The consistency condition requires that a study unambiguously define treatment and that counterfactual outcome for each level of treatment be well-defined.[20,26,27] Many non-experimental studies violated this condition when they tried to estimate the effect of an ill-defined intervention by contrasting the outcome between two groups of participants who happen to differ with respect to some physiological measure (e.g., body mass index, low-density lipoprotein cholesterol).[26,28] Studies violating consistency assumption will have difficulty in achieving the condition of exchangeability and may be of little help to advise public health intervention.[26,28] Therefore, some investigators propose that studies should only estimate the causal effects of exposures that can be manipulated or be hypothetically assigned to a person.[26,27,29]

Specific aims of the dissertation

When applying IPTW to estimate the effect of Glu/Chon on treating OA, we fully realized the complexity of the process of performing IPTW analyses, as well as the importance of various assumptions underlying this method. Therefore, besides applying IPTW to address a real-world study question, we 1) systematically reviewed how recently published

pharmacoepidemiologic studies applied this method and reported findings, and 2) performed a simulation study to assess the impact of various assumptions made in weight construction on the validity and precision of IPTW estimates. We listed below the three specific aims of this dissertation.

Study 1. To systematically review pharmacoepidemiologic studies published in 2012 that used IPTW estimation of MSM to estimate the effect from a time-varying treatment

We extracted information about the type(s) of bias IPTW was used to address, how the identifiability assumptions were assessed, how the weights were constructed and outcome models specified, and whether substantially different results were derived from IPTW method and standard regression models.

Study 2. To explore the impact of various assumptions made during weight construction on the validity and precision of IPTW estimates

Using various simulated scenarios, we assessed the bias and precision of estimates derived from four approaches to constructing weights, including IPTW assuming intention-to-treat, IPTW assuming complex mechanism of treatment assignment, IPTW assuming simple mechanism of treatment assignment, and IPTW assuming invariant confounders.

Study 3. To quantify the extent to which glucosamine and chondroitin relieves symptoms and slows structural progression among persons with radiographic knee osteoarthritis

Using IPTW, we examined the effectiveness of glucosamine and chondroitin in relieving knee symptoms and slowing structural progression among persons with radiographic knee osteoarthritis

Table 1.1. Hypothetical data corresponding to the causal diagram in Figure 1.2

Stratum	No. of patients	$Pr(Pain_1=1)$	$Pr(Glu/Chon_0\|Pain_0)^*$	Weight†	Weighted No. ‡	No. of events in weighted sample
$Pain_0=1$, $Glu/chon_0=1$	40	40%	40/(40+20)=0.67	1/0.67	60	60×40%
$Pain_0=1$, $Glu/chon_0=0$	20	40%	20/(40+20)=0.33	1/0.33	60	60×40%
$Pain_0=0$, $Glu/chon_0=1$	20	20%	20/(20+20)=0.5	1/0.5	40	40×20%
$Pain_0=0$, $Glu/chon_0=0$	20	20%	20/(20+20)=0.5	1/0.5	40	40×20%

* $Pr(Glu/Chon_0|Pain_0)$ is the probability of receiving observed treatment conditional on baseline severity level

† $Weight=1/Pr(Glu/Chon_0|Pain_0)$

‡ Weighted No. = Weight × No. of patients

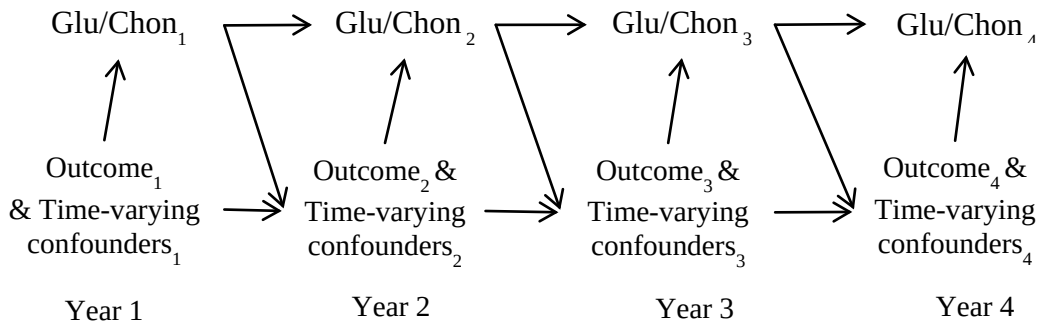

Figure 1.1. Hypothesized causal relationships between glucosamine/chondroitin treatment, study outcomes and potential time-varying confounders

Glu/Chon denotes treatment with glucosamine/chondroitin and the subscript number denotes the follow-up time (year) when the information was measured.

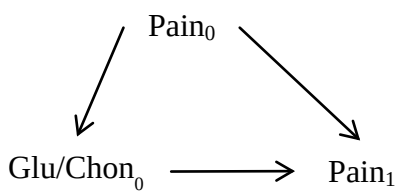

Figure 1.2. A causal diagram with one-time-point treatment

Chapter 2: Application of marginal structural models in pharmacoepidemiologic studies: A systematic review

Abstract

Objectives We systematically reviewed pharmacoepidemiologic studies published in 2012 that used inverse-probability-of-treatment-weighted (IPTW) estimation of marginal structural models (MSM) to estimate the effect from a time-varying treatment.

Methods Potential studies were retrieved through a citation search within Web of Science and a keyword search within PubMed. Eligibility of retrieved studies was independently assessed by at least two reviewers. One reviewer performed data extraction and a senior epidemiologist confirmed the extracted information for all eligible studies.

Results Twenty pharmacoepidemiologic studies were eligible for data extraction. The majority of reviewed studies did not report whether the positivity assumption was checked. Six studies performed intention-to-treat analyses, but none of them reported adherence levels after treatment initiation. Eight studies chose an as-treated analytic strategy, but only one of them reported modeling the multiphase of treatment use. Almost all studies performing as-treated analyses chose the most recent treatment status as the functional form of exposure in the outcome model. Nearly half of the studies reported that the IPTW estimate was substantially different from the estimate derived from a standard regression model.

Conclusions The use of IPTW method to control for time-varying confounding is increasing in medical literature. However, reporting of the application of the technique is variable and suboptimal. It may be prudent to develop best practices in reporting complex methods in epidemiologic research.

Introduction

A time-varying confounder is a time-varying risk factor for the study outcome which brings about changes in the treatment use under study.[30] In the presence of time-varying confounders that are influenced by previous treatment, standard regression models may produce biased estimate of the total treatment effect.[19,31] To obtain unbiased estimate in this situation, Robins et al. proposed the inverse probability weighted (IPTW) estimation of marginal structural models (MSM).[19,31] As the name indicates, IPTW estimation attempts to control for confounding through assigning each participant a weight. The weight is proportional to the inverse probability of receiving observed treatment given the time-varying confounders and previous treatment history. The weights are then used to create a pseudo-population, in which participants receiving treatment and those not receiving treatment are balanced over the time-varying confounders but the relationship between treatment and outcome is not changed.[19]

After publication of the seminal papers on MSM, methodological studies have provided detailed insights regarding the types of bias this method handles well,[17,32] the assumptions under which consistent causal effects can be identified,[23, 25, 26] and the appropriate ways of constructing weights and building outcome models.[20, 33-35] IPTW estimation has been increasingly used in medical research, possibly due to the straightforward interpretation of the parameters derived from MSM.[35] Indeed, from 2000 to October 2009 Suarez et al. noted a 15-fold increase in the number of studies using this approach.[36]

Despite the increase in studies using IPTW, the extent to which these studies conform to the recommendations proposed by methodological studies remains unknown. The purpose of this study was to systematically review pharmacoepidemiologic studies in which IPTW was used to estimate the effect from a time-varying treatment. Based on information abstracted from these

studies, we hope to provide a broader context for scientists considering using this approach through discussing the scenarios under which IPTW method is preferred, appropriate procedures of conducting IPTW analyses and contents which are critical to report when using IPTW in medical literature.

Methods

This study did not require ethics approval as no human subjects were involved.

Selection of articles

Our goal was to retrieve all pharmacoepidemiologic studies published in 2012 that used IPTW to estimate effect from a time-varying treatment. To achieve this, we used two search strategies. First, using the Web of Science database, we retrieved all published studies citing any one of the seminal papers on MSM.[19,20,31,37] Second, in case we missed any relevant studies which did not cite these seminal papers, we also conducted a keyword search within PubMed. To improve the methodological rigor of our search strategy, we worked with a research librarian and developed the following keyword search algorithm: (marginal structural model*) OR (“marginal structural Cox model”) OR (“inverse probability” AND ("weight" OR "weighted" OR "weights" OR "weighting")) OR (inverse weight*). The following types of studies or publications were excluded from the review: (1) methodological or simulation studies, (2) studies assessing effect from a point-treatment, i.e., a treatment that was assumed invariant in the study period; (3) non-pharmacoepidemiologic studies, i.e., studies not focusing on pharmaceuticals, biologics, or medical devices as primary exposure; (4) letters, meeting abstracts, review articles, and editorials.

We followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines for this review.[38] After excluding duplicate records, titles and abstracts of the remaining articles were assigned to two independent reviewers. Studies with titles and abstracts judged relevant by at least one reviewer underwent full-text review. Any discrepancy in eligibility judgment was resolved through discussion between the reviewers. One reviewer (SY) performed data extraction and a senior epidemiologist (KLL) confirmed the extracted information for all eligible studies.

Information abstraction

The following sections provide a brief description and rationale of each element of IPTW method we chose to include in our data collection process. In particular, we extracted information about the type(s) of bias IPTW was used to address, how the identifiability assumptions were assessed, how the weights were constructed and outcome models specified, and whether substantially different results were derived from IPTW method and standard regression models.

Type of bias

As illustrated by Hernán et al.,[17] compared to standard regression models, MSM has the advantages of eliminating bias from two sources when estimating the effect from a time-varying treatment. First, through applying inverse-probability-of-treatment weighting, IPTW can control for the time-varying confounding while avoiding two types of bias that may arise in analyses with standard regression models.[17] The first type of bias occurs when the time-varying confounder is simultaneously a confounder and intermediate variable. Conditioning analysis on

such a variable (as performed in standard regression models) will block the indirect effect from previous treatment on study outcome that is mediated by this variable.[19] Another type of bias (called collider-stratification bias[39] or selection bias[17]) occurs in standard regression models when the time-varying confounder is a common effect (i.e., a collider) of previous treatment and an unmeasured risk factor for the study outcome. Conditioning analysis on this time-varying confounder induces a non-causal relationship between previous treatment and the unmeasured risk factor, which introduces bias in the effect estimate of previous treatment use.[17]

Second, through applying inverse-probability-of-censoring weighting, MSM can control for selection bias from informative censoring.[17,32] Our review focused on the use of inverse-probability-weighting for handling selection bias from artificially censoring participants with treatment noncompliance, e.g., discontinuing the treatment under study or switching to an ineligible treatment.[32] Bias may be introduced when this artificial censoring depends on treatment history and also risk factors for the study outcome.[40] Under certain conditions (discussed below), IPTW can eliminate this bias by simulating a pseudo-population, in which all participants complete the follow-up but the effect of treatment on study outcome is the same as in the unweighted study population.[41]

Identifiability assumptions

There are three conditions or assumptions, under which consistent causal effects can be identified from non-experimental data: no uncontrolled confounding, consistency and positivity.[20, 21] Consistency is the assumption that an individual's potential (or counterfactual) outcome under the observed treatment is precisely the observed outcome.[27] Because consistency

is often considered a reasonable assumption when estimating effects from medical treatments[26], we did not extract information on this assumption.

When there are confounders (time-invariant or time-varying) that are not measured or measured with error, the IPTW estimates will be biased by uncontrolled confounding. We looked for information about whether studies qualitatively discussed the susceptibility of their findings to uncontrolled confounding and whether they performed sensitivity analyses to test the robustness of their results when substantial uncontrolled confounding was suspected.

The positivity assumption states that each treatment level occurs with some positive probability at every level of observed confounders in the study population.[24,25] For example, this assumption is violated when all (or almost all) patients with a specific contraindication (which is also a risk factor for the study outcome) are untreated with the medication under study. Among patients with the contraindication, the probability of receiving treatment will be zero (or close to zero), and the inverse probability will be inestimable (or a very large number). The disproportionate reliance on the experience of a few unusual individuals (i.e., treated patients with the contraindication) in the weighted population can result in imprecise and biased effect estimate.[25] Thus, we extracted information about (1) whether studies reported that the positivity assumption was checked, (2) how the positivity assumption was evaluated, and (3) how violations of the assumption were handled (if detected).

Constructing weights

The validity of IPTW estimates depends on correct construction of weights.[19, 21] There are two types of weights--unstabilized and stabilized. The unstablized weight is calculated as the inverse of conditional probability of receiving observed treatment given the history of time-

varying confounders and previous treatment history (called weight denominator).[19] The stabilized weight can be calculated as the product of the conditional probability of receiving observed treatment given baseline confounders and previous treatment history (called weight numerator) and the unstabilized weight. The stabilized weight is generally recommended because it can yield estimates with greater precision compared to the unstabilized weight.[19] The conditional probability of receiving observed treatment (for weight numerator and denominator) is often estimated with a regression model (i.e., treatment model).

When non-compliance after treatment initiation is low, an observational intention-to-treat (ITT) analysis with IPTW has been recommended.[19, 20, 37] Specifically, this strategy assumes that once a participant initiates treatment, the participant will remain on treatment for the remainder of the study period. This assumption simplifies the process of estimating the probability of receiving observed treatment history, because only one model is needed to estimate the probability of treatment onset.[42] In addition to ITT analyses, analogous to data analysis of a clinical trial, a non-experimental study can perform per-protocol and as-treated analyses using IPTW.[40] In a per-protocol analysis, a comparison is made only among those who adhere to the treatment under study and patients are censored when they deviate from the initial treatment. In an as-treated analysis, individuals are classified according to the treatment they receive during the follow-up rather than the treatment they initiate, and patients who stop or switch the treatment are also included in the analysis.

We extracted information about the analytic strategy each study adopted, and how they specified the treatment models for the weight numerator and denominator. For studies not assuming ITT, we assessed whether or not the authors modeled the multiphase of treatment use (e.g., treatment initiation, continuation, etc.) and how this was done.

Outcome model building

After weights are constructed, a weighted regression model (i.e., outcome model) is typically fit to estimate the effect of treatment on the outcome.[20] All the variables included in the treatment model for the weight numerator should also be included in the outcome model, because they are not balanced between treated and untreated participants in the weighted population and thus can still bias the estimate.[20] Substantive expertise should drive the selection of the functional form of exposure in the outcome model.[42,43] For instance, under the assumption of a linear relationship between treatment duration and study outcome, studies can specify exposure as the total duration of previous treatment use, and the estimate then quantifies the effect from each additional time unit (e.g., one month) of treatment;[19,40] studies performing ITT analyses can also specify exposure with an indicator for treatment initiation (yes or no) to estimate the average effect of initiating treatment in the follow-up period.[40] In this review, we assessed what covariates were included in the outcome model and how they specified the functional form of exposure.

Discrepancy between IPTW estimates and standard regression estimates

The review by Suarez et al. reported that more than half of the studies using IPTW method yielded an estimate substantially different from that produced by standard regression models.[36] However, the review did not provide information about how studies discussed reasons for such discrepancy. In this review, we assessed whether studies found a substantial difference in estimates between the two methods and further extracted information about how studies

explained the discrepancy when it was noted. We considered a difference "substantial" if the difference was more than 20% of the IPTW estimate.[36]

Results

Figure 2.1 depicts the process of identifying studies eligible for the review. We retrieved 164 and 137 studies from citation search in Web of Science and keyword search in PubMed, respectively. After excluding duplicate studies (n=66), methodological or simulation studies (n=92), review studies (n=9), studies not focusing on a health-related outcome (n=12) or not using IPTW (n=7), studies assessing effect from a point-treatment (n=66), and non-pharmacoepidemiologic studies (n=26), we had 23 pharmacoepidemiologic studies which applied IPTW to estimate effect from a time-varying treatment. Among these 23 studies, three used IPTW to evaluate effects from dynamic treatment regimens.[44-46] Considering that weight construction for estimating effects from dynamic regimens is different from that for static regimens,[47] we excluded these studies from the review. Data extraction was performed on the remaining 20 studies.[48-67]

Table 2.1 shows a brief description of the study design, primary exposure and outcome and potential time-varying confounders. Three studies compared treatments that were randomized to participants.[48,58,64] However, they performed analyses as if data were collected from a non-experimental design, so we included them in the review. Half of the 20 studies assessed benefits or risks from antiretroviral therapy among HIV-infected patients[51,53,56,57,59,63,67] or risk of HIV transmission from contraceptive use;[54,60,62] five studies focused on treatment or prevention of cardiovascular diseases;[48-50,52,64] two studies assessed treatments for chronic kidney disease;[55,61] and there was one study assessing the effect of treatment for a protein metabolism

disorder,[65] schizophrenia,[66] and breast cancer,[58] respectively. The primary outcome of most studies was mortality (n=7) or first occurrence of a pre-specified event (n=12), and one study considered a repeated-measure outcome.[67] With the exception of two studies,[52,58] all reviewed studies provided information on the time-varying confounders.

In Table 2.2, the type of bias IPTW addressed and details regarding the assumptions of positivity and no uncontrolled confounding are described for each study. Eleven studies used IPTW owing to concerns that standard regression models might eliminate indirect effects mediated by time-varying confounders, five studies used IPTW to deal with bias from the artificial censoring of noncompliance, and five studies did not provide further details other than stating that IPTW was used because of "concerns of time-varying confounding". The majority of studies did not report whether the positivity assumption was checked. Four studies truncated weights and one study trimmed weights to alleviate the impact of potential positivity violation. Most studies discussed qualitatively the susceptibility of their findings to uncontrolled confounding, but none reported performing formal sensitivity analyses to assess robustness of the results to uncontrolled confounding.

Table 2.3 includes information on the construction of weights and specification of outcome models. Six studies performed ITT analyses, three performed per-protocol analyses and eight performed as-treated analyses. None of the studies assuming ITT reported adherence levels after treatment initiation. The three "per-protocol" studies censored patients when they discontinued the treatment under study, and estimated the probability of treatment continuation (i.e., being uncensored) separately from treatment initiation. One of the eight "as-treated" studies modeled current treatment use stratified by previous treatment status.

One study did not use stabilized weights, four did not report whether stabilized weights were used, and eight reported using stabilized weights but did not describe how it was done. The remaining studies reported stabilizing weights with unconditional probability of receiving observed treatment, or conditional probability given baseline covariates and previous treatment or given baseline covariates only. For the weight denominator, twelve studies estimated the conditional probability given baseline covariates, time-varying confounders and previous treatment, three did this given baseline covariates and time-varying confounders and two adjusted for baseline covariates plus "follow-up period" or baseline covariates only. Four studies selected variables in the treatment model for weight denominator based on a statistical criterion. Two studies included covariates with statistically significant associations with the study outcome and subsequent treatment use. One study included factors significantly associated with the study outcome only. One study used a stepwise procedure to select the treatment model which maximized Akaike information criterion.

Regarding the functional form of exposure in the outcome model, studies performing ITT and per-protocol analyses included an indicator of treatment initiation and the initial treatment status, respectively. Almost all studies performing as-treated analyses included only the most recent treatment status in the outcome model.

