Volume Title: Measuring and Modeling Health Care Costs

Volume Author/Editor: Ana Aizcorbe, Colin Baker, Ernst R. Berndt, and David M. Cutler, editors

Volume Publisher: University of Chicago Press

Volume ISBNs: 978-0-226-53085-7 (cloth); 978-0-226-53099-4 (e-ISBN)

Volume URL: http://www.nber.org/books/aizc13-1

Conference Date: October 18-19, 2013

Publication Date: February 2018

Chapter Title: A Cautionary Tale in Comparative Effectiveness Research: Pitfalls and Perils of Observational Data Analysis

Chapter Author(s): Armando Franco, Dana P. Goldman, Adam Leive, Daniel McFadden

Chapter URL: http://www.nber.org/chapters/c13104

Chapter pages in book: (p. 55 – 80)

2

A Cautionary Tale in Comparative Effectiveness Research
Pitfalls and Perils of Observational Data Analysis

Armando Franco, Dana P. Goldman, Adam Leive, and Daniel McFadden

2.1 Introduction

Comparative effectiveness research (CER) has become increasingly important for payers and policymakers as health care costs continue to grow rapidly. Such research is usually based on the results of randomized controlled trials (RCTs). However, determining whether the "blue pill" or the "red pill" is more effective (and for whom) can be time consuming, challenging, and expensive. Serious but rare side effects may be missed by an underpowered RCT, making surveillance important. Moreover, there may be interest in comparing the benefits and costs of competing drug treatments outside of a clinical trials setting; the populations studied in trials are almost certainly not representative of all patients who will ultimately consume the drug, and trial participants may also behave differently than people "in the real world."

Armando Franco is a health care data scientist at the University of California, Berkeley. Dana P. Goldman is the Leonard D. Schaeffer Chair and Distinguished Professor at the University of Southern California and a research associate of the National Bureau of Economic Research. Adam Leive is assistant professor of public policy and economics at the University of Virginia. Daniel McFadden is a professor of the graduate school at the University of California, Berkeley, the Presidential Professor of Health Economics at the University of Southern California, and a research associate of the National Bureau of Economic Research.

This research was supported by the Behavioral and Social Research program of the National Institute on Aging (grants P01AG033559 and RC4AG039036), with additional support from the E. Morris Cox Fund at the University of California, Berkeley. We are grateful to David Meltzer and participants at the 2013 NBER/CRIW Conference on Measuring and Modeling Health Care Costs for helpful comments. We thank Patricia St. Clair for her support of the data construction effort and Florian Heiss and Joachim Winter for helpful comments and discussions on an earlier draft. For acknowledgments, sources of research support, and disclosure of the authors' material financial relationships, if any, please see http://www.nber.org/chapters/c13104.ack.

Conducting CER using observational data presents a potential solution to some of these problems. Perhaps the most promising source of observational data comes from insurance claims. Claims data generally include large sample sizes that allow for more precise estimates of treatment effects than those possible through RCTs. The greater statistical power of claims data may also permit the detection of rare events not possible with RCTs, such as side effects or interactions with other drugs. Moreover, some side effects may occur after the conclusion of an RCT evaluation. The longer time frame of some claims data is thus another reason why claims data may be particularly well suited to identify drug risks. In addition, RCTs are very expensive compared to accessing observational data.

The Food and Drug Administration (FDA)'s Mini-Sentinel Project is perhaps the most prominent example of pharmacovigilance. Using insurance data on roughly 100 million patients and 2.9 billion prescription drug fills, the project seeks rapid dissemination of safety issues associated with drugs and adverse-reported events. Of course, the lack of randomization is the trade-off for the greater statistical power and detailed information on past medical histories and drug consumption patterns available with claims data. Accordingly, making valid inferences between treatments becomes much harder and there is a greater need for empirical methods to focus on causality.

The purpose of this chapter is to discuss some of the key methodological issues involved in using claims data to conduct CER. Using Medicare claims data from Parts A, B, and D between 2006 and 2009, we discuss the inherent challenges in using claims data and illustrate these issues by analyzing angiotensin II receptor blockers (ARBs) drugs for hypertension. We first document sample contamination observed in the claims—substantial crossover between therapies and discontinuation of hypertension treatment. We then discuss the implications of such sample contamination for CER. We employ two methods to deal with the nonrandom treatment assignment. First, we assume that physicians may have underlying propensities to prescribe ARBs, conditional on observed patient characteristics, and we examine the relationship between ARB prescription propensity and our outcomes at the physician level. Our rationale is that if physicians have underlying propensities to prescribe certain hypertension drugs but patients cannot observe such propensities, then we may view the initial prescription as random. Our second approach is to instrument for individual treatment choice using relative price differences between ARBs and substitute hypertension drugs.

Our evaluation of drug treatment effectiveness focuses on two outcomes: stroke and cancer. Stroke can result from uncontrolled high blood pressure. In general, RCTs have found little evidence of any difference in strokes between ARB users and users of other hypertension drugs, with some indication of fewer strokes among certain groups of ARB users (Wang, Franklin, and Safar 2007; Dahlof et al., 2002; Strauss and Hall 2009). Our second outcome is cancer, which was flagged by the FDA in 2010 as a potential

adverse effect of ARBs. Despite the FDA later determining no link between cancer and ARBs, the safety of ARBs is still debated internally within the FDA (Burton 2013). A key objective of our study is to use observational data to investigate potential side effects or rare events not easily detected in RCTs, especially when consensus of such effects is lacking. We examine strokes in an attempt to validate our methods to assess the relationship between ARBs and cancer. In particular, if we can replicate the results of RCTs for strokes, we would be more confident that the association between ARBs and cancer can be interpreted as causal. This approach is similar to how RCTs are sometimes used in the policy evaluation literature to test the out-of-sample validity of structural econometric models (Todd and Wolpin 2006).

Using these strategies to identify treatment effects, we find mixed evidence that ARBs lead to higher cancer rates and some evidence that ARBs lead to higher stroke rates compared to other hypertension drugs. The increase in strokes associated with ARBs is contrary to evidence from RCTs that demonstrate, if anything, a modest reduction in strokes. As an additional falsification test, we rerun our analysis with a diagnosis of pain as the dependent variable—under the assumption that there should be no relationship between pain and choice of antihypertensive. However, we find that ARBs are associated with more pain diagnoses and the magnitudes of the effects are often larger than those for our main outcomes, possibly due to omitted variable bias from individual-level socioeconomic factors. Combined, these results suggest the relationship between ARBs and cancer should not be interpreted as causal.

The news is not all bad, though. While we document some pitfalls in using observational data to conduct CER, our results also suggest value to using relative price as an instrument for drug treatments, given how well our relative price measure predicts drug use. The remainder of the chapter is organized as follows. Section 2.2 provides background on ARBs and their possible link with cancer. Section 2.3 discusses sample selection and sample contamination, which occurs when people either discontinue treatment or switch treatments. Selection into treatment is discussed in section 2.4. Two robustness checks are presented in section 2.5. Section 2.6 compares our findings with those of RCTs. We briefly conclude in section 2.7.

2.2 Background on Hypertension, ARBs, and Cancer Risk

Hypertension is clinically defined as having either high levels of systolic blood pressure (above 140 millimeters of mercury) or diastolic blood pressure (over 90 millimeters of mercury). There is no single cause for hypertension; blood pressure levels are affected by the levels of water, salt, and hormones in the body as well as the condition of the kidneys, nervous system, and blood vessels. As people age, their blood vessels become stiffer, which increases blood pressure. Other risk factors include obesity, diabetes,

smoking, and being an African American.[1] The major health consequences of hypertension are stroke and heart disease.

