NBER WORKING PAPER SERIES

THE IMPACT OF INFORMATION TECHNOLOGY ON THE DIFFUSION OF NEW
PHARMACEUTICALS

Kenneth J. Arrow
Kamran Bilir
Alan T. Sorensen

The Impact of Information Technology on the Diffusion of New Pharmaceuticals
Kenneth J. Arrow, Kamran Bilir, and Alan T. Sorensen
NBER Working Paper No. 23257
March 2017
JEL No. I10,O33

ABSTRACT

Do information differences across U.S. physicians contribute to treatment disparities? This paper uses a unique new dataset to evaluate how changes in physician access to a decision-relevant drug database affect prescribing decisions. Our results indicate that doctors using the reference have a significantly greater propensity to prescribe generic drugs, are faster to begin prescribing new generics, and prescribe a more diverse set of products. Notably, physicians using the reference database are not faster to prescribe new branded drugs. Given that a new generic drug resembles its branded equivalent clinically, these results are consistent with database users responding primarily to the increased accessibility of non-clinical information such as drug price and insurance formulary data; the results also suggest improvements to physician information access have important aggregate implications for the costs and efficiency of medical care. We address possible selection effects in physician types by relying on within-doctor variation and an instrument for adoption timing that is based on the marketing strategy of the drug reference firm.

Kenneth J. Arrow
Department of Economics
Stanford University
579 Serra Mall
Stanford, CA 94305

Kamran Bilir
Department of Economics
University of Wisconsin - Madison
1180 Observatory Drive Madison,
WI 53706
kbilir@ssc.wisc.edu

Alan T. Sorensen
Department of Economics
University of Wisconsin, Madison
6454 W.H. Sewell Social Science Bldg
1180 Observatory Drive
Madison, WI 53706
and NBER
sorensen@ssc.wisc.edu

1 Introduction

National health expenditures exceed three trillion dollars annually in the United States, account for nearly twenty percent of U.S. GDP, and are to a considerable extent publicly funded.[1] Yet, research by the Dartmouth Atlas Project and Cooper et al (2015) finds substantial, systematic disparities in both the extent of health spending and the quality of medical care across U.S. regions, including threefold per-capita expenditure gaps resulting from inefficient variations in care—differences consistent neither with patient preferences nor with underlying medical conditions.[2] These findings imply significant gains could be achieved by improving efficiency in low-peforming regions, but this requires first identifying the specific mechanisms that *cause* treatment disparities. Among the many potential mechanisms that have been proposed, which include supply, demand, regulatory, and pricing differences, perhaps the most important and intriguing is that disparities result from a lack of uniformity in physicians' information about available therapies.[3] The possibility that information problems underlie observed treatment disparities has inspired calls for the expanded use of medical decision aids, but the difficulties inherent to measuring information differences have led to a paucity of systematic evidence on their actual importance.[4]

To contribute to our understanding of how physicians' information access impacts treatment choices, we assemble and evaluate a unique, new dataset in which treatment decisions and access to a decision-relevant database are observed for a panel of over 125,000 individual doctors during 132 consecutive months.[5] Our data cover the universe of U.S. physicians and capture Medicare patients and the privately insured as well as those with Medicaid or no coverage.[6] To meet the data demands of empirically identifying the influence of information on treatment choices, which requires a medical context with significant information dynamics, we consider prescribing decisions for cholesterol drugs during January 2000–December 2010, during which 12 nationwide product innovations occurred.[7] The differential response

---

[1]See https://www.cms.gov/nationalhealthexpenddata.

[2]The Dartmouth Atlas Project has documented healthcare disparities for Medicare patients over decades (Wennberg et al 1996, Gawande 2009, and Chernew et al 2009); Cooper et al (2015) finds related disparities among the privately insured population. See also Wennberg and Wennberg (2003), Dartmouth Atlas (2007).

[3]For example, Skinner (2012).

[4]See Phelps (2000), Wennberg and Wennberg (2003), and Arrow (1969).

[5]The provider of the database is a leading U.S. point-of-care medical applications firm that chose to remain unnamed in this study. For a description of all major drug references, see Ventola et al (2014).

[6]These latter categories are new relative to existing evidence on healthcare variations; see note #2 above.

[7]Cholesterol drugs are very widely used. The Centers for Disease Control and Prevention estimate that approximately 71 million U.S. adults suffer from chronic hypercholesterolemia and dyslipidemia, conditions in which abnormal levels of cholesterol or lipids are present in the bloodstream. These conditions are associated with heart disease, heart attack risk, and premature death; accordingly, sales of cholesterol therapies accounted for over $18 billion U.S. dollars in 2011 (Ledford 2013). See also Mozaffarian, et al (2014). Drug introductions are listed in Table 1.

across physicians to these repeated drug introductions is essential to our analysis, because it allows us to isolate the influence of information on treatment decisions, while accounting for physician characteristics that simultaneously affect both prescribing and database adoption.

Our results indicate physicians that begin using the drug reference database significantly increase their likelihood of prescribing a generic drug relative to brand-name therapies, and thus increase their generic prescription share—one of the key efficiency metrics emphasized in the Dartmouth Atlas in documenting prescription drug variations.[8] Database users are also faster to begin prescribing newly released generics, an effect notably absent for new branded drugs. These findings suggest database users may be responding primarily to the increased salience of non-clinical information included in the database—including whether a particular drug is currently covered by a patient's insurance plan, and plan-specific pricing— as a generic drug and its branded equivalent share essentially identical clinical attributes. We find that treatment differences decline significantly more across database users than across non-users during the sample period, while the actual diversity of a user's own prescribing increases on average following adoption. Access to detailed information about competing treatments thus appears to raise efficiency and reduce disparities, but importantly, these effects do not appear to come at the expense of patient-sensitive decision making.

Because access to the drug database is not randomly assigned—doctors choose whether and when to subscribe— identifying the causal effects of database access is challenging. Prescription patterns of subscribing doctors may look different from those of non-subscribers not due to any effects of the database itself, but instead due to differences in the types of doctors who choose to subscribe. With this challenge in mind, our analysis relies heavily on within-doctor variation: rather than estimating effects by comparing database users to non-users, we focus on comparisons of a doctor's own prescriptions before versus after she begins using the database. To account for the possibility that dynamic prescribing determinants may be correlated with adoption timing within a location, such as changes in drug advertising, our main specifications also include zipcode-month effects; we verify that our main results also hold in the presence of doctor-specific trends. We further evaluate specifications that treat the timing of doctors' reference adoption as endogenous. Our approach is guided by observing that the network of reference users has grown primarily through a 'word-of-mouth' marketing strategy, whereby users tell friends and colleagues about its value. Accordingly, we use neighboring physicians' lagged reference adoption choices as an instrument that is plausibly independent of patient-specific prescription decisions. Taken together, these specifications support a causal interpretation of our main results—i.e., that using the database caused a small but statistically significant change in prescribing patterns.

Consistent with the regional disparities documented in the Dartmouth Atlas, our data

---

[8]See Munson et al (2013).

reveal substantial prescribing variation across the universe of individual U.S. physicians, particularly with respect to generics and new drug adoption. Doctors differ widely in generic prescription shares (mean 56.4 percent, standard deviation 24.3 percent) in December 2010, and span the full range from no generics (5th percentile) to only generics (95th percentile).[9] We find that these generic shares are strongly and positively correlated within physician across patient insurance types (e.g. private vs. Medicare), suggesting patient cost sharing is unlikely to explain the observed heterogeneity in prescribing patterns. Moreover, with respect to drug adoption, some U.S. physicians begin prescribing a newly-approved cholesterol drug immediately, while others delay for a year or more before prescribing it, a pattern strongly evident even among new generic drugs. Six months after the introduction of generic lovastatin, for example, the molecular equivalent of Mevacor, the generic version accounted for only 83 percent of the drug's prescriptions; by contrast, the generic share reached essentially 100 percent by December 2010. This delayed substitution is evident for each of the generic entrants we observe, contributing to wide differences across doctors in the overall prescription share of generic drugs, and to large corresponding gaps in cost.[10]

Our empirical analysis indicates that some of this observed prescription heterogeneity is explained by differences across doctors in information access. Our most conservative estimates indicate that, after obtaining database access, a physician user increases the likelihood of prescribing a new generic drug within its first market year by 1.4 percent; after correcting for the potential endogeneity of adoption timing, this rises to a 6.4 percent increase. Regarding diversity, database users increase the number of unique drugs prescribed each month by a modest but highly significant 0.030 drugs, reducing the prescription Herfindahl-Hirschman index by 0.001 points. In line with both results, users increase the monthly likelihood of prescribing new and old generic drugs by 1.6 and 2.4 percent, respectively, after database adoption, while reducing the likelihood of prescribing a new branded drug by 0.42 percent. Our back-of-the-envelope calculations suggest the resulting increase in users' generic prescription shares contributes to substantial aggregate cost savings, which amount to approximately $1 billion annually for prescription drugs alone.[11]

This paper is related to an extensive literature documenting wide healthcare disparities across U.S. regions, including the Dartmouth Atlas (e.g. Wennberg et al 1996) and its analysis of prescription drug use among Medicare patients (Munson et al 2013), and Cooper et al (2015) for the privately insured. We contribute to this work by first documenting prescription disparities for the universe of U.S. prescribers and patients with all insurance types within a major therapeutic area. Second, relying on the unusual level of detail and coverage in the dataset we have assembled, we identify a highly significant link between observed dispari-

---

[9]Table 3.
[10]See Section 2.4 and Section 7.
[11]See Section 7.

ties and a specific mechanism—physician information differences—that we find is partially responsible for these disparities. While our data are broader with respect to physician and patient coverage, our empirical strategy is demanding (our dataset includes over 200 million observations) and we therefore focus on a single clinical area while Dartmouth Atlas and Cooper et al (2015) cover a comprehensive set of treatments; aggregating our physician-level data to Dartmouth Atlas regions, we nevertheless find that locations with high generic prescription shares in our dataset also have high generic shares for Medicare overall (correlation 44.4 percent), as well as lower per-capita medical spending for prescription drugs (correlation 23.9 percent) and non-drug healthcare (correlation 12.4 percent)—patterns that underscore the highly systematic nature of U.S. disparities in care, and suggest the potential value of extending our physician-level analysis to other clinical settings.[12]

Our paper is also closely related to work aimed at evaluating the impact of information technology on economic decisions and outcomes.[13] Agents' electronic information access can affect productivity (Solow 1987) and has been specifically shown to improve performance in emergency healthcare delivery (Athey and Stern 2002). However, in routine medical contexts the evidence is less clear: Dranove et al. (2014) finds that the adoption of electronic medical records (EMR) raised hospital costs on average, with an important exception—adopting locations with an abundance of industrial I.T. did in fact experience cost declines. Our results complement this latter finding in that the medical decision-support tool we study is standardized, likely to a greater extent than EMR, yet we observe that both the intensity and impact of its use differ substantially across physicians in the data. In particular, we find that the efficiency impact of database use is systematically larger among adopters using the database intensively to search for information about the cholesterol drugs we study.

In finding that use of an information database tilts prescribing away from branded drugs and toward generics, our results contribute to important work highlighting the influence of information on tastes for generic products. Bronnenberg et al (2015) find that relatively informed buyers are more likely to choose a generic version, for example when purchasing an over-the-counter drug, suggesting consumer misinformation contributes to the brand premium for health products. Our results add nuance to this finding, suggesting that even among highly trained and educated U.S. physicians, access to current product information including pricing *increases* the propensity to prescribe a generic and decreases that for branded drugs. We find that the impact of database access is systematically larger for

---

[12]Our data do not include individual patient characteristics, precluding a direct extension of Munson et al (2013) to non-Medicare patients, as well as a quantitative welfare analysis. This aspect of our dataset further precludes estimating a model featuring prescription dynamics within each patient-physician pair, as in Crawford and Shum (2005) or Dickstein (2015).

[13]See, for example, Attewell (1992); Bresnahan and Greenstein (1996); Black and Lynch (2001); Bresnahan, Brynjolfsson, and Hitt (2002); Brynjolfsson and Hitt (2003); Hubbard (2003); Forman, Goldfarb, and Greenstein (2005); Bloom et al. (2009); Bloom, Sadun, and Van Reenen (2012); Agha (2014).

physicians located far from the information frontier, and that dynamics in the product space (drug entry) may be important in explaining our results, as database users are also faster to begin prescribing a newly-introduced generic version. In finding that physicians' information access affects decisions made on behalf of patients, and that prescribing is highly correlated within a physician across patients regardless of insurance coverage, our results are further aligned with Brot-Goldberg et al (2015) and Cutler et al (2015), which find evidence that physician preferences are key in explaining treatment decisions.

Our results add to the literature examining the determinants of new medical technology diffusion. Classic work by Coleman, Katz, and Menzel (1957, 1996) finds that new pharmaceutical products diffuse unevenly across medical practitioners: physicians that interact more frequently with other physicians are more likely to adopt early. Relatedly, Skinner and Staiger (2007) provide evidence that certain U.S. states have a systematic tendency to adopt early across technology types as varied as beta blockers and hybrid corn (Griliches 1957). We find that physician access to a digital database also speeds new drug adoption—but only for generics; and, to account for local differences in the tendency to adopt both drugs and the database early, which could reflect general differences in unobserved factors such as drug advertising, we emphasize specifications that include both physician and zipcode-by-month fixed effects.[14] In focusing on individual-level drug adoption, our work is also closely related to Crawford and Shum (2005) and Dickstein (2015), who estimate models of physician learning, and Agha and Molitor (2015), who study the diffusion of cancer drugs.[15]

More broadly, our analysis complements research on general theories of technology diffusion featuring agents with imperfect information. Such theories can be shown to explain large existing differences in productivity across locations (Solow 1956, Arrow 1969, Parente and Prescott 1994, Comin and Hobijn 2004) as identified in Klenow and Rodriguez-Clare (1997) and Casselli and Coleman (2006), for example.[16] We introduce a unique dataset in which a sequence of technology adoption decisions is clearly observed at the individual level for the universe of U.S. prescribers, allowing our study to speak both to micro-level mechanisms driving diffusion and to the aggregate consequences of these mechanisms.

The rest of the paper is organized as follows. Section 2 describes the data used in our analysis. Section 3 describes a simple model of prescription choice and our estimation framework. Section 4 presents the empirical results, Sections 5 and 6 discuss interpretation, and Section 7 concludes.

---

[14]We also consider the influence of local differences in mandatory substitution regulations that could be particularly important for explaining generic diffusion in the data, and find that the effects of database access are evident among physicians practicing both within and outside states with a mandatory substitution law.

[15]See also Escarce (1996), who studies physicians' decisions to adopt a surgical technology.

[16]The idea that underlying heterogeneity across agents could influence technology diffusion also relates our work to neoclassical models of technology adoption, e.g. David (1966) and Manuelli and Seshadri (2014).

2 Data and Descriptive Evidence

Evaluating the influence of information access on new pharmaceutical drug diffusion requires detailed measures of drug innovations and individual prescribers' treatment decisions, information usage, and characteristics. We introduce each of these measures below and go on to describe physicians' prescribing of new and existing pharmaceutical drugs.

2.1 U.S. Innovations in Chronic Hypercholesterolemia and Dyslipidemia Therapy

At the start of our sample period in January 2000, HMG-CoA reductase inhibitors (statins) were understood to be the most effective pharmaceutical therapies for hypercholesterolemia, and there were five such products available: Lescol, Lipitor, Mevacor, Pravachol, and Zocor.[17] The most common non-statin used to treat high cholesterol was Niaspan, which is also included in our sample. Thereafter, twelve new cholesterol or lipid control therapies were introduced, including new formulations, combinations, and versions.[18] These include three new molecular entities, Crestor, Lovaza, and Zetia; three generic versions, lovastatin (Mevacor), pravastatin (Pravachol), and simvastatin (Zocor); two new formulations, Altoprev (extended-release Mevacor) and Lescol XL (extended-release Lescol); and four new drug combinations, Advicor (extended-release niacin and Mevacor), Pravigard PAC (aspirin and Pravachol), Vytorin (Zetia and Zocor), Simcor (extended-release niacin and Zocor). Each new therapy received nationwide approval by the U.S. Food and Drug Administration (FDA) on a known, drug-specific date (Table 1). All products are described in Appendix A.1.

