NIH Public Access
Author Manuscript
Int J Biostat. Author manuscript; available in PMC 2009 August 3.

Published in final edited form as:
Int J Biostat. 2007 ; 3(1): 14.

Preference-based instrumental variable methods for the estimation of treatment effects: assessing validity and interpreting results

M. Alan Brookhart, Ph.D. and **Sebastian Schneeweiss, M.D., Sc.D.**
Division of Pharmacoepidemiology and Pharmacoeconomics, Department of Medicine, Brigham and Women's Hospital & Harvard Medical School, Boston, MA

Abstract

Observational studies of drugs and medical procedures based on administrative data are increasingly used to inform regulatory and clinical decisions. However, the validity of such studies is often questioned because available data may not contain measurements of many important prognostic variables that guide treatment decisions. Recently, approaches to this problem have been proposed that use instrumental variables (IV) defined at the level of an individual health care provider or aggregation of providers. Implicitly, these approaches attempt to estimate causal effects by using differences in medical practice patterns as a quasi-experiment. Although preference-based IV methods may usefully complement standard statistical approaches, they make assumptions that are unfamiliar to most biomedical researchers and therefore the validity of such analyses can be hard to evaluate. Here, we propose a simple framework based on a single unobserved dichotomous variable that can be used to explore how violations of IV assumptions and treatment effect heterogeneity may bias the standard IV estimator with respect to the average treatment effect in the population. This framework suggests various ways to anticipate the likely direction of bias using both empirical data and commonly available subject matter knowledge, such as whether medications or medical procedures tend to be overused, underused, or often misused. This approach is described in the context of a study comparing the gastrointestinal bleeding risk attributable to different non-steroidal anti-inflammatory drugs.

Keywords

pharmacoepidemiology; health services research; causal inference; outcomes research; unmeasured confounding; instrumental variables

1 Introduction

Observational studies of prescription medications and other medical interventions based on secondary administrative data are increasingly used to inform regulatory and clinical decisions. However, the validity of such studies is often questioned because the available data may not contain measurements of many important prognostic variables that guide treatment decisions such as lab values (e.g., serum cholesterol levels), clinical variables (e.g., weight, blood pressure), aspects of lifestyle (e.g., smoking status, eating habits), and measures of cognitive and physical functioning (Walker, 1996). This problem is believed to be particularly acute in

Address correspondence to: M. Alan Brookhart, Ph.D., Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital, 1620 Tremont Street, Suite 3030, Boston, MA 02120.

NIH-PA Author Manuscript
NIH-PA Author Manuscript
NIH-PA Author Manuscript

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

studies of intended effects because of the very strong correlation between treatment choice and disease risk.

The method of instrumental variables (IV) provides one potential approach to this problem.[1] Instrumental variables often arise in the context of a natural or quasi-experiment and permit the bounding and estimation of causal effects even when important confounding variables are unrecorded. Informally, an IV is a variable that is predictive of the treatment under study but completely unrelated to the study outcome other than through its effect on treatment. An IV can be thought of as a factor that induces random variation in the treatment under study. Despite their potential to address a fundamental and pervasive problem in observational studies of treatment effects, applications of IV methods in medical research are rare, presumably because plausible IVs have been difficult to find.

In recent work, instrumental variables defined at the level of the geographic region (Wen and Kramer, 1999; Brooks et al, 2003; Stuckel et al, 2007), hospital or clinic (Johnston, 2000; Brookhart, 2007), and individual physician (Korn and Baumrind, 1998; Brookhart et al, 2006; Wang et al, 2005) have been proposed or applied in medical outcomes research. Implicitly, these studies have attempted to estimate causal effects by assuming that a) providers (or groups of providers) differ in their use of the treatment under study; b) patients select or are assigned to providers independently of the provider's patterns of use of the treatment, and c) a provider's use of the treatment is unrelated to use of other medical interventions that might influence the outcome. We call such IVs "preference-based instruments" since they are derived from the assumption that different providers or groups of providers have different preferences or treatment algorithms dictating how medications or medical procedures are used. Although these approaches may reduce confounding in certain circumstances, they depend on strong assumptions that are unfamiliar to most clinical researchers and are therefore hard to evaluate. Furthermore, in some circumstances the treatment effects identified by such instruments can be difficult to interpret.

We attempt to illuminate these important practical issues by describing a theoretical framework that can be used to explore the sensitivity of the standard IV estimator to different types of violations of IV assumptions. This framework assumes the existence of a single unmeasured dichotomous variable that can be both a confounder and a source of treatment effect heterogeneity. We consider how empirical data and subject matter knowledge can be used within this framework to anticipate the direction and magnitude of bias of the standard IV estimator that results from violations of IV assumptions. We also consider how general knowledge about medical practice, such as whether medications tend to be overused, underused, or potentially misused, can help interpret the target of estimation (i.e., the IV estimand). We illustrate these ideas in the setting of an IV analysis of the effect of non-steroidal anti-inflammatory drugs (NSAIDs) and upper gastrointestinal (GI) bleeding risk. Finally, we consider how IV studies of other prescription medications and medical interventions may be more or less sensitive to IV assumptions in studies using preference-based instruments.

2 Motivating Example: Short-term effects of non-steroidal anti-inflammatory treatment assignment on risk of GI complication

We illustrate the ideas described in this paper in the context of a study that we conducted that employed an instrumental variable defined at the level of the prescribing physician (Brookhart et al, 2006). Our study attempted to assess the risk of GI toxicity among new users of non-selective NSAIDs compared with new users of the COX-2 selective NSAIDs (coxibs). This

[1] See Angrist et al, 1996; Greenland, 2000; Martens et al 2006, and Hernán and Robins, 2006 for overviews of IV methods.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

example both illustrates the use of a preference-based instrumental variable and the difficulty of estimating intended treatment effects using administrative data.

As background, coxibs are generally thought to have greater GI tolerability than non-selective NSAIDs; therefore, confounding arises in comparative studies of NSAIDs as a result of the selective prescribing of coxibs to patients who are at elevated risk of GI complications, such as patients with a history of smoking, alcoholism, obesity, or peptic ulcer disease. Because many of these variables are poorly measured or completely unrecorded in typical pharmacoepidemiologic databases, studies comparing the GI risks of different NSAIDs would be expected to understate any protective effect of coxibs. Indeed, several observational studies have been unable to attribute any GI-protective effect to the coxibs (Laporte at al, 2003).