Table 2.4 shows crude estimates, and estimates from IPTW and standard regression models for the associations between primary study exposure and outcome listed in Table 2.1. The last column contains information about whether the IPTW estimate was substantially different from the standard regression estimate for any association assessed in the study, as well as how the study explained any noted discrepancies. Fourteen studies reported results from both methods and a substantial difference was found in six studies. Among studies reporting a

substantial difference, three did not discuss reasons for the discrepancy, two considered IPTW method correctly estimated the indirect effects from previous treatments, and one considered IPTW method controlled for "confounding by indication".

We summarized the review results of the 20 studies in Table 2.5.

Discussion

Our review supports the notion that studies using IPTW to deal with time-varying confounding continue to diffuse in the medical literature. In 2012, 49 studies used IPTW to estimate the effect from a time-varying exposure on a health-related outcome. After reviewing 20 pharmacoepidemiologic studies, we found that the majority lacked sufficient details to evaluate the appropriateness of the application of the method. Most studies did not report that the positivity assumption was checked, and more than half did not report the type of weights (stabilized or unstabilized) applied or how the weights were stabilized. Furthermore, we found that more studies performed as-treated analyses than ITT analyses, but few of these studies considered the multiphase of treatment use in the process of weight construction and almost all chose the most recent treatment status as the functional form of exposure in the outcome model.

Assessment of positivity assumption. Surprisingly, the majority of reviewed studies did not report whether they checked the positivity assumption. The IPTW method is more sensitive to positivity violations than standard regression models.[20,25] Studies using simulated[68] and empirical[69] data have demonstrated that positivity violations could result in substantial bias and imprecision in IPTW estimates. Estimated stabilized weights with the mean far from one or with very extreme values can be indicative of non-positivity.[20] Thus, a thorough examination of the weight distribution is essential for checking the positivity assumption.[20,36] However, a "well-

behaved" weight distribution (i.e., with mean close to one and moderate range) is not sufficient to ensure the absence of positivity violations.[25,70] Thus, Cole et al. recommended assessing the robustness of IPTW estimates with weights truncated at certain percentiles (e.g., 99th, 95th and 90th) as sensitivity analyses.[20]

Assessment of uncontrolled confounding. Although it was difficult to judge the adequacy of control for confounding in the reviewed studies without knowledge in the specific datasets and subject areas, we did find that some studies reported adjusting for "follow-up period" as the only time-varying confounder or adjusting for only baseline covariates. If time-varying disease risk factors that cause changes in treatment use are not correctly measured and appropriately adjusted for, the IPTW estimates will be biased. When substantial uncontrolled confounding is suspected, sensitivity analyses have been recommended to assess the robustness of the IPTW estimates.[23,71] To perform such sensitivity analyses, investigators need to specify a plausible function form which quantifies the direction and magnitude of uncontrolled confounding.[23,71]

ITT analyses. When non-adherence after treatment initiation is minimal, an ITT analysis may be preferred to as-treated analysis in terms of simplifying the weight construction and controlling for confounding.[42,44,47] The ITT assumption simplifies the process of constructing weights, in that the treatment models only need to estimate the probability of treatment initiation. More importantly, for studies performing ITT analyses, the assumption of no uncontrolled confounding is satisfied as long as confounders for treatment onset are correctly measured and specified in the treatment model for weight denominator. This assumption may be viable for many pharmacoepidemiologic studies using healthcare database, because "information used by physicians to make a decision to initiate treatment is often captured in the database".[44] However, the ITT estimate merely measures the effect of treatment initiation instead of effect from actual

treatment.[41] High levels of non-adherence after treatment initiation may drive the ITT estimate away from the true treatment effect.[42,72] For this reason, studies performing ITT analyses should report adherence measures for each treatment arm so that findings can be interpreted under appropriate consideration of the observed adherence patterns.[73]

As-treated analyses. Instead of estimating the effect of treatment initiation, we found that more studies performed as-treated analyses. Validity of "as-treated" estimates relies on the extent to which the study correctly models the relationships between confounders and the multiphase of treatment use.[42] Because it is very likely that the influence of time-varying confounders on initiating a treatment is different from their impact on continuing or resuming the treatment, separate models for different treatment regimens are often needed for adequate control for confounding. However, when information on time-varying confounders that predict treatment changes after initiation is not well-recorded in the data sources or when the number of participants following each specific regimen is small, a correct estimation of the multiphase of treatment use will be difficult, if not impossible.[42,44,47] In sum, when choosing between an ITT and an as-treated analytic strategy, investigators need to take into account adherence levels after treatment initiation and availability of information on the time-varying confounders that predict treatment changes during the study period.

Weight construction. Stabilized weights can generate estimates with greater precision than unstabilized weights and thus are recommended in data analyses.[37] However, we still found that four studies did not report whether they used stabilized weights and one study used unstablized weights. It's unknown to us why unstabilized weights were chosen. Regarding variable selection for treatment model in the weight denominator, we found that most studies chose covariates based on substantive knowledge, while four studies used some statistical

criterion to select covariates significantly associated with treatment use and/or study outcome. A simulation study by Lefebvre et al. found that the performance of IPTW method could be improved when the confounders and risk factors of outcome were included in the treatment model, whereas including pure predictors of treatment use (i.e., not confounders) led to biased and highly variable estimates, particularly in the context of small samples.[33] These findings are consistent with the recommendation for variable selection for building propensity scores.[74] Therefore, an advisable strategy in building treatment model for weight denominator may be to include variables considered to be direct risk factors for the outcome.

Functional form of exposure in outcome models. Almost all studies performing as-treated analyses included only the most recent treatment status in the outcome model. Most of these studies chose IPTW method instead of standard regression models owing to the concerns that standard models would eliminate the indirect effect from previous treatments mediated by the time-varying confounders. This may imply that these studies were interested in estimating the effects from both recent and previous treatments. However, when treatment use is intermittent, including only the most recent exposure status in the outcome model will not correctly capture the effect from previous treatments. Furthermore, when the weights are stabilized with previous treatment history but only the indicator of most recent treatment use is included in the outcome model, the estimate may also be a biased one for the recent treatment effect, because the status of previous treatment is not balanced between recently treated and un-treated patients and thus may bias the estimate.[75] Finally, if only the most recent treatment effect is biologically plausible and is the focus of the study, standard regression models adjusting for time-varying confounders and previous treatment history can also produce unbiased estimate,[37,70]

even though there is disagreement regarding the difference in precision between estimates derived from IPTW and standard regression models.[76,77]

Discrepancy in estimates from IPTW and standard regression models. Similar to the previous review,[36] we found nearly half of the studies, which provided estimates from both methods, reported that IPTW estimates were substantially different from the standard regression estimates adjusted for time-varying confounders. Unfortunately, half of the studies reporting a substantial difference did not discuss reasons for the discrepancy. As mentioned in the section *Type of bias*, the discrepancy can be attributed to the correct estimation of total treatment effect or avoidance of collider-stratification bias by the IPTW method, especially if the direction of discrepancy is consistent with the hypothesized relationships between exposure, outcome and time-varying confounders. In addition, the difference can be due to control for selection bias from informative censoring if censoring weights are incorporated in IPTW analyses.

Non-uniform treatment effects. Several studies also noticed that substantial discrepancy in estimates could arise in the presence of covariates (or a summary of covariates like propensity score) which strongly predict treatment use and are also strong effect modifiers.[69,78,79] Compared to the standard regression models, the IPTW method gives much more weights to the covariate strata within which treatment status is almost completely determined by the covariates.[80,81] If the effect sizes in these strata differ dramatically from other strata, the IPTW estimates will be substantially different from the standard regression estimates.[81] The non-uniform treatment effects across the covariates (or the propensity score) can be due to violation of positivity,[69] unmeasured confounding[78] or true effect-measure modification. When unmeasured confounding or positivity violation is the cause of non-uniformity, the IPTW estimate will be biased and weight truncation or propensity score trimming should be applied to

ameliorate the impact.[69,78] In summary, when substantially different estimates are derived from IPTW and standard regression models, investigators should take into account these alternative explanations before being assured that IPTW method generates unbiased estimates.

Our review has some limitations. First, we included only pharmacoepidemiologic studies published in 2012. The findings may not be representative of all publications using IPTW to deal with time-varying confounding. Second, the reporting practices of published studies may be influenced by journals' requirements. Authors are reporting their findings given strict word limitations and as such may have limited space to provide details on these facets of the application of the method. Nevertheless, with complex methods such as IPTW, such reporting is necessary to evaluate the extent to which the method has been appropriately applied.

In summary, the use of IPTW estimation is increasing in the medical literature. Given the variable and suboptimal reporting of the application of the technique, it may be prudent to develop best practices in reporting complex methods in epidemiologic research and for journal editors to consider adopting such reporting guidelines.

Table 2.1. General description of pharmacoepidemiologic studies published in 2012 and eligible for the systematic review.

Reference	Study design	Exposure [&]	Outcome [&]	Time-varying confounders
Cook et al.[48]	Randomized controlled trial	Aspirin vs. no-treatment	CVD or CVD-related mortality	CVD risk factors, intermediate CVD events
Desai et al. [49]	Cohort	Candesartan vs. losartan	Mortality	Hospitalization
Gerhard et al. [50]	Cohort	Aggressive vs. conventional antihypertensive therapies	CVD or mortality	Blood pressure
Gsponer et al. [51]	Cohort	Switching to second-line ART vs. first-line ART	Mortality	CD4 cell count
Haukka et al. [52]	Cohort	Statins vs. no-treatment	Mortality	Not reported [†]
HCV working group of COHERE [53]	Cohort	Hepatitis C treatment vs. no-treatment	Mortality	CD4 cell count, HIV RNA level, platelet counts, alanine aminotransferase levels
Heffron et al. [54]	Cohort	Hormonal contraceptive vs. no-treatment	HIV infection	Pregnancy, unprotected sex
Hernández et al. [55]	Cohort	ACEI/ARB vs. no-treatment	Graft failure	Smoking, concurrent medication use
HIV-CAUSAL Collaboration [56]	Cohort	Nevirapine vs. efavirenz	Mortality	CD4 cell count, HIV RNA level, AIDS
HIV-CAUSAL Collaboration [57]	Cohort	ART vs. no-treatment	Tuberculosis	CD4 cell count, HIV RNA level, AIDS
Jin et al. [58]	Randomized controlled trial	Letrozole vs. no-treatment	Cancer recurrence	Not reported [*]
Kalayjian et al. [59]	Cohort	Tenofovir+ ritonavir-boosted protease inhibitor vs. efavirenz/nevaripine	Chronic kidney disease	CD4 cell count, viral load
McCoy et al. [60]	Cohort	Injectable hormonal contraceptive vs. no-treatment	HIV infection	Sexual behavioral risk, condom use, sexually transmitted infections

Miller et al. [61]	Cohort	Low dose vs. high dose paricalcital	Mortality	Parathyroid hormone, phosphorus, calcium
Morrison et al. [62]	Cohort	Oral contraceptive vs. non-hormonal use	HIV infection	Sexual behavioral risk, condom use, genital symptoms
Scherzer et al. [63]	Cohort	Tenofovir vs. no-treatment	Proteinuria	CD4 cell count, viral load, lipids, diabetes, hypertension
Shinozaki et al. [64]	Randomized controlled trial	Atorvastatin vs. no-treatment	CVD	Lipid profiles, HbA1c, blood pressure, BMI
Terrier et al. [65]	Cohort	Corticosteroid + rituximab vs. corticosteroid alone	Renal and immunological response	Vasculitis manifestations
Tiihonen et al. [66]	Cohort	Benzodiazepine vs. no-treatment	Mortality	Concurrent medication use
Young et al. [67]	Cohort	Tenofovir + ritonavir-boosted lopinavir vs. renofovir+efavirenz	eGFR	HIV-infection, diabetes, hypertension, hepatitis B or C infection, eGFR, CD4 cell count, virological failure

ACEI: angiotensin-converting enzyme inhibitors; ARB: angiotensin II receptor blockers; ART: antiretroviral therapy; CVD: cardiovascular diseases; eGFR, estimated glomerular filtration rate.

[&] Only the primary study exposure and outcome were reported in this table. "No-treatment" means not using the treatment under study.

[†] This study did not describe any specific substantial time-varying confounders for which adjustment was needed.

[*] This study used inverse probability of censoring weighting to deal with treatment crossover. Probability of treatment crossover was estimated based on baseline characteristics. Time-varying confounders were not mentioned.

Table 2.2 Type of potential bias and examination of identifiability assumptions

Reference	Type of potential bias addressed	Positivity assessed	Weight truncated or trimmed	Uncontrolled confounding discussed
Cook et al.[48]	Bias from blocking mediated effect	Mean: 1.01 Median (Inter Quartile Range): 1.00 (0.97-1.01)	Weight truncation at 0.01th and 99.99th percentiles	Yes
Desai et al. [49]	Bias from blocking mediated effect; Selection bias owing to artificial censoring	Mean (Standard Deviation): 1.00 (0.06)	Not reported	Yes
Gerhard et al. [50]	Bias from blocking mediated effect	Not reported	Not reported	Yes
Gsponer et al. [51]	Bias from blocking mediated effect	Not reported	Not reported	Yes
Haukka et al. [52]	No details provided†	Not reported	Not reported	Yes
HCV working group of COHERE [53]	Bias from blocking mediated effect	Not reported	Not reported	Yes
Heffron et al. [54]	No details provided†	Mean (range): 1.07 (0.82-1.34)	Weight truncation at 1st and 99th percentiles	Yes
Hernández et al. [55]	Bias from blocking mediated effect	Not reported	Not reported	Yes
HIV-CAUSAL Collaboration [56]	Selection bias owing to artificial censoring	Not reported	Weight truncation at 99th percentile	Yes
HIV-CAUSAL Collaboration [57]	Bias from blocking mediated effect	Mean: 1.04	Weight truncation at 10	Yes
Jin et al. [58]	Selection bias owing to artificial censoring	Not reported	Not reported	No
Kalayjian et al. [59]	Selection bias owing to artificial censoring	Not reported	Not reported	Yes
McCoy et al. [60]	Bias from	Not reported	Not reported	Yes

	blocking mediated effect			
Miller et al. [61]	Bias from blocking mediated effect	Not reported	Weight Trimming at 10	Yes
Morrison et al. [62]	No details provided†	Not reported	Not reported	Yes
Scherzer et al. [63]	Bias from blocking mediated effect	Not reported	Not reported	Yes
Shinozaki et al. [64]	Bias from blocking mediated effect	Not reported	Not reported	Yes
Terrier et al. [65]	No details provided†	Not reported	Not reported	No
Tiihonen et al. [66]	No details provided†	Not reported	Not reported	Yes
Young et al. [67]	Selection bias owing to artificial censoring	Not reported	Not reported	No

† If studies reported "using IPTW to control for time-varying confounding" without further specification of relationships between treatment, time-varying confounders and outcomes.

Table 2.3. Specification of treatment models and outcome models

Reference	Analytic strategy* Adherence level	Multiphase of treatment use modeled	Variables in weight numerator / Stabilized	Variables in weight denominator/ Covariates selection	Covariates in outcome model	Functional form of exposure
Cook et al.[48]	As-treated 73% stayed on initial treatment	Yes. Current use modeled by status of previous use	Baseline confounders, previous treatment	Baseline confounders, time-varying confounders, previous treatment	Baseline confounders	Most recent exposure
Desai et al. [49]	As-treated Not reported	No	Baseline confounders, previous treatment	Baseline confounders, time-varying confounders, previous treatment	Baseline confounders	Not reported
Gerhard et al. [50]	Intention to treat Not reported	Not applicable	Not reported /Yes	Baseline confounders, time-varying confounders, previous treatment	Not reported	Indicator of treatment initiation
Gsponer et al. [51]	Intention to treat Not reported	Not applicable	Not reported /Yes	Baseline confounders, time-varying confounders, previous treatment / Stepwise selection based on Akaike information criterion	Baseline confounders	Indicator of treatment "initiation"†; Time to treatment "initiation"†
Haukka et al. [52]	As-treated Treatment use covered 73% of study period	No	Baseline confounders	Baseline confounders, follow-up time	Not reported	Most recent exposure
HCV working group of COHERE [53]	Intention to treat Not reported	Not applicable	Not reported /Yes	Baseline confounders, time-varying confounders, previous treatment	Not reported	Indicator of treatment initiation

Heffron et al. [54]	As-treated 52.0% stayed on treatment	No	Not reported /Yes	Baseline confounders, time-varying confounders	Baseline confounders	Most recent exposure
Hernández et al. [55]	As-treated >85% stayed on treatment	No	Not reported / Not reported	Baseline confounders, time-varying confounders / Variables significantly associated with outcome	Not reported	Most recent exposure?
HIV-CAUSAL Collaboration [56]	Intention to treat Not reported	Not applicable	None/No	Baseline confounders, time-varying confounders, previous treatment	Baseline confounders	Initial treatment
HIV-CAUSAL Collaboration [57]	Intention to treat Not reported	Not applicable	Not reported / Yes	Baseline confounders, time-varying confounders, previous treatment	Baseline confounders	Indicator of treatment initiation; Cumulative exposure
Jin et al. [58]	Per-protocol 31% stayed on initial treatment	Yes. Treatment initiation and "treatment crossover" was considered separately	Not reported / Not reported	Baseline confounders, previous Treatment / Variables significantly associated with outcome and treatment crossover	Not reported	Initial treatment
Kalayjian et al. [59]	Per-protocol 64% stayed on initial treatment	Yes. Treatment initiation and discontinuation modeled separately	Not reported / Not reported	Baseline confounders, time-varying confounders, previous treatment	Not reported	Initial treatment

McCoy et al. [60]	As-treated 51.6% stayed on treatment	No	Unconditional probability of receiving observed treatment	Baseline confounders, time-varying confounders, previous treatment	Not reported	Most recent exposure
Miller et al. [61]	As-treated Not reported	No	Not reported / Yes	Baseline confounders, time-varying confounders, previous treatment	Not reported	Most recent exposure?
Morrison et al. [62]	As-treated 64.4% stayed on treatment	No	Not reported / Yes	Baseline confounders, time-varying confounders / Covariates significantly associated with outcome and treatment use and also predicted by past treatment use	Baseline confounders	Most recent exposure
Scherzer et al. [63]	Not reported	Not reported	Not reported / Yes	Not reported	Baseline confounders	Cumulative exposure; "Ever exposure"
Shinozaki et al. [64]	Intention to treat Not reported	Not applicable	Baseline confounders	Baseline confounders, time-varying confounders, previous treatment	Baseline confounders	Indicator of treatment initiation
Terrier et al. [65]	Not reported	Not reported	Baseline confounders	Not reported	Not reported	Most recent exposure
Tiihonen et al. [66]	Not reported	Not reported	Not reported/ Not reported	Not reported	Not reported	Not reported
Young et al. [67]	Per-protocol Not reported	Yes. Treatment initiation and	Unconditional probability of receiving	Baseline confounders, time-varying confounders,	None	Initial treatment

		discontinuation modeled separately	observed treatment	previous treatment		

* If the study stated that "modeling the probability of receiving observed treatment at each time visit", we assumed that the study was not making the assumption of intention-to-treat.
† "Initiation" refers to switching to the second-line therapy after treatment failure with first-line therapy.