There are a variety of drugs used to treat hypertension. In this chapter, we compare ARBs to other common classes of treatment. In some models, we compare ARBs to angiotensin-converting enzyme (ACE) inhibitors alone, since these drugs represent the closest substitutes, with both operating through the effect of angiotensin (a peptide hormone, angiotensin causes vasoconstriction and also releases aldosterone, both of which lead to an increase in blood pressure). Drug classes we analyze and the mechanism by which they affect blood pressure are summarized below:

- Angiotensin-II Receptor Blockers (ARBs): relaxes blood vessels by blocking the action of angiotensin II.
- ACE inhibitors: prevents the formation of angiotensin II.
- Beta blockers: blocks the effects of the hormone epinephrine, leading the heart to beat more slowly.
- Diuretics: removes salt and water from the body by inducing the kidneys to put more salt into urine, thereby decreasing pressures on artery walls.
- Calcium channel blockers: widens and relaxes blood vessels through influencing the muscle cells in the walls of arteries.
- Other antihypertensives (e.g., vasodilators): opens blood vessels by preventing muscles from tightening and by stopping the walls in the arteries from narrowing.

In July 2010, the FDA issued a safety alert in response to a meta-analysis by Sipahi et al. (2010) suggesting a possible risk of cancer associated with use of ARBs (Food and Drug Administration 2010). The meta-analysis used data on 60,000 patients and found a small but statistically significant increase in new cancer cases among ARB users: 7.2 percent compared to 6.0 percent.[2] The authors considered breast, prostate, and lung cancers and grouped all remaining cancers together. Over the next year, the FDA pursued further analysis based on 156,000 patients enrolled in RCTs. In June 2011, the FDA released its finding that ARBs do not pose a greater risk of cancer relative to other hypertension drugs (Food and Drug Administration 2011).

1. For more background information on risk factors, see http://www.nhlbi.nih.gov/health/health-topics/topics/hbp/.

2. Cancer was not a prespecified endpoint in several of the trials analyzed by Sipahi et al. (2010). This suggests the difference in cancer deaths observed may have resulted from differential effects of drug use on cancer detection. In particular, ARBs may cause more side effects that prompt a diagnostic workup, which ultimately reveal the presence of cancer, even though there is no causal biological mechanism between ARBs and cancer. We investigated this possibility by calculating prevalence rates of major diagnostic cancer tests among people taking different drugs. To keep the comparison as clean as possible, we also only examined one-year incident cases—people who did not take any hypertension drug in 2006 and began taking one in 2007—for those on monotherapy (i.e., treatment with a single drug). We did not find evidence of higher rates of diagnostic cancer exams among ARB users compared to other classes of drugs, adjusting for differences in age and sex across drug classes.

However, by some accounts, the debate remains unresolved. In May 2013, the *Wall Street Journal* reported on dissent within the FDA, where a senior FDA regulator conducted additional analysis with individual-level trial data that estimated an increased cancer risk of over 20 percent among patients taking ARBs (Burton 2013). The research was rebuked by top officials within the FDA, but this rare internal dispute illustrates the lack of consensus on the side effects of ARBs.

The presence of many alternative drug options to treat hypertension increases the value of understanding side effects related to ARBs. If there were no other viable treatments, then rational patients should be willing to accept more risk of potential side effects. But with a plethora of alternative therapy choices, the (expected) benefits of ARBs may not be worth the risk of cancer or other side effects. Although data is currently unavailable to determine how the FDA's 2010 warning affected drug use, it seems likely that determining within a shorter time frame that ARBs do not cause cancer would generate important benefits to patients. The following sections of the chapter describe our attempts to analyze these issues using observational data.

2.3 Sample Selection and Contamination

Our sample is constructed from individual claims data from Medicare parts A, B, and D between 2006 and 2009. We examine enrollees in stand-alone prescription drug plans (PDPs) only because complete Medicare claims for enrollees in Part C are unavailable. Hypertension cases are classified by use of at least one drug commonly used to treat high blood pressure. Patients taking hypertension drugs for less than thirty days are excluded from our sample. We use the word "treatment" to refer to prescription drugs, but recognize there are other forms of treatment for hypertension, such as exercise and dieting. Ignoring unobservable activities like exercise will only be problematic for our results to the extent that these activities vary differentially across drug classes. One way such differential variation could occur is if certain drugs, due to higher prices, are consumed mainly by patients with higher incomes or education levels and such patients also exercise more often. The inability to control for individual-level socioeconomic factors is an important limitation of our study and an issue to which we return in the discussion.

2.3.1 Sample Selection (Left Censoring)

It is common for patients to take hypertension drugs before age sixty-five, when most beneficiaries become eligible for Medicare. We refer to those already on hypertension drugs when they are first observed in claims as "prevalent cases." For these patients, claims data do not permit the researcher to observe the duration of current treatment or patterns of past treatments prior to age sixty-five. Clearly, this unobservability is problematic for classifying the presence and intensity of drug consumption. An alternative to

this left-censoring problem is to restrict the sample to patients enrolled in 2006 who initiate hypertension treatment in 2007. We refer to this group as "incident cases" with a "one-year window." Since hypertension is a chronic condition, incident cases provide a cleaner comparison between ARBs and other drug classes because patients are likely first-time drug users.

Left censoring is a serious analytical problem for evaluating drug treatments, although not necessarily a serious empirical problem. With left censoring, a beneficiary's prior history in terms of both drug consumption and health outcomes is unobserved. Using external information on incident cases that can be linked to Medicare claims, such as the Health and Retirement Study (HRS), to impute "back-dated" information for prevalent cases is not an attractive option, because the real analytical danger is the unobservability of switching between drug classes, to be discussed below. To the extent this unobserved switching is correlated with unobserved health status, using prevalent cases creates analytic problems for researchers that are difficult to surmount. Nevertheless, left censoring should not be an empirical problem because the number of incident cases will grow over time with additional waves of data.

2.3.2 Contamination Bias

An RCT has a very powerful instrument (randomization) that has a strong effect on treatment assignment. Even in RCTs, however, patients often change treatment as their diseases get managed in the trial. In an observational study, the goal is to classify patients based on patterns of drug consumption, and to group patients with similar histories of drug consumption together into pseudotreatment and control arms. For reasons of interpretability and statistical power, it would be ideal to have a small number of treatment and control groups. However, there are several challenges to a precise assignment of such groups. If there were only two competing therapies, classification would be relatively simple with only three possible combinations: drug 1 alone, drug 2 alone, or both drug 1 and drug 2. However, dimensionality problems quickly arise when more than a few drugs can be taken possibly in combination. Moreover, the order in which drugs are taken further complicates analysis. Some drugs may generally be taken as first-line therapy, while others are prescribed as therapies of "last resort." Indeed, this is the case with hypertension; diuretics are often prescribed first, in line with the Joint National Committee's recommendation, and ARBs are more often prescribed after the patient has tried other drug therapies. By including more therapy groups, we trade off ease of interpretation and greater statistical power against contamination bias resulting from heterogeneity within any single group.

How serious is this problem? We find that most patients discontinue their initial drug treatment within the first year. Table 2.1 illustrates that over two-thirds of patients stop their initial treatment within one year for both prevalent and incident cases—where discontinuation is defined as having a

Table 2.1 Adherence of initial treatment by drug class

Therapy	Initial treatment	On treatment through one year	Percent of initial users	On treatment through two years	Percent of initial users
		Prevalent cases, 2006–2009			
ACE inhibitors	608,641	78,932	13	51,973	9
ARBs	300,583	32,357	11	20,193	7
Diuretics	809,900	130,917	16	98,432	12
Calcium channel blockers	474,429	63,824	13	41,692	9
Beta blockers	1,010,950	219,443	22	179,434	18
Other antihypertensives	48,119	7,809	16	5,327	11
		Incident cases, 2007			
ACE inhibitors	152,561	40,033	26	36,814	24
ARBs	56,005	11,111	20	9,903	18
Diuretics	181,630	54,289	30	50,413	28
Calcium channel blockers	91,490	23,144	25	20,849	23
Beta blockers	264,639	84,836	32	78,407	30
Other antihypertensives	9,216	2,423	26	2,144	23

gap in prescription coverage of more than thirty days. However, conditional on maintaining treatment through the first year, most continue through the second year. This pattern suggests that side effects for a subset of patients or heterogeneity in treatment response may drive adherence patterns. It also suggests that the (selected) groups of patients who have maintained initial therapy for one year might make for adequate treatment and control groups.