Although a physician’s choice of NSAID relies strongly on an assessment of a patient’s underlying GI risk, NSAID prescribing is also thought to depend on individual physician preference.(Solomon et al, 2003; Schneeweiss et al, 2005). The possibility that physicians strongly differ in their preference for different NSAIDs suggests that an instrumental variable defined at the level of the prescribing physician could be used to estimate NSAID treatment effects.

2.1 Study Population and Data

Our study was based on 37,842 new NSAID users drawn from a large population-based cohort of Medicare beneficiaries who were eligible for a state-run pharmaceutical benefit plan. State medical license numbers from the pharmacy claims were used to identify the prescribing physician (Brookhart et al, 2007). From the Medicare and pharmacy claims we extracted a treatment assignment X (X=1 if a patient was placed on a coxib, X=0 otherwise), a set of measured covariates C, and an outcome Y indicating an upper GI bleed within 60 days of initiating an NSAID.

We have proposed to use an instrumental variable defined at the level of the prescribing physician. One approach to implementing this would be to use individual physician indicator variables as instrumental variables. This approach would essentially use the proportion of NSAID prescriptions for coxibs as a measure of a physician’s preference for coxibs. Such an approach was implicitly used in studies that have used hospitals (Johnston, 2000) and geographic regions (Brooks et al, 2003) as instrumental variables. In our study of NSAIDs, however, the study period was an era of aggressive marketing and active debate about the safety and effectiveness of coxibs and non-selective NSAIDs. Therefore, we sought an instrumental variable that would allow preference to change. We opted to use the type of the most recent NSAID prescription initiated by each physician as an instrument, i.e., we defined the instrumental variable Z to be equal to 1 if the physician’s most recent new NSAID prescription was for a coxib and zero otherwise. Pharmacy claims that occurred on the same day were randomly ordered.

We justified the use of this variable by assuming that Z was effectively randomly assigned to patients, so that patient characteristics were unrelated to Z, and also that Z was related to Y only through its relationship with X, the choice of NSAID type. We also assumed that $E[X|Z] \neq E[X]$, so that Z predicts X. In the following section we formalize these assumptions in terms of moment assumptions in a structural (counterfactual) model.

3 The Method of Instrumental Variables

We describe our instrumental variable approach using the potential (counterfactual) outcome framework of Rubin (1974). This approach requires that for each subject there exist two counterfactual (potential) outcomes, Y_1 and Y_0, that correspond to the outcomes we would

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

observe if a patient were treated with coxibs or non-selective NSAIDs, respectively. For these outcomes we assume the following model (called a structural model):

$$Y_x=\alpha_0+\alpha_1 x+\epsilon_x \tag{1}$$

where x is an assigned rather than an observed treatment, (x = 1 if the assigned treatment is a coxib, x = 0 otherwise), ϵ_x is an error specific to the assigned treatment, and $E[\epsilon_x] = 0$ for $x \in \{0,1\}$. The average treatment effect (ATE) in the population is expressed as $E[Y_1 - Y_0] = \alpha_1$.

Under the consistency assumption, which states that the observed outcome is indeed a counterfactual outcome, the observed data are linked to the potential outcome through the relation

$$Y=X(Y_1)+(1-X)(Y_0). \tag{2}$$

Substituting the terms from the structural model (1) into the relation (2) allows us to write the observed outcome as a function of the structural model parameters and error terms:

$$Y=\alpha_0+\alpha_1 X+\epsilon_0+X(\epsilon_1-\epsilon_0). \tag{3}$$

The term $\alpha_0 + \epsilon_0$ reflects an individual patient's outcome if treatment were withheld, but everything else were to remain the same about the patient and concomitant treatments. The term $\epsilon_1 - \epsilon_0$ represents the added benefit or harm beyond α_1 that an individual patient receives from treatment. This term captures a patient's unique response to treatment and allows for treatment effect heterogeneity.

In our setting, the term ϵ_0 represents both patient characteristics that are related to baseline prognosis as well as other concomitant treatments that a patient might receive from the physician that could affect the outcome. If Z has an independent relation with Y, either through its association with patient characteristics or concomitant treatments, then $E[\epsilon_0|Z] \neq 0$. For the remainder of the paper, we equate the assumption $E[\epsilon_0|Z] = 0$ with the exclusion restriction of Angrist et al (1996).

Traditional instrumental variables approaches in econometrics assume that treatment effects are constant, so $\epsilon_1 = \epsilon_0$ for all patients. When this assumption and the exclusion restriction hold, it is straightforward to see that

$$\alpha_1=\frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]}.$$

Thus, the parameter α_1 can be estimated by plugging in sample quantities for the conditional expectations

$$\widehat{\alpha}_{IV}=\frac{\widehat{E}[Y|Z=1]-\widehat{E}[Y|Z=0]}{\widehat{E}[X|Z=1]-\widehat{E}[X|Z=0]}. \tag{4}$$

This is the standard instrumental variable estimator or Wald estimator. For this to be a consistent estimator of α_1, we need to assume that one patient's counterfactual outcomes are not affected by the treatment assignment of other patients. This along with the consistency

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

assumption compose the so-called stable unit value treatment assumption (SUTVA) of Rubin (1986). We also need to assume that the instrument is associated with receiving treatment, so that $E[X|Z = 1] - E[X|Z = 0] \neq 0$.

If treatment effects are heterogeneous, an additional assumption is required to meaningfully interpret the standard IV estimator. One such assumption states that the correlation between the received treatment and an individual's response to it (as measured on a linear scale) is the same across levels of Z., i.e., $E[X(\epsilon_1 - \epsilon_0)|Z] = E[X(\epsilon_1 - \epsilon_0)]$. If this holds, the standard IV estimator will be consistent for the average treatment effect in the population.