Table 2.4. Primary exposure-outcome association[†] and discrepancy in IPTW estimates and standard regression estimates

Reference	Crude Hazard Ratio (95% Confidence Interval)	IPTW Hazard Ratio[*] (95% Confidence Interval)	Standard regression Hazard Ratio[*] (95% Confidence Interval)	Discrepancy found Reason discussed
Cook et al.[48]	1.00 (0.89-1.14)	0.93 (0.81-1.07)	0.96 (0.84-1.09)	Yes Correction of blocked mediated effect
Desai et al. [49]	Adjusted for baseline covariates: 0.89 (0.7-1.06)	0.79 (0.42-1.50)	Not reported	Not applicable
Gerhard et al. [50]	Adjusted for baseline covariates: 0.96 (0.87-1.07)	0.81 (0.71-0.92)	Not reported	Not applicable
Gsponer et al. [51]	0.52 (0.20-1.35)	0.25 (0.09-0.72)	Not reported	Not applicable
Haukka et al. [52]	NR	0.42 (0.37-0.47)	0.39 (0.37-0.40)	No
HCV working group of COHERE [53]	0.50 (0.35, 0.71)	0.72 (0.43-1.21)	Not reported	Not applicable
Heffron et al. [54]	1.73 (0.95-3.15)	1.84 (0.98-3.47)	1.98 (1.06-3.68)	No
Hernández et al. [55]	0.77 (0.49-1.21)	0.82 (0.52-1.32)	0.80 (0.51-1.26)	No
HIV-CAUSAL Collaboration [56]	1.46 (1.21-1.76)	1.59 (1.27-1.98)	1.38 (1.13-1.68)	Yes Not reported
HIV-CAUSAL Collaboration [57]	Adjusted for baseline covariates: 0.81 (0.67-0.97)	0.56 (0.44-0.72)	1.03 (0.86-1.24)	Yes Not reported
Jin et al. [58]	0.68 (0.56-0.83)	0.52 (0.45-0.61)	0.58 (0.47-0.72)	Not reported as such
Kalayjian et al. [59]	Not reported	3.35 (1.40-8.02)	1.34 (0.75-2.40)	Yes Not reported
McCoy et al. [60]	1.32 (1.00-1.74)	1.34 (0.75-2.37)	1.37 (1.01-1.85)	No
Miller et al. [61]	Not reported	1.26 (1.19-1.35)	1.07 (1.01-1.14)	Yes Confounding by indication
Morrison et al. [62]	0.89 (0.55-1.44)	0.84 (0.51-1.39)	0.88 (0.49-1.30)	No
Scherzer et al. [63]	Adjusted for baseline covariates: 1.30 (1.22-1.37)	1.24 (1.17-1.32)	1.34 (1.25-1.45)	No

Shinozaki et al. [64]	Adjusted for baseline covariates: 0.65 (0.30-1.40)	0.48 (0.19-1.16)	0.75 (0.34-1.63)	Yes Correction of blocked mediated effect
Terrier et al. [65]	Not reported	3.7 (1.3-10.6)	Not reported	Not applicable
Tiihonen et al. [66]	1.61 (1.06-2.45)	1.80 (1.02-3.20)	1.91 (1.13-3.22)	No
Young et al. [67]	Beta coefficient of exposure term from linear model: -4.6 (-8.6 to -0.5)	Beta coefficient of exposure term from linear model: -2.6 (-7.3 to 2.2)	Not reported	Not applicable

† Primary exposure and outcome are listed in Table 2.1.
* Adjusted for potential time-varying confounders.

Table 2.5. A summary of review results of the 20 pharmacoepidemiologic studies applying IPTW method in 2012

Elements of IPTW method	No. of studies (percent[#])
Types of bias IPTW was used to address	
Blocking mediated effects by time-varying confounders	11 (55)
Collider-stratification bias	0
Selection bias due to artificial censoring	5 (25)
Assessment of identifiability assumptions	
Discussed qualitatively uncontrolled confounding	17 (85)
Performed sensitivity analyses of uncontrolled confounding	0
Reported the weight distribution	4 (20)
Reported truncating or trimming extreme weights	5 (25)
Analytic strategy	
Intention-to-treat analysis	6 (30)
Per-protocol analysis	3 (15)
As-treated analysis	8 (40)
Weight construction	
Reported use of stabilized weights	15 (75)
Described how weights were stabilized	7 (47[†])
Described covariates in the treatment model for weight denominator	17 (85)
Modeled the multiphase of treatment use	1 (12.5[&])
Functional form of exposure in outcome models	
Indicator of treatment initiation or initial treatment use	9 (100[^])
Most recent treatment use	7 (100[*])
Discrepancy in estimates between IPTW and standard regression	
Discussed reasons for the substantial discrepancy	3 (50[§])

[#] The denominator is 20 unless indicated otherwise.
[†] The denominator is 15 studies which reported using stabilized weights.
[&] The denominator is 8 studies performing as-treated analyses.
[^] The denominator is 9 studies performing intention-to-treat or per-protocol analyses. One study performing intention-to-treat analyses also specified cumulative exposure as an alternative.
[*] The denominator is 7 studies performing as-treated analyses which provided information on the functional form of exposure.
[§] The denominator is 6 studies reporting substantial difference.

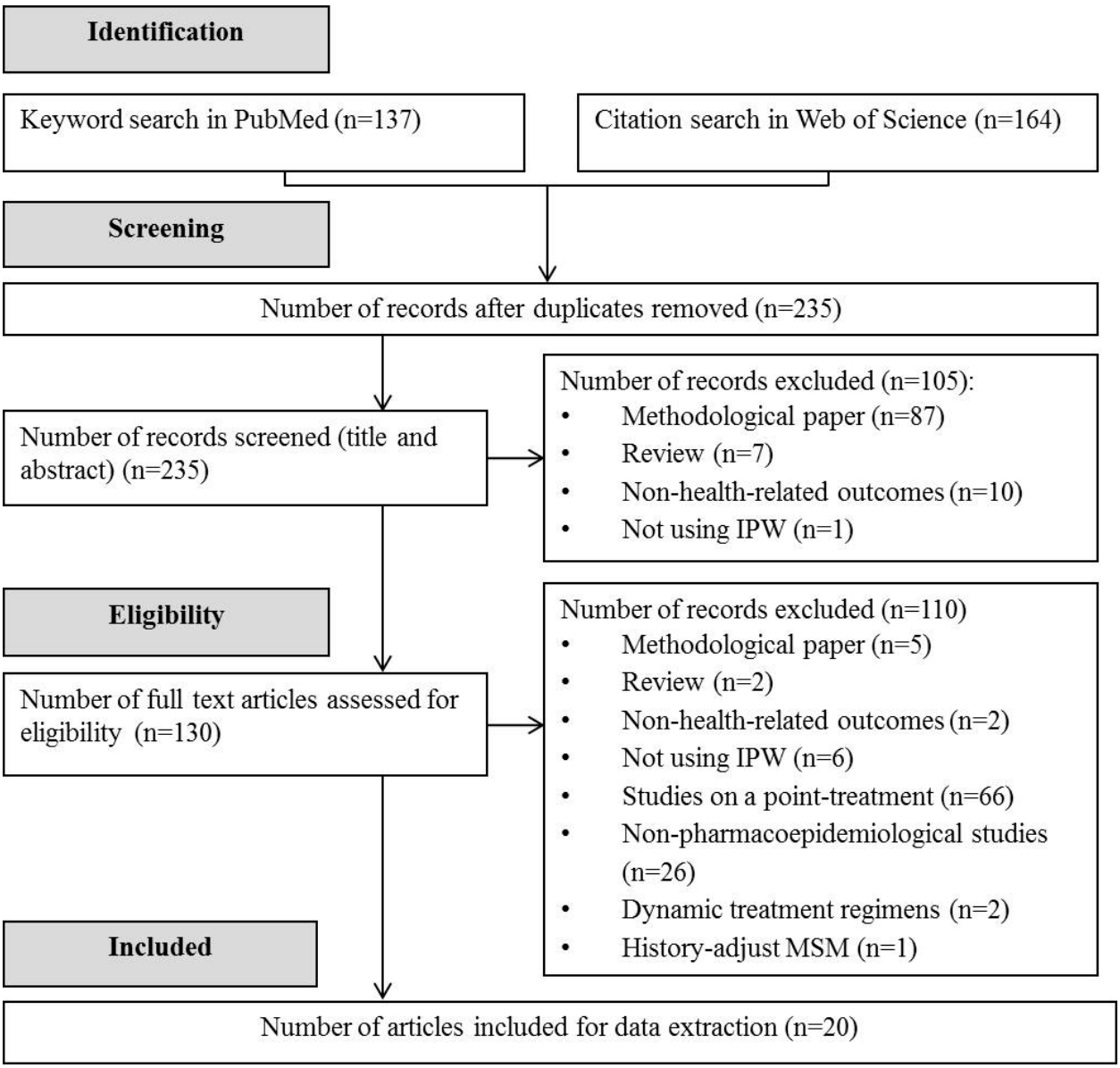

Figure 2.1. Identification of pharmacoepidemiolgoical studies using IPTW to deal with time-varying confounding in 2012

Chapter 3: The choice of analytic strategies in inverse-probability-of-treatment-weighted analysis: A simulation study

Abstract

Objectives To explore the impact of several assumptions made during weight construction on the validity and precision of estimates derived from inverse-probability-of-treatment-weighted analysis (IPTW). In particular, we compared the performance of 1) IPTW assuming intention-to-treat; 2) IPTW assuming complex mechanism of treatment assignment; 3) IPTW assuming simple mechanism of treatment assignment; and 4) IPTW assuming invariant confounders.

Methods We simulated data assuming a non-experimental design and aimed to quantify the effect of statin on lowering low-density lipoprotein cholesterol (LDL-C). Overall, 324 scenarios were simulated with parameter values varied on effect size, sample size, adherence level, probability of treatment initiation, and associations between LDL and treatment initiation and continuation. Effect estimates were derived from four IPTW approaches, and bias and precision of the estimates were evaluated.

Results IPTW estimates assuming intention-to-treat were biased towards to the null when there was non-null treatment effect and non-adherence after treatment initiation. For each one percent decrease in treatment adherence experienced by the sample, the bias in the average treatment effect increased by one percent. Compared to analyses assuming a simple mechanism of treatment assignment or invariant confounders, IPTW analyses that took into account the complex mechanism of treatment assignment generated unbiased estimates without sacrificing precision.

Conclusions Bias in IPTW estimate that assumes an intention-to-treat depends on the level of adherence after treatment initiation. Studies performing ITT analyses should report adherence measures after treatment initiation so that findings can be interpreted under appropriate consideration of the observed adherence patterns. Studies attempting to estimate the actual effect

of a time-varying treatment need to take into account the complex mechanism of treatment assignment in weight construction.

Introduction

Inverse probability of treatment weighted (IPTW) estimation of marginal structural models (MSM) has been increasingly used to adjust for time-varying confounding in pharmacoepidemiologic studies.[82] Unlike conventional methods, IPTW reduces confounding through assigning a weight to each participant, which is proportional to the inverse of conditional probability of receiving the observed treatment given confounders.[17,19] In the presence of time-varying confounders that are influenced by previous treatment, IPTW can adjust for the confounding without blocking the mediated effect or introducing selection bias.[17,19]

The validity of IPTW method relies on the correct estimation of the conditional probability of receiving observed treatment.[19, 21] Many studies applying IPTW have made an observational intent-to-treat assumption.[82] Specifically, this means that once treatment is initiated, patients are assumed to stay on that treatment for the remaining study period.[42] The advantages to invoking this assumption are that doing so simplifies the weight construction process and the assumption of no uncontrolled confounding.[42,44,82] However, for almost all studies assessing effect of a medication on a health-related outcome, the intention-to-treat assumption is violated to some degree.[83] In routine clinical practice in the United States, around one third to one half of the patients do not take medications as prescribed by their doctors.[84] When non-adherence is substantial, ITT analyses may estimate the effect of initiating a treatment, rather than the actual treatment effect.[85]

Our previous review reported that many studies using IPTW chose an as-treated analytic strategy,[82] i.e., they categorized patients according to the treatment actually received by patients during the study period.[40,85] Different from ITT analyses, as-treated analyses attempted to estimate the effect from actual treatment.[85] To correctly estimate the actual effect from a time-

varying treatment, investigators need to take into account the complex mechanism of treatment assignment in the process of weight construction.[42] For instance, several applications of IPTW have demonstrated that the relationships of confounders to initiating the treatment under study were different from continuing the treatment.[48,86] However, our review found that few studies performing as-treated analyses actually considered the complex mechanism of treatment assignment during weight construction.[82]

To our knowledge, no previous study has evaluated the impact of adopting different analytic strategies on the validity and precision of the IPTW estimates. The objectives of this study were to (1) compare the performance of several commonly-used analytic approaches to constructing weights and (2) explore the impact of several study characteristics, e.g., adherence level, prevalence of treatment use and magnitude of confounding, on the performance of different analytic approaches.

Methods

Data generation

We generated data assuming a non-experimental design and attempted to answer a hypothetical study question: "What is the effect of 12-week treatment with statin on lowering low-density lipoprotein cholesterol (LDL-C)?" We chose this question because the efficacy of statin on lowering LDL-C has been established[87, 88] and patterns of statin use among the population were extensively studied.[89,90,91,92,93] These studies provided the parameters to generate data close to reality, so the simulation results could be applied to real-world situations.[94]

Data were generated based on the casual diagram shown in Figure 3.1. In this diagram, *LDL* denotes levels of LDL-C, *A* indicates use of statin, and the subscripts 0, 1, 2 respectively represent baseline, 6 weeks and 12 weeks after baseline. The hypothetical study population

included 1,000 patients who were newly-diagnosed with hypercholesterolemia and failed to control LDL-C through therapeutic lifestyle changes.

At t_0, LDL_0 was simulated from a normal distribution, with mean 130 mg/dL and standard deviation 35 mg/dL.[95,96,97] For simplicity, we assumed the probability of initiating treatment at t_0, i.e., mean of A_0, only depended on levels of LDL_0. Specifically, we simulated A_0 from a binomial distribution with its mean generated from the following formula:

$$logit(Pr(A_0=1|LDL_0))=\alpha_0+log(OR_{LDL\text{-}Initiation})\times LDL\_Level_0 \qquad (1)$$

where LDL_Level_0 was 0 if LDL_0 was less than 160 mg/dL, 1 if LDL_0 was between 160 and 190 mg/dL, and 2 if LDL_0 was ≥190 mg/dL.[87] $OR_{LDL\text{-}Initiation}$ was set at 1.5 based on the literature that higher LDL levels were associated with greater probability of initiating statin treatment.[90,98] α_0 was set at 0.3146, so that 60% of the study participants initiated treatment at t_0.[89]

At t_1, LDL was on average reduced by 30% from LDL_0 among those who initiated treatment at t_0, and remained unchanged among those who did not.[88,99] A random error was added to LDL_1 so that its standard deviation was around 40 mg/dL. Statin use A_1 was generated separately for those who did not initiate treatment at t_0 (i.e., $A_0=0$) and those who did (i.e., $A_0=1$). Among those with $A_0=0$, A_1 was generated in the same way as A_0 using formula (1) expect that A_1 was determined by levels of LDL_1 instead of LDL_0. For those with $A_0=1$, we assumed the probability of continuing treatment at t_1 (defined as adherence level in our study) depended on the reduction in LDL from t_0 to t_1. Specifically, among those with $A_0=1$, A_1 was simulated from a binomial distribution with its mean generated from the following formula:

$$logit(Pr(A_1=1|\ A_0=1,\ LDL_0,\ LDL_1))=\gamma_0+\ log(OR_{LDL\text{-}Continuation})\times LDL\_Red \qquad (2)$$

where LDL_Red was 0 if reduction in LDL was less than 30% of LDL_0, and 1 if the reduction was greater than 30% of LDL_0. Reduction by 30% of LDL_0 was the average change in LDL from

t_0 to t_1 among those with A_0=1, so *LDL_Red* was actually a dummy variable with value 1 indicating an above-average reduction in LDL. $OR_{LDL\text{-}Continuation}$ was set at 1.5, so that patients with above-average reduction in LDL had 50% higher odds of continuing statin treatment compared to those with below-average reduction.[92,93] γ_0 was set at 0.6528, so that 70% of those with A_0=1 continued treatment at t_1.[91,92]

At t_2, we assumed that among patients who initiated treatment at t_1 (i.e., with A_0=0 and A_1=1), LDL on average decreased by 30% from LDL_1, which was the same effect size we specified for treatment initiation at t_0 on LDL_1; among patients continuing treatment at t_1 (i.e., with A_0=1 and A_1=1), LDL on average decreased by 14.3% from LDL_1, which corresponded to a total decrease of 40% from LDL_0 after 12-week of treatment[88,99]; among those discontinuing treatment at t_1 (i.e., with A_0=1 and A_1=0), LDL on average increased by 42.9% from LDL_1, i.e., rebounded to the baseline LDL level.[100,101] A random variation was added to LDL_2 so that its standard deviation was around 45 mg/dL. Based on these specifications, the true effect size of 12-week treatment with statin was 130 mg/dL×(-40%), i.e., -52 mg/dL.

To assess the performance of different analytic approaches under various scenarios, besides the basic-case scenario described above, we also generated alternative scenarios with parameter values varied on the probability of treatment initiation, $OR_{LDL\text{-}Initiation}$, $OR_{LDL\text{-}Continuation}$, adherence level, effect size and sample size. As shown in Table 3.1, in alternative scenarios, we generated data with 10% of the patients starting treatment at t_0 and t_1, $OR_{LDL\text{-}Initiation}$ or $OR_{LDL\text{-}Continuation}$ equal to 3, adherence level equal to 50%, 60%, 80%, 90% or 100%, effect size equal to 0 or -26, and sample size equal to 200 or 20,000.

Analytic approaches

We analyzed the simulated data using IPTW based on four different approaches to constructing the weight: 1) IPTW assuming intention-to-treat (ITT-IPTW); 2) IPTW assuming complex mechanism of treatment assignment (Complex-IPTW); 3) IPTW assuming simple mechanism of treatment assignment (Simple-IPTW); and 4) IPTW assuming invariant confounders (Invar-IPTW). For the Complex-IPTW, Simple-IPTW, and Invar-IPTW approaches, we conducted as-treated analyses.[40] Complex-IPTW acknowledged that the impact of confounders on initiating a treatment was different from their impact on continuing the treatment, whereas Simple-IPTW assumed that confounders had same impact on initiating and continuing the treatment. Invar-IPTW assumed that time-varying confounders remained unchanged during the follow-up period. Our review did find some studies performing Invar-IPTW analyses, perhaps because they did not collect the time-varying information on potential confounders.[82]

The weight construction process of the methods is described in Table 3.2. In analyses with all four methods, patient-specific weights were first estimated separately at t_0 and t_1, which were the unconditional probability of receiving observed treatment divided by the conditional probability of receiving observed treatment given confounders.[19, 20, 37] A patient's final weight was the product of his/her weights at t_0 and t_1.[19, 20, 37] In addition, we assumed all methods correctly recognized the mechanism of treatment initiation at t_0 (i.e., $Pr(A_0|LDL\_Level_0)$), and thus shared the same process of weight construction at t_0. The differences among methods were in the way of estimating probability of treatment use at t_1. ITT-IPTW, Complex-IPTW and Simple-IPTW correctly modeled the probability of treatment initiation at t_1 among patients with $A_0=0$, i.e., $Pr(A_1=a_1| LDL\_Level_1)$, but they differed in estimating the conditional probability of continuing treatment at t_1 among patients with $A_0=1$. ITT-IPTW assumed that the probability of

continuing treatment was 1, and thus the weight was 1 at t_1; Complex-IPTW assumed that treatment continuation depended on reduction in LDL, which was consistent with the true data generation process; Simple-IPTW assumed that, same as treatment initiation, treatment continuation depended on levels of LDL. As for Invar-IPTW, it deviated even further from the true data degeneration process than Simple-IPTW because it used LDL_0 to predict treatment initiation and continuation at t_1.