However, polytherapy also poses a problem. Table 2.2 documents that conditional on not discontinuing initial treatment in the first year, between 14 and 26 percent of patients take at least one other hypertension drug at some point. Between 9 and 15 percent of such patients take at least two other drugs. So not only do people often discontinue their initial treatments, but those who adhere often take multiple treatments concurrently.

Roughly one-third of incident cases are on the same, single monotherapy throughout the sample period (results not shown). The majority of these patients are on either beta blockers or ACE inhibitors. Among combinations of drugs, beta blockers with diuretics are the most common. However, the mix of drugs taken varies widely. We find rates of polytherapy are similar to those cited in the 2003 JNC report, which documents more than two-thirds of patients require at least two drugs to control hypertension.

Other studies find greater rates of adherence than we document in table 2.1, however. A meta-analysis by Matchar et al. (2008) finds one-year adherence rates for ARBs and ACE inhibitors range between 40 and 60 percent. Relaxing our restriction that patients must not discontinue treatment for more than thirty days to be considered adhering to ninety days brings our

Table 2.2 Combination therapy by drug class

Total number of therapies ever taken, incident cases 2007–2008	On initial treatment through one year	On one plus other drugs during first year	Percent of one-year users	On two plus other drugs during first year	Percent of one-year users
ACE inhibitors	78,932	17,692	22	11,033	14
ARBs	32,357	6,060	19	4,912	15
Diuretics	130,917	34,060	26	16,810	13
Calcium channel blockers	63,824	10,360	16	5,596	9
Beta blockers	219,443	51,699	24	26,947	12
Other antihypertensives	7,809	1,088	14	765	10

estimates closer to other studies, but they are still at least 10 percentage points lower across drug classes. We suspect that the main driver behind our higher rates of discontinuation is the greater cost sharing under Part D. Simple preliminary analysis reveals that hypertension use decreases in the "doughnut hole," and this is consistent with more general work on drug consumption in Part D plans by Joyce, Zissimopolous, and Goldman (2013).

As an example of how various drugs are used in sequence, figure 2.1 displays the usage rates of drug treatments among people who ever take an ARB. Over 40 percent of ARB users take another drug before starting ARBs, with most taking either diuretics or beta blockers. The figure clearly reveals that ACE inhibitors substitute for ARBs, as is clinically indicated. There is also evidence that diuretics tend to complement ARB use. Finally, many patients who discontinue ARB use subsequently take another drug.

The multitude and timing of drug consumption patterns issues raise the question of how to measure drug consumption in empirical models. We follow two different approaches. The simplest is to classify patients as using a drug if they have ever had at least two fills of the drug, even if they have previously taken other antihypertensives. We term this group "ever users." The second way is to classify patients based on the initial drug therapy prescribed. This represents an intent-to-treat approach. In both of these approaches, treatment is measured as an indicator function.[3] One might be tempted to restrict attention to incident cases who maintain monotherapy for at least twelve months. Doing so, however, would imply throwing away 71 percent of the observed incident cases and 84 percent of all cases (including prevalent hypertension), and this case deletion obviously occurs in a nonrandom way.

3. A third approach is to model the cumulative exposure to the drug, using a function with an exponential rate of decay. We experimented with this approach by running a series of survival models measuring duration until cancer or stroke instead of using linear IV regressions. The results were qualitatively similar.

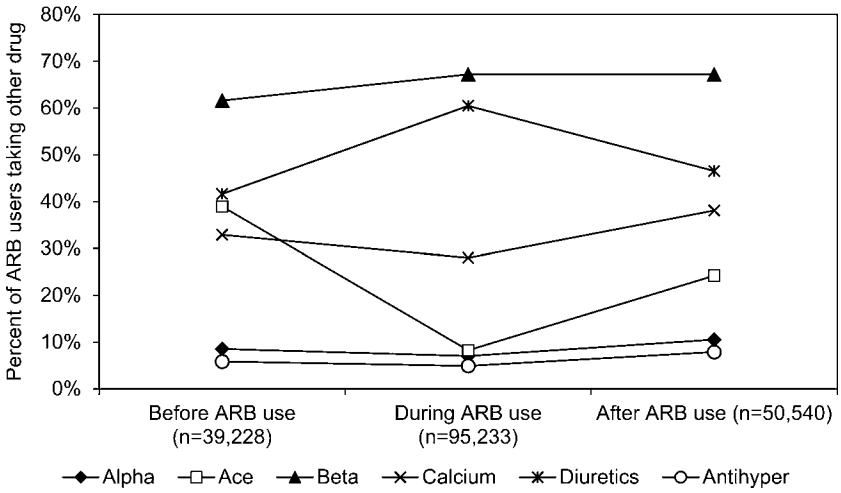

Fig. 2.1 Empirical sequencing of drug use for ARB users

Some of this can be seen in tables 2.3 and 2.4, which present descriptive statistics among "ever users" for prevalent and incident cases, respectively. These unconditional means reveal important differences by drug class. Cancer rates are lowest among ARB and ACE users. The fact that there is much less variation in cancer rates by class among incident users suggests that past history may be very important to determining outcomes. For example, the difference between the highest and lowest cancer rates among incident cases is 33.4 per 10,000, but is 81.8 per 10,000 among prevalent cases. Death rates are considerably lower for ARB users in both prevalent and incident cases.

2.4 Selection into Treatment

One of the fundamental challenges to using observational data for CER is that treatment assignment is not random. Instead, drug treatment is a decision made between the physician and patient. The decision is likely based on many characteristics of the patient, some of which may be unobserved and may also affect cancer risk. We pursue two approaches to deal with selection into treatment: (a) examine how physician propensities to prescribe ARBs are correlated with health outcomes, and (b) estimate linear instrumental variables (IV) regressions using relative price to predict treatment choice. These approaches differ conceptually in the power they ascribe to each side of the physician-patient relationship; the first approach implicitly assumes the physician has control over which drug the patient takes and has latent preferences for prescribing certain drugs. By using the out-of-pocket (OOP) cost the patient pays for drugs as an instrument, the second approach implicitly treats the patient as the decision maker and price as the key factor

Table 2.3 **Descriptive statistics, prevalent cases**

Therapy		Age	Sex (1 = female, 0 = male)	Cancer (per 10,000)	Stroke (per 10,000)	Death (per 10,000)	Avg. no. chronic conditions
ARBs	mean	77.5	0.722	182.0	1,818.4	526.7	4.99
(N = 184,067)	s.d.	7.4	0.448	1,336.9	3,857.1	2,233.7	2.30
ACE inhibitors	mean	77.5	0.647	187.2	1,907.1	736.1	4.78
(N = 406,100)	s.d.	7.6	0.478	1,355.5	3,928.6	2,611.3	2.33
Beta blockers	mean	78.1	0.677	251.0	1,947.2	920.8	5.03
(N = 648,429)	s.d.	7.6	0.468	1,564.3	3,959.9	2,891.4	2.27
Calcium channel blockers	mean	78.2	0.725	251.6	1,981.9	840.4	4.93
(N = 496,334)	s.d.	7.7	0.447	1,566.1	3,986.4	2,774.5	2.31
Other antihypertensives	mean	78.7	0.759	263.8	2,897.2	1,084.8	5.59
(N = 85,437)	s.d.	7.7	0.428	1,602.6	4,536.4	3,109.9	2.34
Diuretics	mean	78.4	0.722	272.2	1,821.3	1,392.4	4.94
(N = 909,615)	s.d.	7.9	0.448	1,627.3	3,859.5	3,461.9	2.34