Alternatively, Imbens and Angirst (1994) and Angrist et al (1996) established that under a monotonicity assumption, the standard IV estimator is consistent for the average effect of treatment among the sub-population of patients termed the "compliers" (Angrist et al, 1996) or "marginal patients" (Harris and Remler, 1998). These are patients whose treatment status depends on the level of the instrument. This parameter is termed the local average treatment effect (Angrist et al, 1996). In the setting of a placebo-controlled RCT with non-compliance in which the instrumental variable is the treatment arm assignment, the compliers are the patients who always take their assigned treatment. Monotonicity requires that the IV deterministically affect treatment in one direction. In the RCT example, 'monotonicity' asserts that there are no patients who would do the opposite of what they were assigned.

In the setting of preference-based instrumental variables, the utility of IV approaches based on a monotonicity assumption is unclear, in part because the concept of a marginal patient type is problematic. For example, a certain type of patient may be treated 95% of the time by physicians with $Z = 1$ and 5% of the time by physicians with $Z = 0$, whereas another patient-type may be treated 52% of the time by physicians with $Z = 1$ and 48% of the time by physicians with $Z = 0$. Both patients are technically "marginal," as their treatment status is affected by the instrument; however, clearly patients of the second type are less likely to have their treatment status influenced by the physician that they see and are therefore likely to be down-weighted in an instrumental variable estimate. See Hernán and Robins (2006) for a discussion of a deterministic monotonicity assumption for preference-based instruments and other theoretical issues concerning preference-based IVs. See Wooldridge (1997) and Heckman, Urzua, Vytlacil (2006), for additional discussion of treatment effect heterogeneity.

In the following sections, we propose a structural model that represents both treatment effect heterogeneity and violations of the structural instrumental variable assumptions. This model is used to explore how violations of assumptions and treatment effect heterogeneity may bias the traditional IV estimator relative to average effect of treatment in the population.

3.1 A structural model for sensitivity analysis

We extend the structural model (1) by introducing a single unobserved dichotomous variable U that could represent a pre-treatment risk factor for the outcome, a concomitant treatment assigned by the physician, or treatment effect modifier on the risk difference scale. Our new model for the counterfactual Y_x is given by

$$Y_x=\alpha_0+\alpha_1 x+\alpha_2 U+\alpha_3 U x+\epsilon_x, \qquad (5)$$

with $E[\epsilon_x|U] = 0$ for $x \in \{0,1\}$. The average treatment effect for those with $U = 0$ is given by $E[Y_1 - Y_0|U = 0] = \alpha_1$ and the average effect of treatment among those with $U = 1$ is $E[Y_1 - Y_0|U = 1] = \alpha_1 + \alpha_3$. The average treatment effect in the whole population is given by $E[Y_1 - Y_0] = \alpha_1 + \alpha_3 E[U]$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Under the consistency assumption, we can re-write the observed Y as a function of the structural model parameters and errors

$$Y=\alpha_0+\alpha_1 X+\alpha_2 U+\alpha_3 XU+\epsilon_0+X(\epsilon_1-\epsilon_0).$$

We assume that $E[\epsilon_x|X,U] = 0$ for $x \in \{0,1\}$, so that the parameters of (5) could be consistently estimated by least-squares if both X and U were observed. Given this assumption, we can see that by iterated expectations $E[\epsilon_0 + X(\epsilon_1 - \epsilon_0)|X] = 0$. Therefore,

$$E[Y|X=1]-E[Y|X=0]=\alpha_1+\alpha_2(E[U|X=1]-E[U|X=0])+\alpha_3 E[U|X=1].$$

This expression tells us that a crude estimate of the treatment effect based on differences in means between treatment groups (e.g., a risk difference for a dichotomous outcome) is inconsistent for the average treatment effect in the population if U is not mean independent of X.

To evaluate the IV estimand, we further assume that $E[\epsilon_0|Z] = 0$, so that the instrument can be related to the observed outcome only through its effect on X or association with U; and also that $E[X(\epsilon_1 - \epsilon_0)|Z] = E[X(\epsilon_1 - \epsilon_0)]$ so that there is no relevant treatment effect heterogeneity beyond that generated by U. Under these assumptions, the standard IV estimand can be written as

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]}=\alpha_1+\gamma_1+\gamma_2$$

where

$$\gamma_1=\alpha_2\frac{E[U|Z=1]-E[U|Z=0]}{E[X|Z=1]-E[X|Z=0]}$$

and

$$\begin{aligned}\gamma_2&=\alpha_3\frac{E[XU|Z=1]-E[XU|Z=0]}{E[X|Z=1]-E[X|Z=0]}\\&=\alpha_3\frac{E[X|Z=1,U=1]E[U|Z=1]-E[X|Z=0,U=1]E[U|Z=0]}{E[X|Z=1]-E[X|Z=0]}.\end{aligned}$$

So the asymptotic bias in IV estimator relative to the average effect of treatment in the population is given by

$$\begin{aligned}BIAS(\widehat{\alpha}_{\mathrm{IV}})&=\alpha_2\frac{E[U|Z=1]-E[U|Z=0]}{E[X|Z=1]-E[X|Z=0]}\\&+\alpha_3\left\{\frac{E[X|Z=1,U=1]E[U|Z=1]-E[X|Z=0,U=1]E[U|Z=0]}{E[X|Z=1]-E[X|Z=0]}-E[U]\right\}\end{aligned}$$

By considering the above expressions, we can understand how violations of the instrumental variable assumptions and treatment effect heterogeneity due to a single binary covariate can bias the IV estimand relative to the average effect of treatment in the population. In the following sections, we will illustrate these ideas in the context of our study of NSAIDs outlined earlier.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

4 Application of Methods

In table 1, we give the distribution of patient-level GI risk factors across levels of the received treatment. The third column gives the prevalence difference between levels of the exposure and 95% confidence limits (reported in percentage points). The variables in the rows of this table were selected because they were either GI risk factors that could be related to both the outcome and NSAID treatment assignment, or they were recent/concomitant GI-related treatments that could be associated with the exposure or instrument. This table reveals that patients prescribed coxibs were older, more likely to be female, and more likely to have a history of GI hemorrhage and peptic ulcer disease. These patients are also more likely to have recently used warfarin and glucocorticoids, medications that increase the risk of GI hemorrhage. Coxib users are also more likely to have recently used GI protective drugs, suggestive of unmeasured GI problems. This table is consistent with our expectation that coxib users tend to be at greater baseline risk of GI complications. In table 2, we give the distribution of patient-level GI risk factors across levels of the instrument. This table parallels table 1, except that the columns are defined by levels of the instrument rather than the received treatment. We find that the imbalance of GI risk factors and concomitant/recent treatments has been greatly reduced and, in most cases, reduced to statistical insignificance. However, there is some evidence of a weak association between *Z* and certain GI risk factors. This could be be due to specialist physicians seeing sicker patients and also being more likely to prescribe coxibs, or possibly patients who are at greater GI risk are seeking out physicians who are more likely to prescribe coxibs.