Probability of receiving treatment given LDL was estimated with logistic regression models. For instance, the conditional probability of initiating treatment at t_1 given levels of LDL_1 was estimated using the following logistic model among patients with A_0=0:

$$logit(A_1=1 \mid LDL\_Level_1, A_0=0)=\eta_0+ \eta_1\times LDL\_Level_1 \qquad (3)$$

For those with A_1=0, the probability of receiving observed treatment was 1 minus the predicted probability derived from model (3).

After weights were constructed, the second step in each approach was to fit a weighted structural model to estimate the effect of statin on LDL_2. Except using different weights in different approaches, we used the same linear structural model in all four methods as follows:

$$LDL_2= \beta_0+ \beta_1\times A_{11}+ \beta_2\times A_{01}+ \beta_3\times A_{10}+ \varepsilon \qquad (4)$$

where A_{11} indicates statin use at both t_0 and t_1, A_{10} and A_{01} respectively indicates statin use only at t_0 and t_1. Because ITT-IPTW assumed that no patients discontinued the treatment once they initiated it, A_{11} represents treatment initiation at t_0, A_{01} represents treatment initiation at t_1 and A_{10} is always 0 in ITT-IPTW analyses. The primary parameter of interest in this study is β_1, which estimates the difference between LDL after the study population was treated with statin for 12 weeks and LDL when none of the population was treated with statin.[19]

Assessment of model performance

Under each scenario, we simulated 2,000 datasets, and with each dataset we performed analyses with the four approaches described above. Each analytic method generated 2,000 estimates under every scenario. We evaluated the validity of different methods using percentage bias, which was calculated as the difference between the average of 2,000 estimates and the true effect size, divided by the true effect size.[94] To compare the precision of estimates derived from different methods, we calculated the standard deviation of the 2,000 estimates under each scenario.[94]

Results

Overall, 324 scenarios were simulated with parameter values varying on effect size (3 options), sample size (3 options), adherence level (6 options), probability of treatment initiation (2 options), and associations between LDL and treatment initiation and continuation (3 options). Performance of the four analytic approaches under various scenarios is shown in Tables 3.3 and 3.4. In these tables, we show results for scenarios with β, n, $Pr(A_0=1)$, $OR_{LDL\text{-}Continuation}$ and $OR_{LDL\text{-}Initiation}$ set at the basis-case values, as well as scenarios in which we changed one parameter at a time while keeping all others at their basic values. To fully illustrate the impact of non-adherence on the performance of different approaches, we reported results for scenarios with all 6 adherence levels.

Table 3.3 shows the simulated bias in estimates from the four analytic approaches. When there was no treatment effect, the ITT-IPTW estimates were close to the true effect size regardless of adherence levels. When the true effect was non-null and the adherence level was less than 100%, ITT-IPTW estimates were biased towards the null. Bias in ITT-IPTW estimates

was not influenced by probability of treatment initiation, levels of confounding, sample size or effect size (results for scenarios with an effect size of -26 were not shown but similar to those with an effect size of -52). Instead, the extent of bias in ITT-IPTW estimates was linearly correlated with levels of non-adherence: one percent increase in non-adherence was associated with approximately one percent increase in the bias in ITT-IPTW estimates.

Complex-IPTW estimates were close to the true effect regardless of effect size or choices of other parameter values. When the sample size was 200 or $OR_{LDL\text{-}Initiation}$ was 3, Complex-IPTW estimates were biased upward for less than 2%. Under all other scenarios shown in Table 3.3, IPTW estimates were biased less than 0.5%. Simple-IPTW estimates were biased downwards under all scenarios except several scenarios with a sample size of 200 or $OR_{LDL\text{-}Initiation}$ of 3. This downward bias became more apparent when $OR_{LDL\text{-}Continuation}$ was 3 or as the level of non-adherence increased. Invar-IPTW estimates were biased upward for most scenarios shown in Table 3.3. This upward bias became stronger when $OR_{LDL\text{-}Initiation}$ was 3, but became less so when $OR_{LDL\text{-}Continuation}$ was 3 or the level of non-adherence increased.

The empirical standard errors of estimates derived from the four approaches are shown in Table 3.4. Under scenarios with no treatment effect, standard errors of ITT-IPTW estimates increased when only 10% of the population initiated treatment, or $OR_{LDL\text{-}Continuation}$ was 3 (data not shown), but did not depend on levels of adherence. When the treatment effect was non-null, standard errors of ITT-IPTW estimates increased along with the levels of non-adherence, and this relationship was also observed for estimates derived from other three approaches.

Compared to ITT-IPTW estimates, Complex-IPTW estimates had larger standard errors when there was no treatment effect, but smaller standard errors when the treatment effect was non-null. Compared to Complex-IPTW estimates, Simple-IPTW estimates had slightly larger

standard errors under all scenarios except those with no treatment effects, and Invar-IPTW estimates had slightly larger standard errors under all scenarios except those with a sample size of 200 or $OR_{LDL\text{-}Initiation}$ of 3.

Discussion

Under the realistically constructed scenarios in this study, we demonstrated that ITT-IPTW estimates were biased towards to the null when there was non-null treatment effect and non-adherence after treatment initiation. Interestingly, the extent of bias in ITT estimates appeared solely depending on the level of non-adherence. One percent decrease in adherence level among the study sample was associated with one percent increase in the bias in ITT estimates. IPTW analyses that ignored the complex mechanism or the time-varying nature of confounders were biased. IPTW analyses that took into account the complex mechanism of treatment assignment generated approximately unbiased estimates without sacrificing precision.

IPTW assuming a simple mechanism of treatment assignment failed to correctly model the relationship between LDL and treatment continuation. As such, the negative confounding bias could not be fully-controlled in the weighted population. As expected, this uncontrolled confounding bias became more apparent when the impact of LDL on treatment continuation became stronger (i.e., $OR_{LDL\text{-}Continuation}$=3). Similarly, the weight construction process in IPTW assuming invariant confounders did not correctly model the relationships of LDL to both treatment initiation and continuation. As such, the uncontrolled confounding biased the estimates. As the impact of LDL on treatment initiation increased, a positive bias became more dominant, and as the impact of LDL on treatment continuation increased, a negative bias became more dominant.

Conventional wisdom suggested that IPTW considering the complex mechanisms of treatment would generate estimates that were more valid but less precise than either ITT or Simple-IPTW approaches.[42] However, we found that estimates derived from a method using IPTW with the complex mechanisms of treatment actually had smaller standard errors than ITT-IPTW estimates when the treatment effect was non-null. The standard error of an IPTW estimate probably depends on the variation of the constructed weights,[20] and variance of study exposure and mean squared error of the outcome model (if it is a linear regression model).[102] Because the additional incorporation of the probability of treatment continuation in the weight construction process in Complex-IPTW analyses, weights in Complex-IPTW had a larger variance than those in ITT-IPTW analyses. This explains the finding that under scenarios with no treatment effect, the Complex-IPTW estimates had larger standard errors than ITT-IPTW estimates. However, when there was non-null treatment effect, standard errors of the ITT-IPTW estimate were probably inflated by the increased mean squared error due to the misspecification of study exposure in the outcome model.[102]

It is well-known that ITT estimates are unbiased when there is no treatment effect but biased towards the null when the effect is non-null.[85] However, to our knowledge, this was the first study that explored the relationship of the extent of bias in ITT-IPTW estimates in relation to levels of non-adherence and patterns of confounding. Under the causal structure assessed in our study, after the analyses appropriately controlled for confounding for treatment initiation, for each one percent decrease in treatment adherence experienced by the sample, the bias in the average treatment effect increased by one percent. This finding has important implications for future studies performing ITT-IPTW analyses. First, the bias in ITT-IPTW estimates is non-trivial even if the adherence level is high. For instance, when the adherence level was as high as

90%, the ITT estimates underestimated the treatment effect by ~10%. This underestimation may be especially problematic for drug safety studies, because the ITT analysis may miss the harmful medication effects.[85] Second, given the dependence of ITT estimates on adherence levels, we recommend that studies performing ITT analyses should report adherence measures after treatment initiation. If provided, findings can be interpreted under appropriate consideration of the observed adherence patterns.[73] Our previous review found that few studies performing ITT-IPTW analyses actually did this.[82]

Our study demonstrated the necessity of taking into account the different relationships between confounders and different treatment regimens in the process of weigh construction. Besides the realistic relationships between LDL-C and statin initiation and continuation illustrated in this simulation, the phenomenon of complex treatment assignment was also noted by other studies.[48,86] For instance, when using IPTW to estimate the effect of aspirin on preventing cardiovascular disease, Cook et al. found that, potential confounders such as occurrence of angina and transient ischemic attacks were negatively associated with continuing treatment with aspirin, but positively correlated with starting aspirin.[48] To correctly perform Complex-IPTW analyses, substantive knowledge regarding the relationships between potential confounders and different treatment regimens (e.g., initiation, continuation, and resumption, etc) should guide the specification of treatment models during weight construction. Furthermore, the findings that IPTW estimates with time invariant confounders were biased due to uncontrolled confounding emphasized the importance of collecting information on time-varying factors that predict the study outcomes and also bring about changes in treatment use.[85]

To our knowledge, this was the first study that explored the impact of various assumptions made in weight construction on the validity and precision of IPTW estimates.

Besides simulating data that mimicked a real-world situation, we generated a total 324 scenarios that varied in parameter values for a range of study characteristics. However, several limitations must be considered. First, we simulated scenarios with treatment use varying only at two time points. A real-world longitudinal study will likely involve more time points. If so, it is likely that treatment assignment mechanisms become more complex than what we simulated. The extent to which our findings would be generalizable to more complex scenarios remains unknown. Second, we simulated a continuous variable as the study outcome. Whether or not our findings extend to different types of outcomes such time-to-event outcomes or categorical outcomes needs to be explored.

In conclusion, under a range of simulated scenarios, we demonstrated that IPTW estimates assuming intention-to-treat were biased towards the null when there was non-null treatment effect and the adherence after treatment initiation was not 100%. This bias was linearly correlated with non-adherence levels. Studies attempting to estimate the actual effect of a time-varying treatment on a continuous outcome variable should take into account the complex mechanism of treatment assignment in the process of weight construction.

Table 3.1. Parameter values used for data generation

Parameter	Meaning	Basic-case scenario	Alternative scenarios
$Pr(A_0=1)$	Probability of starting statin treatment at baseline or time 1	60% [89]	10%
$OR_{LDL\text{-}Initiation}$	Odds ratio of starting statin treatment comparing a higher level LDL to a lower level LDL	1.5 [90]	3
$Pr(A_1=1\|A_0=1)$	Probability of continuing statin treatment at time 1 among those on treatment at baseline, i.e., adherence level	70% [91,92]	50%, 60%, 80% 90%, 100%
$OR_{LDL\text{-}Continuation}$	Odds ratio of continuing statin treatment comparing above-average reduction in LDL to below-average reduction	1.5 [92,93]	3
β	Effect size	52 [87,88]	0, 26
n	Sample size	1000	200, 20000

Table 3.2. Approaches of constructing weights

Modeling approach	Weight construction
ITT-IPTW: Marginal structural models assuming intention-to-treat	At t_0: $w_0=Pr(A_0=a_0)/Pr(A_0=a_0\mid LDL\_Level_0)$ At t_1: If $A_0=0$, $w_1= Pr(A_1=a_1)/Pr(A_1=a_1\mid LDL\_Level_1)$; If $A_0=1$, $w_1=\mathbf{1}$ Final weight: $w_{final}= w_0\times w_1$
Complex-IPTW: Marginal structural models assuming complex mechanism of treatment assignment	At t_0: $w_0=Pr(A_0=a_0)/Pr(A_0=a_0\mid LDL\_Level_0)$ At t_1: If $A_0=0$, $w_1^*= Pr(A_1=a_1)/Pr(A_1=a_1\mid LDL\_Level_1)$; If $A_0=1$, $w_1^*= \mathbf{Pr(A_1=a_1)/Pr(A_1=a_1\mid LDL\_Red)}$ Final weight: $w_{final}^*= w_0\times w_1^*$
Simple-IPTW: Marginal structural models assuming simple mechanism of treatment assignment	At t_0: $w_0=Pr(A_0=a_0)/Pr(A_0=a_0\mid LDL\_Level_0)$ At t_1: If $A_0=0$, $w_1^{**}= Pr(A_1=a_1)/Pr(A_1=a_1\mid LDL\_Level_1)$; If $A_0=1$, $w_1^{**}= \mathbf{Pr(A_1=a_1)/Pr(A_1=a_1\mid LDL\_Level_1)}$ Final weight: $w_{final}^{**}= w_0\times w_1^{**}$
Invar-IPTW: Marginal structural models assuming invariant confounders	At t_0: $w_0=Pr(A_0=a_0)/Pr(A_0=a_0\mid LDL\_Level_0)$ At t_1: If $A_0=0$, $w_1^{***}= \mathbf{Pr(A_1=a_1)/Pr(A_1=a_1\mid LDL\_Level_0)}$; If $A_0=1$, $w_1^{***}= \mathbf{Pr(A_1=a_1)/Pr(A_1=a_1\mid LDL\_Level_0)}$ Final weight: $w_{final}^{***}= w_0\times w_1^{***}$

$Pr(A_t=a_t)$: unconditional probability of receiving observed treatment at time t.

LDL_Level_t was 0 if LDL_t was <160, 1 if $160 \le LDL_t \le 190$, and 2 if LDL_t was ≥190.

LDL_Red was 0 if LDL_0-LDL_1 was ≤ 30% of LDL_0, and 1 if LDL_0-LDL_1 was greater than 30% of LDL_0.

Bold texts highlight the differences in weight construction among approaches.

Table 3.3. Simulated bias in four analytic approaches with marginal structural models under various scenarios

β	$Pr(A_0=1)$	OR_{LDL} -Initiation	OR_{LDL}- Continuation	n	$Pr(A_1=1 \mid A0=1)$ (%)	ITT-IPTW		Complex-IPTW*		Simple-IPTW*		Invar-IPTW*	
						$\hat{\beta}$	Bias (%)	$\hat{\beta}$	Bias (%)	$\hat{\beta}$	Bias (%)	$\hat{\beta}$	Bias (%)
0	60%	1.5	1.5	1000	100	0.1	--						
0	60%	1.5	1.5	1000	90	0.1	--	0.1	--	0	--	1.8	--
0	60%	1.5	1.5	1000	80	0.1	--	0.1	--	-0.2	--	1.4	--
0	60%	1.5	1.5	1000	70	0.1	--	0.1	--	-0.4	--	1.0	--
0	60%	1.5	1.5	1000	60	0.1	--	0.1	--	-0.6	--	0.7	--
0	60%	1.5	1.5	1000	50	0.1	--	0.2	--	-0.7	--	0.3	--
-52	60%	1.5	1.5	1000	100	-51.9	-0.2						
-52	60%	1.5	1.5	1000	90	-46.5	-10.6	-51.9	-0.2	-52.2	0.3	-50.2	-3.5
-52	60%	1.5	1.5	1000	80	-41.2	-20.8	-51.9	-0.2	-52.5	0.9	-50.5	-2.9
-52	60%	1.5	1.5	1000	70	-35.9	-31	-51.9	-0.2	-52.8	1.5	-50.8	-2.3
-52	60%	1.5	1.5	1000	60	-30.6	-41.2	-51.9	-0.2	-53.0	2.0	-51.1	-1.7
-52	60%	1.5	1.5	1000	50	-25.4	-51.2	-51.9	-0.3	-53.3	2.5	-51.4	-1.1
-52	**10%**	1.5	1.5	1000	100	-52	-0.1						
-52	**10%**	1.5	1.5	1000	90	-46.6	-10.5	-52	0	-52.3	0.6	-51.9	-0.1
-52	**10%**	1.5	1.5	1000	80	-41.2	-20.7	-52	-0.1	-52.6	1.1	-52.2	0.4
-52	**10%**	1.5	1.5	1000	70	-35.9	-31	-52	0	-52.9	1.7	-52.5	1.0
-52	**10%**	1.5	1.5	1000	60	-30.6	-41.2	-52	-0.1	-53.2	2.3	-52.8	1.5
-52	**10%**	1.5	1.5	1000	50	-25.4	-51.2	-51.9	-0.1	-53.5	3.0	-53.1	2.1
-52	60%	**3**	1.5	1000	100	-51.3	-1.3						
-52	60%	**3**	1.5	1000	90	-45.9	-11.7	-51.3	-1.3	-51.6	-0.8	-46.5	-10.5
-52	60%	**3**	1.5	1000	80	-40.6	-22	-51.3	-1.3	-51.9	-0.2	-46.8	-9.9
-52	60%	**3**	1.5	1000	70	-35.3	-32.2	-51.3	-1.3	-52.2	0.4	-47.2	-9.3
-52	60%	**3**	1.5	1000	60	-30	-42.3	-51.3	-1.3	-52.5	0.9	-47.5	-8.7
-52	60%	**3**	1.5	1000	50	-24.8	-52.3	-51.3	-1.3	-52.7	1.4	-47.8	-8.1
-52	60%	1.5	**3**	1000	100	-51.9	-0.2						
-52	60%	1.5	**3**	1000	90	-46.2	-11.1	-51.9	-0.2	-52.6	1.2	-50.7	-2.6

-52	60%	1.5	**3**	1000	80	-40.7	-21.8		-51.9	-0.2		-53.3	2.6		-51.5	-1.0
-52	60%	1.5	**3**	1000	70	-35.2	-32.4		-51.9	-0.2		-54.1	4.0		-52.3	0.6
-52	60%	1.5	**3**	1000	60	-29.8	-42.7		-51.9	-0.2		-54.9	5.5		-53.1	2.2
-52	60%	1.5	**3**	1000	50	-24.5	-52.8		-51.9	-0.3		-55.6	6.9		-53.9	3.7
-52	60%	1.5	1.5	**200**	100	-51.2	-1.5									
-52	60%	1.5	1.5	**200**	90	-45.9	-11.8		-51.2	-1.5		-51.5	-1.0		-49.5	-4.8
-52	60%	1.5	1.5	**200**	80	-40.5	-22.1		-51.2	-1.5		-51.8	-0.3		-49.8	-4.2
-52	60%	1.5	1.5	**200**	70	-35.2	-32.3		-51.2	-1.5		-52.2	0.4		-50.1	-3.6
-52	60%	1.5	1.5	**200**	60	-29.9	-42.5		-51.2	-1.5		-52.5	0.9		-50.4	-3.0
-52	60%	1.5	1.5	**200**	50	-24.7	-52.4		-51.1	-1.7		-52.8	1.5		-50.7	-2.5
-52	60%	1.5	1.5	**20,000**	100	-52	0									
-52	60%	1.5	1.5	**20,000**	90	-46.6	-10.4		-52	0		-52.3	0.5		-50.3	-3.3
-52	60%	1.5	1.5	**20,000**	80	-41.3	-20.6		-52	0		-52.5	1.0		-50.6	-2.7
-52	60%	1.5	1.5	**20,000**	70	-36	-30.9		-52	0		-52.8	1.6		-50.9	-2.1
-52	60%	1.5	1.5	**20,000**	60	-30.7	-41		-52	0		-53.1	2.1		-51.2	-1.5
-52	60%	1.5	1.5	**20,000**	50	-25.5	-51		-52	0		-53.4	2.6		-51.5	-0.9

$\bar{\beta}$ is the mean of effect estimates from 2,000 trials.