Table 2.4 **Descriptive statistics, incident cases**

Therapy		Age	Sex (1 = female, 0 = male)	Cancer (per 10,000)	Stroke (per 10,000)	Death (per 10,000)	Avg. no. chronic conditions
ARBs	mean	77.4	0.669	119.7	1,509.6	315.0	4.33
(N = 23,143)	s.d.	7.3	0.470	1,087.4	3,580.2	1,746.6	2.29
ACE inhibitors	mean	77.4	0.587	124.5	1,613.5	452.3	4.19
(N = 63,763)	s.d.	7.4	0.492	1,108.9	3,678.5	2,078.0	2.29
Beta blockers	mean	78.3	0.615	144.2	1,705.1	620.5	4.61
(N = 104,901)	s.d.	7.5	0.487	1,192.2	3,760.8	2,412.5	2.26
Calcium channel blockers	mean	78.4	0.660	153.1	1,773.3	543.0	4.43
(N = 64,267)	s.d.	7.6	0.474	1,227.8	3,819.5	2,266.1	2.31
Other antihypertensives	mean	79.2	0.680	148.9	2748.4	842.4	5.06
(N = 9,268)	s.d.	7.8	0.467	1,211.0	4,464.6	2,777.6	2.36
Diuretics	mean	78.7	0.663	148.6	1,567.7	932.3	4.45
(N = 130,572)	s.d.	7.9	0.473	1210.1	3635.9	2907.5	2.36

influencing her decision. These approaches thus attempt to identify the effect of ARBs on health outcomes along different margins.

2.4.1 Physician Propensity to Prescribe ARBs

The rationale behind our first approach using physician propensities is to view the initial physician-patient match as random. More precisely, if physicians have underlying propensities to prescribe certain hypertension drugs, conditional on patient characteristics, but patients cannot observe such propensities and thus do not choose physicians based on them, then we may view the initial prescription as random. In this sense, physicians with a greater propensity to prescribe ARBs are analogous to the randomly assigned treatment group of an RCT. (We fully recognize that many patients may shop for doctors in certain clinical circumstances, thereby violating this assumption. However, in the case of antihypertensive prescribing, such an assumption seems more plausible.)

To examine physician-prescribing decisions, we limit our sample to include only initial therapy choice and do not allow for switching or adding therapies. To fully capture a physician's prescribing tendencies, the sample is restricted to physicians with at least thirty patients on hypertension drugs. Our final data set for this analysis is composed of 1,176,311 patients and 25,477 physicians, with the average physician treating forty-six patients for hypertension.

Following the theoretical model derived by Chandra and Staiger (2011), we model physician's propensity θ to prescribe ARBs based on the fact that some physicians might have an underlying tendency to prescribe ARBs. We regress whether a patient receives ARBs against her chronic conditions, basic demographics, and the physician's propensity effect. Here θ is assumed to be a normally distributed random effect with mean μ and σ^2. The functional form of $F(.)$ is taken to be logistic. We estimate the mixed-effects model and recover estimates of both μ and σ and use them to construct a posterior distribution of the estimated θ for each physician, which we then use to regress against death and cancer rates.

2.4.2 Results

Figure 2.2 plots the cancer rate for each physician's set of hypertension patients against the physician's propensity to prescribe ARBs. Our measure of cancer includes breast, prostate, lung, colorectal, and endometrial cancers, which account for the large majority of cancer deaths. Here, the diameter of each circle represents the number of patients with at least one claim to that particular physician. There appears to be a slight positive relationship between ARB use and cancer, which is statistically significant at the 1 percent level based on the large sample size. Perhaps more interestingly, figure 2.3 displays a stronger negative relationship between death rates and

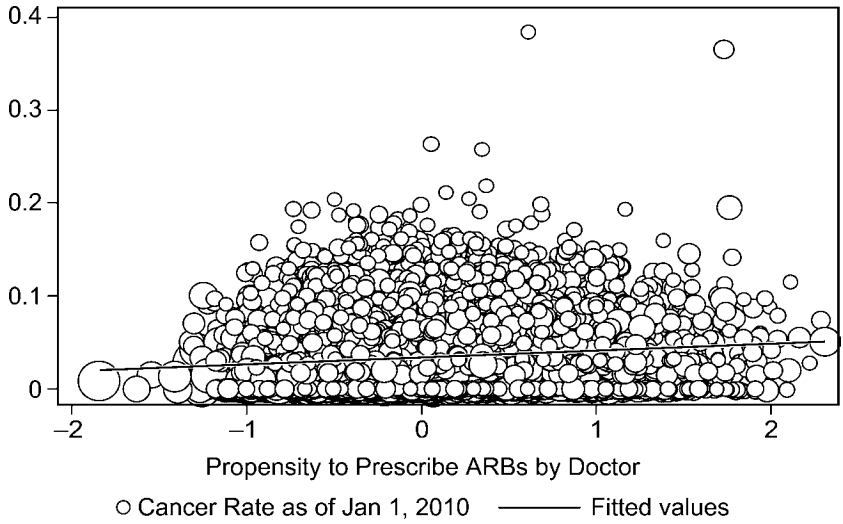

Fig. 2.2 Unconditional relationship between cancer rate and propensity to prescribe ARBs

Note: For 25,000 prescribers with 30–500 patients: events from 2006 to 2008. Beta = .007 and significant at the 1 percent level.

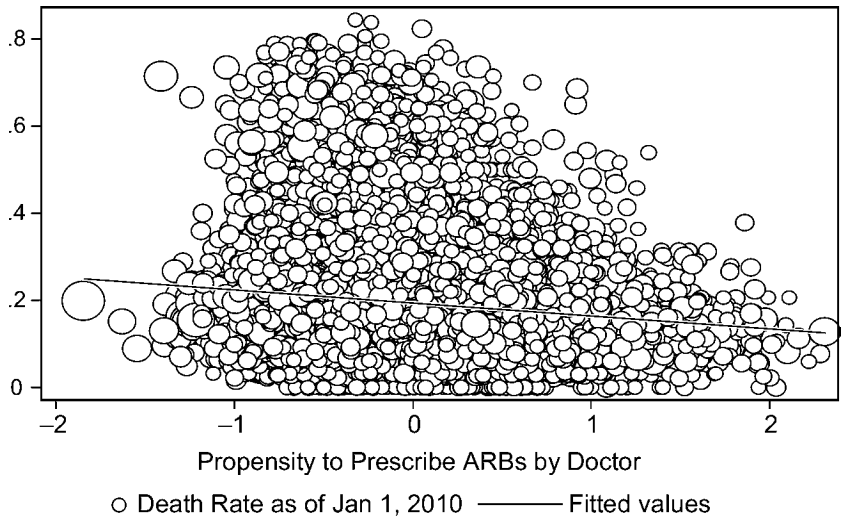

Fig. 2.3 Unconditional relationship between death rate and propensity to prescribe ARBs

Note: For 25,000 prescribers with 30–500 patients: events from 2006 to 2008. Beta = −.03 and significant at the 1 percent level.