The instrumental variable approach requires that the instrument is related to the exposure. In our study, we found that $\hat{E}[X|Z = 1] - \hat{E}[X|Z = 0] = 22.8\%$. Therefore, within our population, seeing a physician who most recently prescribed a coxib was associated with an absolute increase of 22.8% in a patient’s probability of receiving a coxib. The instrument was also related to the outcome. Seeing a physician whose previous new NSAID prescription was a coxib decreased a patient’s probability of a 60-day GI complications by 0.21%.

Using these statistics, we can evaluate the standard IV estimator:

$$\widehat{\alpha}_{IV}=\frac{\widehat{E}[Y|Z=1]-\widehat{E}[Y|Z=0]}{\widehat{E}[X|Z=1]-\widehat{E}[X|Z=0]}=\frac{-0.0021}{0.2280}=-0.0092,$$

which suggests a risk reduction of approximately 1 event per 100 patients treated with coxibs.

However, several important questions about this result remain unanswered. To what extent could this estimate be biased by a residual association between an unmeasured GI risk factor and the IV? In what direction would this bias be expected to operate? How might treatment effect heterogeneity lead to further bias in this estimator relative to the average effect of treatment in the population?

In the following sections we consider how our sensitivity analysis framework may shed some light on these issues. To simplify exposition and to facilitate intuition, we consider two scenarios: one in which the exclusion restriction is violated, but the average effect of treatment does not vary with the unmeasured variable *U*; and another in which the exclusion restriction holds, but the average treatment effect varies with *U*.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

4.1 Scenario 1: Average treatment effect does not vary with *U*, but the exclusion restriction is violated

If we assume that the average effect of treatment is the same across levels of U, then α_3 is zero, and thus the bias in the OLS estimator is given by

$$BIAS(\widehat{\alpha}_{OLS})=\alpha_2(E[U|X=1]-E[U|X=0]),$$

the difference in the prevalence of the risk factor between treatment groups multiplied by the excess risk of the outcome among patients with $U = 1$.

The violation of the exclusion restriction tells us that $E[U|Z] \neq E[U]$. Therefore, the asymptotic bias in the IV estimator is given by

$$BIAS(\widehat{\alpha}_{IV})=\alpha_2\frac{E[U|Z=1]-E[U|Z=0]}{E[X|Z=1]-E[X|Z=0]}.$$

The term $E[U|Z = 1] - E[U|Z = 0]$ is the difference in the prevalence of the risk factor between levels of the instrument. The total bias in the IV estimator is this difference multiplied by the excess risk of the outcome among patients with $U = 1$ divided by the strength of the instrument. This expression illustrates the importance of instrument strength – as the IV gets weaker the denominator gets smaller and the bias term increases without bound. Thus, even a minor violation of the exclusion restriction can lead to substantial bias if the instrument is weak (Bound et al, 1995; Small and Rosenbaum, in press).

For the IV to have less asymptotic bias than OLS,

$$\frac{E[U|Z=1]-E[U|Z=0]}{E[U|X=1]-E[U|X=0]}<E[X|Z=1]-E[X|Z=0].$$

In other words, the difference in the prevalence of U between levels of Z relative to the difference in the prevalence of U between levels of X must be less than the strength of the instrument.

Although U is assumed to be an unmeasured variable, the plausibility of this condition can be explored by using the measured variables as proxies for U. In column 3 of table 1, we report the difference in prevalence between treatment groups for all measured risk factors and the variables capturing concomitant and recent medication use. In column 3 of table 2, we report the difference in prevalence between groups defined by the instrument for the same risk factors. In table 3, we report the ratio of these two statistics, which we term the *prevalence difference ratio* (PDR). We would like for the ratio of these imbalances to be less than the strength of the instrument (i.e., about 23%), particularly for those variables clearly related to the outcome.

Only age ≥ 75 and recent warfarin use had a statistically significant association with the IV. Furthermore, most of the PDRs from table 3 were smaller than strength of the IV in this study. Three variables, however, raise some concerns. First, although recent glucocortidoid use was not significantly associated with the IV, its PDR was 32% (larger than the desired 23%). Possibly physicians who use more coxibs may be slightly more likely to co-prescribe oral steroids or are seeing more patients with inflammatory conditions, such as rheumatoid arthritis. Secondly, recent use of warfarin remained significantly associated with the instrument and its PDR was 27% (also slightly more than the desired 23%). Finally, the PDR associated with a history of cardiac-related hospitalizations or procedures was 56%, although it was not

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

significantly associated with the instrument. It should be noted that a history of cardiovascular problems is itself not clearly a GI risk factor, but we included it in our analysis because it may be correlated with other GI risk factors such as obesity and smoking status.

Instrumental variable methods can make statistical adjustments for these potentially problematic measured covariates; however, the residual associations between the instrument and several observed variables suggest that the instrument may be associated with unmeasured variables. In particular, the associations between the instrument and two different treatment modalities, (warfarin and glucocorticoids) suggest that physicians who frequently prescribe coxibs may practice differently, e.g., prescribe different medications. Short-term GI illness, however, is a specific outcome and there are relatively few ways that a physician could directly affect it. Those ways can be measured reasonably well in health care utilization data, e.g., co-prescribing of GI protective drugs or medication that cause GI problems. Associations between the IV and other treatment modalities may be more concerning in studies of outcomes such as all cause mortality that a physician could potentially influence in many ways.

In the present analysis, because the PDR for most variables, including history of major GI problems, was less than 23%, we expect that other unmeasured GI risk factors may have a similarly weak association with the IV. This is speculative, but if true, the IV approach should have decreased confounding bias relative to OLS caused by violations of the exclusion restriction.