While these 18 products are therapeutic substitutes, in that they aim at a similar clinical endpoint—cholesterol or trigliceride reduction—they are only imperfect substitutes: each product features distinctive characteristics relevant for the prescribing decision. First, many but not all cholesterol therapies are pure statins, which act to reduce cholesterol synthesis in the liver by inhibiting a specific coenzyme; these include Lescol (fluvastatin), Lipitor (atorvastatin), Mevacor (lovastatin), Pravachol (pravastatin), Zocor (simvastatin), Crestor (rosuvastatin), Altoprev (extended-release lovastatin), and Lescol XL (extended-release fluvastatin). Other products rely on different mechanisms of action: Zetia (ezetimibe), for example, is distinct in that it achieves cholesterol reduction by reducing intestinal absorption of cholesterol. A second distinction involves therapeutic intensity. High doses of Lipitor and Crestor are more effective at lowering low-density lipoprotein (LDL) cholesterol than

---

[17]Cannon et al (2004).

[18]To ensure adequate coverage in the data, we consider all cholesterol therapies introduced by December 2008, but not those introduced after this date. For the same reason, our analysis excludes Baycol, a drug that was available in January 2000 but withdrawn from the market in August 2001.

alternatives (Law et al 2003). Side effects are also relevant; evidence suggests, for example, that high doses of Lipitor and Crestor may increase the incidence of adverse reactions, while combination therapies such as Vytorin may in certain cases be more appropriate care for patients with severe cholesterol abnormalities (Kastelein et al 2008).

More subtly, clinical evidence suggests the benefits and risks associated with statins are heterogeneous across patients; randomized-controlled trials (RCTs) indicate, for example, that the benefits of statin use are higher for patients with diabetes, negligible among those with prior heart failure, and vary with age; risks and side-effects also vary with statin intensity, age, weight, comorbidities, and so on (Brooks et al 2014). Adding to this, patients with 'complex' attributes are often underrepresented in RCTs, raising clinical uncertainty and, accordingly, the likelihood that patient preferences—including willingness to suffer side effects and to pay for medications—may influence the prescribing choice (Brooks et al 2014).

Physicians' decisions about which drugs to prescribe are further affected by the evolution of clinical information as new trials are completed—particularly head-to-head studies aimed at establishing the relative efficacy of one drug therapy over another.[19] These ongoing changes in clinical evidence, combined with an expanding set of available products and the accompanying evolution in prices and insurance coverage (e.g. Duggan and Scott Morton 2010), suggest that physicians may turn to drug references that help to ensure patient-specific prescription decisions are based on accurate information.

2.2   Prescriptions by U.S. Physicians

To measure physicians' prescribing of new and existing therapies aimed at cholesterol and lipid control over time, we use physician-level prescription data for the 18 drugs described above from the IMS Health Xponent database. These data are provided at a monthly frequency by drug during the period January 2000 through December 2010, and cover each of the 280,622 U.S. physicians associated with at least ten cholesterol-drug prescriptions during January to December 2010; this low threshold for inclusion implies that our dataset captures essentially the universe of U.S. cholesterol drug prescriptions during this period. For each product and month, we observe the number of prescriptions written by each physician and filled through a U.S. pharmacy. Beginning in January 2006, the data also include information on the method of payment used to fill each prescription (Medicaid, Medicare

---

[19]For example, an RCT completed in 2004 demonstrated that for patients with severe cholesterol abnormality, the incrementally larger reductions achieved by Lipitor resulted in fewer deaths and major coronary events relative to patients taking Pravachol (Cannon et al 2004). Another such study released in 2008 found that, while Vytorin achieved larger cholesterol reductions than simvastatin, the two drugs were observably identical when it came to the thickness of arterial plaque buildup (atherosclerosis); adding to this, a second study in 2008 found a positive association between Vytorin and cancer (Rossebo 2008) that was later reversed (Cannon et al 2015).

Part D, Cash, or Commercial Third-Party Insurance). Importantly, each physician in the dataset is identified by a unique medical education number, name (first name, last name, middle name), and location (a five-digit U.S. zipcode). These identifiers enable us to match individual prescribers with their observed pharmaceutical information technology use.

To ensure that our sample includes only those physicians actively prescribing cholesterol drugs during the entire sample period, we restrict attention to the 128,043 physicians that prescribe ten or more statins both during January to December 2000, and during January to December 2010; this allows us to abstract from potential differences in prescribing that may surround a physician's entry into or exit from medical practice, and also ensures that we have adequate data on database adopters' pre-adoption and post-adoption prescribing patterns. The final prescription dataset includes over 200 million observations (132 months $\times$ 128,043 physicians $\times$ up to 18 drugs). Summary statistics appear in Table 2, and additional details regarding data assembly and the Xponent database appear in Appendix A.2.

2.3   Drug Information Access by U.S. Physicians

To construct an index for the extent of physicians' pharmaceutical information access, we use novel physician-level data from the private firm that owns and operates a prominent electronic reference for pharmaceutical products. The data include a monthly indicator for whether a U.S. physician is a registered user of the reference database, and this suggests the database is widely used: by December 2010, 45.1 percent of sample physicians had established an individual database account (Table 2). The data also include information about registered physicians' actual use of the reference during the sample period: we observe a lower bound on the number of times a physician looks up a cholesterol drug using the database. This proxy is 3.83 on average, and the data indicate that, while 24.2 percent of physicians are registered database users in the average month, only 13.1 percent of physicians use the database to look up one of the cholesterol drugs considered in our study. It is for this latter group of physicians that database access is likely to be relevant to the cholesterol-drug prescribing outcomes we consider; in Section 5, we thus consider whether the observed intensity of database use explains variation in its impact on prescribing.

The drug reference contains information that is, in principle, relevant for improving the match between patient characteristics and available pharmaceutical products. At any point in time, the drug reference contains detailed information about each available U.S. FDA-approved medication. This information is obtained from the medical literature, specialist recommendations, clinical guidelines, manufacturer labeling, standard medical references, and FDA drug safety alerts and is updated continually; the results of this ongoing research are condensed into drug-specific monographs that may be accessed through the electronic database interface. Beyond standard clinical information such as contraindications, cautions,

adverse reactions, safety, monitoring, and pharmacology, the reference monographs also include a set of additional variables for each product that may affect prescribing decisions. Specifically, the monographs include retail pricing and formulary status information for each drug, drug interaction information, FDA warnings, and off-label and pediatric usage guidelines. Each physician customizes the tool with respect to formularies, selecting those relevant to their decision needs; it is then straightforward for any drug to check copay tiers, formulary alternatives, generic substitutions, criteria for prior authorizations, and quantity limits, facets of a formulary that are subtle but often have significant consequences for patient costs. The database includes separate entries for each branded product and each generic, based on product-specific information such as available formulations, dosing, indications, manufacturer, and pricing. The database is updated to reflect both the current set of products and formulary details, as well as the current state of knowledge regarding drug characteristics and clinical practice. Importantly, information for new drugs becomes available around the time the drug is released by the FDA for commercial prescription.

Because the drug reference combines available information into a single, current monograph rather than contributing new or proprietary drug information, it is best viewed as a tool that makes it convenient for physicians to quickly access condensed clinical, insurance, and pricing information about a drug. Doctors commonly use the reference to check dosages, contraindications, and coverage details, but rely on other sources, such as medical journals or more encyclopedic references, for information such as a drug's results in clinical trials.

For our study, it is critical to understand what drives database adoption. Figure 1 indicates that use of the reference database during the sample period is not random, but differs according to observable doctor characteristics.[20] Throughout the sample period, physicians are more likely to have adopted the database if they had graduated from a top-ranked U.S. medical school (Panel A) and had graduated recently (Panel B); males are also more likely to adopt (Panel C). Doctors in obstetrics and gynecology (Panel E), and those practicing in the U.S. South (Panel F) appear systematically slower to adopt the database.

Presently, new physician adopters tend to learn about the reference while in medical school. However, because the physicians in our dataset had all completed medical school before the database became available, their adoption decisions are more likely to have been driven by marketing or peer effects. Documents filed along with the reference firm's initial public offering state that its marketing strategy was, in fact, an informal 'word-of-mouth' approach, and that throughout the sample period, the network of reference users grew over time primarily through users telling friends and colleagues about its value. The filings state that this strategy had been both highly effective and inexpensive relative to the alternative

---

[20]Physician characteristics were obtained from the Centers for Medicare and Medicaid Services Physician Compare database, and were matched based on physician first name, last name, and five-digit zipcode.

of hiring a dedicated sales force.[21] Thus, while our data indicate that physicians are visibly idiosyncratic in their adoption timing, a doctor is much more likely a user if a high share of other doctors in her zipcode are also users—consistent with the firm's reported marketing strategy. Moreover, only 16 percent of the variation in the time to adoption is explained by zipcode fixed effects, indicating that within-zipcode dynamics are quite important.

Regarding adoption motives, the main reason doctors cite for registering is convenience: database use reportedly yields meaningful time savings. By contrast, it is unlikely that price was an important factor in physicians' adoption decisions. Access was always available through a free version of the database application, which included the core drug reference tools (e.g. dosage lookups) that are relevant to our study. Additional features were available with a paid subscription, but the annual fee for this version was low, never above $200.

It should also be noted that database adoption appears to be mostly an individual decision, even for doctors in group practices. Large clinics and physician groups sometimes purchase site licenses for institution-level access to the database as part of broader I.T. initiatives; however, some of the benefits of using the database require individual registration, and most doctors therefore have individual accounts even if their group or clinic has a site license. Nevertheless, to check whether doctors practicing in groups tend to synchronize their database adoption, which would suggest the influence of a group-level adoption decision, we used the 2014 CMS Physician Compare database to identify doctors who were likely working in the same practice during our sample period. Among over 7,000 groups we identified, just 38 were ones in which all doctors in the group adopted the database at the same time. In light of this, it seems unlikely that site-level access or group adoption decisions are primary drivers of the physician-level database use we consider.

2.4 Descriptive Evidence

The data provide suggestive indications that incomplete information may affect physicians' prescribing as well as the rate and extent of new product diffusion, depicted in Figure 2. Consider the statistics presented in Table 3, which quantify differences in prescribing across U.S. physicians for the class of cholesterol medications evaluated. The statistics in Panel A provide evidence for the December 2010 cross section. It is apparent that the pronounced variation in cholesterol-drug prescribing previously found among Medicare patients (e.g. Munson et al 2013, Brooks et al 2014) is also present within the overall population both across zipcodes (columns 5–8) and individual physicians (columns 1–4). The share of prescriptions accounted for by generics ranges from zero to one in column 4; moreover, while the average physician prescribes a generic in 56.4 percent of cases, the standard deviation is also large

---

[21]By contrast, sales force marketing is standard for new pharmaceuticals. See Datta and Dave (2013).

(24.3 percent), and spans the full range from zero (5th percentile physician) to 100 percent (95th percentile). The relative heterogeneity across doctors is even wider for the specific drugs described in columns 1 through 3, and across all columns, the idiosyncratic behavior of individual physicians appears to be important: for the vast majority of zipcodes, there is substantial within-zipcode variation in generic shares across local physicians (Table A.1).

Even if physicians were perfectly informed, variation in prescribing could result from an uneven distribution of patient characteristics. For example, Lipitor is a high-intensity statin that may be preferable for patients with a severe cholesterol abnormality, the incidence of which may cluster geographically. Similarly, risk-averse patients may prefer an established drug over a new one—even if the new drug is simply a new generic version—if they perceive the quality of a new product as uncertain relative to another option. Such underlying patient heterogeneity may partially explain the slow and incomplete diffusion of new drug varieties, which is apparent for each new drug except simvastatin (Figure 2 and Table 1, column 3). It may also explain why a substantial fraction of the variation across doctors observed in columns 1–4 remains even after aggregating to the zipcode level (columns 5–8).

Unobserved patient heterogeneity likely explains some of this variation in prescribing, but columns 1, 2, and 3 indicate that additional factors are also likely present. Specifically, these columns assess within-physician changes in the prescription of new generic products. The advantage of focusing on these columns is that it is possible to compare prescribing of a branded product with its molecularly-equivalent generic, two distinct drugs that have no relevant clinical differences. And, by examining changes over time in the generic share of molecule-specific prescriptions—e.g the share of simvastatin plus Zocor prescriptions that are accounted for by generic simvastatin—it is possible to determine whether stable patient heterogeneity is likely to be the only explanation for variations in care. For each of the three generic drug introductions (lovastatin, pravastatin, and simvastatin), the data indicate that physicians differ in their use of generics in the short run, six months after generic entry, and that substitution toward generics is initially incomplete at this point (Panel B). By contrast, in the long run, physicians differ substantially less: nearly complete substitution toward generics is observed for each of the three products (Panel C).[22] This pattern of *delayed substitution* between two molecularly equivalent products strongly suggests factors other than time-invariant patient heterogeneity contribute to prescribing differences among cholesterol drugs, and is consistent with the influence of information frictions.

Beyond cost implications, these same factors may impede the diffusion of new non-generic therapies, with consequences for health outcomes. The data indicate that physicians are slow to begin prescribing new molecular entities, new drug combinations, and new dosage

---

[22]By December 2010, physicians had broadly switched away from prescribing Mevacor, Pravachol, and Zocor. However, six months after each respective patent expired, generic prescribing was far less prevalent for each molecule, though the generic version was in each case already substantially less expensive.

forms—branded products not facing generic competition. Figure 2 shows that diffusion curves differ considerably across new drugs; Figure 3 plots the gradual diffusion of Crestor across U.S. zipcodes; and Table 1 describes how the time lag in months between a drug's approval and its initial prescription varies across U.S. physicians for each drug introduction. The average physician delays prescribing a new drug for 20.3 months among the new products considered in our analysis; the standard deviation is even larger (22.1 months), and this adoption lag ranges from zero to 122 months, indicating some physicians adopt immediately and others had yet to adopt the first new drug by the end of our sample period (Table 2).

Unlike the Dartmouth Atlas and Cooper et al (2015) studies, which cover a comprehensive set of treatments and analyze regional differences in the cost and quality of care, we analyze behavior at the physician level and focus on the specific clinical decision of which cholesterol drug to prescribe. However, it is nevertheless useful to ask whether the patterns we observe for this context are consistent with the broader treatment patterns documented by the Atlas. Aggregating our physician-level data to Hospital Referral Regions (HRRs) and comparing against data from the Dartmouth Atlas project, we find that locations with high generic prescription shares in our dataset also have high generic shares for Medicare overall (correlation 44.4 percent), as well as lower per-capita medical spending for prescription drugs (correlation 23.9 percent) and non-drug healthcare (correlation 12.4 percent). As noted above, these patterns underscore the highly systematic nature of U.S. disparities in care, and suggest that the disparities in physicians' prescribing of cholesterol drugs may reflect some of the same factors that drive disparities in treatment decisions more broadly.

3   Empirical Strategy

In this section we provide a conceptual framework indicating how physician information and prescribing outcomes may be related. We describe the model implications and restrictions that guide our approach to estimating the treatment impact of physicians' database access.

3.1   Conceptual Framework

Consider a baseline model in which physician i faces a period-t choice over which drug to prescribe for each of her patients $n = 1, 2, ..., N_{it}$. Like other economic studies of prescribing decisions, suppose that physician i makes this decision for each patient by selecting the single drug $j \in \{1, 2, ..., J_t\}$ available at t that maximizes patient utility according to physician-i information.[23] Specifically, suppose that the true utility derived by patient n from drug j at t is $u_{njt} \equiv \theta_{jt} + V_{njt}$, which combines the quality of drug j that is both known at t and

---

[23]See, for example, Dickstein (2015), Crawford and Shum (2006). Unlike these papers, we do not observe patient-level information; this precludes estimating a model of learning within each patient-physician pair.

common across patients (θ_{jt}) with the quality of j that is unknown and partially specific to patient n (V_{njt}). The first of these terms (θ_{jt}) thus captures the accepted wisdom at t about the efficacy, costs, side effects, and so on of drug j for the average patient, while the second reflects novel information that may, in part, be relevant to the match between j and patient n. In particular, suppose that V_{njt} combines two terms: $V_{njt} \equiv v_{jt} + \epsilon_{njt}$, where v_{jt} is a drug-specific value—a revision to accepted wisdom about the quality of drug j—and where ϵ_{njt} reflects the quality of the match between patient n and drug j. We assume the physician is only partially informed about V_{njt}, to a degree indexed by a parameter ϕ_{it}; she bases her prescribing decision on a partial observation of u_{njt} given by

$$\hat{u}_{njt} \equiv \theta_{jt} + (1 - e^{-\phi_{it}})V_{njt} = \theta_{jt} + (1 - e^{-\phi_{it}})\,(v_{jt} + \epsilon_{njt}). \tag{1}$$

Physicians with a higher value of ϕ_{it} in (1) are more responsive to information about drug quality that is not commonly known at t (v_{jt}), and about the patient-specific match (ϵ_{njt}). In particular, (1) implies physicians with no special information ($\phi_{it} = 0$) are insensitive to V_{njt} and thus prescribe the same drug—that with the highest θ_{jt}—for all patients, while physicians who are fully informed ($\phi_{it} \to \infty$) respond to V_{njt} perfectly.