Bias (%)=$[\sum(\hat{\beta}-\beta)/\beta\times100\%]/2000$.

* Estimates from Complex-IPTW, Simple-IPTW and Invar-IPTW were same as ITT-IPTW estimates in scenarios with 100% adherence.

-- When the true effect size was zero, the percentage of bias could not be calculated.

Table 3.4. Simulated standard errors of estimates from four analytic approaches of marginal structural models under various scenarios

β	$Pr(A_0=1)$	$OR_{LDL\text{-}Initiation}$	$OR_{LDL\text{-}Continuation}$	n	$Pr(A_1=1\|A0=1)$ (%)	ITT-IPTW	Complex-IPTW*	Simple-IPTW*	Invar-IPTW*
0	60%	1.5	1.5	1000	100	3.10			
0	60%	1.5	1.5	1000	90	3.10	3.15	3.12	3.25
0	60%	1.5	1.5	1000	80	3.10	3.22	3.17	3.30
0	60%	1.5	1.5	1000	70	3.10	3.29	3.22	3.37
0	60%	1.5	1.5	1000	60	3.10	3.37	3.28	3.47
0	60%	1.5	1.5	1000	50	3.10	3.49	3.38	3.60
-52	60%	1.5	1.5	1000	100	2.97			
-52	60%	1.5	1.5	1000	90	3.05	2.99	2.99	3.09
-52	60%	1.5	1.5	1000	80	3.13	3.03	3.04	3.13
-52	60%	1.5	1.5	1000	70	3.22	3.08	3.09	3.18
-52	60%	1.5	1.5	1000	60	3.26	3.12	3.15	3.24
-52	60%	1.5	1.5	1000	50	3.27	3.18	3.23	3.34
-52	**10%**	1.5	1.5	1000	100	3.25			
-52	**10%**	1.5	1.5	1000	90	3.89	3.40	3.41	3.38
-52	**10%**	1.5	1.5	1000	80	4.32	3.55	3.60	3.56
-52	**10%**	1.5	1.5	1000	70	4.59	3.84	3.92	3.84
-52	**10%**	1.5	1.5	1000	60	4.75	4.06	4.21	4.11
-52	**10%**	1.5	1.5	1000	50	4.89	4.41	4.60	4.46
-52	60%	**3**	1.5	1000	100	6.36			
-52	60%	**3**	1.5	1000	90	6.38	6.38	6.39	4.68
-52	60%	**3**	1.5	1000	80	6.39	6.41	6.41	4.71
-52	60%	**3**	1.5	1000	70	6.44	6.43	6.44	4.75
-52	60%	**3**	1.5	1000	60	6.48	6.45	6.46	4.80
-52	60%	**3**	1.5	1000	50	6.47	6.47	6.49	4.87
-52	60%	1.5	**3**	1000	100	2.96			
-52	60%	1.5	**3**	1000	90	3.06	2.98	2.99	3.09
-52	60%	1.5	**3**	1000	80	3.17	3.03	3.04	3.14
-52	60%	1.5	**3**	1000	70	3.23	3.07	3.08	3.18
-52	60%	1.5	**3**	1000	60	3.27	3.12	3.14	3.25
-52	60%	1.5	**3**	1000	50	3.30	3.22	3.23	3.35
-52	60%	1.5	1.5	**200**	100	7.54			
-52	60%	1.5	1.5	**200**	90	7.78	7.58	7.59	7.48
-52	60%	1.5	1.5	**200**	80	7.92	7.69	7.71	7.57
-52	60%	1.5	1.5	**200**	70	8.05	7.78	7.78	7.65
-52	60%	1.5	1.5	**200**	60	8.09	7.85	7.88	7.77
-52	60%	1.5	1.5	**200**	50	8.13	8.03	8.11	7.96
-52	60%	1.5	1.5	**20,000**	100	0.67			
-52	60%	1.5	1.5	**20,000**	90	0.69	0.67	0.67	0.71
-52	60%	1.5	1.5	**20,000**	80	0.71	0.68	0.68	0.71
-52	60%	1.5	1.5	**20,000**	70	0.72	0.69	0.69	0.72
-52	60%	1.5	1.5	**20,000**	60	0.73	0.70	0.71	0.74
-52	60%	1.5	1.5	**20,000**	50	0.74	0.72	0.72	0.75

Standard error is the standard deviation of the estimates from 2,000 trials.

*Standard errors of estimates from Complex-IPTW, Simple-IPTW and Invar-IPTW were same as those of ITT-IPTW estimates in scenarios with 100% adherence.

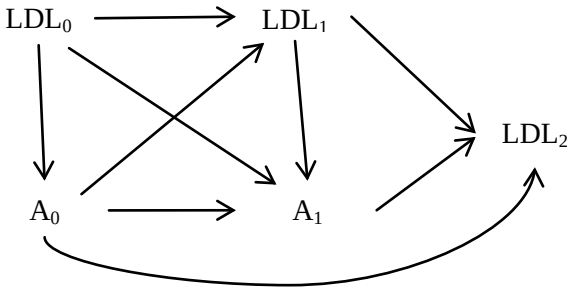

Figure 3.1. The causal diagram guiding data generation.

LDL denotes levels of low-density lipoprotein cholesterol, A indicates use of statin medication, and the subscripts 0, 1, 2 represent baseline, 6 weeks and 12 weeks after baseline, respectively.

Chapter 4: Long-term effects of glucosamine and chondroitin on treating knee osteoarthritis: An analysis with marginal structural models

Abstract

Objectives: The purpose of this study was to estimate the long-term effectiveness of glucosamine and chondroitin in relieving knee symptoms and slowing disease progression among patients with knee osteoarthritis (OA).

Methods: The 4-year follow-up data from Osteoarthritis Initiative were analyzed. We used a "new-user" design, for which only participants who were not using glucosamine/chondroitin at baseline were included in analyses (n=1,625). Cumulative exposure was calculated as the number of visits when participants reported use of glucosamine/chondroitin. Knee symptoms were measured with WOMAC Pain, Stiffness and Physical Function, and structural progression of OA was measured with joint space width (JSW). Sociodemographic characteristics and indices of disease severity were considered as potential confounders. To take into account that the indices of disease severity may be simultaneously confounders and intermediate variables, we used marginal structural models to estimate the long-term treatment effects.

Results: During the study period, 18% of the participants initiated treatment with glucosamine/chondroitin and 4% reported use at all assessments. After adjustment for potential confounders with marginal structural models, we found no clinically significant differences between users at all assessments and never-users of glucosamine/chondroitin in WOMAC Pain: 0.68 (95% CI: -0.16 to 1.53); WOMAC Stiffness: 0.41 (95% CI: 0 to 0.82); WOMAC Function: 1.28 (95% CI: -1.23 to 3.79); or JSW: 0.11 (95% CI: -0.21 to 0.44).

Conclusions: Long-term use of glucosamine/chondroitin did not appear to relieve symptoms or modify disease progression among radiographically confirmed patients with OA. Our findings are consistent with the results from recent long-term clinical trials.

Introduction

Osteoarthritis (OA) is the most common type of arthritis and a leading cause of pain and physical disability in older adults.[1] Although currently no effective remedies for OA exist, clinical guidelines recommend both pharmacological and non-pharmacological therapies to relieve symptoms.[103] In the United States, glucosamine and chondroitin are two dietary supplements that are commonly used among patients with OA.[8] Both glucosamine and chondroitin are essential components of the proteoglycans in normal cartilage and were purported to provide substrate for the biosynthesis of proteoglycans.[104] In vitro and animal studies suggest that glucosamine and chondroitin simulate the synthesis of proteoglycans and inhibit the synthesis of proteolytic enzymes that lead to the premature breakdown of cartilage.[105,106]

Despite the biologic plausibility, evidence regarding the efficacy of glucosamine and chondroitin in relieving OA symptoms and modifying structural progression is not established. Several meta-analyses which pooled results from existent randomized clinical trials that assessed symptomatic benefits reported substantial heterogeneity in findings across studies.[11, 107-109] Differences in study quality, preparation of the interventions, industry involvement and study size may have explained the observed heterogeneity.[11, 107-109] Large-scale trials with high quality and little connection to industry often reported a much smaller effect size of symptoms relief than earlier small industry-funded studies.[107,108] Regarding the efficacy of glucosamine and chondroitin in modifying disease progression, several meta-analyses reported small to moderate effect sizes and studies with longer intervention periods demonstrated a stronger effect of glucosamine on slowing joint space narrowing than studies with a shorter treatment period.[110,111]

The purpose of this study was to quantify the effectiveness of glucosamine and chondroitin in relieving OA symptoms and modifying structural progression. This study is warranted for several reasons. First, in the United States, glucosamine and chondroitin are almost always sold in a combination pill.[112] Despite the extensive research on single treatment with glucosamine or chondroitin, studies of the combined treatment are sparse.[113] Second, to our knowledge, the longest studies were three-year trials conducted more than a decade ago in Europe and supported by one pharmaceutical company.[12,13] The Osteoarthritis Initiative (OAI) provides a unique opportunity to examine the long-term effectiveness of glucosamine and chondroitin on treating OA, because it administered comprehensive measurements on treatment use and changes in knee symptoms and joint structure for up to four years.[16] Third, efficacy evidence of a treatment derived from clinical trials is often limited in generalizability because they often use strict study protocols and highly selected patients and are typically conducted at large medical centers.[114] Non-experimental studies, on the other hand, can provide clinicians and patients with a more realistic expectation for treatment benefits in real-world environments.[114] We are aware of one non-experimental study which assessed the impact of glucosamine and chondroitin on slowing structural progression.[115] Our study used different designs and analytic methods and extended their work by assessing both symptoms relief and reduction in structural progression.

Methods

The Institutional Review Boards of the University of Massachusetts Medical School and the Memorial Hospital of Rhode Island approved this study.

Data source and study sample

This study used publicly available data from the OAI (http://oai.epi-ucsf.org/). From 2004 to 2006, four study sites (i.e., Baltimore, MD; Columbus, OH; Pittsburgh, PA; and Pawtucket, RI) enrolled 4,796 residents who had established or were at high risk for developing knee OA.[16] The detailed OAI protocol can be found elsewhere.[16] Follow-up information for up to four years was used (the dataset version numbers are 0.2.2, 1.2.1, 3.2.1, 5.2.1, and 6.2.2). Inclusion/exclusion criteria are shown in Figure 4.1. We included OAI participants with radiographic knee OA at enrollment, i.e., those with a Kellgren-Lawrence (K-L) grade of 2 or greater in at least one knee (n=2,539).

To improve study validity, we used a "new-user" design,[116] for which only participants not reporting use of glucosamine or chondroitin at baseline were included in analyses (n=1,731). From this group, we then identified two samples: 1) for analysis of symptoms and 2) for the analysis of structural changes. For the analysis of symptoms, we also excluded participants with missing data on key confounders at baseline (n=44) and those missing exposure or outcome data at year 1 of the study (n=62). When analyzing structural progression, we excluded persons with following characteristics: 1) end-stage OA (i.e., K-L grade 4) or primary joint space narrowing in the lateral tibiofemoral compartment at baseline (n=150); 2) missing measures of joint space width (JSW) or JSW measures with a poor alignment of the tibial plateau at baseline (n=169); or 3) missing key confounders at baseline (n=37) or exposure or outcome at year 1 (n=262). The remaining 1,625 participants with 4,264 person-visits contributed to the analyses of symptoms, and 1,113 participants with 2,367 person-visits were included in analyses of structural changes.

Exposure definition

Use of glucosamine and chondroitin was defined based on self-reported information. At baseline and annual follow-up visits, participants were asked "During the past 6 months, did you use the following health supplements for joint pain or arthritis?" with separate questions for glucosamine and chondroitin sulfate use. We considered a participant taking glucosamine or chondroitin if he/she reported using it for at least 4 days per week, and not taking the supplement if they reported not using it or using it for less than 4 days per week. Throughout the study period, ~90% of the participants taking either one of the supplements were taking both concurrently. So at each visit we defined use of glucosamine/chondroitin as taking either of these supplements. To estimate the long-term treatment effects, we calculated the cumulative exposure by summing the number of visits when participants reported using glucosamine/chondroitin.

Assessment of OA symptoms

If both knees had radiographic OA, we used measurements from the knee with more severe pain at baseline. OA symptoms and function were assessed annually with Western Ontario and McMaster Universities Arthritis Index (WOMAC) (the Likert version 3.1). WOMAC measures three separate domains: Pain (5 items), Stiffness (2 items), and Physical Function (17 items).[117] Each scale item uses a range of 5 Likert responses, ranging from '0=none' to '4=extreme'. Responses to items in each domain were summed to produce subscale score, which ranges from 0~20 for Pain, 0~8 for Stiffness and 0~68 for Physical Function. Larger WOMAC scores represent worse symptoms or knee-related function.

Assessment of JSW

If both knees had radiographic OA, we used measures from the knee with narrower space width in the medial tibiofemural joint at baseline. Joint structural progression was measured with changes in medial JSW during follow-up from baseline. All participants at baseline and annual visits had bilateral standing knee x-rays obtained in posterior anterior projection with knees flexed to 20-30 degrees and feet internally rotated 10 degrees.[16] Longitudinal measurements of JSW from serial knee x-rays were conducted through a customized software tool, which automatically delineated the margin of the femoral condyle and the tibial plateau.[118] Multiple JSWs were measured at fixed locations along the joint. An anatomical coordinate system, which extended from the medial end (x=0) to the lateral end (x=1) in the joint space, was defined to facilitate an objective determination of measurement location. We chose the JSW measure at x=0.25 (in the medial compartment) because it was demonstrated to have best responsiveness to changes.[119]

At each assessment of JSW, the distance from tibial plateau to tibial rim closest to femoral condyle was measured to indicate knee positioning.[120] To take into account the potential error in JSW measurement due to poor knee positioning at a single visit or inconsistent positioning between visits, we did not use the JSW measures (i.e., considered them missing) if the plateau to rim distance was larger than 6.5 mm or change in this distance between visits was greater than 2 mm.[120] Among the 431 persons who were excluded at baseline (n=169) and year 1 (n=262) due to not having a valid measure of exposure or JSW (shown in Figure 4.1), 355 (82%) were excluded due to a poor or inconsistent knee positioning at JSW measurement. An additional 181 persons had inconsistent knee positioning for ≥1 assessment at year 2 to year 4. The probability of having a potentially erroneous JSW measure at following assessments was comparable among users (10.7%) and non-users (11.4%) of glucosamine/chondroitin.

Measurement of potential confounders

The following variables were considered potential confounders: sociodemographics, clinical characteristics of OA, indices of general health status, body mass index (BMI), and use of treatments other than glucosamine/chondroitin. Our previous work has shown that use of glucosamine and chondroitin was more prevalent among older adults, women, non-Hispanic Whites, individuals with higher education or higher income.[8] Income was defined as personal family income for the last year, including all sources such as wages, salaries, social security and retirement benefits.

OAI administered comprehensive measurements on participants' clinical characteristics, including knee alignment,[121] symptom-related multi-joint OA,[122] K-L grade,[122] and history of having a knee injury or surgery.[123] Knee alignment was measured with goniometer, and varus or valgus deformity was recorded if malalignment was found. We considered symptom-related multi-joint OA present if participants had OA symptoms in at least two joints other than knee.[124] Information was also collected on history of having a knee injury that limited ability to walk for at least two days, and history of having knee surgery including arthroscopy, ligament repair or meniscectomy.

The 12-item Short-Form Health Survey (SF-12) provided an assessment of general health status.[125] Answers to the 12 questions were combined to generate Physical and Mental Component Summary scores, which range from 0 to 100, with higher scores indicating better health status. BMI has been reported as a risk factor for OA progression due to its potential local biomechanical effect and systemic metabolic effect.[126] We calculated BMI from measured

height and weight [weight (kg)/height (m^2)]. Participants with a BMI less than 25 were defined as having normal weight, 25 and less than 30 as overweight, and 30 and over obese.

Trained interviewers obtained information about use of other arthritis treatments, including conventional medications and complementary and alternative medicine (CAM). At each visit, separate dummy variables were generated to indicate use of acetaminophen, non-steroidal anti-inflammatory agents (NSAIDs) and opioids in the past 30 days. Use of acetaminophen and NSAIDs included use of prescriptions and/or over-the-counter medications. Use of CAM, which was surveyed at baseline and year 2, covered therapies commonly used in the United States, including alternative medical systems, mind-body interventions, manipulation and body-based methods, energy therapies and biologically based therapies.[127]

Sociodemographics and history of a knee surgery were considered invariant and all other potential confounders were considered time-varying during the study period. For participants missing information on the time-varying variables, we imputed missing values with the last observation carried forward.[128]

Statistical analyses

We first described baseline sociodemographic and clinical characteristics of study participants by status of glucosamine/chondroitin use at year 1. When estimating the long-term effects of glucosamine/chondroitin use, we chose marginal structural models (MSMs) as the primary analytic method because we hypothesized that the data structure involved time-varying confounders that were influenced by previous treatments.[19] Figure 4.2 depicts the hypothesized relationships between glucosamine/chondroitin use, study outcomes, and potential time-varying confounders. Previously measured study outcomes and time-varying confounders may be

simultaneously confounders and intermediate variables. For instance, when studying WOMAC Pain as the outcome, the Pain score measured at the previous visit can be a potential confounder because it correlates with Pain score measured at current visit and patients with more severe pain are more likely to use glucosamine/chondroitin.[8] Furthermore, if glucosamine/chondroitin is effective in relieving pain (which is a hypothesis tested in our study), the previously measured Pain score lies on the causal path from prior treatment use and currently measured WOMAC Pain. If so, standard regression models adjusting for previous pain severity will produce biased estimates of the long-term treatment effects.[17]

MSMs rely on inverse probability weighting to adjust for time-varying confounding.[19] At each visit, we estimated the conditional probability of receiving observed treatment with glucosamine/chondroitin given baseline characteristics and time-varying confounders (including WOMAC subscale, K-L grade, SF-12 subscales, BMI, knee alignment, prior incidence of knee injury, use of analgesics and CAMs) that were measured at the same visit as use of glucosamine/chondroitin. For each specific WOMAC outcome, we adjusted for only the same previously-measured subscale as potential confounder. When analyzing JSW, we adjusted for previously measured WOMAC Pain because we found it a stronger correlate with treatment use than Stiffness and Physical Function.[8]

The inverse of the conditional probability was stabilized with the conditional probability of receiving observed treatment given baseline covariates. Conditional probabilities in numerator and denominator were estimated with logistic regression models (i.e., treatment models).[19] To take into account that associations of confounders to treatment initiation may be different from their associations to treatment continuation, we fit treatment models stratified by previous treatment status.[42] Specifically, the treatment models estimated the probability of

initiating treatment among those not using treatment at previous visit and the probability of continuing the treatment among those reporting use at previous visit.