Table 2.5 **Simulations of reducing prescriber propensity**

Cancer	No. predicted cancer cases	Percent of total	Decreased no. of cancer cases	Percent of cancer cases
No restrictions	39,530	3.36	n/a	n/a
Restrict those with propensity in the top 75% to 25%	35,822	3.05	3,708	10.35
Restrict those with propensity greater than 2 std. dev.	39,455	3.35	75	0.19
Restrict those with propensity greater than 1 std. dev.	39,096	3.32	434	1.10
Death	**No. predicted deaths**	**Percent of total**	**Increased no. of deaths**	**Percent of deaths**
No restrictions	227,415	19.33	n/a	n/a
Restrict those with propensity in the top 75% to 25%	242,489	20.61	15,074	6.22
Restrict those with propensity greater than 2 std. dev.	227,718	19.36	303	0.13
Restrict those with propensity greater than 1 std. dev.	229,179	19.48	1,764	0.77

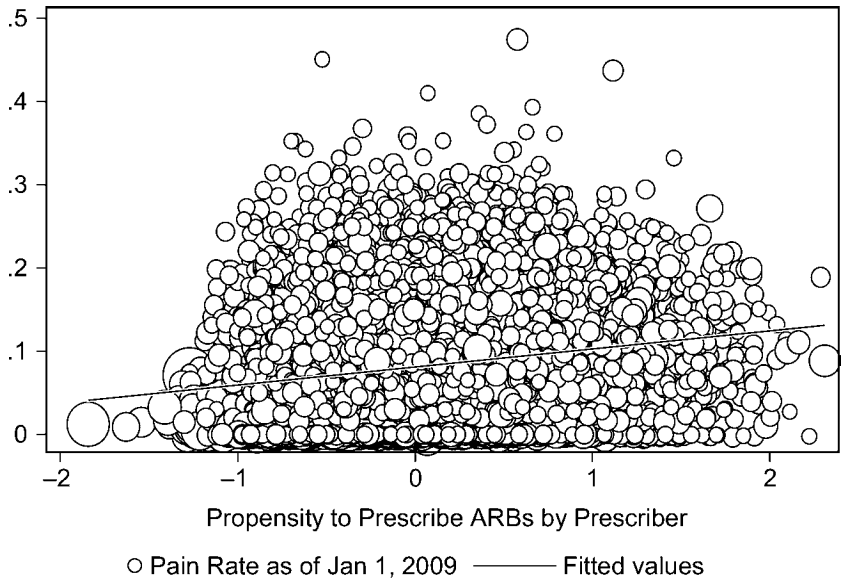

Fig. 2.4 Unconditional relationship between pain rate and propensity to prescribe ARBs

Note: For 25,000 prescribers with 30–500 patients: events from 2006 to 2008. Beta = .022 and significant at the 1 percent level.

the propensity to prescribe ARBs. This may be due to an omitted variable, or it may be explained by competing risks; ARBs may increase cancer rates but reduce overall death rates due to fewer occurrences of some other disease(s).

Taking these estimates as "true" causal estimates, the natural question is how many cancer cases and deaths would be avoided (or incurred) if physicians lowered their ARB-prescribing tendencies. As an illustration, table 2.5 displays the results of several simulations. If only the physicians in the right tail of the distribution were to reduce their prescribing rates to two standard deviations above the mean, the changes in outcomes are modest. If physicians above the 25th percentile reduced their prescribing rates to those of the 25th percentile, the implications are, not surprisingly, more dramatic. Doing so would decrease cancer rates by more than 10 percent and increase death rates by 6 percent. These are sizable numbers: saving 3,700 cancer cases through less ARB prescribing involves sacrificing 15,000 lives—hardly an attractive trade-off.

We do not, however, believe these are causal. When we replicate these scatterplots using pain as the dependent variable in the regressions—an outcome we assume is clinically independent of ARB prescribing—we actually find a stronger relationship between pain and ARB use than that for cancer (figure 2.4).

This finding unfortunately raises questions about the validity of using physician treatment propensities to identify CER models, at least in the case of antihypertensives.

2.4.3 Linear IV Models

Our second approach follows more standard, economic techniques to deal with causality using observational data: instrumental variables regressions. We use the ratio of the average OOP cost of ARBs to the average OOP cost of ACE inhibitors at the regional level using cost-sharing information at the plan level. In particular, we calculate the average OOP based on the copayment or the coinsurance rate of the plan multiplied by the average total cost of ARBs and ACE inhibitors at the regional level. We choose to examine ARBs with ACE inhibitors alone for a cleaner comparison since these two drugs represent the closest substitutes for one another. There is little reason to believe that the price difference between ARBs and ACE inhibitors at the regional level should be correlated with health outcomes.

The ARBs are nearly always more expensive than ACE inhibitors. In fact, only 10 of 10,087 county-years have cheaper ARBs. On average, ARBs are over 4.5 times as expensive as ACE inhibitors, and the standard deviation for this ratio is 1.5. As will be shown in the first-stage regressions below, there is still enough variation for price to serve as a good predictor of ARB use.[4]

Our first-stage regression is a probit of treatment on relative price, a vector of chronic conditions (diabetes, heart disease, heart failure, depression, Alzheimer's disease, glaucoma, ischemic heart disease, chronic obstructive pulmonary disease, pelvic/hip fracture, osteoporosis, rheumatoid arthritis/osteoarthritis, cataracts, and chronic kidney disease), age, sex, county-level socioeconomic factors (percent home ownership, education levels, percent African American, average income, percent below poverty line, unemployment rate, percent married, and percent foreign born, all among adults), and state fixed effects. The second stage is a linear probability model where the dependent variable is an indicator of the health outcome in the two years following the first hypertension prescription. The first hypertension prescription is calculated as the first ARB fill among ARB users or the first fill for another drug class among non-ARB users. We adjust standard errors in all regressions by clustering at the county level.

Our rationale for examining the incidence of outcomes up to two years posttreatment is to adjust the duration of exposure for the sequencing of drug classes. As described earlier, some treatments, such as ARBs, are often initiated after trying other therapies first. This mechanically reduces the amount of time spent on ARBs compared to other drugs over the time

4. Other percentiles of the distribution of relative price are as follows: 1st percentile—2.1; 25th percentile—3.7; 75th percentile—5.5; and 99th percentile—9.6.

Table 2.6 **First-stage IV regressions**

	Prevalent cases		Incident cases	
	All drug users (1)	ACE or ARB monotherapy only (2)	All drug users (3)	ACE or ARB monotherapy only (4)
	Treatment definition: Patient ever used drug			
Price	−0.039	−0.035	−0.024	−0.039
	(−60.77)	(−16.83)	(−12.36)	(−7.51)
F-stat	520.99	31.35	55.09	7.75
N	1,406,463	95,386	153,904	19,466
	Treatment definition: Intent to treat			
Price	−0.018	−0.035	−0.020	−0.039
	(−33.02)	(−16.83)	(−11.87)	(−7.51)
F-stat	271.5	31.55	40.35	7.75
N	1,406,469	95,386	153,904	19,466

Notes: Cluster-adjusted robust *T*-statistics in parentheses. Regressions also include indicators for age, sex, diabetes, heart disease, heart failure, depression, Alzheimer's disease, glaucoma, ischemic heart disease, chronic obstructive pulmonary disease, pelvic/hip fracture, osteoporosis, rheumatoid arthritis/osteoarthritis, cataracts, and chronic kidney disease, state fixed effects, and the following county-level variables: percent home ownership, percent with high school degree, percent with some college or two-year degree, percent with four-year college degree or beyond, percent African American, average income, percent below poverty line, unemployment rate, percent married, and percent foreign born, all among adults.

period we observe, and so there is also less time to be diagnosed with cancer or stroke.[5] We only examine patients with a full two-year window after they begin antihypertensives. Since our data extends to December 31, 2009, we exclude anyone beginning hypertension treatment in 2008 or later. By imposing a standard level of follow-up across all drug classes, this improves the comparability of different drugs even when some are routinely prescribed first.

