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

ABSTRACT

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."

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. 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 paper 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 examine enrollees in standalone Medicare prescription drug plans (PDPs). Hypertension patients are classified by use of at least one drug commonly used to treat high blood pressure. As part of our evaluation of drug treatment effectiveness, we focus on stroke—which can result from uncontrolled high blood pressure —and cancer, both of which were flagged by the FDA in 2010 as a potential adverse effect of ARBs.

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. Another challenge for researchers is left censoring; it is common for patients to take hypertension drugs before age 65, when most beneficiaries become eligible for Medicare. We discuss analytical and empirical approaches to overcome this unobservability of prior drug use.

We employ two methods to deal with the non-random 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. Using these two 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.

As a 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 anti-hypertensive. 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. This suggests the relationships between ARBs and cancer or stroke 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. Such methods help to overcome the non-random treatment assignment, which is perhaps the key shortcoming of observational data.