If we assume that the ϵ_{njt} follow an i.i.d. Type-1 Extreme Value distribution, it is straightforward in this simple setup to show that the probability physician i prescribes drug j for patient n at t depends on the information index ϕ_{it} as follows

$$p_{jt}(\phi_{it}) = \frac{\exp\left\{\frac{\theta_{jt}}{1 - e^{-\phi_{it}}} + v_{jt}\right\}}{\sum_{k=1}^{J_t} \exp\left\{\frac{\theta_{kt}}{1 - e^{-\phi_{it}}} + v_{kt}\right\}}$$

and that, accordingly, the probability P_{ijt} that drug j is prescribed by physician i at least once during period t is

$$P_{ijt}(\phi_{it}) \equiv P\{X_{ijt} > 0\} = 1 - P\{X_{ijt} = 0\} = 1 - (1 - p_{jt}(\phi_{it}))^{N_{it}} \tag{2}$$

where X_{ijt} is the number of physician-i prescriptions written for drug j during period t.[24] Moreover, starting from the introduction date t_j^0 of a new drug j, the expected number of periods T_{ij} that lapse before drug j is prescribed at least once by physician i is

---

[24]Qualitatively identical results hold under more general assumptions regarding the distribution of ϵ_{njt}; the Type-1 Extreme Value assumption is thus imposed here only for expositional simplicity. A realistic alternative would be to allow for persistence in the ϵ_{njt} draws, reflecting that the chronic nature of cholesterol and lipid disorders implies physicians often treat the same patient for multiple successive periods. In our analysis to follow, we thus consider the possibility that the prescription outcomes we evaluate are persistent.

$$E[T_{ij}] = \sum_{t=t_j^0}^{\infty} (t - t_j^0) P_{ijt}(\phi_{it}) \prod_{s=t_j^0}^{t-1} (1 - P_{ijs}(\phi_{is}))$$

$$= \sum_{t=t_j^0}^{\infty} (t - t_j^0) \left(1 - (1 - p_{jt}(\phi_{it}))^{N_{it}}\right) \prod_{s=t_j^0}^{t-1} (1 - p_{js}(\phi_{is}))^{N_{is}} \qquad (3)$$

which also depends on ϕ_{it}, as does the expected number of unique drugs M_{it} prescribed by physician i during t,

$$E_t[M_{it}] \equiv E_t \left[\sum_{j=1}^{J_t} 1\{X_{ijt} > 0\} \right] = \sum_{j=1}^{J_t} P_{ijt}(\phi_{it}) = \sum_{j=1}^{J_t} \left(1 - (1 - p_{jt}(\phi_{it}))^{N_{it}}\right). \qquad (4)$$

Within this framework, we regard the electronic database as a technology that increases a physician's ϕ_{it}, which is otherwise unobserved. The database is continuously updated, so users of the database are more likely aware of any new information about the drug, including price changes, new warnings, or new results about its efficacy for different patient types. And the database allows doctors to look up a drug's current formulary status for a specific patient's insurance plan, so database users should also be more responsive to differences in, and changes in, match quality across patient-drug pairs. From (2), if database use indeed increases ϕ_{it}, it impacts the probability drug j is prescribed: whether P_{ijt} increases or decreases for drug j depends on the distribution of V_{njt} across drugs j and patients n. In general, P_{ijt} will increase for drugs with high values of v_{jt} relative to other drugs; alternatively, if all $v_{jt} = 0$, an increase in ϕ_{it} raises P_{ijt} for all drugs (due to the ϵ_{njt}) except for that with the highest θ_{jt}. Similarly, (3) implies the expected number of periods that pass before drug j is prescribed declines in ϕ_{it} whenever P_{ijt} increases in ϕ_{it}. That is, if a permanent increase in ϕ_{it} causes a permanent increase in P_{ijt} for drug j, then it also causes a decrease in T_{ij}. The impact of an increase in ϕ_{it} on the number of distinct drugs prescribed depends on the sum of derivatives $\partial P_{ijt}(\phi_{it})/\partial \phi_{it}$ across drugs j in (4). If a higher ϕ_{it} implies increased sensitivity to patient-specific match quality ϵ_{njt}, for example, P_{ijt} would increase for most drugs j, and diversity of prescribing would also rise.[25]

It is important to note that doctors who regularly prescribe cholesterol medications will be aware of most drugs' clinical attributes. However, if patient-specific economic details such as the pricing and formulary status of a drug evolve substantially over time, or if news about negative drug interactions and other adverse reactions emerges only gradually, doctors may prefer to look up these drug attributes prior to writing a prescription. For newer, less

---

[25]See also Berndt et al (2015).

familiar drugs, doctors may also be inclined to look up details like dosage, and it is for these drugs that one may expect differences between u_{njt} and $\hat{u}_{njt}$ to be particularly relevant.

3.2 Econometric Model

One natural approach to evaluating the influence of increases in ϕ_{it} due to database adoption would be to directly estimate equations derived from the conceptual model above. However, the model indicates it is important to control not only for the number of drugs J_t and i's prescribing intensity N_{it}, but also for unobserved drug quality θ_{jt} and unobserved determinants of ϕ_{it}. Given the size of the data set, handling the nonlinearity implied by (2) in the presence of multiple sets of fixed effects is computationally infeasible. We therefore estimate the effects of database use through specifications that are guided by the model, but linear.

In particular, we consider three main linear specifications corresponding to each of the three observable outcomes discussed above: the new-drug adoption lag T_{ij}, prescription diversity M_{it}, and prescription probabilities P_{ijt}. Our estimation approach does not impose the restrictions that link P_{ijt} with T_{ij} and M_{it} in the model; as a result, comparing our estimates across these outcomes is qualitatively informative regarding the fit of the model.

We first assess the time lapse T_{ij} between the initial market release of drug j and its first prescription by physician i as in Coleman, Katz, and Menzel (1957) using the following equation

$$P\{T_{ij} \leq 12\} = \eta_i + \eta_{zj} + \beta Z_{ij} + \delta N_{it(j)-1} + \varepsilon_{ij}, \tag{5}$$

where T_{ij} is measured in months, $P\{T_{ij} \leq 12\}$ is the probability that j is prescribed by i within twelve months of release, and where η_{zj} and η_i are zipcode-drug and physician fixed effects, respectively. Z_{ij} indicates whether doctor i has database access at the time drug j is first introduced, and $N_{it(j)-1}$ is i's total prescription volume for cholesterol drugs in the month preceding j's introduction. We also estimate (5) with closely-related alternative dependent variables including $P\{T_{ij} \leq 6\}$, $P\{T_{ij} \leq 24\}$, and T_{ij}.[26]

Both (3) and (5) are expressed at the doctor-drug level and, but for the functional-form differences mentioned above, the two equations are connected. The estimating equation essentially considers ϕ_{it} to be a function of database use (Z_{it}), physician fixed effects (η_i), and zipcode-month fixed effects (η_{zt}) reflecting local changes in access to information. We arrive at (5) by noting that the evolution of time is, in doctor-drug space, marked by the sequential introduction of each new drug j, and that the η_{zj} we therefore include take the place of η_{zt} while also accounting for the drug quality effects θ_{jt} in (3). In addition, like (3),

---

[26]With T_{ij} as the dependent variable, it is necessary to address truncation—which is more pronounced for drugs introduced late in the sample period. To apply a uniform truncation rule, a significant number of observations must be omitted, hence T_{ij}, though more direct, is not our preferred dependent variable.

(5) includes i's prescription volume $N_{it(j)-1}$, and J_t is absorbed by the η_{zj}.

Equation (5) is estimated on the subset of drugs first introduced during the sample period. Finding that the coefficient of interest β is positive would indicate that when a physician obtains database access, she significantly increases her likelihood of prescribing a new drug within its first year, relative to before access began. This would be consistent with database use increasing ϕ_{it} and the probability P_{ijt} of prescribing a new drug j at t: $\partial P_{ijt}(\phi_{it})/\partial\phi_{it} > 0$. Notice that the inclusion of physician fixed effects implies that β is identified using within-doctor variation over time: i may be a database user at the time drug j is first introduced, but may not yet be a user upon the introduction of drug j'. These effects are important if stable, unobserved physician characteristics determine both physician-specific database use Z_{ij} and the rate of drug adoption T_{ij} (e.g. early adopters). Possible common unobserved random shocks that are local and correlated with database adoption are further accounted for by clustering standard errors at the zipcode level.

Nevertheless, even with these fixed effects and clusters, there could be time-varying factors such as local technology adoption propensities (Skinner and Staiger 2007) or pharmaceutical advertising that jointly determine, or are correlated with, both physician i's database use and her rate of new drug adoption. The η_{zj} in (5) partially address this by accounting for differences across zipcode-drug pairs in doctors' average first-prescription timing, which in this context would be correlated with Z_{ij}. However, if the omitted factor is idiosyncratic across physicians, even within a zipcode, then $\mathrm{Cov}(Z_{ij}, \varepsilon_{ij}) \neq 0$ and (5) will fail to yield a consistent estimate of β. We return to this in describing our instrumental-variables estimates in Section 5.

Second, building from (4), we consider the possibility that information access could affect physician i's knowledge about the match quality between drug j and patient n, inducing better-informed physicians to prescribe a more diverse set of products than less-informed peers. To assess this possibility, we determine the number of unique drug products $M_{it} \equiv \sum_{j \in \mathcal{J}_t} 1\{X_{ijt} > 0\}$, where $1\{X_{ijt} > 0\}$ is an indicator for whether physician i writes at least one prescription for drug j during month t, and evaluate the following specification

$$M_{it} = \eta_i + \eta_{zt} + \beta Z_{it} + \delta N_{it-1} + \varepsilon_{it}, \tag{6}$$

where all variables are as defined above. M_{it} is low when the prescriptions of physician i are concentrated within a narrow subset of products during month t, and is high when prescribing is diverse. Finding that β is positive in (6) would thus indicate that database access is associated with higher product diversity among physician i's prescriptions; notice that, through (4), this occurs only when the period-t prescription probability P_{ijt} rises more, on average, than it falls—that is, when $\sum_{j=1}^{J_t} \partial P_{ijt}(\phi_{it})/\partial\phi_{it} > 0$. Setting aside functional forms, the connection between (6) and (4) again rests on the idea that the information index

16

ϕ_{it} is a function of database use, physician fixed effects, and zipcode-month fixed effects. We control for N_{it-1} directly, and for J_t through η_{zt}. Standard errors are clustered by zipcode-month to allow for local unobserved shocks correlated with database adoption Z_{it}.

We also estimate (6) replacing M_{it} with the Herfindahl-Hirschman index (HHI_{it}) as an alternative dependent variable; an advantage of this alternative is that it simultaneously captures both intensive- and extensive-margin effects of information on prescribing. Notice that β in (6) is again identified using within-physician variation over time in information access Z_{it}. The zipcode-month fixed effects further help to account for changes over time in unobserved, location-specific determinants of prescribing diversity; these are particularly important if patient characteristics—such as insurance coverage, mandatory substitution laws, patient preferences, and disease severity—or other local factors evolve in ways that affect prescribing and are correlated with measured physician technology adoption.

Finally, we evaluate directly the impact of information access on P_{ijt}, the probability that drug j is prescribed by doctor i at t. It is of particular interest to understand how database users' P_{ijt} values across new and old drugs j change after database adoption. Moreover, because new patent-protected products differ from new generics in both cost and novelty, access to the database may also tilt prescribing based on the patent status of a product. We thus evaluate whether physicians using the database are more or less likely to prescribe a product of a given type using the following specification

$$
\begin{aligned}
\text{P}\{X_{ijt} > 0\} = \eta_{jt} + \eta_i + \eta_{zt} + \delta N_{it-1} &+ [\beta_0\, Gen_j + \beta_1 (1 - Gen_j)] \times Z_{it} \times New_{jt}^\tau \\
&+ [\beta_2\, Gen_j + \beta_3 (1 - Gen_j)] \times Z_{it} \times (1 - New_{jt}^\tau) + \varepsilon_{ijt},
\end{aligned}
\tag{7}
$$

where $\text{P}\{X_{ijt} > 0\}$ is the probability that physician i writes at least one prescription for drug j during month t, Gen_j is an indicator that is equal to 1 if product j is a generic variety, and New_{jt}^τ indicates whether drug j is within τ months of its initial approval for U.S. sale. The main coefficients of interest β_0, β_1, β_2, and β_3 jointly capture the association between database use Z_{it} and prescribing propensity for both new drugs (β_0, β_1) and established products (β_2, β_3), where finding $\beta_0 > 0$ would indicate that database users are more likely to prescribe a given drug j that is both new (within τ months of initial market release) at t and generic, relative to other physicians. Similarly, finding that $\beta_1 > 0$ would indicate that database users are more likely to prescribe a new, branded product j. Note that the estimates of (7) have implications for T_{ij} and M_{it} through (3) and (4) above.

Equation (7) includes three sets of fixed effects, in line with (2). Drug-month effects η_{jt} account for the average perceived quality of drug j across physicians at t (θ_{jt}), which may depend on factors such as drug potency and side effects known at t, as well as the average expected pharmacy price at t. As with our other estimating equations, we further

include physician fixed effects η_i that absorb any individual characteristics affecting ϕ_{it} or the prescribing propensity such as location, patient composition, and physician age, education, and medical specialty. The coefficients of interest $\boldsymbol{\beta}$ are thus identified primarily from within doctor variation over time in information access Z_{it}. Zipcode-month fixed effects η_{zt} absorb any dynamic, location-specific determinants of ϕ_{it} or prescribing that may be correlated with physicians' database use; standard errors are adjusted for clustering at the zipcode-month level, and we control for N_{it-1} directly.

4 Main Results

4.1 Time to First Prescription

We begin by evaluating the relationship between a physician's database use and whether she adopts a new drug j within a year of its release. The model indicates that physician i is faster to begin prescribing j if she is a database user, for any new drug j satisfying $\partial P_{ijt}(\phi_{it})/\partial\phi_{it} > 0$. For these drugs, we thus expect a positive coefficient on Z_{ij} (Database$_{ij}$), where Z_{ij} takes a value of 1 if physician i has access to the drug reference database at the time a new drug j receives approval for sale in the U.S. market, and is otherwise zero.[27]

Estimates of (5) appear in Table 4. Columns 1 and 2 support the idea that database users are more likely, on average, to begin prescribing newly-approved drugs early, within their first year. The estimated coefficient on Z_{ij} is positive and highly significant in both columns, suggesting users are 1.91 percentage points more likely than non-users to write their initial prescription for a new drug within its first year (column 1); the estimate changes to 0.89 percentage points if we include physician fixed effects (column 2).[28] The data also confirm that doctors with large prescription volumes N_{it-1} are also significantly faster to begin prescribing a new drug, consistent with the model's qualitative predictions.

Columns 3 and 4 assess potential differences between new brand-name and new generic drugs. With prescriber fixed effects in column 4, we find that the estimated effect for generics (Database$_{ij} \times$ Generic$_j$) remains positive and significant, while that on branded products is indistinguishable from zero. Specifically, physicians using the database are 2.13 percentage points more likely to prescribe a new generic within its initial year, but are no faster in the case of new branded drugs.[29] That database use may tilt prescribing toward faster generic

---

[27]Throughout Section 4, Database$_{ij}$ (Z_{ij}) is the Database and Use Indicator described in Table 2.