The first-stage regressions in table 2.6 show that price is a strong predictor of ARB use. For both definitions of treatment, we run four sets of regressions that divide the population based on drug use. The first column includes all patients who are prevalent cases, representing the largest number of beneficiaries. The second column includes prevalent cases who either only take ARBs or only take ACE inhibitors. This dramatically reduces sample size, given the popularity of combination therapy. The third column includes all incident cases, comprising slightly more than 10 percent of prevalent cases.

5. We ran regressions with the number of days without the outcome as the dependent variable along the lines of Basu et al. (2007), but this does not get around the issue of drug therapy sequencing that is prevalent in our data. As part of this alternative analysis, we classified incident hypertension cases using shorter time windows to test whether later initiation of ARBs drove our findings, but we did not find support for this hypothesis.

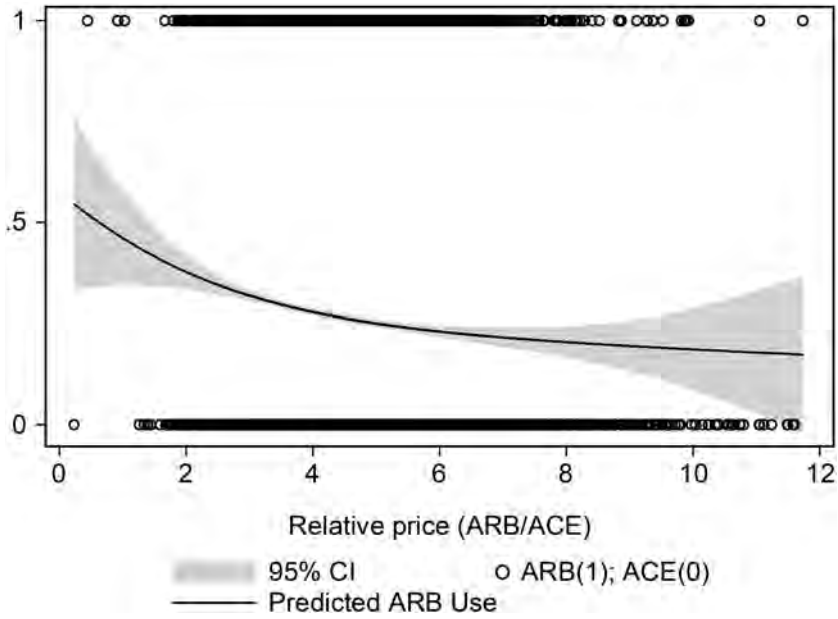

Fig. 2.5 Unconditional relationship between ARB use and relative price

The fourth column is the smallest sample and includes only incident ARB or ACE monotherapy users. The comparisons between ARB users and the control group thus become progressively "cleaner" moving from left to right. The second and fourth columns are the same between the two treatment definitions, because the ACE monotherapy group is restricted to never be on ARBs.

Figure 2.5 graphically displays the results of the first-stage regressions for incident cases. The binary indicator for treatment (0 or 1) for each patient is plotted against the relative price of ARBs to ACE inhibitors. The downward sloping curve reveals that as ARBs become more expensive, more patients take ACE inhibitors.

There appears to be little evidence that ARBs lead to cancer based on the second-stage regressions. Table 2.7 presents the results using the "ever use" treatment definition. The point estimate on predicted treatment is negative and statistically significant among all prevalent cases (column [1]), with ever using ARBs decreasing the probability of cancer within two years by 1.2 percent. However, the estimate becomes positive and statistically significant among monotherapy users (column [2]), increasing the probability of cancer by 2.8 percent. Among incident cases, the point estimates on predicted treatment are positive, but not statistically significant. So as the sample becomes "cleaner," the evidence that ARBs are associated with cancer becomes weaker. Using the intent-to-treat definition as shown in table 2.8, the evidence on a link between ARB use and cancer remains weak. Under both treatment definitions, there is

Table 2.7 **Second-stage IV regressions: Ever-user treatment definition**

	Prevalent cases		Incident cases	
	All drug users (1)	ACE or ARB monotherapy only (2)	All drug users (3)	ACE or ARB monotherapy only (4)
		Cancer		
ARB treatment	−0.012	0.0288	0.0057	0.0273
	(−1.90)	(2.97)	(0.43)	(1.73)
N	1,406,463	95,386	153,904	19,466
		Stroke		
ARB treatment	0.0154	0.0268	0.0234	0.0501
	(1.65)	(2.46)	(1.47)	(2.71)
N	1,406,469	95,386	153,904	19,466

Notes: Cluster-adjusted robust *T*-statistics in parentheses. Regressions also include indicators for age, sex, diabetes, heart disease, heart failure, depression, Alzheimer's disease, glaucoma, ischemic heart disease, chronic obstructive pulmonary disease, pelvic/hip fracture, osteoporosis, rheumatoid arthritis/osteoarthritis, cataracts, and chronic kidney disease, state fixed effects, and the following county-level variables: percent home ownership, percent with high school degree, percent with some college or two-year degree, percent with four-year college degree or beyond, percent African American, average income, percent below poverty line, unemployment rate, percent married, and percent foreign born, all among adults.

more consistent evidence that ARB users have more strokes, although it is fair to question whether a 5 percent significance level is an appropriate threshold given the large sample size. Nonetheless, we should expect fewer or no difference in the number of strokes based on the results of RCTs, and our opposite finding casts further doubt on the validity of our IV methods.

2.5 Robustness Checks

2.5.1 Subsample Analysis: Healthy Patients

Healthy patients serve as a first robustness check. Such patients arguably represent cleaner treatment and control groups than the full sample that includes people with a variety of health conditions, since healthy patients likely also have fewer unobserved conditions that may be correlated with both ARB use and the outcomes. We classify healthy patients as beneficiaries without any of the thirteen Chronic Conditions Data Warehouse (CCW) chronic conditions measured, which comprises 28 percent of the full sample.

Price is still statistically significant in the first-stage IV regressions as shown in appendix table 2A.1. However, the *F*-statistic is lower than in the baseline regressions and below ten in three of the four specifications. This may be due to the smaller sample size or suggest that chronic conditions are important to explain treatment patterns. In the second-stage regressions, the

Table 2.8 Second-stage IV regressions: Intent-to-treat treatment definition

	Prevalent cases		Incident cases	
Cancer	All drug users (1)	ACE or ARB monotherapy only (2)	All drug users (3)	ACE or ARB monotherapy only (4)
ARB treatment	−0.0139	0.0288	−0.002	0.0273
	(−1.26)	(2.97)	(−0.13)	(1.73)
N	1,406,469	95,386	153,904	19,466
	Prevalent cases		Incident cases	
Stroke	All drug users (1)	ACE and ARB users only (2)	All drug users (3)	ACE and ARB users only (4)
Treatment	0.041	0.0268	0.0249	0.0501
	(2.40)	(2.46)	(1.32)	(2.71)
N	1,406,469	95,386	153,904	19,466

Notes: Cluster-adjusted robust *T*-statistics in parentheses. Regressions also include indicators for age, sex, diabetes, heart disease, heart failure, depression, Alzheimer's disease, glaucoma, ischemic heart disease, chronic obstructive pulmonary disease, pelvic/hip fracture, osteoporosis, rheumatoid arthritis/osteoarthritis, cataracts, and chronic kidney disease, state fixed effects, and the following county-level variables: percent home ownership, percent with high school degree, percent with some college or two-year degree, percent with four-year college degree or beyond, percent African American, average income, percent below poverty line, unemployment rate, percent married, and percent foreign born, all among adults.

estimates on predicted treatment are lower than in the baseline regressions and very imprecise. To the extent that the healthy subsample produces more similar control and treatment groups, the IV regressions in appendix tables 2A.2 and 2A.3 suggest that there is little meaningful or statistically significant difference between ARBs and other hypertension drugs.