4.2 Scenario 2: Exclusion restriction holds, but average treatment effect varies with *U*

In this section, we assume that the exclusion restriction holds, so $E[U|Z = 1] - E[U|Z = 0] = 0$, but the average treatment effect varies with U, so that $\alpha_3 \neq 0$. Here we imagine U to be an unmeasured patient risk factor that is a source of treatment effect heterogeneity (rather than a concomitant treatment).

In these assumptions, the asymptotic bias in the IV estimator is given by

$$BIAS(\widehat{\alpha}_{IV}) = \alpha_3 E[U]\left[\frac{E[X|Z=1, U=1] - E[X|Z=0, U=1]}{E[X|Z=1] - E[X|Z=0]} - 1\right]. \tag{6}$$

The denominator is the strength of the instrument in the whole population. The numerator is the strength of the instrument among people with $U = 1$, e.g., those with a particular unmeasured GI risk factor. From this expression we can make two immediate observations for the situation in which the average treatment effect varies with U. First, if the strength of the instrument is the same in both groups defined by U, then the bias is zero. Second, if the instrument is not predictive of exposure among patients with $U = 1$, i.e., $E[X|Z = 1,U = 1] - E[X|Z = 0,U = 1] = 0$, then the bias is equal to $-\alpha_3 E[U]$. In other words, the IV estimates the average effect of treatment among people with $U = 0$. This would be the case if a patient with the risk factor were equally likely to be treated by either type of physician. One extreme example of this would be patients who are always treated or never treated.

Next we consider how subject-matter knowledge of medical practice patterns can be used to anticipate the magnitude and direction of this bias term.

Bias in the IV estimator when medications or procedures are overused in the population under study—During the period of our study, coxibs were thought to be generally overused, more likely to be prescribed to patients who did not need them than to be withheld from patients who needed them (Desmet et al, 2006). Let U denote an unmeasured GI risk factor (e.g., smoking status) that is observed by the physician and could modify the

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

effect of coxib exposure. If coxibs are overused, then patients who have an indication for coxib exposure, e.g., those who smoke, are likely to get a coxib from either type of physician. The additional people being treated by physicians with $Z = 1$ are those who are less likely to benefit from a coxib. Under these assumptions, $0 \le E[X|Z = 1,U = 1] - E[X|Z = 0,U = 1] < E[X|Z = 1] - E[X|Z = 0]$, so the bias is bounded as follows

$$-\alpha_3 E[U] \le BIAS(\widehat{\alpha}_{IV}) < 0.$$

So, the IV estimator is over-weighting the effect of treatment in the low-risk group.

To better understand this bias, consider an extreme example: all patients who could benefit from a coxib would get one regardless of the physician's preference. In this case, Z will have no marginal association with the outcome ($E[Y|Z = 1] - E[Y|Z = 0] = 0$), and the IV estimand will be zero. Here, the IV is reflecting the effect of treatment in a population of patients who would not benefit from treatment with coxibs and thus underestimates the average effect of coxib exposure in the larger population.

Bias in the IV when medications or procedures are underused in the population under study—In many cases, medications and medical procedures are thought to be underused, in that they are not given to many patients who might benefit from them. One well-known example is bone resorption agents that are used to treat osteoporosis. If these medications are underused, we would expect the instrument to be more strongly related to treatment among those with clinically evident osteoporosis (e.g., low bone mineral density test results or history of osteoporotic fractures) than among the entire population. If U indicated a risk factor for a fracture, we anticipate that $E[X|Z = 1, U = 1] - E[X|Z = 0, U = 1] > E[X|Z = 1] - E[X|Z = 0]$ and therefore $BIAS(\hat{\alpha}_{IV} > 0$. Here, the IV estimator is extrapolating the treatment effect of bone resorption agents in a high-risk group to the rest of the population. If treatment is more effective in high-risk patients and the instrument is also stronger within this group, treatment effect heterogeneity would lead preference-based IV estimators to exaggerate the protective effect of medications or procedures.

Bias in the IV estimator when medications or procedures are misused in the population under study—In some cases, medications or medical procedures may be misused, in the sense that they may be given to patients with specific contraindications, or necessary follow-up tests are not performed after patients have been started on a medication or have received a procedure. Preference-based IV studies of drugs or procedures that are commonly misused can be subject to counter-intuitive biases.

For example, consider a study that compares the safety of metformin and other oral antihyperglycemic drugs used to treat Type II diabetes. Metformin is contraindicated in patients with decreased renal function or liver disease, as it can cause lactic acidosis, a potentially fatal side effect. We speculate that physicians who infrequently use metformin will be less likely to understand its contraindications and would therefore be more likely to misuse it. Let U be an indicator of decreased renal function or liver disease. If our hypothesis is true, $E[X|Z = 1,U = 1] - E[X|Z = 0, U = 1]$ will be negative. In other words, physicians with $Z = 1$ are less likely than physicians with $Z = 0$ to prescribe metformin to patients with a specific contraindication. In this case, a preference-based instrumental variable could make metformin appear to prevent lactic acidosis, as patients of physicians with $Z = 1$ are at lower risk of being inappropriately treated with metformin.

Empirically evaluating the magnitude and direction of bias due to treatment effect heterogeneity—The results from this section suggest that we can look for evidence

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

of bias due to treatment effect heterogeneity using observed data. The expression (6) for the bias depends on the strength of the instrument within the sub-population defined by $U = 1$ relative to the strength of the instrument in the entire population. Because U is a variable that is assumed to be unobserved, we propose to use measured factors as proxies for U. If the strength of the instrument varies strongly across different subgroups defined by observed factors, it is reasonable to be concerned that the instrument strength may vary across subgroups defined by unobserved variables and therefore that the IV estimator may be biased for the average effect of treatment in the population under study.

For example, we have speculated that, consistent with other research, coxibs are likely to be overused in our study population, given to many patients who might not need them (Desmet et al, 2006). To evaluate this assertion, we determine whether the strength of the instrument is in the whole population is different from the the strength of the instrument within specific subgroups. If coxibs are overused, we would expect the IV to be weaker within subgroups defined by strong GI risk factors.