[28]For comparison, we replicate Table 4 using $\ln T_{ij}$ as the outcome variable. Truncation in T_{ij} is addressed by limiting the duration of analysis to a window of 54 months following each new drug introduction and omitting Simcor; 54 months is the time span between the release of simvastatin and December 2010. The results are qualitatively similar to those in Table 4, and imply that database users begin prescribing a newly-approved drug 1.2 months or 0.83 months sooner, on average, than non-users. We also consider P$\{T_{ij} \leq 6\}$ and P$\{T_{ij} \leq 24\}$ as alternatives, and find results for each that are qualitatively identical to those in Table 4.

[29]Notice that the zipcode-drug fixed effects ensure that this result is not explained by differences in local

adoption has potentially significant aggregate cost implications given the size of the market for cholesterol therapies and the chronic nature of the condition they treat.[30] Moreover, because generic drugs share identical clinical attributes with branded versions, these results strongly suggest database users may be responding to the increased salience of non-clinical information—in particular, price and insurance formulary data.

While the estimates in columns 1–4 rely on the full sample of U.S. physicians, these results may be sensitive to differences across database users and non-users in the evolution of prescription outcomes. To better isolate the within-doctor impact of database use, columns 5 and 6 restrict the physician sample, including only those that both register for and use the drug database during the sample period. Doctors that have yet to adopt the database by December 2010 are thus omitted. With this restriction, the estimates in column 6 suggest, as in column 4, that database users are indeed significantly faster to begin prescribing new generics, but show no significant effects on the adoption of new branded drugs.

Across all columns, zipcode-drug fixed effects absorb variation across locations and over time in a) access to other drug information sources (e.g. advertising), b) physicians' tendency to adopt new technology, c) patient characteristics affecting the price or match quality of drug j, and d) competition. Any component of factors a) through d) that is stable over time is further captured by physician fixed effects in columns 2, 4, and 6.

4.2 Prescription Diversity

To evaluate the impact of physicians' database use on the diversity of prescribing, Table 5 provides estimates of (6) for two outcome variables: the number of unique drugs M_{it} prescribed by physician i at t (columns 1 and 3), and the associated Herfindahl-Hirschmann index, HHI_{it} (columns 2 and 4). The coefficient of interest is on Database_{it} (Z_{it}), which takes a value of 1 if physician i has access to the drug database at t, and is otherwise zero.

In the model, prescribing diversity increases when a physician adopts the drug database if the prescription probability P_{ijt} rises more, on average, than it falls—that is, if adoption induces an increase in ϕ_{it} and if $\sum_{j=1}^{J_t} \partial P_{ijt}(\phi_{it})/\partial \phi_{it} > 0$. The results in columns 1–4 are strongly consistent with this. In column 1, the estimated coefficient on physician-i database access Z_{it} is indeed positive and highly significant, and indicates that database users prescribe, on average, 0.123 additional unique drug varieties each month relative to a non-user.[31] In column 2, we find that HHI_{it} is also strongly responsive to database adoption: the estimate -0.006 is negative and highly significant, indicating that database users' prescribing

---

mandatory substitution laws. We nevertheless provide additional results regarding the effects of substitution laws on prescribing outcomes in Section 5.

[30]See Section 7.

[31]Because M_{it} is a count variable, we re-estimate the coefficients in columns 1 and 3 using both Poisson and Negative Binomial estimators, and find qualitatively identical results.

is substantially less concentrated across drugs j. Consistent with the model, we also find that prescription diversity increases significantly in a physician's monthly prescription volume.

In columns 3–4, physicians differ only in their respective drug reference adoption dates, yet the estimates are qualitatively identical to those in columns 1–2: prescription diversity increases significantly when physicians adopt the drug database. The estimates in columns 3–4 are considerably smaller than the full-sample results, however. This suggests the possible influence of unobserved factors, correlated with Z_{it}, that influence prescription diversity. We discuss this concern in more detail in Section 5 below.

4.3   Prescription Propensity

We report estimates of the propensity equation (7) in Table 6, again for all physicians (column 1) and for the sample restricted to eventual database users (column 2). The outcome variable $1\{X_{ijt} > 0\}$ is binary, indicating whether physician i writes at least one prescription for drug j at t, and the main coefficients of interest β_0 and β_1 correspond to Database$_{it}$ (Z_{it}) and its interaction with indicators for new generic and branded drugs, respectively.[32] We define j to be new while it is within $\tau = 24$ months of its initial market approval.

As described above, the model predicts that the coefficients β_0, β_1, β_2, and β_3 could be positive or negative depending on the distribution across drugs in the unobserved quality terms v_{jt} and ϵ_{njt} about which physicians learn as ϕ increases. P_{ijt} will tend to increase in ϕ_{it} for drugs with high v_{jt} values, but will generally fall if v_{jt} is relatively low. Whether database adoption increases or decreases prescribing for a particular drug type is thus an empirical question. It is, however, one to which we can partially predict the answer, given the model and results in Tables 4 and 5 above. First, Table 4 indicates physicians using the database are faster to begin prescribing new generics, but mildly slower to adopt new branded drugs; if the model is correct, this suggests, through (3), that $\beta_0 > 0$ and $\beta_1 < 0$. Second, prescription diversity increases with database adoption in Table 5; in (4), this implies that the sum across β_0, β_1, β_2, and β_3, weighted by the number of drugs per category, is positive.

Across both columns in Table 6, it is clear that physicians with database access are substantially more likely to prescribe generic products, regardless of vintage: $\beta_0 > 0$ and $\beta_2 > 0$. Specifically, these estimated coefficients are positive and highly significant, indicating the likelihood physician i prescribes a new generic drug is 1.62 percentage points higher when she has access to the database, and 2.40 percentage points higher for an older generic product in the restricted sample (column 2). The coefficients on branded drugs are much smaller

---

[32]The estimates in Table 6 span January 2000 through December 2010. To manage the computational burden of the sample size, we include only observations in January 2000 and each subsequent June and December in estimating the coefficients. The results are not qualitatively or quantitatively sensitive to this sample size reduction.

in magnitude, but still precisely estimated: users are 0.15 percentage points more likely to prescribe an old, branded product ($\beta_3 > 0$), and are 0.42 percentage points *less* likely to prescribe a new, branded product ($\beta_1 < 0$). Viewed through the lens of the model, the estimated coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ are thus exactly aligned with the results in Table 4. Moreover, the coefficients in Table 6 are consistent with the diversity estimates in Table 5. Accounting for the distribution of drug types in the data, the weighted sum across the coefficients $\boldsymbol{\beta}$ in column 2 is always positive. The predictions of the model, in light of the estimates in Tables 4 and 5, are thus strongly consistent with the results in Table 6.

Importantly, the estimates in Table 6 provide additional suggestive evidence regarding the mechanism through which database adoption may influence prescribing. That the impact of database access is apparent not only for new, but also older products suggests that there could be important reasons for doctors to continually reference the drugs in question—possibly to learn about aspects of a drug that are either time-varying or patient specific. This ongoing process of information acquisition could thus be an important factor in explaining why a significant fraction of a new drug's diffusion occurs beyond its first two years (Figure 2), a feature of technology diffusion that has fascinated economists for decades (Manuelli and Seshadri 2014). Second, the estimated coefficients of interest are again consistent with responsiveness to economic information—price and formulary status—the inclusion of which is a distinctive feature of the drug database we consider. In particular, users tilt prescribing away from new, branded products—for which prices are generally high—and strongly toward generic products, for which prices are low. That the estimated increases in generic propensity (β_0, β_2) are so large in magnitude relative to those for branded products further suggests the response to price information is economically important—which is not only surprising, but also indicates database adoption could therefore have significant implications for aggregate prescription costs.

5   Interpreting the Results

The results in Tables 4 through 6 above indicate that physicians using the point-of-care reference begin prescribing new generics sooner than non-users, prescribe a more diverse set of drugs than non-users, and more often prescribe a generic drug. Because database adoption is not randomly assigned, however, a key question that remains is whether a causal interpretation of our results is supported by the data. Our fixed-effects approach helps rule out certain alternative explanations including cases in which 'early-adopters' or physicians facing intense local advertising exposure begin both using the database and prescribing a new drug sooner than other physicians—for reasons unrelated to the actual impact of database use. A limitation of this approach arises when database adoption is either correlated with, or

involves selection on, time-varying, idiosyncratic physician characteristics that are relevant to the prescribing outcomes we consider.

In the case of advertising, for example, if an individual physician's database adoption timing were to be correlated with the idiosyncratic change in her own exposure to detailing—that is, the component not accounted for by zipcode-month effects—our estimates of the impact of adoption could be confounded with the influence of drug advertising. Without granular data on detailing, it is not possible to know whether there are meaningful shocks to detailing efforts that differ substantially across physicians within a U.S. zipcode-month for the drugs we consider. It is important to note, however, that, our results indicate database adopters are primarily quicker to prescribe new generic drugs, which are not advertised.[33]

Relatedly, it is possible in Table 5 that a physician's decision to adopt the database is partially determined by the underlying, idiosyncratic rate of increase in her prescribing diversity: naturally, adopting the database today could be a more attractive option for a physician who anticipates prescribing a wider range of products in the future, than for a physician in the opposite situation.[34] In this section we describe additional results in order to clarify the degree to which a causal interpretation is warranted and to determine which alternative interpretations can be ruled out.

5.1 Endogenous Database Adoption

One approach to handling the endogeneity of database adoption is to find an instrument that generates quasi-random variation in a physician's database adoption decision, and to estimate the impact of information access relying on variation in this instrument. We have considered three such instruments: 1) a measure of location-year specific hospital I.T. use from Dranove et al (2014), and 2) a measure of location-year specific high-speed internet penetration, and 3) the doctor-month specific share of other local physicians that have adopted the reference database; all three are factors that could influence doctors' database adoption decisions while being plausibly unrelated to choices over which anti-cholesterol drugs to prescribe. While we find that the first two instruments are only weak predictors of database adoption, resulting in second-stage estimates highly sensitive to small specification changes, the third instrument is a robust predictor of database adoption.

We therefore reassess the results in Tables 4, 5, and 6 using this third instrument. Estimates appear in Tables 7, 8, and 9, respectively. The logic underlying the first stage is

---

[33]The data studied by Larkin et al (2017) show that detailing is almost exclusively done for branded drugs.

[34]While U.S. physicians always face the same set of drugs approved for prescription, and in that sense do not differ in changes to the range of products available, their exposure to such changes may differ due to the potentially distinct characteristics of the specific patients they treat. To the extent that these distinctions are time-varying and correlated with database adoption, even across doctors practicing in the same zipcode-month, our fixed effects estimator could yield biased estimates.

identical across all three specifications. Consider Table 5: the first stage corresponding to a version of (6) that replaces η_{zt} with η_t is

$$Z_{it} = \alpha \, S_{z(i)t-1} + \nu_i + \nu_t + \xi N_{it-1} + u_{it}, \tag{8}$$

where $S_{z(i)t-1}$ is the share of physicians, excluding i, that are located in i's zipcode $z(i)$ and are database users at $t-1$.[35] The identification restriction is that the instrument is conditionally uncorrelated with the error term in (6): $\text{Cov}(S_{z(i)t-1}, \varepsilon_{it} | \eta_i, \eta_t, N_{it-1}) = 0$.

Most explanations linking neighbors' database adoption decisions would suggest $\alpha > 0$ in (8). One possible mechanism is that neighboring physicians are likely to share information about tools and techniques that improve professional performance; alternatively, physicians interacting locally may simply observe a peer accessing the database, and may decide to adopt on that basis. As discussed above, the reference firm's public statements indicate that such informal peer effects were the most important driver of database adoption among doctors during the sample period. Regarding excludability, it does not seem likely that the database adoption decisions of physicians neighboring i would directly affect i's own prescription decisions, as patient medical information is privacy-protected by law.[36]

Whether the bias in our baseline estimates is upward or downward hinges on the relationship between the omitted factor and database adoption (Z_{it}). One possibility is that a physician makes an unobserved decision to move closer to the efficiency frontier by adopting new generic drugs quickly, prescribing them more often in place of a branded drug, and increasing the influence of patients' diverse characteristics on prescription choice—and that as a result of this unobserved decision, the doctor now finds it profitable to begin using the database to assist her increasingly complex prescription choices. This would suggest $\text{Cov}(Z_{it}, \varepsilon_{it}) \geq 0$ in (6) and a corresponding upward bias in our baseline estimate of β. On the other hand, if time-constrained physicians choose among multiple sources of drug information, those adopting the reference we consider (with $Z_{it} = 1$) may do so at the expense of relying on a substitute resource that could affect prescribing similarly.[37] This would tend

---

[35]Notice that this instrument is not valid in the presence of zipcode-month fixed effects η_{zt}. With η_{zt}, identification through the instrument comes from comparing two doctors i and i' in the same zipcode and month that have different adoption shares $S_{z(i)t-1}$ and $S_{z(i')t-1}$. However, such a difference arises only when exactly one of the two has adopted the database. But in this case, the instrument for i directly reflects Z_{it-1}, which is correlated with Z_{it} and thus also with the prescription outcome M_{it}. We therefore replace the η_{zt} (or η_{zj}) with η_t (or η_j) in the instrumented specifications.

[36]There are channels through which $S_{z(i)t-1}$ could be correlated ε_{it} in the second-stage equation. Changes in neighbors' database adoption could reflect changes in the proximity of their location to the technology frontier, possibly due to advertising; this could manifest itself not only in a high early rate of database adoption but also a high early rate of new drug adoption. Neighboring physicians that share information about database use could also share information about new drugs, affecting prescribing even if patient-specific details are not discussed. Such cases are better addressed by the baseline specification (6), which includes zipcode-month fixed effects that absorb local changes of these types.

[37]For example, relevant alternatives include the Micromedex and UpToDate Lexicomp databases.

to result in $\text{Cov}(Z_{it}, \varepsilon_{it}) \leq 0$ in (6), placing downward pressure on the estimate of β.

The estimates in Tables 7 through 9 are largely in line with the latter interpretation. Based on our preferred second-stage estimates in Table 7, column 6, database users are 6.10 percentage points more likely to begin prescribing a new generic within its initial year, relative to a non-user, with no significant effects among new branded drugs. The magnitude of the effect for generics is larger than in Table 4, suggesting a bias toward zero for this coefficient; the coefficients for branded products are statistically indistinguishable. In Table 8, column 4 similarly suggests that the true impact of database adoption on diversity is an order of magnitude larger than in Table 5; however, this result is limited to prescription concentration HHI_{it}, with the effect on M_{it} insignificant in Table 8, column 3. Table 9 suggests that Table 6 understates the impact of the database on generic prescribing by a factor of four for new generics (β_0) and a factor of ten for old generics (β_2).[38] However, database adopters are less likely to prescribe old branded drugs ($\beta_3 < 0$). Each table includes the first-stage estimates and reports the F statistic, which in every case is substantially higher than its weak-instrument threshold value.

Overall, the IV estimates reported in Tables 7–9 confirm a pattern of database impacts that is similar to the corresponding OLS estimates, but with larger estimated magnitudes. We view these as a set of robustness checks, with results that lend credibility both to the qualitative effects estimated above, and to a causal interpretation of these effects. Nevertheless, it is worth noting that our leave-out mean instrument relies on variation in group composition that in many applications leads to small-sample bias from weak instruments, and that could confound interpretation in certain cases (Angrist 2014). Regarding the former, we have confirmed the strength of the instrument, aided by the fact that the instrument varies over 13,000 zipcodes $\times$ 131 months = 1.7 million observations, across which there is sufficient heterogeneity for identification. Regarding the latter concern, that groups with high database adoption rates could also have different prescribing tendencies due to factors other than actual database access, three observations are useful. First, such an unobserved factor would need to cause a correlation between zipcodes' database adoption timing and increased generic prescribing *in the absence* of effects on brand-name drug adoption; second, provided this unobserved factor varies by zipcode-month, then it is reassuring that the baseline estimates including these fixed effects are qualitatively similar (Tables 4–6); third, we find that the results in tables 7–9 are essentially unchanged when defining the instrument at a broader geographic unit (three digit zipcodes).

---

[38]In the interest of space, first-stage estimates for Table 6 appear in Appendix Table A.2.