2.5.2 Falsification Test: Pain

As another falsification check, we rerun our IV regressions with a diagnosis of pain within one year of starting hypertension treatment as the dependent variable. Since hypertension drugs should have little impact on the diagnosis of pain, the magnitude of any association between ARBs and pain should be lower if the effects with cancer and stroke are real. We consider ICD-9 codes for sprains and strains (excluding ankle and back) and open wounds (excluding head wounds) in diagnosing pain. The ARBs are associated with less pain as shown in appendix table 2A.4, but the magnitude of the coefficient estimates are similar to the results for cancer and stroke, and in some cases larger.[6]

6. In case pain followed or preceded a stroke, we also recoded any diagnosis of pain to 0 within a one-month window of a stroke diagnosis. The results were similar as reported in table 2.8.

As the samples become cleaner, the estimates become smaller in magnitude and less precise. We find a similar pattern if we use the subsample of healthy patients to estimate the incidence of pain (results not shown). Perhaps these results are due to the omitted variable bias, such as individual-level socioeconomic factors, which may drive both treatment choice and the number of office visits (and thus diagnoses) of a patient. Since ARBs are more expensive than other antihypertensives (there are no generic ARBs), one might speculate that higher-income patients are more likely to take ARBs. There is also evidence that higher-income, white Medicare beneficiaries have fewer hospital discharges than lower-income beneficiaries (Gornick et al. 1996; Gornick 2003). So if higher-income patients are also less likely to receive a pain diagnosis due to fewer hospital admissions, then the omission of individual income biases our estimates downward. Overall, our falsification test fails, casting doubt on the validity of the IV regressions of cancer and stroke.

2.6 Comparison with Randomized Controlled Trials

The results of RCTs and IV regressions are both relevant for policy, but measure different quantities. The RCTs measure the average treatment effect (ATE), while IV regressions measure the local average treatment effect (LATE). Importantly, the parameter in IV regressions is identified only by the subgroup of observations affected by changes in the instrument (price in our example). This implies that IV regressions are only useful for drawing inferences to people who are affected by price changes. For example, the LATE tells us nothing about people who would never consider changing drug treatments because of side effects. From the perspective of evaluating drug safety, the ATE is arguably the more relevant quantity than the LATE, since policymakers are interested in the effect of ARBs on all individuals. It is hard to envision many cases where the LATE is more informative for policy than the ATE. Given that the population in RCTs can sometimes be narrowly defined, the ATE estimate may only apply to a group with select covariates, whereas the LATE calculated from observational data on a wider population may be more informative about treatment effects among individuals with different levels of covariates (e.g., age, sex, medical history, etc.). However, that is a shortcoming in the construction of small, tightly defined RCTs, not with the ATE per se, and calls for expanding the population of RCTs or conducting numerous RCTs on different subpopulations.

Although RCTs are viewed (rightly) as the gold standard in evaluation, there may still be unobserved behavioral changes between treatment and control groups that bias results. In RCTs, for example, individuals may not always comply with the therapy assigned to them. In comparing the treatment and control groups, assuring that both groups comply at the same rates is critical to obtaining unbiased estimates (Hamilton 2001). The implication for using observational data is that researchers should also compare groups that are most likely to comply with the therapy prescribed. Additionally,

compliance may also depend on whether an individual believes to be assigned to the treatment or control arms of an RCT. Malani (2006) builds a model demonstrating the importance of placebo effects, where individuals believing to be assigned to the treatment arm are more likely to comply. Using data on the probabilities of assignment to the treatment group of various RCTs, he finds empirical support for this model. So behavior changes within RCTs may be just as important as behavioral patterns in observational data. Furthermore, even when a study includes both randomized and self-selected observational data for the same population, economic and statistical models using self-selected data may fail to replicate the results of RCTs (Goldman, Leibowitz, and Buchanan 1998).

2.7 Conclusion

This chapter highlights some of the perils and pitfalls of using observational data for CER. We document that not only is the lack of randomization a problem, but the existence of competing therapies and prevalence of polytherapy also poses challenges to researchers. To deal with sample selection problems, we restrict our sample to monotherapy users and incident hypertension cases. While this allows us to sidestep the unobservability of prior drug use, it comes at the price of a sample that is not only small, but also not representative of all drug users. This partly defeats one of the key assets of observational data, which is the potential for greater representativeness than RCTs.

Our empirical approaches to tackle nonrandom treatment assignment are strong conceptually, albeit unsuccessful. Our first approach of using physician propensities is an innovative solution to initial therapy choice, but does not pass our falsification test using pain. Our second approach using conventional IV methods finds price to be a strong predictor of treatment, but our results are often sensitive to the sample analyzed. Overall, we find little evidence from our IV regressions that ARBs are associated with cancer and weak evidence that ARBs are positively related to strokes. The latter result contradicts the findings from RCTs, and thus indicates our empirical approach is likely not valid. In addition, the fact that estimates from our pain regressions are often larger than estimates for cancer or stroke also suggest our IV estimates are not causal.

One might argue with the exogeneity of both our instrument and physician-prescribing propensities in purging selection bias. For example, perhaps unobserved patient attributes affect plan choice and thus ARB prices through copayments. And copayments (for non-ARB utilization) may affect cancer detection through moral hazard. As evidence of this pattern, Meeker et al. (2011) find first-dollar coverage increases utilization of common cancer screens—lipid screens, Pap smears, mammograms, and fecal occult blood tests—relative to plans with cost sharing. So even if patients do not choose

a plan based on ARB copayments, correlation between ARB copayments and copayments for other services would make prices endogenous to cancer rates. Our use of regional ARB prices attempts to deal with such issues, but correlation between regional costs and cancer would still bias our results. In terms of physician propensities, unobserved patient attributes might affect the choice of physician and also be correlated with attributes of other patients in the region, which in turn affects physician propensities. In short, selection could contaminate our results if patients are neither randomly assigned to physicians with high propensity to prescribe ARBs nor randomly assigned to insurance with a low price for ARBs. And despite our best efforts to use observational data for causal inference, we certainly cannot rule out the possibility of such bias.

While claims data, in principle, offer several advantages to evaluating drug treatments over RCTs, researchers must be careful to deal with left-censoring, contamination bias, and selection into treatment. By illustrating these pitfalls with the case of ARBs for hypertension, our chapter provides a cautionary tale for researchers interested in using claims data for CER.

Appendix

Table 2A.1 **First-stage regressions: Healthy subpopulation**

	Prevalent cases		Incident cases	
	All drug users (1)	ACE or ARB monotherapy only (2)	All drug users (3)	ACE or ARB monotherapy only (4)
	Treatment definition: Patient ever used drug			
Price	−0.038	−0.019	−0.032	−0.038
	(−21.47)	(−3.95)	(−6.53)	(−3.03)
F-stat	60.11	6.69	9.26	2.22
N	175,346	14,667	22,057	3,259
	Treatment definition: Intent to treat			
Price	−0.014	−0.019	−0.023	−0.038
	(−9.69)	(−3.95)	(−5.51)	(−3.03)
F-stat	44.43	6.69	7.01	2.22
N	175,348	14,667	22,057	3,259

Notes: Cluster-adjusted robust *T*-statistics in parentheses. Regressions also include indicators for age and sex, state fixed effects, and the following county-level variables: percent home ownership, percent with high school degree, percent with some college or two-year degree, percent with four-year college degree or beyond, percent African American, average income, percent below poverty line, unemployment rate, percent married, and percent foreign born, all among adults.