Table 4 presents the strength of the instrument within various strata defined by measured variables. We observed little variation in the strength of the IV within measured variables. Consistent with our hypothesis, the IV was slightly weaker within strata of the strongest GI risk factors, e.g., history of GI hemorrhage, peptic ulcer disease, or GI protective drug use. However, none of these differences reached statistical significance. Interestingly, the instrument was the weakest among those who recently used warfarin, an anticoagulant well-known to increase the risk of catastrophic hemorrhagic events. The weakness of the instrument among recent warfarin users suggests that many physicians try to avoid prescribing non-selective NSAIDs to this group because of their bleeding risk.

5 Interpretation of preference-based IV estimands in the presence of treatment effect heterogeneity

We suggested earlier that the concept of a “marginal” patient may not be clear when using preference-based instruments, as patients can be marginal to differing degrees. We extend the results of the previous section to describe the target of IV estimation in the setting of preference-based IV methods and treatment effect heterogeneity.

First, suppose that the study population can be decomposed into a set of $\kappa + 1$ mutually exclusive groups of patients with common clinical and lifestyle characteristics. Patient membership in these groups is denoted with the set of indicators $\mathbf{S} = \{S_1, S_2, \cdots, S_\kappa\}$. Group membership is observed by the clinician but is unrecorded in the data file. We generalize our structural model (3) as follows

$$Y_x=\alpha_0+\alpha_1 x+\alpha_{\mathbf{2}}^T\mathbf{S}+\alpha_{\mathbf{3}}^T\mathbf{S}x+\epsilon_x, \tag{7}$$

with $E[\epsilon_x|\mathbf{S}] = 0$. Now $\alpha_{\mathbf{2}}$ and $\alpha_{\mathbf{3}}$ are vectors of coefficients. The average effect of treatment in the population is given by $\alpha_1 + \alpha_3{}^T E[\mathbf{S}]$. We assume that $E[\epsilon_0|Z] = 0$ and that there is no relevant heterogeneity beyond S, so that $E[\epsilon_0 - X(\epsilon_1 - \epsilon_0)|Z] = E[\epsilon_0 - X(\epsilon_1 - \epsilon_0)]$.

Assuming that $E[\mathbf{S}|Z] = E[\mathbf{S}]$, we can extend the results from the previous section to show that

$$\frac{E[Y|Z=1]-E[Y|Z=0]}{E[X|Z=1]-E[X|Z=0]}=\alpha_1+\sum_{j=1}^{k}\alpha_{3,j}E[S_j]w_j.$$

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

The estimated treatment effect turns out to be a “weighted average” of treatment effects in different sub-groups, where the weights are given by

$$w_j=\left[\frac{E[X|Z=1,S_j=1]-E[X|Z=0,S_j=1]}{E[X|Z=1]-E[X|Z=0]}\right],$$

and thus could be negative or have absolute values greater than one.

The interpretation of the weights follows from the previous discussion of treatment effect heterogeneity. If the instrument is stronger in sub-group *j* than in the whole population, then weight for that sub-group is increased. If the instrument is weaker in sub-group *j* than it is in the whole population, then the sub-group effect is weighted down. If the effect of the instrument is reversed in sub-group *j*, e.g., in the case of contraindications, the weight will be negative. Lastly, if the IV has no effect in a particular group, than the weight is zero and the effect of treatment in that sub-group is not reflected in the IV estimator.

In our example, we see relatively little variation in the strength of the instrument across levels of observed patient risk factors. As expected, the instrument appears to be slightly weaker among patients with strong GI risk factors, suggesting that preference is slightly weaker among these patients, as more physicians will treat these patients with coxibs rather than NSAIDs. To the extent that this practice pattern holds across unmeasured risk factors, our IV estimate will be slightly conservative, perhaps down-weighting the contributions of these groups by around 20%.

For another example, we consider the case of statins, cholesterol-lowering drugs that are thought to be substantially underused, in that they are not given to many patients who might benefit from them (Majumdar et al, 1999). Suppose we are doing a typical study using health care claims data to assess the effectiveness of statins in a population at-risk of an acute coronary event. Using health care claims data, we have attempted to identify a study population consisting of people with at least one cardiovascular risk factor, e.g., patients with a diagnosis of hypertension, unstable angina, myocardial infarction, diabetes, or hypercholesterolemia. In this population there is still considerable variation in underlying risk. We speculate that those at greatest risk, e.g., those who smoke, are overweight, and who have a history of myocardial infarction, will be treated with statins by many physicians. Therefore the contribution of the treatment effects in the highest-risk group could be down-weighted. Similarly, those at lowest risk may be treated by few physicians of either type, and their contribution to the IV estimate would also be down-weighted. The instrument may be the strongest among patients at moderate risk, so the IV estimate may tend to reflect the effect of treatment in these patients.

In the case of statins, which are relatively safe with few contraindications, there are not likely to be pathologies that would lead a small sub-group to have a very large negative weight. Finally, we note that one can explore the likely magnitude and directions of these weights empirically. As we proposed earlier, by assessing the strength of the instrument within sub-groups, researchers can gather evidence about important differences in practice patterns between physicians with $Z = 1$ and those with $Z = 0$. For example, to explore the plausibility of our hypothesis about statin prescribing, we could examine the strength of the IV across a range sub-groups of varying degrees of cardiovascular risk according to the observed variables.

6 Discussion

We have discussed issues related to the validity and interpretation of studies using preference-based instrumental variables that are defined at the level of a health care provider or an

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

aggregation of providers (e.g., hospital or geographic region). We illustrate various ways that observed variables can be used as proxies for unobserved confounders to anticipate the direction of bias due to violations of IV assumptions. Using these variables, we provided a benchmark to assess whether the IV approach is likely to reduce confounding bias relative to a conventional estimate of treatment effect. We have also described how one can use observed variables and subject matter knowledge to anticipate the direction of bias in a standard instrumental variable estimator due to treatment effect heterogeneity.

The ideas discussed in this paper were presented in the context of a study of the short-term risk of GI bleeding among elderly new users of non-selective, non-steroidal anti-inflammatory drugs. The analysis based on the methods described herein suggested that, in the absence of treatment effect heterogeneity, violations of the exclusion restriction may have caused our IV estimate to slightly underestimate the average effect of coxib treatment in the population. This is due primarily to the the IV having a weak positive association with some GI risk factors and recent use of medications that can increase GI risk. To the extent the measured variables are reasonable proxies for the unmeasured variables, our analysis suggested that the bias in the IV is likely to be smaller than the bias in a conventional analysis.