5.2 Timing the Impact of Database Adoption

For an important class of alternative explanations, physicians either adopt the reference database in response to pre-existing changes in prescribing, or are influenced by omitted, dynamic factors that simultaneously affect both prescribing and database adoption. A symptom that would likely appear where these influences are active is adopters who exhibit a trend toward the predicted outcomes, even before accessing the database. To account for this possibility, we consider variants of (6) and (7) that include a full set of physician-specific time trends. The results appear in Tables 10 and 11. Considering the estimates in Table 10, it is clear that including doctor trends reduces the estimated impact of database adoption in the overall sample (columns 1 and 2) relative to the baseline estimates in Table 5; however, the impact *increases* within our preferred specifications (columns 3 and 4) that include only eventual database adopters. Relative to Table 6, the results in Table 11 are essentially identical. Accounting for physician-specific trends thus leaves unchanged the main qualitative implications of Tables 5 and 6.

5.3 The Intensity of Database Use

The model proposes a specific mechanism: physicians' prescription outcomes are influenced by database access because the information obtained through the database is important, yet otherwise unknown. If this proposed mechanism is correct, then if we observe physicians using the database to search for cholesterol drugs with different intensities, those searching more intensely should have correspondingly larger prescription responses. There could certainly be other explanations linking search intensity with the prescribing outcomes we consider, but this evaluation is informative nevertheless: if intensity and response are *not* linked in the data, it would strongly suggest the proposed mechanism is invalid.

With these considerations in mind, we make use of a key feature of the data that allows us to observe not only a physician's database registration date, but also a proxy for the extent of her drug-month specific database use, conditional on adoption. The variable is unfortunately not an exact lookup count, as the database company changed the way it maintained lookup data over time; given this, we aggregate the lookup proxy into a coarser measure reflecting a lower bound for each physician's overall intensity of database use. We then divide the sample of physicians into three groups based on this lower bound, and reassess the results in Tables 4, 5, and 6 allowing different coefficients for each intensity group: low-intensity users are database adopters for whom the lower bound on cholesterol drug lookups is zero; the high-intensity group includes doctors whose total lookup proxy is in the top decile. The results appear in Tables A.3, A.4, and A.5 in the Appendix.

The estimates in Table A.3, columns 1 and 2, reveal that the impact of database adop-

tion is systematically and monotonically increasing in the intensity of usage. In column 1, the low-intensity users are 0.32 percentage points more likely to adopt new drugs within one year of release, while high-intensity users are 3.41 percentage points more likely to do so; both coefficients are strongly significant, but reveal meaningful heterogeneity across database adopters in its prescribing impact. Column 2 reveals a similar pattern of increasing impact, and column 6 confirms the finding in Table 4 that database use systematically speeds adoption of new generic products.

Analogous results obtain in Tables A.4 and A.5. The estimates in Table A.4 indicate that diversity increases with database adoption only for high-intensity users; Table A.5 suggests that the highest-intensity users are the most responsive to database access in switching away from branded drugs and toward generic products. Taken together, these results are consistent with the mechanism we propose in the model.

5.4 Mandatory Substitution Laws

To encourage cost savings, many U.S. states impose regulations mandating generic substitution where available; in most cases, such laws have been in force since the 1970s (Grabowski and Vernon 1979) and were thus in effect during the sample period. This means that pharmacists dispense the generic version of a drug even if the physician prescribed the branded version. Since our IMS data are collected from pharmacies and are based on prescriptions dispensed, a possible concern is that the patterns in our data reflect pharmacists' behavior rather than physicians' prescribing behavior—for example if the implementation of mandatory substitution laws differs across states and over time in a manner correlated with physicians' adoption of the drug database.

On this point it is important to note that all of our baseline results include zipcode-drug or zipcode-month fixed effects that absorb any impact of mandatory substitution laws on the prescription outcomes we consider—even where these laws may be correlated with physicians' database adoption choices. Nevertheless, to further check the robustness of our results to this potential concern, we re-estimate the regressions reported in Table 4, replacing the zipcode-drug fixed effects with a set of drug fixed effects, and splitting the sample between doctors in states with vs. without mandatory substitution laws. We also consider triple-interacted specifications in which the effect of database adoption is allowed to depend on whether the doctor is in a mandatory substitution state. The results, shown in Table A.6, indicate that database adoption is associated with faster adoption of new generics regardless of whether the physician is in a mandatory substitution state.

5.5   Pharmaceutical Innovation

Physicians practicing in locations known for pharmaceutical innovation may have access to frontier knowledge regarding pharmaceutical development and pricing, limiting the potential for the reference database we observe to influence prescribing decisions. Within the conceptual framework outlined in section 3, proximity to the frontier could imply physicians have initially high ϕ_{it} parameters that are either minimally or not responsive to database use. If so, database use has little potential to affect prescribing. The zipcode-drug or zipcode-month fixed effects included in the baseline specifications account for the innovativeness of a physician's local environment. To assess whether location-specific differences in innovativeness impact the mechanism, we therefore replace these effects with either drug or month fixed effects, and use patent data from the NBER U.S. Patent Citations Data File (Hall, Jaffe, and Trajtenberg 2001) to measure the number of pharmaceutical patents granted between 1975 and 1999 by zipcode. We then re-evaluate the adoption-lag specifications in Table 4 separately for two samples corresponding to the top and bottom five percent across zipcodes based on the number of pharmaceutical patents granted. We also consider interacted specifications that account for differences in pharmaceutical patenting across locations.

The estimates appear in Table A.6, and suggest that physicians plausibly located near the knowledge frontier—that is, physicians in zipcodes among the top five percent by drug patenting—indeed respond to drug information differently than their more distant peers. Specifically, the estimates indicate that while use of the database in the least-innovative locations is associated with a larger impact on the likelihood of new drug adoption within one year (columns 2 and 6) relative to the most innovative locations (columns 1 and 5). Considering the full sample, columns 3 and 7 both indicate that the database speeds generic adoption on average, but has especially pronounced effects among the least-innovative locations that are likely to be far from the information frontier. Columns 4 and 8 confirm this result using a continuous measure of local patenting. Innovative areas adopt generics more quickly regardless of database adoption, but on average, a physician using the database in these locations is significantly less responsive to the information in terms of new generic drug adoption. The database nevertheless has an independent effect, speeding the adoption of new generic drugs regardless of patenting.

5.6   Other Robustness Checks

We re-estimate the results involving the count variable M_{it} in Table 5 using Poisson or Negative Binomial regressions, and we estimate specifications involving the binary variables in Tables 4 and 6 using logistic fixed effects regressions. We reevaluate Table 6 including doctor-drug fixed effects and zipcode-month-drug type fixed effects. To allow for persistence

in prescription outcomes, we also control for the first lag of each outcome in Tables 5 and 6.[39] Each of these robustness checks reveals qualitatively similar results.[40]

Finally, a physician's decision to prescribe a generic drug may be related to the insurance coverage of her patient population. We therefore evaluate split-sample estimates based on whether physicians receive a high or low share of Medicare and Medicaid patients, relative to the privately insured; separately, we repeat this split-sample analysis, distinguishing physicians based on whether a high or low share of their patients pay for prescriptions with cash. In both cases, we find negligible differences across groups.

6   Physician Heterogeneity

As a final point, we consider whether the data are broadly consistent with the idea that incomplete information contributes to disparities in prescribing behavior across physicians. If indeed these disparities partially reflect systematic informational differences, then physicians sharing access to a common source of drug information, like the reference database in our study, should tend to exhibit observable homogenization relative to other doctors. We consider this possibility using a simple approach. We first assign each physician to one of two groups based on her database registration status in December 2010. Then, within each group, we measure the extent of prescribing heterogeneity across physicians: specifically, we determine the vector of prescription shares for each prescriber i in December 2010, and then compute the Euclidean distance between this physician-i vector and the average vector of prescription shares among physicians in her group (database users or non-users).

These within-group similarity measures are reported in Table A.8, Panel A. The prescription shares of database users are indeed more homogeneous than those of non-users. The average Euclidean distance between the physician-i prescription vector and the group-specific average is 0.152 for users and 0.176 for non-users, and the difference (-0.0236) is highly statistically significant. Importantly, note that database users prescribe a significantly more diverse set of products than non-users, as shown above in Table 5; the relative homogeneity of database users' prescribing patterns thus does not imply a loss of variation in therapies generally. Rather, the result implies that physicians who are connected to the same information source resemble each other more closely in spite of the fact that they tend to prescribe a more diverse set of drugs.

Of course, the fact that database users' prescribing patterns are less heterogeneous could reflect selection rather than any causal effect of information access. Indeed, Panel B of Table A.8 shows that *eventual* database users' prescribing exhibited greater homogeneity than non-

---

[39]Structural persistence could arise in the presence of persistent patient specific match quality ϵ_{njt} terms in the model, given the chronic nature of the relevant medical condition.

[40]Detailed results available on request.

users' even in January 2000, before anyone was using the database. But, the changes by group between 2000 and 2010, summarized in Panel C, indicate that while a) both groups (users and non-users) exhibit homogenization over time, with the average Euclidean distance declining by 0.052 for users and by 0.040 for non-users, b) the difference in differences is also highly significant—i.e., significantly more within-group homogenization is observed among database users than among non-users, even when controlling for physician fixed effects.[41]

Our data cannot definitively say whether the faster convergence for database users was directly caused by the database: unlike the regression analyses reported in Section 4, the results described in Table A.8 are based on across-group comparisons rather than within-doctor comparisons, so the stronger trend toward homogenization among adopters could reflect other characteristics that are correlated with the decision to adopt. Nevertheless, we view these results as suggestive of the idea that database use could reduce disparities in care—an idea that merits exploration in future research.

7 Conclusion

This paper has empirically examined how physicians' prescribing decisions are affected by access to a drug reference database at the point of care. Using a novel dataset that includes prescription choices and drug reference use for over 125,000 individual U.S. physicians, we find that after adopting the database, users increase the likelihood of prescribing a generic drug, are faster to begin prescribing a newly-released generic, and yet also significantly increase the diversity of products prescribed each month. These results are consistent with the predictions of a simple, incomplete-information model of prescription choice, and are robust across specifications that control for physician and location-month unobserved prescribing determinants and that treat the timing of physicians' database adoption as endogenous.

While the magnitude of database users' estimated shift toward generic drugs is modest at the prescriber level, the implied aggregate impact on drug spending is economically significant. U.S. pharmacies filled approximately 170 million cholesterol drug prescriptions in 2010, for example, a year during which roughly 45 percent of sample physicians were users of the drug reference database, and during which the average price difference between branded and generic cholesterol-drug prescriptions was around \$94.[42] If 45 percent of these prescriptions correspond to database users, and if users' generic shares increase by even half

---

[41]For clarity, Panel C reports coefficients from a least-squares regression of the Euclidean distance to the mean D_{it} for doctor i at $t = \{$January 2000, December 2010$\}$ on an indicator I_{2010} for December 2010, its interaction $Z_{2010} \times I_{2010}$ with an indicator $Z_{i,2010}$ for physician-i database access in December 2010, and physician fixed effects.

[42]Assuming each prescription is for a standard 30-day supply, this estimated price difference based on Marketscan data for December 2009 is conservative; CMS data indicate the price gap corresponding to Medicare and Medicaid patients is substantially wider.

a percentage point—approximately the magnitude of the measured effect of database use in our data—the implied annual cost savings of database usage would exceed $35 million for cholesterol drugs alone.[43] If effects of the same magnitude apply to all drug classes, the implied savings would be on the order of $1 billion annually.

More importantly, our study speaks to policy debates regarding the efficiency of U.S. healthcare provision, particularly those concerning unwarranted disparities in the observed cost and quality of medical care (Wennberg et al 1996) including that involving prescription drugs (Munson et al 2013). Our results provide new, systematic evidence that information differences contribute significantly to treatment variation across U.S. physicians, and suggest that connecting physicians to common, high-quality information sources has the potential to meaningfully increase the efficiency of health care delivery.

REFERENCES

[1] **Agha, Leila.** 2014. "The Effects of Health Information Technology on the Costs And Quality of Medical Care." *Journal of Health Economics*, vol. 34, pp. 19–30.

[2] **Agha, Leila and David Molitor.** 2015. "The Local Influence of Pioneer Investigators in Technology Adoption: Evidence from New Cancer Drugs." *Review of Economics and Statistics, forthcoming.*

[3] **Angrist, Joshua D.** 2014. "The Perils of Peer Effects." *Labour Economics*, vol. 30, pp. 98–108.

[4] **Arrow, Kenneth J.** 1969. "Classificatory Notes on the Production and Transmission of Technological Knowledge." *American Economic Review, Papers and Proceedings*, vol. 59, no. 2, pp. 29–35.

[5] **Athey, Susan, and Scott Stern.** 2002. "The Impact of Information Technology on Emergency Health Care Outcomes." *RAND Journal of Economics*, vol. 33, no. 3, pp. 399–432.

[6] **Attewell, Paul.** 1992. "Technology Diffusion and Organizational Learning: The Case of Business Computing." *Organizational Science*, vol. 3, no. 1, pp. 1–19.

[7] **Berndt, Ernst R., Robert S. Gibbons, Anton Kolotilin, and Anna Levine Taube.** 2015. "The Heterogeneity of Concentrated Prescribing Behavior: Theory and Evidence from Antipsychotics." *Journal of Health Economics*, vol. 40, pp. 26–39.

[8] **Black, Sandra E., and Lisa M. Lynch.** 2001. "How To Compete: The Impact of Workplace Practices and Information Technology on Productivity." *Review of Economics and Statistics*, vol. 83 no. 3, pp. 434–445.

---

[43]To determine the relationship between physicians' generic prescription share and database use, we estimate a version of equation (7) that replaces M_{it} with the share of physician-i prescriptions in month t that are accounted for by generics. The estimated coefficient on Z_{it} is 0.0061 (standard error 0.00021).

[9] **Bloom, Nicholas, Luis Garicano, Raffaella Sadun, and John Van Reenen.** 2009. "The Distinct Effects of Information Technology and Communication Technology on Firm Organization." National Bureau of Economic Research Working Paper 14975.

[10] **Bloom, Nicholas, Raffaella Sadun, and John Van Reenen.** 2012. "Americans Do IT Better: US Multinationals and the Productivity Miracle." *American Economic Review*, vol. 102, no. 1, pp. 167–201.

[11] **Bresnahan, Timothy F., Erik Brynjolfsson, and Lorin M. Hitt.** 2002. "Information Technology, Workplace Organization, and the Demand for Skilled Labor: Firm-Level Evidence." *Quarterly Journal of Economics*, vol. 117, no. 1, pp. 339–76.

[12] **Bresnahan, Timothy F., and Shane Greenstein.** 1996. "Technical Progress and Co-invention in Computing and in the Uses of Computers." Brookings Papers on Economic Activity Microeconomics: pp. 1–8.

[13] **Bronnenberg, Bart J., Jean-Pierre Dubé, Matthew Gentzkow, and Jesse M. Shapiro.** 2015. "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium." *Quarterly Journal of Economics*, vol. 130, no. 4, pp. 1669–1726.

[14] **Brooks, John M., Elizabeth A. Cook, Cole G. Chapman, et al.** 2014. "Geographic Variation in Statin Use for Complex Acute Myocardial Infarction Patients." *Medical Care*, vol. 52, no. 3 Suppl 2, pp. S37-S44.

[15] **Brot-Goldberg, Zarek, Amitabh Chandra, Ben Handel, Jonathan T. Kolstad.** 2015. "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." National Bureau of Economic Research Working Paper 19320.

[16] **Brynjolfsson, Erik, and Lorin M. Hitt.** 2003. "Computing Productivity: Firm-Level Evidence." *Review of Economics and Statistics*, vol. 85, no. 4, pp. 793–808.

[17] **Cannon, Christopher P., Michael A. Blazing, Robert P. Giugliano, et al.** 2015. "Ezetimibe Added to Statin Therapy After Acute Coronary Syndromes." *New England Journal of Medicine*, vol. 372, no. 25, pp. 2387–2379.

[18] **Cannon, Christopher P., Eugene Braunwald, Carolyn H. McCabe, et al.** 2004. "Intensive Versus Moderate Lipid Lowering with Statins After Acute Coronary Syndromes." *New England Journal of Medicine*, vol. 350, no. 15, pp. 1495–1504.

[19] **Caselli, Francesco, and Wilbur John Coleman.** 2001. "Cross-Country Technology Diffusion: The Case of Computers." *American Economic Review*, vol. 91, no. 2, pp. 328–335.