Patients were excluded from analyses at the first occurrence of loss-to-follow-up, undergoing total knee replacement, or missing information on glucosamine/chondroitin use or outcome, whichever came first. To address the potential bias from informative dropout, we incorporated inverse-probability-of-censoring weighting in analyses.[19] At each visit from year 2 to year 4, "censoring" status was categorized as follows: 1) not censored; 2) censored due to illness/death/total knee replacement; 3) censored due to refusal to participate/loss of contact/missing exposure or outcome. Censoring weights were calculated in the same way as treatment weights, except that multinomial logistic models were used to estimate the probability of having observed censoring status and that current treatment use was added in the censoring models.[19] The final weights were the products of visit-specific treatment weights and censoring weights.[19] To ameliorate the impact of potential positivity violations, we truncated the final weights at the 99th percentile.[20]

After weights were constructed, weighted linear models (i.e., outcome models) were fit to estimate the relationships between cumulative exposure to glucosamine/chondroitin up to previous visit and changes in WOMAC scores and JSW measured at current visit.[19] In addition to the cumulative treatment use, baseline variables were also included in these outcome models.[19] We fit the outcome models using the GENMOD procedure in SAS (with an "independent" correlation structure and using "robust" standard errors).[37] Under the assumptions of no unmeasured confounding and correct specifications of the treatment and outcome models, the MSM estimates represent the causal effects of using glucosamine/chondroitin for 1, 2 and 3 years on WOMACs and JSW among the study population.[19] Previous validation studies [129-133]

suggest the minimal clinically-important improvement ranged from -4.6 to -1.2 for WOMAC Pain, -1.5 to -0.5 to for WOMAC Stiffness, -9.9 to -4.1 to for WOMAC Physical Function, and 0.2 to 0.5 mm for JSW.

We also compared the MSM estimates with the estimates derived from analyses with generalized estimating equations (GEE). In GEE analyses, we adjusted for baseline and time-varying confounders in the model and chose the working correlation structure that maximized the quasi-likelihood information criterion.[134] We hypothesized that, if there is treatment effect that is mediated by the time-varying confounders, GEE estimates would be smaller in magnitude than the MSM estimates because GEE analyses cannot correctly estimate such mediated effect.[19]

Results

Characteristics of study sample

Table 4.1 shows the baseline characteristics of the 1,625 participants included in analyses of WOMACs by status of glucosamine/chondroitin use at year 1. Overall, 43.6% were aged ≥65 years, 58.0% were women, 72.9% were non-Hispanic White and 37.8% had K-L grade 3 or 4. Ten percent of non-users at baseline initiated glucosamine/chondroitin at year 1. Compared to non-initiators of glucosamine/chondroitin at year 1, initiators tended to be younger, have higher education attainment and higher income, and were more likely to use other CAM and have a BMI ≥25 kg/m^2 and valgus deformity at baseline. Similar trends were found in the study sample for the JSW analyses.

Predictors of glucosamine/chondroitin use

Around 18% of participants initiated glucosamine/chondroitin during the study period. Among these initiators, 22.8% reported treatment use at all assessments (other than baseline) and 38.4% discontinued the treatment at a later assessment. Table 4.2 shows the correlates of initiating and continuing glucosamine/chondroitin treatment. Older adults were less likely to initiate treatment, but more likely to stay on treatment once they initiated it. Longitudinally, participants were less likely to initiate treatment (comparing year 3 and year 2 with year 1), but more likely to continue the treatment (comparing year 3 with year 2). Being overweight, having K-L grade 3/4 and using NSAIDs were correlates of treatment initiation, while use of other CAM methods and acetaminophen was associated with both initiating and continuing treatment.

When analyzing WOMAC Pain as outcome, the mean of final weights was 1.00, and the maximum value was 5.92 and 99th percentile was 1.83. Final weights for analyses of Stiffness and Function had similar distributions. The final weights for analyzing JSW had a mean of 0.99 and ranged from 0.13 to 4.92.

Effects of glucosamine/chondroitin on treating knee OA

As shown in the top section in Table 4.3, after adjustment for potential confounders with MSMs, compared to participants who never reported previous use of glucosamine/chondroitin, those reporting use for three, two and one assessments had on average 0.68 points increase (95% CI: -0.16 to 1.53), 0.12 points decrease (95% CI: -0.71 to 0.48) and 0.28 points increase (95% CI: -0.08 to 0.65) in WOMAC Pain, respectively. In terms of WOMAC Stiffness and Function, the average differences in changes from baseline between participants using the treatment at all assessments and never-users were 0.41 (95% CI: 0 to 0.82) and 1.28 (95% CI: -1.23 to 3.79), respectively. The bottom section in Table 4.3 shows the estimates of treatment effects on JSW.

After adjustment for confounders with MSMs, compared to never-users, those who reported previous use for three, two and one assessments had on average 0.11mm wider (95% CI: -0.21 to 0.44), 0.14mm wider (95% CI: -0.07 to 0.35) and 0.03mm narrower (95% CI: -0.16 to 0.10) in medial JSW, respectively.

Discussion

Following a large sample of participants with knee OA who were "naïve" to treatment with glucosamine/chondroitin, we found that around 18% initiated the treatment and 4% reported use at all assessments during the study period. Age, BMI levels, K-L grade and use of other treatments were important correlates of initiating and/or continuing glucosamine/chondroitin treatment. After adjustment for potential confounders with MSMs, we found that treatment with glucosamine/chondroitin for three years did not appear to bring about relief in symptoms or retardation of disease progression. Analyses with GEE yielded similar results as the MSM analyses.

Our data relating to symptomatic effects are consistent with recent systematic reviews[11,107] on single treatment with glucosamine or chondroitin and with independent long-term clinical trials on combination treatment with both supplements.[135,136] The recently updated Cochrane review concluded that clinical trials with adequate allocation concealment did not demonstrate a superiority of glucosamine over placebo for pain or physical function.[11] Likewise, a recent meta-analysis found that large-scale clinical trials using an intention-to-treat analysis reported minimal or nonexistent symptomatic benefits from chondroitin compared to placebo.[107] Moreover, as far as we know, there are two published long-terms trials, i.e., the Glucosamine/Chondroitin Arthritis Intervention Trial (GAIT) [135] and the Long-term Evaluation

of Glucosamine Sulfate (LEGS) study,[136] which assessed the efficacy of combination treatment with both supplements in treating knee OA. Both studies found that combination treatment did not confer symptomatic benefits compared to treatment with either supplement or with placebo.[135,136] Our data join a growing body of evidence suggesting that glucosamine/chondroitin has no impact on relieving OA symptoms.

With respect to the effect of glucosamine/chondroitin on structural progression, our findings are consistent with some,[14] but not all of the literature.[115,136] The GAIT study reported a difference of 0.028mm in joint space narrowing between the combination treatment group and the placebo group and concluded no benefits of modifying disease progression from combination treatment.[14] On the contrary, the LEGS study found that the difference in joint space narrowing was 0.10mm after two-year follow-up, which was in favor of the combination treatment and was marginally statistically significant (p=0.046). This absolute reduction in joint space narrowing was comparable to that found in our study. Considering that the smallest detectable change in JSW measures was 0.2 mm,[132] this reduction may be trivial.

Moreover, we are aware of another non-experimental study by Martel-Pelletier et al., which was based on OAI participants and assessed use of glucosamine/chondroitin on slowing OA progression.[115] Using a different study design and analytic approaches, our study confirms their finding that combination use of glucosamine and chondroitin does not have an impact on slowing joint space narrowing. However, their study reported that glucosamine and chondroitin reduced loss of cartilage volume in some subregions of the tibiofemoral joint assessed with MRI.[115] We interpret this conclusion cautiously. Over 60 comparisons were conducted to compare cartilage volume loss in different subregions of the knee joint between users and non-users of glucosamine/chondroitin, but no adjustments for multiple comparisons were made. Once

Bonferroni corrections were applied,[137] none of the comparisons would have been statistically significant.

Our study has some merits that are worth mentioning. First, we used a new-user design by excluding participants using glucosamine/chondroitin at baseline. A new-user design is considered a gold standard in pharmacoepidemiologic studies due to the well-recognized advantages of studying initiators of treatments.[73] In particular, a new user design can avoid the selection of prevalent users who are responsive to the treatment and thus prevent overestimating the treatment benefits.[73,116] In addition, a new-user design can avoid bias from adjusting for confounders that may be affected by previous treatments in prevalent users.[73,116] Second, we used MSMs to estimate causal effects by adjusting for time-varying confounders which may also be intermediate variables and by controlling for bias from potential informative dropout.[17,19] GEE models generally produce associative effects and may estimate causal effects under very stringent assumptions, including the assumption that time-varying confounders are not influenced by prior treatments.[17,19] GEE adjusts for time-varying confounders through conditioning analysis on these covariates and thus eliminates any indirect effect from prior treatments that are mediated by the time-varying confounders.[17,19] Unlike GEE, MSM adjusts for time-varying confounders through assigning weights to participants and thus is capable of estimating overall treatment effects, if they exist.[17,19]

Notwithstanding, our findings must be considered with limitations in mind. First, there may be misclassification in use of glucosamine/chondroitin. Treatment use was assessed annually, and it is likely that participants were on and off the treatment during the intervals of assessments. If this misclassification was non-differential, we would have underestimated the treatment effects. Moreover, we do not have information on treatment dosage or the extent of

purity of the supplements. The supplements evaluated in our study were likely over-the-counter products, which have been reported to be different from those tested in clinical trials in terms of quality, strength, and composition.[138] Finally, despite that OAI administered comprehensive measurement on the disease severity that might affect patients in seeking treatment and that these indices were adjusted to deal with the potential confounding by indication, we could not rule out the possibility that our findings may still be biased by unmeasured confounding.

In summary, long-term use of glucosamine/chondroitin as dietary supplements did not appear to relieve symptoms or modifying disease progression among radiographically confirmed OA patients. Our findings are consistent with the results from recent long-term clinical trials and support the latest guidelines for OA treatment which recommend against using the nutritional supplements of glucosamine and chondroitin.[103]

Table 4.1. Baseline characteristics by use of glucosamine/chondroitin at year one among persons with radiographic knee OA (n=1,625)

Characteristics	Glucosamine/ chondroitin initiators (n=165)	Non-initiators of glucosamine/ chondroitin (n=1,460)	Total
Age (years)	*Percentage*		
<65	60.0	56.0	56.4
65-74	32.7	31.6	31.7
≥75	7.3	12.4	11.9
Women	57.0	58.1	58.0
Ethnicity/Race			
Non-Hispanic White	72.1	73.0	72.9
Non-Hispanic Black	23.0	24.1	24.0
Other	4.9	3.0	3.1
Education			
High school or less	19.4	20.8	20.6
Some college	20.6	27.0	26.3
College graduate	17.0	20.8	20.4
Graduate school	43.0	31.4	32.6
Income ($)			
<25,000	17.6	17.5	17.5
25,000 - 50,000	23.0	29.5	28.9
>50,000	59.4	53.0	53.7
KL grade 3 or 4	37.6	37.8	37.8
Symptom-related multi-joint OA	49.1	49.4	49.4
Use of non-steroidal anti-inflammatory agents	42.4	35.4	36.1
Use of acetaminophen	12.1	13.9	13.7
Use of opioids	6.1	6.1	6.1
Use of complementary and alternative medicine	35.8	24.4	25.5
History of knee injury	37.0	37.7	37.6
History of knee surgery	73.3	70.8	29.0
Body Mass Index (kg/m^2)			
<25	8.5	15.7	15.0
25 - <30	42.4	37.5	38.0
≥30	49.1	46.8	47.0
Knee alignment			
Normal	24.9	26.5	26.3
Varus	24.2	28.0	27.6

Valgus	50.9	45.5	46.0
	Mean (Standard Deviation)		
WOMAC Pain	4.2 (3.9)	3.8 (4.1)	3.9 (4.1)
WOMAC Stiffness	2.2 (1.8)	2 (1.8)	2 (1.8)
WOMAC Physical Function	12 (11.8)	12 (13)	12 (12.9)
SF-12 Physical Component Score	48.1 (8.8)	47.6 (9.6)	47.7 (9.5)
SF-12 Mental Component Score	54.6 (7.9)	53.3 (8.5)	53.4 (8.4)
Medial joint space width (mm)*	5.2 (1.3)	5.2 (1.2)	5.2 (1.2)

*Based on information on 1,113 participants included in JSW analyses, among which 107 reported initiating glucosamine/chondroitin at year 1.

Table 4.2. Correlates* of glucosamine/chondroitin use in the three-year follow-up period among persons with radiographic knee OA

Correlates of treatment use	Adjusted odds ratios§ (95% CI) of "initiating" treatment	Adjusted odds ratios§ (95% CI) of "continuing" treatment
Baseline characteristics		
Age: 75 vs <65 years	0.63 (0.40-0.99)	2.31 (0.69-7.74)
Age: 65-74 vs <65 years	0.71 (0.53-0.95)	2.27 (1.17-4.40)
Women vs men	0.89 (0.68-1.17)	0.65 (0.33-1.27)
Black vs White	0.74 (0.53-1.03)	0.67 (0.34-1.32)
Other race vs White	1.02 (0.53-1.97)	0.28 (0.08-1.02)
Graduate education vs High school	1.60 (1.07-2.38)	1.39 (0.59-3.26)
College graduate vs High school	1.21 (0.78-1.86)	0.57 (0.21-1.53)
Some college vs High school	0.91 (0.61-1.36)	1.05 (0.43-2.59)
Income ($): >50 k vs <25k	0.99 (0.65-1.51)	1.60 (0.66-3.92)
Income ($): 25-50k vs <25k	0.93 (0.62-1.41)	0.79 (0.33-1.89)
History of knee surgery	0.78 (0.58-1.05)	0.42 (0.22-0.83)
Time-varying confounders (concurrent)		
Year 3	(*vs Year 1*) 0.44 (0.32-0.60)	(*vs Year 2*) 1.71 (1.01-2.88)
Year 2	(*vs Year 1*) 0.51 (0.39-0.69)	--
Obese vs Normal weight	1.20 (0.82-1.76)	0.85 (0.37-1.93)
Overweight vs Normal weight	1.49 (1.02-2.16)	1.21 (0.55-2.69)
Alignment: Valgus vs Normal	1.02 (0.74-1.40)	0.58 (0.27-1.25)
Alignment: Varus vs Normal	0.86 (0.61-1.21)	0.70 (0.31-1.59)
K-L: 3/4 vs 2	1.37 (1.06-1.78)	1.07 (0.59-1.93)
Multi-joint osteoarthritis	1.15 (0.89-1.50)	0.94 (0.53-1.67)
History of knee injury	0.99 (0.76-1.30)	1.69 (0.92-3.10)
Use of non-steroidal anti-inflammatory agents	1.46 (1.12-1.90)	1.06 (0.57-1.94)
Use of acetaminophen	1.45 (1.00-2.11)	1.42 (0.58-3.50)
Use of opioids	0.67 (0.40-1.13)	0.61 (0.21-1.75)
Use of complementary/alternative medicine	2.20 (1.69-2.89)	2.90 (1.64-5.12)
WOMAC Pain #	1.13 (0.98-1.32)	1.23 (0.86-1.75)
SF-12 Physical Component Score #	1.20 (1.02-1.41)	1.32 (0.94-1.84)
SF-12 Mental Component Score #	1.15 (1.00-1.32)	0.99 (0.75-1.30)

* Correlates in this table were included in treatment models when analyzing WOMAC Pain as the outcome.

§ Adjusted for other variables in this table.

Odds ratios are per one standard deviation changes in WOMAC Pain or SF-12 subscales.

Table 4.3. Estimated effects of glucosamine/chondroitin on treating OA among persons with radiographic knee OA, beta coefficients (95% CI)[*]

Models	Cumulative exposure to glucosamine/chondroitin[#]		
	3	**2**	**1**
WOMAC Pain (Minimally important improvement: -4.6 to -1.2)			
GEE: Crude [§]	0.86 (0.10 to 1.61)	0.14 (-0.33 to 0.60)	0 (-0.36 to 0.37)
GEE: Full-adjusted [§]	0.81 (0.16 to 1.45)	0.07 (-0.30 to 0.45)	0.20 (-0.04 to 0.44)
MSM with truncated weights	0.68 (-0.16 to 1.53)	-0.12 (-0.71 to 0.48)	0.28 (-0.08 to 0.65)
WOMAC Stiffness (Minimally important improvement: -1.5 to -0.5)			
GEE: Crude [§]	0.48 (0.08 to 0.89)	0.09 (-0.17 to 0.34)	0.14 (-0.04 to 0.31)
GEE: Full-adjusted [§]	0.41 (0.04 to 0.79)	0.13 (-0.08 to 0.34)	0.17 (0.05 to 0.30)
MSM with truncated weights	0.41 (0 to 0.82)	0.10 (-0.18 to 0.37)	0.25 (0.06 to 0.43)
WOMAC Function (Minimally important improvement: -9.9 to -4.1)			
GEE: Crude [§]	2.56 (0.64 to 4.48)	1.23 (-0.17 to 2.64)	-0.06 (-1.15 to 1.02)
GEE: Full-adjusted [§]	1.74 (0.03 to 3.46)	0.94 (-0.19 to 2.07)	0.31 (-0.41 to 1.03)
MSM with truncated weights	1.28 (-1.23 to 3.79)	0.24 (-1.45 to 1.94)	0.66 (-0.50 to 1.82)
Joint space width (Minimally important improvement: 0.2 to 0.5)			
GEE: Crude [§]	-0.35 (-0.58 to -0.12)	-0.25 (-0.45 to -0.06)	-0.12 (-0.23 to -0.01)
GEE: Full-adjusted [§]	0.05 (-0.13 to 0.22)	0.04 (-0.08 to 0.15)	-0.03 (-0.09 to 0.03)
MSM with truncated weights	0.11 (-0.21 to 0.44)	0.14 (-0.07 to 0.35)	-0.03 (-0.16 to 0.10)

[*] The reference group includes persons never using glucosamine/chondroitin up to "previous visit".

[§] Generalized estimating equations (GEE) analyses assumed an unstructured correlation matrix. The full-adjusted GEE estimates adjusted for baseline characteristics and time-varying confounders that were measured at the same visit as glucosamine/chondroitin use.

[#] Analyses of WOMAC outcomes and JSW were based on 1,625 persons (4,264 person-visits) and 1,113 persons (2,367 person-visits), respectively.