We also found that treatment effect heterogeneity may have led to some small differences between the IV estimand and the average treatment effect in the population. Empirical data suggest that patients at lower GI risk were slightly more likely to have had their treatment influenced by the instrumental variable. According the framework we described, the contribution of the effect of treatment in these patients may be slightly up-weighted by the instrumental variable estimator. To extent that coxibs may be less effective (on a risk difference scale) in lower-risk patients, treatment effect heterogeneity would have caused the IV estimator to further understate the average protective effect of coxibs in the population.

In the expressions for bias that we have derived, it is assumed that the parameter of interest is the average effect of treatment in the population under study. This parameter is of inherent interest as it is what would be estimated by an RCT conducted in the population. However, in many observational studies of drugs and medical procedures, the population under study may include many patients for whom there is little clinical equipoise (i.e., patients who would be almost always treated or rarely treated). In these settings, other measures of treatment effect may be of greater interest. For example, when many patients are appropriately untreated, one may be more interested in the average effect of treatment on those who did received treatment (the effect of treatment on the treated). When drugs are underused in the population under study, and the IV affects treatment in a small, high-risk segment of the population, the IV estimand is likely to be closer to the effect of treatment in the treated.

Our study is limited primarily by the simplicity of our analytic framework. We have considered bias in the setting of the standard instrumental variable estimator with a single dichotomous instrument, unmeasured covariate, and treatment. For more complex situations involving non-linear models, continuous treatments, and multiple continuous instruments, the analyst will need to use subject-matter expertise to make assumptions about both the model for the treatment choice and outcome. When treatment effects are heterogeneous, the interpretation of the effect estimate can depend on the assumptions one makes about these models. Our results will not immediately apply to these more complex settings.

As with any analysis of observational data, analyses based on preference-based IV methods rely on assumptions that cannot be verified with observed data. In many cases, these assumptions will not completely hold, and IV methods may lead to estimates that are both highly biased and excessively variable. We have outlined an approach that to assess the likely extent of the problem. Further research may reveal additional ways to evaluate the validity of

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

preference-based IV methods or to improve them through study design or statistical innovations.

Acknowledgments

This project was funded under Contract No. 290-2005-0016-I-TO3-WA1 from the Agency for Healthcare Research and Quality (AHRQ), US Department of Health and Human Services (DHHS) as part of the Developing Evidence to Inform Decisions about Effectiveness (DECiDE) program. The authors of this report are responsible for its content. Statements in the report should not be construed as endorsement by AHRQ or DHHS. Dr Brookhart was additionally supported by a career development award from the National Institute on Aging (AG027400). The authors acknowledge helpful comments from the anonymous referees.

References

1. Walker A. Confounding by indication. Epidemiology 1996;7(4):335–336. [PubMed: 8793355]
2. Angrist J, Imbens G, Rubin DB. Identification of causal effects using instrumental variable. J Amer Stat Assoc 1996;91(434):444–455.
3. Greenland S. An introduction to instrumental variables for epidemiologists. Int J Epidemiol 2000;29:722–729. [PubMed: 10922351]
4. Martens EP, Pestman WR, de Boer A, Belitser SV, Klungel OH. Instrumental variables: application and limitations. Epidemiology 2006;17(3):260–267. [PubMed: 16617274]
5. Hernán MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? Epidemiology 2006;17(4):360–372. [PubMed: 16755261]
6. Wen SW, Kramer MS. Uses of ecologic studies in the assessment of intended treatment effects. J Clin Epidemiol 1999;52(1):7–12. [PubMed: 9973068]
7. Brooks JM, Chrischilles EA, Scott SD, Chen-Hardee SS. Was breast conserving surgery underutilized for early stage breast cancer? Instrumental variables evidence for stage II patients from Iowa. Health Serv Res 2003;38(6 Pt 1):1385–1402. [PubMed: 14727779]
8. Stukel TA, Fisher ES, Wennberg DE, Alter DA, Gottlieb DJ, Vermeulen MJ. Analysis of observational studies in the presence of treatment selection bias: effects of invasive cardiac management on AMI survival using propensity score and instrumental variable methods. JAMA 2007;297(3):278–285. [PubMed: 17227979]
9. Johnston SC. Combining ecological and individual variables to reduce confounding by indication: case study-subarachnoid hemorrhage treatment. J Clin Epidemiol 2000;53(12):1236–1241. [PubMed: 11146270]
10. Brookhart, MA. Assessing the safety of recombinant erythropoietin using instrumental variable methods [abstract]. Meeting of the International Biometrics Society; Eastern North American Region; Atlanta, Georgia. 2007.
11. Korn MA, Baumrind E. Clinician preferences and the estimation of causal treatment differences. Statistical Science 1998;13(3):209–235.
12. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Evaluating short-term drug effects using a physician-specific prescribing preference as an instrumental variable. Epidemiology 2006;17(3): 268–275. [PubMed: 16617275]
13. Wang PS, Schneeweiss S, Avorn J, Fischer MA, Mogun H, Solomon DH, Brookhart MA. Risk of death in elderly users of conventional vs. atypical antipsychotic medications. N Engl J Med 2005;353 (22):2335–2341. [PubMed: 16319382]
14. Laporte JR, Ibanez L, Vidal X, Vendrell L, Leone R. Upper gastrointestinal bleeding associated with the use of NSAIDs: newer versus older agents. Drug Saf 2004;27(6):411–420. [PubMed: 15144234]
15. Solomon DH, Schneeweiss S, Glynn RJ, Levin R, Avorn J. Determinants of selective cyclooxygenase-2 inhibitor prescribing: are patient or physician characteristics more important? Am J Med 2003;115(9):715–720. [PubMed: 14693324]
16. Schneeweiss S, Glynn RJ, Avorn J, Solomon DH. A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of first-time prescriptions for COX-2 inhibitors. J Clin Epidemiol 2005;58(1):98–102. [PubMed: 15649677]

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

17. Brookhart MA, Polinski JM, Avorn J, Mogun H, Solomon DH. The medical license number accurately identifies the prescribing physician in a large pharmacy claims dataset. Med Care 2007;45(9):907–910. [PubMed: 17712263]

18. Rubin DB. Estimating causal effects of treatment in randomized and non-randomized studies. Journal of Eductional Psychology 1974;66:688–701.