[20] **Castelli W. P., Garrison R. J., Wilson P. W. F., et al.** 1986. "Incidence of Coronary Heart Disease and Lipoprotein Cholesterol Levels: The Framingham Study." *Journal of the American Medical Association*, vol. 256, no. 20, pp. 2835–2838.

[21] **Chernew, Michael E., Richard A. Hirth, and David M. Cutler.** 2009. "Increased Spending On Health Care: Long-Term Implications For The Nation." *Health Affairs*, vol. 28, no. 5, pp. 1253–1255.

[22] **Coleman, James, Elihu Katz, and Herbert Menzel.** 1957. "The Diffusion of an Innovation Among Physicians." *Sociometry*, vol. 20, no. 4, pp. 253–270.

[23] **Coleman, James, Elihu Katz, and Herbert Menzel.** 1966. *Medical Innovation: A Diffusion Study.* New York: Bobbs-Merrill Co.

[24] **Comin, Diego, and B. Hobijn.** 2004. "Cross-Country Technology Adoption: Making Theory Face the Facts." *Journal of Monetary Economics*, vol. 51, pp. 39–83.

[25] **Cooper, Zack, Stuart Craig, Martin Gaynor, and John Van Reenen.** 2015. "The Price Ain't Right? Hospital Prices and Health Spending on the Privately Insured." Yale University mimeo.

[26] **Crawford, Gregory S., and Matthew Shum.** 2005. "Uncertainty and Learning in Pharmaceutical Demand." *Econometrica*, vol. 73, no. 4, pp. 1137–1173.

[27] **Cutler, David, Jonathan Skinner, Ariel Stern, David Wennberg.** 2015. "Physician Beliefs and Patient Preferences: A New Look at Regional Variation in Health Care Spending." National Bureau of Economic Research Working Paper 19320.

[28] **Datta, Anusua, and Dhaval M. Dave.** 2013. "Effects of Physician-Directed Pharmaceutical Promotion on Prescription Behaviors: Longitudinal Evidence." *Health Economics, forthcoming.*

[29] **David, Paul A.** 1966. "The Mechanization of Reaping in the Ante-Bellum Midwest," Chapter 1 of H. Rosovsky, ed., *Industrialization in Two Systems: Essays in Honor of Alexander Gerschenkron*, New York: Wiley and Sons.

[30] **Dickstein, Michael J.** 2015. "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice." New York University mimeo.

[31] **Dranove, David, Chris Forman, Avi Goldfarb, and Shane Greenstein.** 2014. "The Trillion Dollar Conundrum: Complementarities and Health Information Technology." *American Economic Journal: Economic Policy*, vol. 6, no. 4, pp. 239–270.

[32] **Duggan, Mark, and Fiona Scott Morton.** 2010. "The Effect of Medicare Part D on Pharmaceutical Prices and Utilization." *American Economic Review*, vol. 100, no. 1, pp. 590–607.

[33] **Escarce, Jose.** 1996. "Externalities in Hospitals and Physician Adoption of A New Surgical Technology: An Exploratory Analysis." *Journal of Health Economics*, vol. 15, pp. 715–734.

[34] **Forman, Chris, Avi Goldfarb, and Shane Greenstein.** 2005. "How Did Location Affect the Adoption of the Commercial Internet? Global Village vs. Urban Leadership." *Journal of Urban Economics*, vol. 58, no. 3, pp. 389–420.

[35] **Fuchs, Victor R.** 1974. "Who Shall Live?" Basic Books, New York.

[36] **Gawande, Atul**. 2009. "The Cost Conundrum," *The New Yorker*. June 1.

[37] **Grabowski, Henry G., and John M. Vernon.** 1979. "Substitution Laws and Innovation in the Pharmaceutical Industry." *Law and Contemporary Problems*, vol. 43, no. 1, pp. 43–66.

[38] **Griliches, Zvi.** 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, vol. 25, no. 4, pp. 501–522.

[39] **Hall, Bronwyn H., Adam B. Jaffe, and Manuel Tratjenberg.** 2001. "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools," NBER Working Paper 8498.

[40] **Hubbard, Thomas N.** 2003. "Information, Decisions, and Productivity: On-Board Computers and Capacity Utilization in Trucking." *American Economic Review*, vol. 93, no. 4, pp. 1328–53.

[41] **Kastelein, John J., Fatima Akdim, Erik S. Stroes, et al.** 2008. "Simvastatin With or Without Ezetimibe in Familial Hypercholesterolemia." *New England Journal of Medicine*, vol. 358, no. 14, pp. 1431–1443.

[42] **Klenow, Peter, and Andres Rodriguez-Clare.** 1997. "Economic Growth: A Review Essay." *Journal of Monetary Economics*, vol. 40, pp. 597–618.

[43] **Law, M.R., N.J. Wald, and A.R. Rudnicka.** 2003. "Quantifying Effect of Statins on Low Density Lipoprotein Cholesterol, Ischaemic Heart Disease, and Stroke: Systematic Review and Meta-Analysis." *British Medical Journal*, vol. 326, pp. 1–7.

[44] **Larkin, I., D. Ang, J. Steinhart, M. Chao, M. Patterson, S. Sah, T. Wu, M. Schoenbaum, D. Hutchins, T. Brennan, and G. Loewenstein** 2017. "Quantifying Effect of Statins on Low Density Lipoprotein Cholesterol, Ischaemic Heart Disease, and Stroke: Systematic Review and Meta-Analysis." *Journal of the American Medical Association*, vol. 317, n. 17, pp. 1785-1795.

[45] **Ledford, Heidi.** 2013. "Cholesterol Limits Lose Their Lustre." *Nature*, vol. 494, pp. 410–11.

[46] **Manuelli, Rodolfo, and Ananth Seshadri.** 2014. "Frictionless Technology Diffusion: The Case of Tractors." *American Economic Review*, vol. 104, no. 4, pp. 1368–91.

[47] **Mozaffarian, Dariush, Emelia J. Benjamin, Alan S. Go, et al.** 2015. "Heart Disease and Stroke Statistics—2015 Update." *Circulation*, vol. 131, pp. e29–e322.

[48] **Munson, Jeffrey C., Nancy E. Morden, David C. Goodman, Luca F. Valle, and John E. Wennberg.** 2013. *The Dartmouth Atlas of Medicare Prescription Drug Use*, American Hospital Association Press, *at* http://www.dartmouthatlas.org.

[49] **Parente, Stephen L., and Edward C. Prescott.** 1994. "Barriers to Technology Adoption and Development." *Journal of Political Economy*, vol. 102, no. 2, pp. 298–321.

[50] **Phelps, Charles E.** 2000. "Information Diffusion and Best Practice Adoption." In A. J. Culyer & J. P. Newhouse (Eds.) *Handbook of Health Economics*, pp. 223–264.

[51] **Rossebø, Anne B., Terje R. Pedersen, Kurt Boman, et al.** 2008. "Intensive Lipid Lowering with Simvastatin and Ezetimibe in Aortic Stenosis." *New England Journal of Medicine*, vol. 359, no. 13, pp. 1343–56.

[52] **Skinner, Jonathan, and Douglas Staiger.** 2007. "Technology adoption from hybrid corn to beta-blockers." In *Hard-to-Measure Goods and Services: Essays in Honor of Zvi Griliches.* NBER Chapters, National Bureau of Economic Research, Inc, pp. 545–570.

[53] **Skinner, Jonathan.** 2012. "Causes and Consequences of Regional Variations in Health Care." *Handbook of Health Economics*, vol. 2, pp. 45–93.

[54] **Solow, Robert M.** 1956. "A Contribution to the Theory of Economic Growth." *Quarterly Journal of Economics*, vol. 70, no. 1, pp. 65–94.

[55] **Solow, Robert M.** 1987. "We'd Better Watch Out." *New York Times*, July 12, 36.

[56] **Ventola, C. Lee.** 2014. "Mobile Devices and Apps for Health Care Professionals: Uses and Benefits." *Pharmacy and Therapeutics*, vol. 39, no. 5, pp. 356–64.

[57] **Wennberg, John E. and Megan M. Cooper, eds.** 1996. *The Dartmouth Atlas of Health Care*, American Hospital Association Press, *at* http://www.dartmouthatlas.org.

[58] **Wennberg, David E. and John E. Wennberg.** 2003. "Addressing Variations: Is There Hope For The Future?" *Health Affairs*, December 10, pp. W3 614–617.

Appendix

A.1 Medical Innovation

Innovation in _hypercholesterolemia_ and _dyslipidemia_ therapy: Information about the evolving set of pharmaceutical therapies available for prescription was obtained from the U.S. Food and Drug Administration (FDA) for the period January 2000 through December 2010. Twelve new statin or lipid-lowering drugs, including new formulations, combinations, and versions, introduced during this period and are described below. These include three new molecular entities Crestor, Lovaza, and Zetia; three generic versions lovastatin (Mevacor), pravastatin (Pravachol), and simvastatin (Zocor); two new formulations Altoprev (extended-release Mevacor) and Lescol XL (extended-release Lescol); and four new drug combinations Advicor (extended-release niacin and Mevacor), Pravigard PAC (aspirin and Pravachol), Vytorin (Zetia and Zocor), Simcor (extended-release niacin and Zocor). A description of each drug innovation appears below based on publicly available data including approval letters and administrative, medical, and pharmacological review. Baycol was withdrawn early in the sample period in August 2001 and is thus omitted.

Existing therapies available in January 2000:
1. Lescol (fluvastatin) is a statin marketed by Novartis since its FDA approval as a new molecular entity on December 31, 1993; its patent protection expired in 2012. Like other statins, its mechanism of action is to limit a specific enzyme in the liver, preventing cholesterol synthesis.
2. Lipitor (atorvastatin) is a statin marketed by Pfizer. Its mechanism of action is similar to that of fluvastatin, but unlike other statins, atorvastatin is a synthetic compound. The therapy was approved by the FDA as a new molecular entity on December 17, 1996. Between 1996 and 2012, Lipitor was the best-selling drug globally; its patent expired in November 2011.
3. Mevacor (lovastatin) is the first statin to receive FDA approval. The drug was approved as a new molecular entity on August 31, 1987 for sale in the United States by Merck. The therapy was protected by patents through June 2001.
4. Niaspan (extended-release niacin) is vitamin B_3, or nicotinic acid, and is marketed by Abbott Laboratories. Extended-release niacin was approved for sale in the United States on July 28, 1997.
5. Pravachol (pravastatin) is a statin marketed by Bristol Myers Squibb since its FDA approval on October 31, 1991. In addition to inhibiting cholesterol synthesis, Pravachol also inhibits low-density lipoprotein synthesis. Two clinical trials, each completed in November 2003, suggest Pravachol is outperformed by both Zocor and Lipitor. Patent protection expired in June 2006.
6. Zocor (simvastatin) is a statin marketed by Merck since its FDA approval as a new molecular entity on December 23, 1991. Zocor outperformed Pravachol in its prevention of cholesterol synthesis in a clinical trial completed in November 2003. Patent protection expired in April 2006.

New chemical entities, January 2000–December 2010:
1. Crestor (rosuvastatin calcium) is a new molecular entity approved by the FDA for sale in the United States by Astra Zeneca Pharmaceuticals on August 12, 2003. The molecule acts by reducing intestinal absorption of cholesterol and related phytosterols, and is thereby distinct relative to other statin therapies. The drug was approved for use in treating primary hypercholesterolemia and mixed dyslipidemia (by reducing total-C, LDL-C, and Apo B), and as an adjunct to other lipid-lowering treatments. It was thus approved for use alone or with other statins. A 2008 clinical trial revealed additional evidence supporting the superior performance of Crestor compared with a placebo treatment. Patent protection expires in January 2016.
2. Lovaza (omega-3-acid ethyl esters) is a new molecular entity introduced by Abbott labs and approved by the FDA on November 10, 2004. It was initially introduced under the trade name Omacor. Unlike statins, Lovaza is aimed at reducing tricylerides rather than low-density lipopro-

teins and may thus be combined with a statin as an adjunct therapy. Patent protection expired in September 2012.

3. Zetia (ezetimibe) is a new molecular entity introduced by Schering and approved by the FDA on October 25, 2002 for sale in the United States. The molecule acts by reducing intestinal absorption of cholesterol and related phytosterols, and is thus distinct from statins. The drug was initially approved for use in treating hypercholesterolemia for use alone or with other statins. In January 2008, a clinical trial found Zetia performed poorly compared with other therapies, and it was at that time recommended that Zetia not be prescribed except in cases for which all other cholesterol drugs had previously failed. Patent protection expires in April 2017.

New generic versions, January 2000–December 2010:
1. Lovastatin is the generic equivalent of Mevacor, and was initially approved by the FDA for sale in the United States by Geneva Pharmaceuticals applied on December 17, 2001.
2. Pravastatin is the generic equivalent of Pravachol, and was initially approved by the FDA for sale in the United States by Teva Pharmaceuticals on April 24, 2006.
3. Simvastatin is the generic equivalent of Zocor, and was initially approved by the FDA for sale in the United States by Teva Pharmaceuticals on June 23, 2006.

New dosage forms, January 2000–December 2010:
1. Altoprev (extended-release lovastatin) is a new dosage form and was approved by the FDA on June 26, 2002 for sale in the United States, following a new drug application by Aura Pharmaceuticals, Inc. of March 30, 2001. The approval is for use of Altoprev for lowering cholesterol and LDL-C to target levels along with diet and exercise, to slow the progression of atherosclerosis in patients with coronary heart disease, and to reduce total-C, LDL-C, Apo B, and triclycerides and to increase HDL-C in patients with dyslipoproteinemia. The drug was found to outperform Mevacor (lovastatin). Altoprev is protected by patents though at least December 2017.
2. Lescol XL (extended-release Lescol) is a new dosage form and was approved by the FDA for sale in the United States by Novartis on October 6, 2000. Patent protection expired in 2012.

New drug combinations, January 2000–December 2010:
1. Advicor (Mevacor and extended-release Niacin) is a new drug combination approved by the FDA on December 17, 2001 for sale in the United States by Kos Pharmaceuticals. Advicor was approved for use in treating primary hypercholesterolemia and mixed dyslipidemia in two types of patients: a) those treated with lovastatin who require further triglyceride lowering or HDL raising who may benefit from adding niacin to their regimen, and b) patients previously treated with niacin who require further LDL lowering and may benefit from having lovastatin added to their regimen. Thus, Advicor was not approved as an initial therapy for lowering LDL levels. Moreover, in clinical trials, Advicor was found to perform no better than Mevacor as a first-line agent.
2. Pravigard PAC (Pravachol and aspirin) is a new drug combination approved by the FDA on June 24, 2003 for sale in the United States by Bristol Myers Squibb.
3. Vytorin (Zetia and Zocor) is a new drug combination approved by the FDA for use, along with diet or with other lipid-lowering treatments to reduce total C, LDL-C and raise HDL-C, on July 23, 2004 by MSP Singapore company, LLC. The drug combination was more effective at lowering lipids, but was also associated with more adverse events (both serious and leading to discontinuation) than either monotherapy. In January 2008, a completed clinical trial revealed Zetia, a component of Vytorin, performed poorly relative to other therapies.
4. Simcor (simvastatin and extended-release niacin) is a new drug combination approved by the FDA on February 15, 2008 for sale in the United States by Abbott Laboratories. Like Advicor, Simcor is approved only as a second-line treatment for cases in which the monotherapy is considered to be inadequate.

A.2 Data

U.S. Prescriptions for *Hypercholesterolemia* and *Dyslipidemia* Therapies: Prescription data for U.S. medical practitioners and each of the products described above were obtained from the IMS Health Xponent database. IMS Health draws its prescription data from a large but non-random sample of over 70 percent of U.S. pharmacies. As of the time our data were assembled, Xponent included direct information from over 38,000 retail stores, including approximately 119 mail-service pharmacies and 820 long-term care facilities; this compares with a universe of approximately 57,000 retail pharmacies, 327 mail-service outlets, and 3,000 long-term care facilities. In addition to observing directly dispensed prescription volumes (or "sell-out") for each sample pharmacy, IMS Health observes prescription drug purchase volumes (or "sell-in") for the universe of U.S. pharmacies and drugs—that is, including both sample and non-sample stores. To correct for sampling error and to ensure the data are representative, IMS Health has applied a proprietary procedure that a) combines sell-in and sell-out data for sample pharmacies to determine what ratio of purchased product is actually dispensed for each drug and store, b) uses this ratio (or "projection factor"), appropriately weighted by store type and proximity, to estimate the dispensed volume by drug for any store reporting sell-in but not sell-out volumes. Importantly, this projection and weighting procedure applies only to strictly positive prescription levels, but does not apply to zeros, enabling us to accurately track the initial adoption of new products over time for each physician.