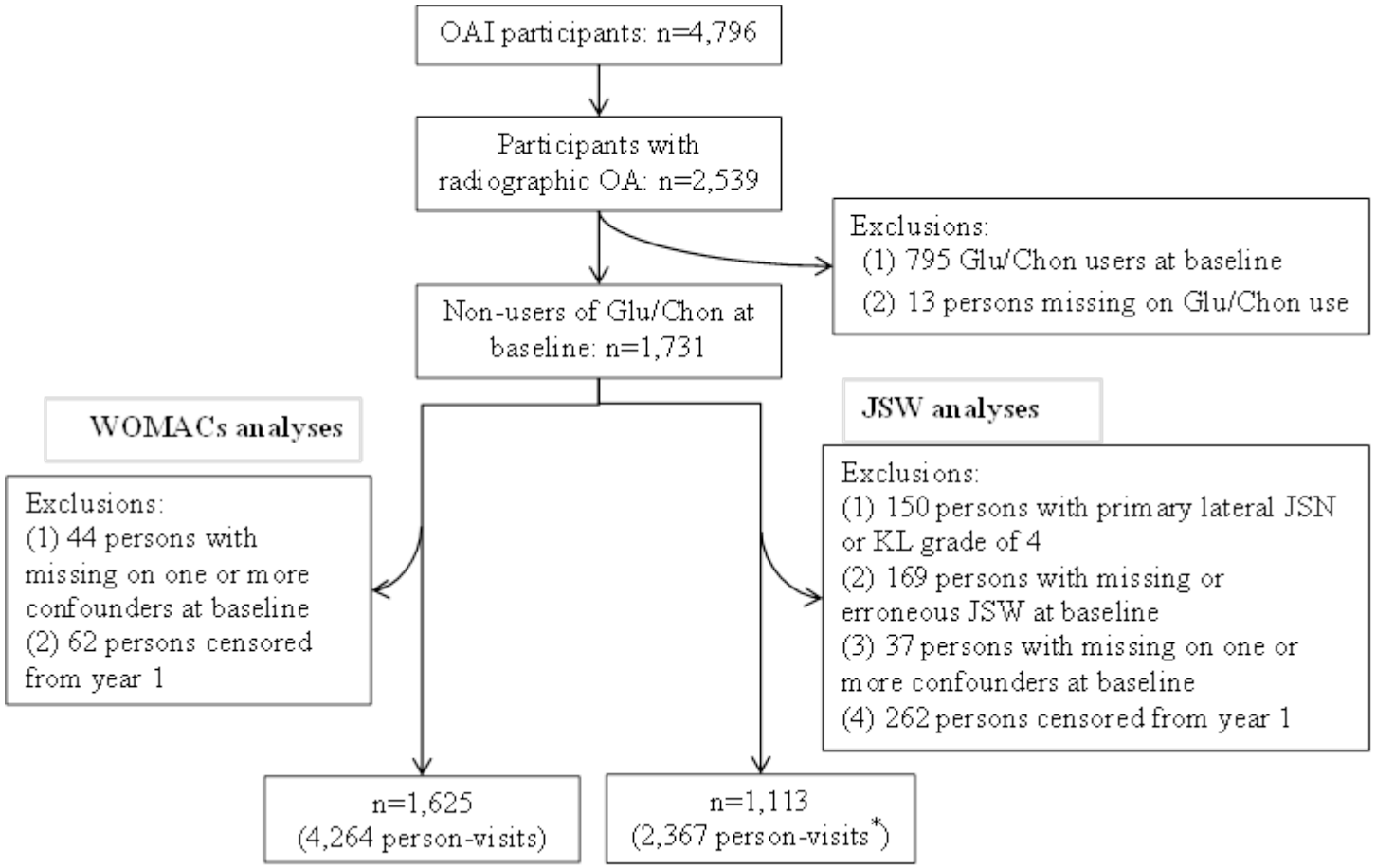

Figure 4.1. Flow-chart of identifying study samples.

Glu/Chon: glucosamine/chondroitin; JSN: joint space narrowing.
**295 persons were further censored at visits from year 2 to year 4 because the JSW measures were missing or invalid due to poor knee positioning.*

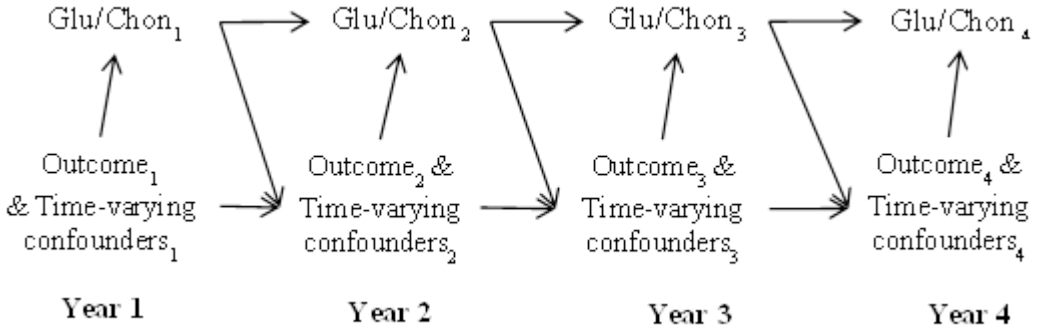

Figure 4.2. Hypothesized causal relationships between glucosamine/chondroitin treatment, study outcomes and potential time-varying confounders.

Glu/Chon denotes treatment with glucosamine/chondroitin and the subscript number denotes the follow-up time (year) when the information was measured.

List of References

List of References

1. Lawrence R, Felson D, Helmick C, Arnold L, Choi H, Deyo R, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part II. *Arthritis Rheumatism* 2008;58(1):26-35.

2. Chisholm-Burns MA, Schwinghammer TL, Wells BG, Malone PM, Kolesar JM, Rotschafer JC,Dipiro JT. *Pharmacotherapy principles and practice, 2rst ed.* McGraw Hill Medical; 2010.

3. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;16(4):494-502.

4. Guccione AA, Felson DT, Anderson JJ, Anthony JM, Zhang Y, Wilson PW, et al. The effects of specific medical conditions on the functional limitations of elders in the Framingham Study. *Am J Public Health* 1994;84(3):351-8.

5. Ettinger WH, Davis MA, Neuhaus JM, Mallon KP. Long-term physical functioning in persons with knee osteoarthritis from NHANES. I: Effects of comorbid medical conditions. *J Clin Epidemiol* 1994;47(7):809-15.

6. Zhang Y, Jordan J. Epidemiology of osteoarthritis. *Clin Geriatr Med* 2010;26(3):355-69.

7. Smith SM, Epstein BJ, Gums JG. Osteoarthritis. In: Chisholm-Burns MA, Schwinghammer TL, Wells BG, Malone PM, Kolesar JM, DiPiro JT. *Pharmacotherapy Principles & Practice*. 2 nd ed. New York, NY: McGraw-Hill; 2010: 997-1009.

8. Lapane KL, Sands MR, Yang S, McAlindon TE, Eaton CB. Use of complementary and alternative medicine among patients with radiographic-confirmed knee osteoarthritis. *Osteoarthritis Cartilage* 2012;20(1):22-8.

9. Poolsup N, Suthisisang C, Channark P, Kittikulsuth W. Glucosamine long-term treatment and the progression of knee osteoarthritis: systematic review of randomized controlled trials. *Ann Pharmacother* 2005;39(6):1080-7.

10. Hochberg MC, Perlmutter DL, Hudson JI, Altman RD. Preferences in the management of osteoarthritis of the hip and knee: results of a survey of community-based rheumatologists in the United States. *Arthritis Care Res (Hoboken)* 1996;9(3):170-6.

11. Towheed TE, Maxwell L, Anastassiades TP, Shea B, Houpt J, Robinson V, et al. Glucosamine therapy for treating osteoarthritis. *Cochrane Database Syst Rev* 2005;(2):CD002946.

12. Reginster JY, Deroisy R, Rovati LC, Lee RL, Lejeune E, Bruyere O, et al. Long-term effects of glucosamine sulphate on osteoarthritis progression: a randomised, placebo-controlled clinical trial. *Lancet* 2001;357(9252):251-6.

13. Pavelk K, Gatterov J, Olejarov M, Machacek S, Giacovelli G, Rovati L. Glucosamine sulfate use and delay of progression of knee osteoarthritis: a 3-year, randomized, placebo-controlled, double-blind study. *Arch Intern Med* 2002;162(18):2113-23.

14. Sawitzke A, Shi H, Finco M, Dunlop D, Bingham C, Harris C, et al. The effect of glucosamine and/or chondroitin sulfate on the progression of knee osteoarthritis: a report from the glucosamine/chondroitin arthritis intervention trial. *Arthritis Rheumatism* 2008;58(10):3183.

15. Rozendaal R, Koes B, van Osch,Gerjo J V M., Uitterlinden E, Garling E, Willemsen S, et al. Effect of glucosamine sulfate on hip osteoarthritis: a randomized trial. *Ann Intern Med* 2008;148(4):268-77.

16. The Osteoarthritis Initiative protocol for the cohort study. Available at: http://oai.epi-ucsf.org/datarelease/docs/StudyDesignProtocol.pdf.

17. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15(5):615-25.

18. Robins JM. Marginal structural models versus structural nested models as tools for causal inference. In: Halloran ME, Berry D, editors. *Statistical Models in Epidemiology: The Environment and Clinical Trials.* New York: Springer-Verlag; 1999. p. 95-134.

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

20. Cole S, Hernán MA. Constructing inverse probability weights for marginal structural models. *Am J Epidemiol* 2008;168(6):656-64.

21. Hernán MA, Robins J. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006;60(7):578-86.

22. VanderWeele T, Hawkley L, Thisted R, Cacioppo J. A marginal structural model analysis for loneliness: implications for intervention trials and clinical practice. *J Consult Clin Psychol* 2011;79(2):225-35.

23. Brumback B, Hernán MA, Haneuse SJPA, Robins J. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. *Stat Med* 2004;23(5):749-67.

24. Westreich D, Cole S. Invited commentary: positivity in practice. *Am J Epidemiol* 2010;171(6):674-7.

25. Petersen M, Porter K, Gruber S, Wang Y, van der Laan MJ. Diagnosing and responding to violations in the positivity assumption. *Stat Methods Med Res* 2012;21(1):31-54.

26. Hernán MA, Taubman SL. Does obesity shorten life? The importance of well-defined interventions to answer causal questions. *Int J Obes* 2008;32 Suppl 3:S8-14.

27. Cole S, Frangakis C. The consistency statement in causal inference: a definition or an assumption?. *Epidemiology* 2009;20(1):3-5.

28. Hernán MA. Invited commentary: hypothetical interventions to define causal effects--afterthought or prerequisite?. *Am J Epidemiol* 2005;162(7):618-20.

29. Holland PW. Statistics and Causal Inference. *Journal of the American Statistical Association* 1986;81(396):945.

30. Platt RW, Schisterman EF, Cole SR. Time-modified Confounding. *Am J Epidemiol* 2009;170(6):687-94.

31. Robins JM. Association, Causation, and Marginal Structural Models. *Synthese* 1999;121(1/2, Statistics and Causation):151-79.

32. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000; 56(3):779-88.

33. Lefebvre G, Delaney JAC, Platt R. Impact of mis-specification of the treatment model on estimates from a marginal structural model. *Stat Med* 2008;27(18):3629-42.

34. Mortimer K, Neugebauer R, van der Laan M, Tager I. An application of model-fitting procedures for marginal structural models. *Am J Epidemiol* 2005;162(4):382-8.

35. Moodie EEM. Risk factor adjustment in marginal structural model estimation of optimal treatment regimes. *Biometrical Journal* 2009;51(5):774-88.

36. Suarez D, Borràs R, Basagaña X. Differences between marginal structural models and conventional models in their exposure effect estimates: a systematic review. *Epidemiology* 2011;22(4):586-8.

37. Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology* 2000;11(5):561-70.

38. Moher D, Liberati A, Tetzlaff J, Altman D. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ.British Medical Journal* 2009;339:b2535-.

39. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology* 2003;14(3):300-6.

40. Danaei G, Rodriguez LA, Cantero OF, Logan R, Hernán MA. Observational data for comparative effectiveness research: An emulation of randomised trials of statins and primary prevention of coronary heart disease. *Stat Methods Med Res* 2013;22(1):70-96.

41. Toh S, Hernán MA. Causal inference from longitudinal studies with baseline randomization. *Int J Biostat* 2008;4(1):22.

42. Platt R, Brookhart MA, Cole S, Westreich D, Schisterman E. An information criterion for marginal structural models. *Stat Med* 2012;32:1383-93.

43. Lee TA, Pickard AS. Exposure definition and measurement. In: Velentgas P, Dreyer NA, Nourjah P, editors. *Developing a protocol for Observational Comparative effectiveness research: A User's Guide* Rockville, MD: Agency for Healthcare Research and Quality; 2013. p. 45-57.

44. Neugebauer R, Fireman B, Roy JA, O'Connor PJ, Selby JV. Dynamic marginal structural modeling to evaluate the comparative effectiveness of more or less aggressive treatment intensification strategies in adults with type 2 diabetes. *Pharmacoepidemiol Drug Saf* 2012; 21:99-113.

45. Shortreed SM, Moodie EEM. Estimating the optimal dynamic antipsychotic treatment regime: evidence from the sequential multiple-assignment randomized Clinical Antipsychotic Trials of Intervention and Effectiveness schizophrenia study. *Journal of the Royal Statistical Society Series C-Applied Statistics* 2012; 61:577-99.

46. Rosthoj S, Keiding N, Schmiegelow K. Estimation of dynamic treatment strategies for maintenance therapy of children with acute lymphoblastic leukaemia: an application of history-adjusted marginal structural models. *Stat Med* 2012; 31(5):470-88.

47. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernán MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *International Journal of Biostatistics* 2010; 6(2):18.

48. Cook NR, Cole SR, Buring JE. Aspirin in the primary prevention of cardiovascular disease in the Women's Health Study: Effect of noncompliance. *Eur J Epidemiol* 2012; 27(6):431-8.

49. Desai RJ, Ashton CM, Deswal A, Morgan RO, Mehta HB, Chen H, et al. Comparative effectiveness of individual angiotensin receptor blockers on risk of mortality in patients with chronic heart failure. *Pharmacoepidemiol Drug Saf* 2012; 21(3):233-40.

50. Gerhard T, Delaney JA, Cooper-Dehoff RM, Shuster J, Brumback BA, Johnson JA, et al. Comparing marginal structural models to standard methods for estimating treatment effects of antihypertensive combination therapy. *BMC Med Res Methodol* 2012; 12:119.

51. Gsponer T, Petersen M, Egger M, Phiri S, Maathuis MH, Boulle A, et al. The causal effect of switching to second-line ART in programmes without access to routine viral load monitoring. *AIDS* 2012; 26(1):57-65.

52. Haukka J, Niskanen L, Partonen T, Lonnqvist J, Tiihonen J. Statin usage and all-cause and disease-specific mortality in a nationwide study. *Pharmacoepidemiol Drug Saf* 2012;21(1):61-9.

53. HCV working group of the Collaboration of Observational HIV Epidemiological Research in Europe (COHERE) in EuroCoord. Effect of hepatitis C treatment on CD4+ T-cell counts and the risk of death in HIV-HCV-coinfected patients: the COHERE collaboration. *Antivir Ther* 2012;17(8):1541-50.

54. Heffron R, Donnell D, Rees H, Celum C, Mugo N, Were E, et al. Use of hormonal contraceptives and risk of HIV-1 transmission: a prospective cohort study. *Lancet Infectious Diseases* 2012; 12(1):19-26.

55. Hernandez D, Muriel A, Abraira C, Perez G, Porrini E, Marrero D, et al. Renin-angiotensin system blockade and kidney transplantation: a longitudinal cohort study. *Nephrology Dialysis Transplantation* 2012; 27(1):417-22.

56. HIV-CAUSAL Collaboration. The effect of efavirenz versus nevirapine-containing regimens on immunologic, virologic and clinical outcomes in a prospective observational study. *AIDS* 2012; 26(13):1691-705.

57. del Amo J, Moreno S, Bucher HC, Furrer H, Logan R, Sterne J, et al. Impact of Antiretroviral Therapy on Tuberculosis Incidence Among HIV-Positive Patients in High-Income Countries. *Clinical Infectious Diseases* 2012; 54(9):1364-72.

58. Jin H, Tu D, Zhao N, Shepherd LE, Goss PE. Longer-term outcomes of letrozole versus placebo after 5 years of tamoxifen in the NCIC CTG MA.17 trial: analyses adjusting for treatment crossover. *J Clin Oncol* 2012; 30(7):718-21.

59. Kalayjian RC, Lau B, Mechekano RN, Crane HM, Rodriguez B, Salata RA, et al. Risk factors for chronic kidney disease in a large cohort of HIV-1 infected individuals initiating antiretroviral therapy in routine care. *AIDS* 2012; 26(15):1907-15.

60. McCoy SI, Zheng W, Montgomery E, Blanchard K, van Der Straten A, de Bruyn G, et al. Oral and injectable contraception use and risk of HIV acquisition among women in the methods for improving reproductive health in Africa (MIRA) study. *AIDS* 2012; 27:1001-1009.

61. Miller JE, Molnar MZ, Kovesdy CP, Zaritsky JJ, Streja E, Salusky I, et al. Administered paricalcitol dose and survival in hemodialysis patients: a marginal structural model analysis. *Pharmacoepidemiol Drug Saf* 2012; 21(11):1232-9.

62. Morrison CS, Skoler-Karpoff S, Kwok C, Chen P, van de Wijgert J, Gehret-Plagianos M, et al. Hormonal contraception and the risk of HIV acquisition among women in South Africa. *AIDS* 2012; 26(4):497-504.

63. Scherzer R, Estrella M, Li Y, Choi AI, Deeks SG, Grunfeld C, et al. Association of tenofovir exposure with kidney disease risk in HIV infection. *AIDS* 2012; 26(7):867-75.

64. Shinozaki T, Matsuyama Y, Iimuro S, Umegaki H, Sakurai T, Araki A, et al. Effective prevention of cardiovascular disease and diabetes-related events with atorvastatin in Japanese elderly patients with type 2 diabetes mellitus: adjusting for treatment changes using a marginal structural proportional hazards model and a rank-preserving structural failure time model. *Geriatr Gerontol Int* 2012; 12 Suppl 1:88-102.

65. Terrier B, Krastinova E, Marie I, Launay D, Lacraz A, Belenotti P, et al. Management of noninfectious mixed cryoglobulinemia vasculitis: data from 242 cases included in the CryoVas survey. *Blood* 2012; 119(25):5996-6004.

66. Tiihonen J, Suokas JT, Suvisaari JM, Haukka J, Korhonen P. Polypharmacy With Antipsychotics, Antidepressants, or Benzodiazepines and Mortality in Schizophrenia. *Arch Gen Psychiatry* 2012; 69(5):476-83.

67. Young J, Schaefer J, Fux CA, Furrer H, Bernasconi E, Vernazza P, et al. Renal function in patients with HIV starting therapy with tenofovir and either efavirenz, lopinavir or atazanavir. *AIDS* 2012; 26(5):567-75.

68. Barber J. Adjusting for Time-Varying Confounding in Survival Analysis. *Sociological Methodology* 2004;34:163-92.

69. Platt R, Delaney JAC, Suissa S. The positivity assumption and marginal structural models: the example of warfarin use and risk of bleeding. *Eur J Epidemiol* 2012;27(2):77-83.