Table 2A.2 **Second-stage IV regressions: Ever-user treatment definition, healthy subpopulation**

	Prevalent cases		Incident cases	
Cancer	All drug users (1)	ACE or ARB monotherapy only (2)	All drug users (3)	ACE or ARB monotherapy only (4)
Treatment	−0.0209	0.0071	0.0068	0.0399
	(−4.00)	(0.51)	(0.41)	(1.27)
N	175,346	14,667	22,057	3,259
	Prevalent cases		Incident cases	
Stroke	All drug users (1)	ACE and ARB users only (2)	All drug users (3)	ACE and ARB users only (4)
Treatment	−0.0067	−0.0006	−0.0022	0.0054
	(−1.88)	(−0.09)	(−0.17)	(0.32)
N	175,346	14,667	22,057	3,259

Notes: Cluster-adjusted robust *T*-statistics in parentheses. Regressions also include indicators for age and sex, state fixed effects, and the following county-level variables: percent home ownership, percent with high school degree, percent with some college or two-year degree, percent with four-year college degree or beyond, percent African American, average income, percent below poverty line, unemployment rate, percent married, and percent foreign born, all among adults.

Table 2A.3 **Second-stage IV regressions: Intent-to-treat treatment definition, healthy subpopulation**

	Prevalent cases		Incident cases	
Cancer	All drug users (1)	ACE or ARB monotherapy only (2)	All drug users (3)	ACE or ARB monotherapy only (4)
Treatment	−0.0076	0.0071	0.0134	0.0399
	(−0.90)	(0.51)	(0.71)	(1.27)
N	175,348	14,667	22,057	3,259
	Prevalent cases		Incident cases	
Stroke	All drug users (1)	ACE and ARB users only (2)	All drug users (3)	ACE and ARB users only (4)
Treatment	0.0068	−0.0006	0.0028	0.0054
	(1.08)	(−0.09)	(0.18)	(0.32)
N	175,348	14,667	22,057	3,259

Notes: Cluster-adjusted robust *T*-statistics in parentheses. Regressions also include indicators for age and sex, state fixed effects, and the following county-level variables: percent home ownership, percent with high school degree, percent with some college or two-year degree, percent with four-year college degree or beyond, percent African American, average income, percent below poverty line, unemployment rate, percent married, and percent foreign born, all among adults.

Table 2A.4 Falsification test: Second-stage regressions for pain

	Prevalent cases		Incident cases	
	All drug users (1)	ACE and ARB users only (2)	All drug users (3)	ACE and ARB users only (4)
		Ever users		
Treatment	−0.0511	−0.0255	−0.0249	−0.0248
	(−5.11)	(−1.95)	(−1.65)	(−1.19)
N	1,406,469	95,386	153,904	19,466
		Intent to treat		
Treatment	−0.0718	−0.0255	−0.0329	−0.0248
	(−4.42)	(−1.95)	(−1.73)	(−1.19)
N	1,406,469	95,386	153,904	19,466

Notes: Cluster-adjusted robust *T*-statistics in parentheses. Regressions also include indicators for age, sex, diabetes, heart disease, heart failure, depression, Alzheimer's disease, glaucoma, ischemic heart disease, chronic obstructive pulmonary disease, pelvic/hip fracture, osteoporosis, rheumatoid arthritis/osteoarthritis, cataracts, and chronic kidney disease, state fixed effects, and the following county-level variables: percent home ownership, percent with high school degree, percent with some college or two-year degree, percent with four-year college degree or beyond, percent African American, average income, percent below poverty line, unemployment rate, percent married, and percent foreign born, all among adults.

References

Basu, A., J. Heckman, N. L. Salvador, and S. Urzua. 2007. "Use of Instrumental Variables in the Presence of Heterogeneity and Self-Selection: An Application to the Treatments of Breast Cancer Patients." *Health Economics* 16 (11): 1133–57.

Burton, Thomas. 2013. "Dispute Flares inside FDA over Safety of Popular Blood-Pressure Drugs." *Wall Street Journal*, May 30.

Chandra, A., and D. O. Staiger. 2011. "Expertise, Underuse, and Overuse in Healthcare." Unpublished Manuscript, March.

Dahlof, B., R. Devereux, S. Kjeldsen, S. Julius, G. Beevers, U. de Faire, F. Fyhrquist, et al. 2002. "Cardiovascular Morbidity and Mortality in the Losartan Intervention for Endpoint Reduction in Hypertension Study (LIFE): A Randomized Trial against Atenolol." *Lancet* 359:995–1003.

Food and Drug Administration (FDA). 2010. "FDA Drug Safety Communication: Ongoing Safety Review of the Angiotensin Receptor Blockers and Cancer." July 15. www.fda.gov/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/ucm218845.htm.

———. 2011. "FDA Drug Safety Communication: No Increase in Risk of Cancer with Certain Blood Pressure Drugs—Angiotensin Receptor Blockers (ARBs)." June 2. www.fda.gov/Drugs/DrugSafety/ucm257516.htm.

Goldman, D., A. Leibowitz, and J. Buchanan. 1998. "Cost-Containment and Adverse Selection in Medicaid HMOs." *Journal of the American Statistical Association* 93 (441): 54–62.

Gornick, M. 2003. "A Decade of Research on Disparities in Medicare Utilization: Lessons for the Health and Health Care of Vulnerable Men." *American Journal of Public Health* 93 (5): 753–59.

Gornick, M., P. Eggers, T. Reilly, R. Mentnech, L. Fitterman, L. Kucken, and B. Vladeck. 1996. "Effects of Race and Income on Mortality and Use of Services among Medicare Beneficiaries." *New England Journal of Medicine* 335:791–99.

Hamilton, B. 2001. "Estimating Treatment Effects in Randomized Clinical Trials with Non-Compliance: The Impact of Maternal Smoking on Birthweight." *Health Economics* 10:399–410.

Joyce, G., J. Zissimopolous, and D. Goldman. 2013. "Digesting the Doughnut Hole." *Journal of Health Economics* 32 (6): 1345–55.

Malani, A. 2006. "Identifying Placebo Effects with Data from Clinical Trials." *Journal of Political Economy* 114 (2): 236–56.

Matchar, D., D. McCrory, L. Orlando, M. Patel, P. Uptal, M. Patwardhan, B. Powers, G. Samsa, and R. Gray. 2008. "Systematic Review: Comparative Effectiveness of Angiotensin-Converting Enzyme Inhibitors and Angiotensin II Receptor Blockers for Treating Essential Hypertension." *Annals of Internal Medicine* 148 (1): 16–29.

Meeker, D., J. Joyce, J. Malkin, S. Teutsch, A. Haddix, and D. Goldman. 2011. "Coverage and Preventive Screening." *Health Services Research* 46 (1): 173–84.

Sipahi, I., S. M. Debanne, D. Y. Rowland, D. I. Simon, and J. C. Fang. 2010. "Angiotensin-Receptor Blockade and Risk of Cancer: Meta-Analysis of Randomised Controlled Trials." *Lancet Oncology* 11 (7): 627–36.

Strauss, M., and H. Hall. 2009. "Angiotensin Receptor Blockers Should Be Regarded as First-Line Drugs for Stroke Prevention in Both Primary and Secondary Prevention Settings: No." *Stroke* 40:3161–62.

Todd, P., and K. Wolpin. 2006. "Assessing the Impact of a School Subsidy Program in Mexico Using Experimental Data to Validate a Dynamic Behavioral Model of Child Schooling." *American Economic Review* 96 (5): 1384–417.

Wang, J., Y. Li, S. Franklin, and M. Safar. 2007. "Prevention of Stroke and Myocardial Infarction by Amlodipine and Angiotensin Receptor Blockers: A Quantitative Overview." *Hypertension* 2007 (50): 181–88.