The data IMS Health provided include prescriptions by 280,622 unique U.S. physicians for each product in each month during January 2000 through December 2010. To avoid studying physicians specialized outside cardiovascular care, we restrict analysis to physicians that prescribe at least some cholesterol products. Specifically, for a physician to be included in the dataset, he or she needs to have written at least ten filled prescriptions for cholesterol therapies during the calendar year 2010. The data provide precise identifying information for each prescribing physician, including the unique, 11-digit American Medical Association Medical Education Number, the first name, last name, and middle name, and the five-digit zipcode corresponding to the medical practice of the physician. From January 2006 through December 2010, the data provide additional detail regarding prescriptions: for each drug, a separate prescription count is observed for each of four payment methods, including Medicare Part D, Fee-for-Service Medicaid, cash, and commercial insurance. In the data, approximately half of dispensed prescriptions for cholesterol drugs correspond to individuals with commercial insurance; 34 percent obtain products through Medicare Part D, ten percent purchase medications with cash, and the remaining six percent are covered by Medicaid.

To prepare the data for analysis, we reshaped the files provided so that each row corresponds to a doctor-drug-month triplet. With guidance from IMS Health, zeros were explicitly introduced in this step for missing observations corresponding to existing products not associated with positive prescriptions in the IMS data. Starting in 2006, we aggregated prescriptions across methods of payment to arrive at a single number of prescriptions written by physician, drug, and month. We combined prescriptions for "Pravastatin" and "Pravastatin SOD", which are the same product, and did likewise for "Lovaza" and "Omacor", which are the same product. We dropped Baycol from the dataset. For some years, due to the projection calculation described above, the prescription variable was not a whole number; with guidance from IMS Health, we rounded the number of prescriptions to the nearest whole number. To abstract from physician entry during the sample period, we impose a sample restriction in addition to that described above: specifically, each physician included must prescribe at least ten cholesterol drugs during the calendar year 2000. Finally, we used information from the U.S. FDA to determine the approval date for each therapy. The first month after this date was determined to be the first month of a drug's market life in the United States. We created indicator variables for drugs that are new corresponding to the first six months of the drug's market

life in the United States, and separately, to the first 24 months of the drug's market life in the United States. We created indicator variables for generic products lovastatin, pravastatin, and simvastatin.

Electronic Database Use for *Hypercholesterolemia* and *Dyslipidemia* Therapies, by U.S. Physicians: We obtained data on individual physicians' information access from the leading U.S. point-of-care medical applications firm. For each physician, we observe the corresponding initial database registration date; this is used to construct the indicator variable Z_{it} that takes on a value of one for registered users, and that is otherwise zero. For each physician-product-month triplet, we also observe a proxy for the number of lookups completed. We use this proxy to construct a physician-specific indicator for database use that is equal to 1 if the doctor records at least one cholesterol drug lookup during the sample period, and that is otherwise zero. During January 2000 through December 2010, the share of sample physicians registered as database users rose from 0.003 to 0.451. Our analysis is thus based on a sample combining a) physicians that first registered during or before the sample period, and b) physicians that registered before the sample period, and c) physicians that never registered. Each physician is identified in the data by a unique, 11-digit American Medical Association Medical Education Number, first name, last name, middle name, and five-digit zipcode. These characteristics form the basis for a merge with the prescription information described above.

Table 1: Descriptive Statistics, U.S. Cholesterol Drug Introductions, January 2000—December 2008

| Drug Name | Release Date | FDA Approval Category | Months to First Prescription, Conditional on Prescription | | Adoption Share December 2010 |
| | | | Mean | St Dev | |
			(1)	(2)	(3)
Lescol XL	October 2000	Dosage form	28.89	23.56	0.620
Advicor	December 2001	Combination	64.77	15.38	0.295
lovastatin	December 2001	Generic version	19.87	22.66	0.923
Altoprev	June 2002	Dosage form	42.62	24.07	0.151
Zetia	October 2002	Molecular entity	15.13	17.34	0.928
Pravigard PAC	June 2003	Combination	7.30	5.94	0.037
Crestor	August 2003	Molecular entity	22.67	21.83	0.923
Vytorin	July 2004	Combination	13.10	13.48	0.891
Lovaza	November 2004	Molecular entity	34.98	17.08	0.659
pravastatin	April 2006	Generic version	7.41	12.30	0.909
simvastatin	June 2006	Generic version	3.05	7.33	0.982
Simcor	February 2008	Dosage form	12.33	9.18	0.230

Notes: This table summarizes the variation across individual U.S. physicians in the initial prescription of twelve new pharmaceutical products, each aimed at controlling blood cholesterol or lipid levels. Each product was approved for sale in the United States on the date indicated. New drug approvals are categorized by the FDA based on whether the product is a new molecular entity, a new drug combination, a new dosage form, or a new generic equivalent. The distribution of initial prescription dates across the set of U.S. physicians that prescribe the drug at least once by December 2010 is described by the mean (1) and standard deviation (2) in months from initial FDA approval to the first prescription filled at a U.S. pharmacy. The share of physicians that prescribe the product at least once by December 2010 (3) ranges from 3.7 percent (Pravigard PAC) to 98.2 percent (simvastatin). Prescription data are from IMS Health.

Table 2: Regression Summary Statistics

Variable	Mean	St Dev	Min	Max
Physician-Drug-Month Level:				
Number of Prescriptions	4.429	12.721	0	700
Indicator for Positive Prescriptions	0.355	0.479	0	1
Physician-Month Level:				
Drug Database Indicator	0.248	0.432	0	1
Drug Database and Use Indicator	0.133	0.340	0	1
Drug Database Other Adoption Share in Zipcode $_{t-1}$	0.131	0.132	0	1
Proxy for Intensity of Database Use	3.829	11.04	0	1268
Number of Unique Drugs Prescribed	5.304	2.775	1	16
Prescription Herfindahl-Hirschman Index (HHI)	0.438	0.223	0.097	1
Prescription Volume	65.79	66.31	1	2503
Physician-Drug Level:				
Months to First Prescription	19.12	21.95	0	122
First Prescription Within Initial Year Indicator	0.352	0.478	0	1
Drug-Month Level:				
Indicator for New Drug, 24 months	0.155	0.363	0	1

General:				
Number of Physicians	128043			
Number of Drugs, January 2000	6			
Number of Drugs, January 2000 - December 2010	18			

Notes: This table summarizes the data on physician-level prescriptions and database access used in the analysis. Statistics correspond to U.S. physicians that prescribe a minimum of ten statin or lipid-lowering products both during January-December 2000 and January-December 2010 and that work in a zipcode hosting three or more prescribing physicians. The Drug Database indicator varies by physician-month and is equal to one for physicians that are registered users of the drug database; Drug Database and Use indicates that a physician both has database access and is observed using it to search for information about at least one of the 18 cholesterol drugs during the sample period. Drug Database Other Adoption Share in Zipcode varies by physician-month and is the fraction, in the previous month, of *other* physicians practicing in the same zipcode for which Drug Database and Use is equal to one. The intensity of use proxy is a lower-bound on the number of physician-specific database queries corresponding to the cholesterol drugs considered in this analysis. Prescription diversity by physician-month is summarized by the number of unique drugs prescribed and the corresponding Herfindahl-Hirschman index. First Prescription Within Initial Year Indicator takes a value of one for doctors that prescribe the new drug within its initial market year and is otherwise zero. Drugs are considered New if within 24 months of market approval by the U.S. Food and Drug Administration. Prescription variables are from IMS Health and database registration data are from a leading U.S. point-of-care medical applications firm.

Table 3: Descriptive Statistics

	Physician Level					Zipcode Level			
Product:	lovastatin	pravastatin	simvastatin	Generic		lovastatin	pravastatin	simvastatin	Generic
Variable	(1)	(2)	(3)	(4)		(5)	(6)	(7)	(8)
Panel A	Share in Total Rx, by Physician					Share in Total Rx, by Zipcode			
Final month, December 2010									
Mean	0.059	0.091	0.414	0.564		0.065	0.098	0.427	0.591
St Dev	0.116	0.135	0.238	0.243		0.083	0.094	0.159	0.161
5th Percentile	0	0	0	0		0	0	0.180	0.336
25th Percentile	0	0	0.258	0.427		0.017	0.041	0.335	0.502
Median	0.016	0.047	0.410	0.581		0.041	0.078	0.423	0.596
75th Percentile	0.066	0.121	0.551	0.719		0.084	0.129	0.515	0.687
95th Percentile	0.271	0.338	0.858	1		0.209	0.259	0.271	0.834
Panel B	Generic Rx Share, by Physician-Molecule					Generic Rx Share, by Zipcode-Molecule			
Six months after generic release									
Molecule-specfic branded drug	Mevacor	Pravachol	Zocor			Mevacor	Pravachol	Zocor	
Mean	0.828	0.820	0.862			0.822	0.827	0.870	
St Dev	0.338	0.279	0.208			0.192	0.197	0.150	
5th Percentile	0	0	0.448			0.491	0.490	0.582	
25th Percentile	1	0.714	0.800			0.737	0.756	0.819	
Median	1	1	0.949			0.861	0.866	0.914	
75th Percentile	1	1	1			1	1	0.977	
95th Percentile	1	1	1			1	1	1	
Panel C	Generic Rx Share, by Physician-Molecule					Generic Rx Share, by Zipcode-Molecule			
Final month, December 2010									
Molecule-specfic branded drug	Mevacor	Pravachol	Zocor			Mevacor	Pravachol	Zocor	
Mean	1.000	0.993	0.997			1.000	0.995	0.998	
St Dev	0.019	0.059	0.028			0.005	0.030	0.011	
5th Percentile	1	1	0.995			1	0.976	0.990	
25th Percentile and above	1	1	1			1	1	1	

Notes: This table describes prescription heterogeneity across U.S. physicians and the U.S. zipcodes they occupy. Panel A describes prescribing in December 2010 across all physicians (columns 1-4), and across U.S. zipcodes (columns 5-8). Panels B and C describe physicians' within-molecule substitution toward generics for lovastatin (column 1), pravastatin (column 2), and simvastatin (column 3); columns 5, 6, and 7 provide analogous statistics by U.S. zipcode. Panel B describes this substitution six months after the generic release in question, while Panel C describes prescribing in the final sample period, December 2010. The upper-left number in Panel A (mean, lovastatin, 0.059) is the average, across physicians, in the fraction of cholesterol drug prescriptions prescriptions in December 2010 that are accounted for by generic lovastatin; the upper-left number in Panel B (mean, lovastatin, 0.828) is the average, across physicians, in the fraction of Mevacor plus generic lovastatin prescriptions that are accounted for by generic lovastatin in October 2002, six months after expiration of the Mevacor patent; the upper-left number in Panel C is the analogous statistic for December 2010. Generic approval dates are from the U.S. Food and Drug Administration; all other variables are from IMS Health.

Table 4: Time to First Prescription of a New Drug, U.S. Physicians, 2000—2010

Dependent Variable:	Indicator for prescription within first year of drug introduction					
	All physicians				Eventual users	
	(1)	(2)	(3)	(4)	(5)	(6)
Database $_{ij}$	0.0191***	0.0089***	0.0136***	0.0015	-0.0049**	-0.0037
	0.0015	0.0020	0.0016	0.0022	0.0019	0.0025
Database $_{ij}$ x Generic $_j$			0.0192***	0.0213***	0.0157***	0.0140***
			0.0023	0.0023	0.0035	0.0036
Prescription Volume $_{it(j)-1}$	0.0075***	0.0034***	0.0075***	0.0034***	0.0069***	0.0029***
	0.0001	0.0000	0.0001	0.0000	0.0002	0.0001
Physician FE	N	Y	N	Y	N	Y
Zipcode-Drug FE	Y	Y	Y	Y	Y	Y
Observations	1513408	1513408	1513408	1513408	313699	313699
R^2	0.4556	0.5690	0.4556	0.5690	0.4876	0.5937

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides least-squares estimates of equation (5) for U.S. physicians' prescription of twelve cholesterol drugs first approved for U.S. sale during January 2000—December 2008 (Table 1). The binary dependent variable captures the time lapse between FDA approval of drug j and physician i's initial prescription of it, taking a value of 1 if initial prescription occurs within a year of FDA approval; specifications are included for the full sample of physicians (columns 1-4) and for the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (columns 5-6). Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access at the time drug j receives FDA approval. Generic indicates the products pravastatin, lovastatin, and simvastatin. Regressions include zipcode-drug (columns 1-6) and physician (columns 2, 4, 6) fixed effects as well as the cholesterol drug prescription volume for physician i in the month prior to drug j's introduction. Results are robust to Poisson and Negative Binomial estimation, and are qualitatively identical when replacing the dependent variable with an indicator for first prescription within six months, an indicator for first prescription within two years, or a continuous measure for the time to first prescription (corrected for censoring). Standard errors clustered by zipcode appear below each point estimate; results are robust to clustering errors by physician.

Table 5: Prescription Diversity, U.S. Physicians, 2000—2010

Dependent Variable:	Number of Unique Drugs	Prescription HHI	Number of Unique Drugs	Prescription HHI
	All physicians		Eventual users	
	(1)	(2)	(3)	(4)
Database $_{it}$	0.1231***	-0.0059***	0.0298***	-0.0012***
	0.0029	0.0003	0.0036	0.0004
Prescription Volume $_{it-1}$	0.0233***	-0.0008***	0.0228***	-0.0007***
	0.0000	0.0000	0.0001	0.0000
Physician FE	Y	Y	Y	Y
Zipcode-Month FE	Y	Y	Y	Y
Observations	15510386	15510386	3013086	3013086
R^2	0.8327	0.6317	0.8572	0.6775

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides least-squares estimates of equation (6) for cholesterol drug prescriptions by U.S. physicians during January 2000 through December 2010, including all sample physicians (columns 1-2) and the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (columns 3-4). The dependent variable in columns 1 and 3 captures the prescription diversity of physician i as the number of unique drugs j that are prescribed by i during month t. The dependent variable in columns 2 and 4 is the prescription Herfindahl-Hirschman index for physician i in month t. Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access in month t; it is otherwise zero. All regressions include physician-specific time trends, physician and zipcode-month fixed effects, and the cholesterol drug prescription volume for physician i in month t-1. Results in columns 1 and 3 are robust to Poisson and Negative Binomial estimation; all columns are robust to including doctor-specific time trends and the first lag of the dependent variable. Standard errors clustered by zipcode-month appear below each point estimate.

Table 6: Prescription Propensity, U.S. Physicians, 2000—2010

	Dependent Variable:	$1\{(\text{prescriptions of drug } j \text{ by } i \text{ at } t) > 0\}$	
		All physicians	Eventual users
		(1)	(2)
Database $_{it}$ x New $_{jt}$ x Generic $_j$		0.0240***	0.0162***
		0.0012	0.0021
Database $_{it}$ x New $_{jt}$ x Branded $_j$		-0.0103***	-0.0042***
		0.0008	0.0010
Database $_{it}$ x Old $_{jt}$ x Generic $_j$		0.0308***	0.0240***
		0.0011	0.0022
Database $_{it}$ x Old $_{jt}$ x Branded $_j$		-0.0015***	0.0015***
		0.0006	0.0007
Prescription Volume $_{it-1}$		0.0013***	0.0013***
		0.0000	0.0000
Physician FE		Y	Y
Zipcode x Month FE		Y	Y
Drug x Month FE		Y	Y
Observations		36238793	7674288
R^2		0.5193	0.5393

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides least-squares estimates of equation (7) for prescriptions of cholesterol drugs by U.S. physicians during January 2000 through December 2010, including all sample physicians (column 1) and the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (column 2). The dependent variable is an indicator for whether the doctor i prescribes drug j during month t. Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access in month t; it is otherwise zero. New is an indicator that is equal to 1 if drug j is within 24 months of its initial approval by the U.S. FDA in month t; drug j is otherwise Old. Generic indicates the products pravastatin, lovastatin, and simvastatin; other drugs are Branded. All regressions include physician, zipcode-month, and drug-month fixed effects, as well as the cholesterol drug prescription volume for physician i in month t-1. Results are robust to logit estimation, and to including doctor-specific time trends, the first lag of the dependent variable, doctor-drug type fixed effects, and zipcode-month-drug type fixed effects. For computational ease, the estimates above rely only on observations in January 2000 and every subsequent June and December. Standard errors clusterd by zipcode-month appear below each point estimate.