19. Rubin DB. Statistics and causal inference. Comment: which ifs have causal answers? J Amer Stat Assoc 1986;81:961–962.

20. Imbens G, Angrist J. Identification and estimation of local average treatment effects. Econometrica 1994;62(2):467–476.

21. Harris KM, Remler DK. Who is the marginal patient? Understanding instrumental variables estimates of treatment effects. Health Serv Res 1998;33(5 Pt 1):1337–1360. [PubMed: 9865223]

22. Wooldridge J. On two-stage least squares estimation of the average treatment effect in a random coefficient model. Economic Letters 1997;56:129133.

23. Heckman JJ, Urzua S, Vytlacil EJ. Understanding instrumental variable models with essential heterogeneity. NBER working paper 2006:12574.

24. Bound J, Jaeger DA, Baker RM. Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J Am Stat Assoc 1995;90:443–450.

25. Small D, Rosenbaum PR. War and Wages: The strength of instrumental variables and their sensitivity to unobserved biases. J Am Stat Assoc. In press

26. De Smet BD, Fendrick MA, Stevenson JG, Bernstein SJ. Over and Under-utilization of cyclooxygenase-2 selective inhibitors by primary care physicians and specialists: the tortoise and the hare revisited. J Gen Intern Med 2006;21:694–697. [PubMed: 16808768]

27. Majumdar SR, Gurwitz JH, Soumerai SB. Undertreatment of hyperlipi-demia in the secondary prevention of coronary artery disease. J Gen Intern Med 1999;14:711–717. [PubMed: 10632815]

NIH-PA Author Manuscript

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

Distribution of GI risk factors and recent/concomitant therapies across levels of the exposure.

Variable	Coxib Users	NSAID Users	Prevalence Difference and 95% CI
U*	***E*[*U**|*X* = 1]**	***E*[*U**|*X* = 0]**	***E*[*U**|*X* = 1] – *E*[*U**|*X* = 0]**
Patient Characteristics			
Female Gender	85.89%	81.11%	4.79% (4.09% – 5.48%)
Age ≥ 75	75.08%	65.28%	9.80% (8.95% – 10.64%)
History of GI Bleed	1.71%	1.11%	0.60% (0.39% – 0.81%)
History of Peptic Ulcer Disease	3.71%	2.41%	1.29% (0.99% – 1.60%)
History of Coronary Hospitalization or Procedure	16.42	14.76	1.67% (1.00% – 2.33%)
Recent/Concomitant Rxs			
Concomitant Use of GI-protective Drugs	5.08%	4.00%	1.08% (0.71% – 1.46%)
Recent Use of GI-protective Drugs	27.34%	20.41%	6.93% (6.16% – 7.70%)
Recent Use of Glucocorticoids	8.73%	7.80%	0.94% (0.44% – 1.44%)
Recent Use of Warfarin	13.25%	6.53%	6.71% (6.19% – 7.23%)

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 2

Distribution of GI risk factors and recent/concomitant therapies across levels of the instrument.

Variable	**Coxib Preference**	**NSAID Preference**	**Prevalence Difference and 95% CI**
U*	***E*[*U**|*Z* =1]**	***E*[*U**|*Z* = 0]**	***E*[*U**|*Z* = 1] – *E*[*U**|*Z* = 0]**
Patient Characteristics			
Female Gender	84.43%	84.13%	0.30% (−0.48% – 1.08%)
Age ≥ 75	72.59%	71.42%	1.17% (0.21% – 2.13%)
History of GI Bleed	1.46%	1.39%	0.06% (−0.19% – 0.32%)
History of Peptic Ulcer Disease	3.25%	3.05%	0.20% (−0.17% – 0.57%)
History of Coronary Hospitalization or Procedure	15.67%	14.98%	0.70% (−0.70% – 1.46%)
Recent/Concomitant Medications			
Concomitant Use of GI-protective Drugs	4.61%	4.58%	0.03% (−0.04% – 0.04%)
Recent Use of GI-protective Drugs	24.52%	24.20%	0.32% (−0.60% – 1.24%)
Recent Use of Glucocorticoids	8.33%	8.03%	0.30% (−0.29% – 0.88%)
Recent Use of Warfarin	11.80%	9.99%	1.81% (1.15% – 2.47%)

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 3

Imbalance of GI risk factors across levels of the instrument relative to imbalance across levels of treatment.

Variable	**Prevalence Difference Ratio**
U*	$\frac{E[U^*\|Z=1]-E[U^*\|Z=0]}{E[U^*\|X=1]-E[U^*\|X=0]}$
Patient Characteristics	
Female Gender	6%
Age ≥ 75	12%
History of GI Bleed	10%
History of Peptic Ulcer Disease	16%
History of Coronary Hospitalization or Procedure	42%
Recent/Concomitant Medications	
Concomitant Use of GI-protective Drugs	3%
Recent Use of GI-protective Drugs	5%
Recent Use of Glucocorticoids	32%
Recent Use of Warfarin	27%

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

Table 4

Strength of the instrument in subgroups defined by observed risk factors

Variable ***U****	**Instrument Strength** $\boldsymbol{E[X \mid U^* = 1, Z = 1]}$ $\boldsymbol{- E[X \mid U^* = 1, Z = 0]}$	**95% CI**
Patient Characteristics		
Full Population	22.8%	21.8 – 23.8%
Female Gender	22.1%	23.6 – 28.8%
Age ≥ 75	23.1%	21.9 – 24.2%
History of GI Bleed	18.2%	10.3 – 26.2%
History of Peptic Ulcer Disease	18.9%	13.5 – 24.3%
History of Coronary Procedure or Hospitalization	22.6%	20.0 – 20.5%
Recent/Concomitant Medications		
Concomitant Use of a GI Protective Drugs	18.7%	14.1 – 23.2%
History of GI Protective Drug Use	20.1%	18.1 – 22.0%
Recent Use of Glucocorticoids	19.7%	16.2 – 23.2%
Recent Use of Warfarin	14.9%	12.1 – 17.7%