70. Dumas O, Le Moual N, Siroux V, Heederik D, Garcia-Aymerich J, Varraso R, et al. Work related asthma. A causal analysis controlling the healthy worker effect. *Occup Environ Med* 2013; 70(9):603-610.

71. Chiba Y. Sensitivity Analysis of Unmeasured Confounding for the Causal Risk Ratio by Applying Marginal Structural Models. *Communications in Statistics-Theory and Methods* 2010; 39(1):65-76.

72. Cain LE, Cole SR. Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident AIDS or death. *Stat Med* 2009; 28(12):1725-38.

73. Setoguchi S, Gerhard T. Comparator selection. In: Velentgas P, Dreyer NA, Nourjah P, editors. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Rockville, MD: Agency for Healthcare Research and Quality; 2013. p. 59-70.

74. Brookhart MA, Schneeweiss S, Rothman K, Glynn R, Avorn J, Stürmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006;163(12):1149-56.

75. Lipkovich I, Mallinckrodt C, Faries D. The challenges of evaluating dose response in flexible-dose trials using marginal structural models. *Pharmaceutical Statistics* 2012;11(6):485-93.

76. Xiao Y, Abrahamowicz M, Moodie EEM. Accuracy of Conventional and Marginal Structural Cox Model Estimators: A Simulation Study. *International Journal of Biostatistics* 2010; 6(2):13-.

77. Bryan J, Yu Z, Van Der Laan MJ. Analysis of longitudinal marginal structural models. *Biostatistics* 2004; 5(3):361-80.

78. Stuermer T, Rothman KJ, Avorn J, Glynn RJ. Treatment Effects in the Presence of Unmeasured Confounding: Dealing With Observations in the Tails of the Propensity Score Distribution-A Simulation Study. *Am J Epidemiol* 2010; 172(7):843-54.

79. Kurth T, Walker AM, Glynn RJ, Chan KA, Gaziano JM, Berger K, et al. Results of multivariable logistic regression, propensity matching, propensity adjustment, and propensity-based weighting under conditions of nonuniform effect. *Am J Epidemiol* 2006; 163(3):262-70.

80. Greenland S, Maldonado G. The interpretation of multiplicative-model parameters as standardized parameters. *Stat Med* 1994; 13(10):989-99.

81. Sturmer T, Rothman KJ, Glynn RJ. Insights into different results from different causal contrasts in the presence of effect-measure modification. *Pharmacoepidemiol Drug Saf* 2006; 15(10):698-709.

82. Yang S, Eaton CB, Lu J, Lapane KL. Application of marginal structural models in pharmacoepidemiologic studies: a systematic review. *Pharmacoepidemiol Drug Saf* 2014; 23(6):560-71.

83. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med* 2005;353(5):487-97.

84. Kripalani S, Yao X, Haynes RB. Interventions to enhance medication adherence in chronic medical conditions: a systematic review. *Arch Intern Med* 2007;167(6):540-50.

85. Hernán MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012;9(1):48-55.

86. Yang S, Eaton CB, McAlindon TE, Lapane KL. Long-term effects of glucosamine and chondroitin on treating knee osteoarthritis: an analysis with marginal structural models. *Arthritis Rheum.* [Revise and Resubmit].

87. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the third report of the National Cholesterol Education Program (NECP) expert panel on detection, evaluation, and treatment of high blood cholesterol in adults (Adult Treatment Panel III). *JAMA* 2001; 285(19):2486-97.

88. Wright JM, Wright JM. Lipid lowering efficacy of atorvastatin. Cochrane Database Syst Rev 2012; 12:CD008226.

89. Centers for Disease Control and Prevention (CDC). Vital signs: prevalence, treatment, and control of high levels of low-density lipoprotein cholesterol--United States, 1999-2002 and 2005-200. *MMWR Morb Mortal Wkly Rep* 2011; 60(4):109-14.

90. Caspard H, Chan AK, Walker AM. Compliance with a statin treatment in a usual-care setting: retrospective database analysis over 3 years after treatment initiation in health maintenance organization enrollees with dyslipidemia. *Clin Ther* 2005; 27(10):1639-46.

91. Simons LA, Ortiz M, Calcino G. Long term persistence with statin therapy -- experience in Australia 2006-2010. *Aust Fam Physician* 2011; 40(5):319-22.

92. Simons LA, Levis G, Simons J. Apparent discontinuation rates in patients prescribed lipid-lowering drugs. *Med J Aust* 1996; 164(4):208-11.

93. Benner JS, Pollack MF, Smith TW, Bullano MF, Willey VJ, Williams SA. Association between short-term effectiveness of statins and long-term adherence to lipid-lowering therapy. *Am J Health Syst Pharm* 2005; 62(14):1468-75.

94. Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. *Stat Med* 2006;25(24):4279-92.

95. Ballantyne CM, Blazing MA, King TR, Brady WE, Palmisano J. Efficacy and safety of ezetimibe co-administered with simvastatin compared with atorvastatin in adults with hypercholesterolemia. *Am J Cardiol* 2004; 93(12):1487-94.

96. Ballantyne CM, Andrews TC, Hsia JA, Kramer JH, Shear C. Correlation of non-high-density lipoprotein cholesterol with apolipoprotein B: effect of 5 hydroxymethylglutaryl coenzyme A reductase inhibitors on non-high-density lipoprotein cholesterol levels. *Am J Cardiol* 2001; 88(3):265-9.

97. Faergeman O, Hill L, Windler E, Wiklund O, Asmar R, Duffield E, et al. Efficacy and tolerability of rosuvastatin and atorvastatin when force-titrated in patients with primary hypercholesterolemia: results from the ECLIPSE study. *Cardiology* 2008;111(4):219-28.

98. Pearson TA, Laurora I, Chu H, Kafonek,S. The lipid treatment assessment project (l-tap): A multicenter survey to evaluate the percentages of dyslipidemic patients receiving lipid-lowering therapy and achieving low-density lipoprotein cholesterol goals. *Arch Intern Med* 2000; 160(4):459-67.

99. Cubeddu LX, Cubeddu RJ, Heimowitz T, Restrepo B, Lamas GA, Weinberg GB. Comparative lipid-lowering effects of policosanol and atorvastatin: A randomized, parallel, double-blind, placebo-controlled trial. *Am Heart J* 2006; 152(5):982.

100. Chen H, Ren J, Xing Y, Zhang W, Liu X, Wu P, et al. Short-term withdrawal of simvastatin induces endothelial dysfunction in patients with coronary artery disease: A dose–response effect dependent on endothelial nitric oxide synthase. *Int J Cardiol* 2009; 131(3):313-20.

101. van der Harst P, Asselbergs FW, Hillege HL, Bakker SJL, Voors AA, van Veldhuisen DJ, et al. Effect of Withdrawal of Pravastatin Therapy on C-Reactive Protein and Low-Density Lipoprotein Cholesterol. *Am J Cardiol* 2007; 100(10):1548-51.

102. Kutner MH, Nachtsheim CJ, Neter J,Li W. *Applied Linear Statistical Models* 5th Edition. McGraw-Hill/Irwin;2005.

103. Hochberg MC, Altman RD, April KT, Benkhalti M, Guyatt G, McGowan J, et al. American College of Rheumatology 2012 recommendations for the use of nonpharmacologic and pharmacologic therapies in osteoarthritis of the hand, hip, and knee. *Arthritis Care Res (Hoboken)* 2012; 64(4):465-74.

104. McCarty MF. The neglect of glucosamine as a treatment for osteoarthritis--a personal perspective. *Med Hypotheses* 1994; 42(5):323-7.

105. Sherman A, Ojeda Correal G, Mena J. Use of glucosamine and chondroitin in persons with osteoarthritis. *PM R* 2012;4(5 Suppl):S110-6.

106. Huskisson EC. Glucosamine and chondroitin for osteoarthritis. *J Int Med Res* 2008;36(6):1161-79.

107. Reichenbach S, Sterchi R, Scherer M, Trelle S, Brgi E, Brgi U, et al. Meta-analysis: chondroitin for osteoarthritis of the knee or hip. *Ann Intern Med* 2007;146(8):580.

108. McAlindon TE, LaValley MP, Gulin JP, Felson DT. Glucosamine and chondroitin for treatment of osteoarthritis: a systematic quality assessment and meta-analysis. *JAMA* 2000; 283(11):1469-75.

109. Vlad SC, LaValley MP, McAlindon TE, Felson DT. Glucosamine for pain in osteoarthritis: why do trial results differ? *Arthritis Rheum* 2007; 56(7):2267-77.

110. Lee YH, Woo JH, Choi SJ, Ji JD, Song GG. Effect of glucosamine or chondroitin sulfate on the osteoarthritis progression: a meta-analysis. *Rheumatol Int* 2010; 30(3):357-63.

111. Hochberg MC. Structure-modifying effects of chondroitin sulfate in knee osteoarthritis: an updated meta-analysis of randomized placebo-controlled trials of 2-year duration. *Osteoarthritis Cartilage* 2010; Suppl 1:S28-31.

112. Felson DT. Chondroitin for pain in osteoarthritis. *Ann Intern Med* 2007; 146(8):611-2.

113. Wandel S, Jüni P, Tendal B, Nüesch E, Villiger P, Welton N, et al. Effects of glucosamine, chondroitin, or placebo in patients with osteoarthritis of hip or knee: network meta-analysis. *BMJ* 2010;16:c4675.

114. Nallamothu BK, Hayward RA, Bates ER. Beyond the randomized clinical trial: the role of effectiveness studies in evaluating cardiovascular therapies. *Circulation* 2008; 118(12):1294-303.

115. Martel-Pelletier J, Roubille C, Abram F, Hochberg MC, Dorais M, Delorme P, et al. First-line analysis of the effects of treatment on progression of structural changes in knee osteoarthritis over 24 months: data from the osteoarthritis initiative progression cohort. *Ann Rheum Dis* 2013; doi: 10.1136/annrheumdis-2013-203906. [Epub ahead of print]

116. Ray W. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* 2003;158(9):915-20.

117. Roos EM, Klässbo M, Lohmander LS. WOMAC osteoarthritis index. Reliability, validity, and responsiveness in patients with arthroscopically assessed osteoarthritis. Western Ontario and MacMaster Universities. *Scand J Rheumatol* 1999;28(4):210-5.

118. Duryea J, Li J, Peterfy CG, Gordon C, Genant HK. Trainable rule-based algorithm for the measurement of joint space width in digital radiographic images of the knee. *Med Phys* 2000;27(3):580.

119. Duryea J, Neumann G, Niu J, Totterman S, Tamez J, Dabrowski C, et al. Comparison of radiographic joint space width with magnetic resonance imaging cartilage morphometry: analysis of longitudinal data from the Osteoarthritis Initiative. *Arthritis Care Research* 2010;62(7):932-7.

120. Central assessment of longitudinal knee X-rays for quantitative JSW. Available at: https://oai.epi-ucsf.org/datarelease/SASDocs/kXR_QJSW_Duryea_descrip.pdf.

121. Felson D, Niu J, Gross KD, Englund M, Sharma L, Cooke TDV, et al. Valgus malalignment is a risk factor for lateral knee osteoarthritis incidence and progression: Findings from MOST and the Osteoarthritis Initiative. *Arthritis Rheumatism* 2012; 65(2):355-62.

122. Chapple C, Nicholson H, Baxter GD, Abbott JH. Patient characteristics that predict progression of knee osteoarthritis: a systematic review of prognostic studies. *Arthritis Care Research* 2011;63(8):1115-25.

123. Cooper C, Snow S, McAlindon TE, Kellingray S, Stuart B, Coggon D, et al. Risk factors for the incidence and progression of radiographic knee osteoarthritis. *Arthritis Rheumatism* 2000;43(5):995-1000.

124. Okma Keulen P, Hopman Rock M. The onset of generalized osteoarthritis in older women: a qualitative approach. *Arthritis Rheumatism* 2001;45(2):183-90.

125. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care* 1996;34(3):220-33.

126. Reijman M, Pols HAP, Bergink AP, Hazes JMW, Belo JN, Lievense AM, et al. Body mass index associated with onset and progression of osteoarthritis of the knee but not of the hip: the Rotterdam Study. *Ann Rheum Dis* 2007;66(2):158-62.

127. Barnes P, Bloom B, Nahin R. Complementary and alternative medicine use among adults and children: United States, 2007. *National Health Statistics Reports (NHSR)* 2008;(12):1-23.

128. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. *Am J Epidemiol* 1995; 142(12):1255-64.

129. Greco NJ, Anderson AF, Mann BJ, Cole BJ, Farr J, Nissen CW, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. *Am J Sports Med* 2010; 38(5):891-902.

130. Angst F, Aeschlimann A, Stucki G. Smallest detectable and minimal clinically important differences of rehabilitation intervention with their implications for required sample sizes using WOMAC and SF-36 quality of life measurement instruments in patients with osteoarthritis of the lower extremities. *Arthritis Rheum* 2001; 45(4):384-91.

131. Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthritis Cartilage* 2007;15(3):273-80.

132. Hunter DJ. Risk stratification for knee osteoarthritis progression: a narrative review. *Osteoarthritis Cartilage* 2009;17(11):1402.

133. Vignon E, Piperno M, Le Graverand MP, Mazzuca SA, Brandt KD, Mathieu P, et al. Measurement of radiographic joint space width in the tibiofemoral compartment of the osteoarthritic knee: comparison of standing anteroposterior and Lyon schuss views. *Arthritis Rheum* 2003; 48(2):378-84.

134. Pan W. Akaike's information criterion in generalized estimating equations. *Biometrics* 2001; 57(1):120-5.

135. Sawitzke A, Shi H, Finco M, Dunlop D, Harris C, Singer N, et al. Clinical efficacy and safety of glucosamine, chondroitin sulphate, their combination, celecoxib or placebo taken to treat osteoarthritis of the knee: 2-year results from GAIT. *Ann Rheum Dis* 2010;69(8):1459-64.

136. Fransen M, Agaliotis M, Nairn L, Votrubec M, Bridgett L, Su S, et al. Glucosamine and chondroitin for knee osteoarthritis: a double-blind randomised placebo-controlled clinical trial evaluating single and combination regimens. *Ann Rheum Dis* 2014; doi: 10.1136/annrheumdis-2013-203954. [Epub ahead of print]

137. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995; 310(6973):170.

138. Russell AS, Aghazadeh-Habashi A, Jamali F. Active ingredient consistency of commercially available glucosamine sulfate products. *J Rheumatol* 2002; 29(11):2407-9.

Vita

Shibing Yang was born on September 26, 1984, in Jiangsu, China. He received his Bachelor of Medicine in Preventive Medicine from Capital Medical University, Beijing, China in 2007, and received a Master of Science in Social Medicine and Healthcare Management from Capital Medical University in 2010.

Honors and Awards

2013, Member of the Honor Society of Phi Kappa Phi
2012, C.C. Clayton Award, Virginia Commonwealth University
2011, ICPE Scholarship, International Society for Pharmacoepidemiology
2011, Graduate Student Travel Grant, Virginia Commonwealth University
2010, First Place Scholarship for Graduate Student, Capital Medical University, China
2007, Honor of Outstanding Graduate, Beijing Municipal Education Committee, China
2003, First Place National Scholarship, Ministry of Education, China

Professional and Research Positions

2010-14 Graduate Research Assistantship, Virginia Commonwealth University, Division of Epidemiology, Department of Family Medicine and Population Health
Mentor: Kate Lapane, Ph.D.

2012 Teaching Assistant, Virginia Commonwealth University, Division of Epidemiology, Department of Family Medicine and Population Health
Professor: Derek Chapman, Ph.D.

2010 Intern at World Health Organization (WHO) China Office, Beijing, China
Supervisor: Chin-Kei Lee, Ph.D.

2007 Graduate Research Assistantship, Capital Medical University, Department of Epidemiology and Biostatistics, China
Mentor: Xiuhua Guo, Ph.D.

Publications

1. Yang S, Eaton CB, Lu J, Lapane KL. Application of marginal structural models in pharmacoepidemiologic studies: a systematic review. *Pharmacoepidemiol Drug Saf*. 2014; 23(6):560-71.
2. Jawahar R, Oh U, Yang S, Lapane KL. Alternative approaches: a systematic review of non-pharmacological treatments for non-spastic and non-trigeminal pain in patients with multiple sclerosis. *Eur J Phys Rehabil Med*. 2014. [Epub ahead of print]
3. Yang S, Dubé CE, Eaton CB, McAlindon TE, Lapane KL. Longitudinal use of complementary and alternative medicine among older adults with radiographic knee osteoarthritis. *Clin Ther*. 2013, 35(11):1690-702.
4. Lapane KL, Yang S, Brown MJ, Jawahar R, Pagliasotti C, Rajpathak S. Sulfonylureas and risk of falls and fractures: a systematic review. *Drugs Aging.* 2013;30(7):527-47.
5. Jawahar R, Oh U, Yang S, Lapane KL. A systematic review of pharmacological pain management in multiple sclerosis. *Drugs.* 2013;73(15):1711-22.
6. Ning Y, Yang S, Evans RK, Stern M, Sun S, Francis GL, Wickham EP. Changes in body anthropometry and composition in obese adolescents in a lifestyle intervention program. *Eur J Nutr*. 2013; 53(4):1093-102.
7. Helou M, Ning Y, Yang S, Irvine P, Bachmann LM, Godder K, Massey G. Vitamin D deficiency in children with cancer. *J Pediatr Hematol Oncol.* 2014; 36(3):212-7.
8. Lapane KL, Yang S, Jawahar R, McAlindon T, Eaton CB. Missed opportunities: Lower CAM use among overweight and obese persons with radiographic knee osteoarthritis. *BMC Complement Altern Med.* 2013;13(1):241.
9. Yang S, Jawahar R, McAlindon TE, Eaton CB, Lapane KL. Racial differences in symptom management approaches among persons with radiographic knee osteoarthritis. *BMC Complement Altern Med.* 2012; 12: 86.
10. Lapane KL, Sands MR, Yang S, McAlindon TE, Eaton CB. Use of complementary and alternative medicine among patients with radiographic-confirmed knee osteoarthritis. *Osteoarthritis Cartilage.* 2012; 20(1): 22-8.
11. Jawahar R, Yang S, Eaton CB, McAlindon TE, Lapane KL. Gender-specific correlates of complementary and alternative medicine use for knee osteoarthritis. *J Womens Health (Larchmt).* 2012; 21(10):1091-9.
12. Guo X, Jia Z, Zhang P, Yang S, Wu W, Sang L, Luo Y, Lu X, Dai H, Zeng Z, Wang W. Impact of mode of transportation on dyslipidemia in working people in Beijing. *Br J Sports Med.* 2009; 43(12): 928-31.
13. Yang S, Eaton CB, McAlindon TE, Lapane KL. The long-term effects of glucosamine and chondroitin on treating knee osteoarthritis: an analysis with marginal structural models. *Arthritis Rheum*. [Revise and resubmit]
14. Lapane KL, Yang S, Driban JB, Liu SH, Dubé CE, McAlindon TE, Eaton CB. Long-term effects of use of prescription non-steroidal anti-inflammatory agents on symptoms and disease progression among patients with radiographically confirmed osteoarthritis of the knee. *Arthritis Rheum.* [Revise and resubmit]