Table 7: Time to First Prescription of a New Drug, Two-Stage Least Squares, U.S. Physicians, 2000—2010

	All physicians				Eventual users	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A Dependent Variable:	Indicator for prescription within first year of drug introduction					
Database $_{ij}$	-0.0022	0.0260**	-0.0247**	0.0023	-0.0133	-0.0032
	0.0102	0.0111	0.0121	0.0120	0.0109	0.0102
Database $_{ij}$ x Generic $_j$			0.0771***	0.0693***	0.0755***	0.0642***
			0.0146	0.0143	0.0163	0.0168
Panel B	First stage for Database $_{ij}$					
Adoption Share in Zipcode $_{it(j)\ -1}$	0.8102***	0.7099***	0.8260***	0.7082***	1.1455***	1.1773***
	0.0025	0.0024	0.0030	0.0026	0.0084	0.0082
Adoption Share in Zipcode $_{it(j)-1}$ x Generic $_j$			-0.0517***	0.0045*	-0.3212***	-0.1399***
			0.0054	0.0026	0.0150	0.0088
Panel C	First stage for Database $_{ij}$ x Generic $_j$					
Adoption Share in Zipcode $_{it(j)\ -1}$			-0.0001	-0.0503***	0.0000	-0.0871***
			0.0016	0.0027	0.0042	0.0068
Adoption Share in Zipcode $_{it(j)-1}$ x Generic $_j$			0.7744***	0.7899***	0.8253***	0.8618***
			0.0029	0.0027	0.0076	0.0072
Drug FE	Y	Y	Y	Y	Y	Y
Physician FE	N	Y	N	Y	N	Y
Prescription Volume $_{it(i)-1}$	Y	Y	Y	Y	Y	Y
Observations	1513408	1513408	1513408	1513408	313699	313699
R^2	0.4339	0.5534	0.4337	0.5533	0.4521	0.5669
First-Stage F Statistic	46646	89203	51292	44601	8526	7598

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides two-stage least squares of equation (5) for U.S. physicians' prescription of cholesterol drugs first approved for U.S. sale during January 2000—December 2008 (Table 1). The binary dependent variable captures the time lapse between FDA approval of drug j and physician i's initial prescription of it, taking a value of 1 if initial prescription occurs within a year of FDA approval; specification are included for the full sample of physicians (columns 1-4) and for the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (columns 5-6). Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access at the time drug j receives FDA approval. The instrument for Database at drug j's introduction is the share of physicians _other than i_ that are located in i's zipcode and have adopted the database by or before the month immediately preceding drug j's introduction. The sample is thus restricted to include only zipcodes with at least three physician prescribers (two physicians other than i). Generic indicates the products pravastatin, lovastatin, and simvastatin. Regressions include drug fixed effects (columns 1-6) and physician fixed effects (columns 2, 4, 6), as well as the cholesterol drug prescription volume for physician i in the month prior to drug j's introduction. Second-stage estimates are in Panel A; first-stage estimates appear in Panels B and C. Standard errors clustered by zipcode appear below each point estimate; results are qualitatively identical when clustering by physician.

Table 8: Prescription Diversity, Two-Stage Least Squares, U.S. Physicians, 2000—2010

	All physicians		Eventual users	
	(1)	(2)	(3)	(4)
Panel A	Number of Unique Drugs	Prescription HHI	Number of Unique Drugs	Prescription HHI
Database $_{it}$	0.4124***	-0.0309***	-0.0869	-0.0428***
	0.0477	0.0056	0.0607	0.0072
Panel B	First stage for Database $_{it}$			
Adoption Share in Zipcode $_{it-1}$	0.1294***	0.1294***	0.1945***	0.1945***
	0.0007	0.0007	0.0024	0.0024
Physician FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Prescription Volume $_{it-1}$	Y	Y	Y	Y
Observations	15510386	15510386	3013086	3013086
R^2	0.8785	0.5812	0.8112	0.5777
First-Stage F statistic	36569	36569	6407	6407

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides two-stage least squares estimates of equation (6) for cholesterol drug prescriptions by U.S. physicians during January 2000 through December 2010, including all sample physicians (columns 1-2) and the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (columns 3-4). The dependent variable in columns 1 and 3 captures the prescription diversity of physician i as the number of unique drugs j that are prescribed by i during month t. The dependent variable in columns 2 and 4 is the prescription Herfindahl-Hirschman index for physician i in month t. Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access in month t; it is otherwise zero. The instrument for Database at t is the share of physicians *other than i* that are located in i's zipcode and have adopted the database by or before month t-1. The sample is thus restricted to include only zipcodes with at least three physicians (two physicians other than i) that prescribe cholesterol drugs. All regressions include physician and month fixed effects, as well as the cholesterol drug prescription volume for physician i in month t-1. Second-stage estimates are in Panel A; first-stage estimates appear in Panel B. Standard errors clustered by zipcode-month are reported below each point estimate.

Table 9: Prescription Propensity, Two-Stage Least Squares, U.S. Physicians, 2000—2010

	Dependent Variable:	$1\{(\text{prescriptions of drug } j \text{ by } i \text{ at } t) > 0\}$	
		All physicians	Eventual users
		(1)	(2)
Database $_{it}$ x New $_{jt}$ x Generic $_j$		0.0258***	0.0506***
		0.0032	0.0070
Database $_{it}$ x New $_{jt}$ x Branded $_j$		-0.0425***	-0.0303***
		0.0020	0.0033
Database $_{it}$ x Old $_{jt}$ x Generic $_j$		0.0547***	0.2279***
		0.0021	0.0132
Database $_{it}$ x Old $_{jt}$ x Branded $_j$		-0.0189***	-0.0210***
		0.0013	0.0026
Physician FE		Y	Y
Drug x Month FE		Y	Y
Prescription Volume $_{it-1}$		Y	Y
Observations		36238793	7674288
R^2		0.5149	0.5262
First-Stage F statistic		7.9e+05	95593

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides two-stage least squares estimates of equation (7) for prescriptions of cholesterol drugs by U.S. physicians during January 2000 through December 2010, including all sample physicians (column 1) and the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (column 2). The dependent variable is an indicator for whether the doctor i prescribes drug j during month t. Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access in month t; it is otherwise zero. The instrument for Database at t is the share of physicians *other than i* that are located in i's zipcode and have adopted the database by or before month t-1. The sample is thus restricted to include only zipcodes with at least three physicians (two physicians other than i) that prescribe cholesterol drugs. New is an indicator that is equal to 1 if drug j is within 24 months of its initial approval by the U.S. FDA in month t; drug j is otherwise Old. Generic indicates the products pravastatin, lovastatin, and simvastatin; other drugs are Branded. All regressions include physician and drug-month fixed effects, as well as the cholesterol drug prescription volume for physician i in month t-1. First stage estimates appear in Appendix Table A.2. Standard errors clustered by zipcode-month are reported below each point estimate.

Table 10: Prescription Diversity, Physician Trends, U.S. Physicians, 2000—2010

Dependent Variable:	Number of Unique Drugs	Prescription HHI	Number of Unique Drugs	Prescription HHI
	All physicians		Eventual users	
	(1)	(2)	(3)	(4)
Database $_{it}$	0.0426***	-0.0037***	0.0350***	-0.0027***
	0.0029	0.0003	0.0034	0.0004
Physician FE	Y	Y	Y	Y
Zipcode-Month FE	Y	Y	Y	Y
Prescription Volume $_{it-1}$	Y	Y	Y	Y
Physician x t trends	Y	Y	Y	Y
Observations	15510386	15510386	3013241	3013241
R^2	0.8785	0.7114	0.8941	0.7484

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides least-squares estimates of a variant of equation (6) that includes a full set of physician-specific time trends. The sample includes all physicians (columns 1-2) and the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (columns 3-4). The dependent variable in columns 1 and 3 captures the prescription diversity of physician i as the number of unique drugs j that are prescribed by i during month t. The dependent variable in columns 2 and 4 is the prescription Herfindahl-Hirschman index for physician i in month t. Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access in month t; it is otherwise zero. All regressions include physician and zipcode-month fixed effects, as well as the cholesterol drug prescription volume for physician i in month t-1. Standard errors clustered by zipcode-month appear below each point estimate.

Table 11: Prescription Propensity, Physician Trends, U.S. Physicians, 2000—2010

|	Dependent Variable:	1{(prescriptions of drug j by i at t) > 0}	
		All physicians	Eventual users
		(1)	(2)
Database $_{it}$ x New $_{jt}$ x Generic $_j$		0.0266**	0.0158***
		0.0013	0.0020
Database $_{it}$ x New $_{jt}$ x Branded $_j$		-0.0089***	-0.0051***
		0.0008	0.0012
Database $_{it}$ x Old $_{jt}$ x Generic $_j$		0.0351***	0.0240***
		0.0012	0.0022
Database $_{it}$ x Old $_{jt}$ x Branded $_j$		0.0019***	0.0013*
		0.0006	0.0007
Physician FE		Y	Y
Zipcode x Month FE		Y	Y
Drug x Month FE		Y	Y
Prescription Volume $_{it-1}$		Y	Y
Physician x t trends		Y	Y
Observations		36238793	7674288
R^2		0.5277	0.7179

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides least-squares estimates of a variant of equation (7) that includes a full set of physician-specific time trends. The sample includes all physicians (column 1) and the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (column 2). The dependent variable is an indicator for whether the doctor i prescribes drug j during month t. Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access in month t; it is otherwise zero. New is an indicator that is equal to 1 if drug j is within 24 months of its initial approval by the U.S. FDA in month t; drug j is otherwise Old. Generic indicates the products pravastatin, lovastatin, and simvastatin; other drugs are Branded. All regressions include physician, zipcode-month, and drug-month fixed effects, as well as the cholesterol drug prescription volume for physician i in month t-1. For computational ease, the estimates above rely only on observations in January 2000 and every subsequent June and December. Standard errors clustered by zipcode-month appear below each point estimate.

Figure 1: Drug Database Diffusion Curves, U.S. Physicians, January 2000—December 2010

Panel A — By school rank
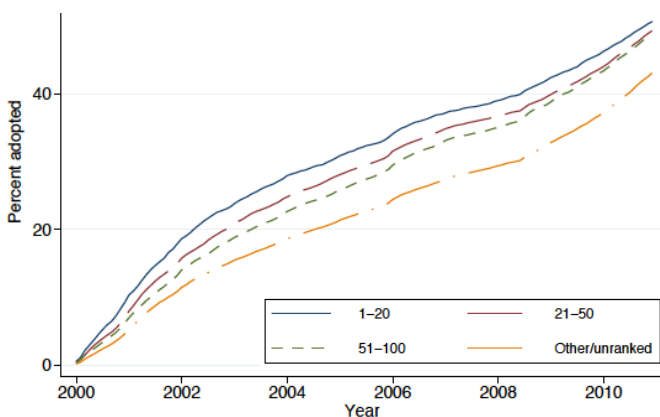
Panel B — By graduation year
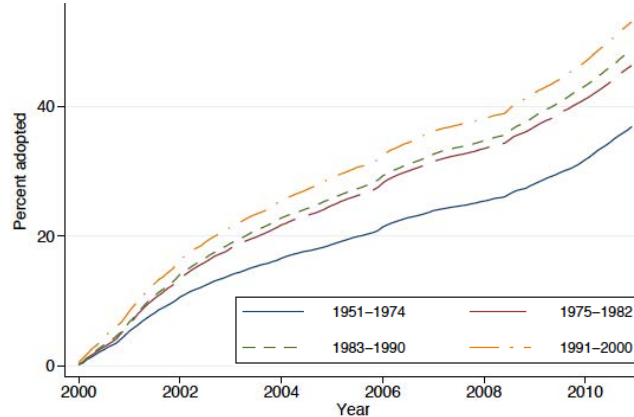
Panel C — By sex
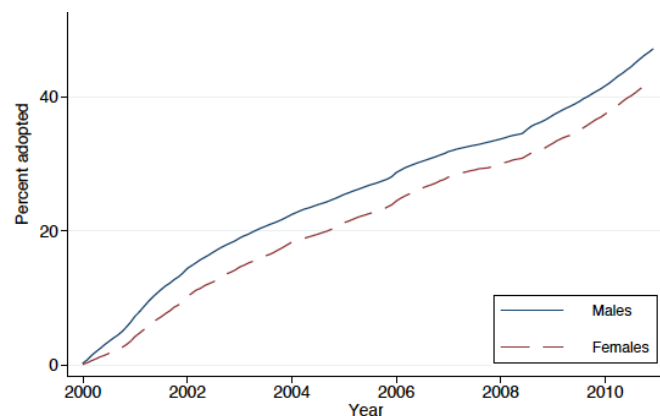
Panel D — By monthly prescription volume
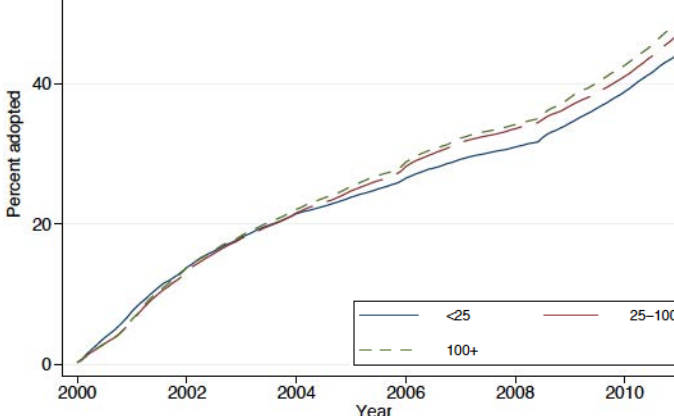
Panel E — By medical specialty
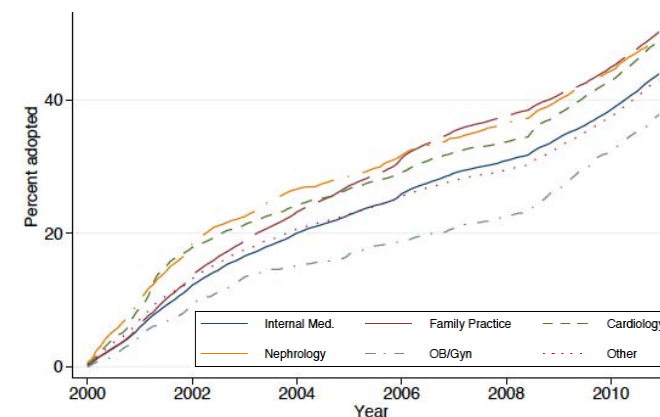
Panel F — By U.S. region
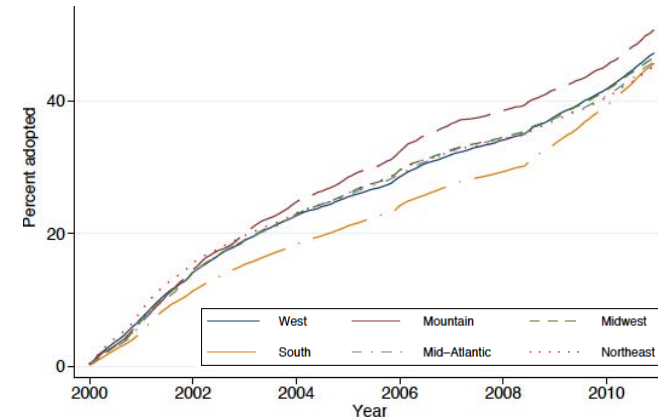
Notes: This figure plots the fraction of the approximately 67,000 sample U.S. physicians included in the CMS Physician Compare database that are also registered users of the electronic drug reference database by the date indicated, and shows the extent to which adoption rates differ across physicians according their observable characteristics. Database registration data are from the drug reference database firm. Medical school rank is determined based on data from the U.S. News and World Report service, and all other variables are from the CMS Physician Compare database.

Figure 2: Drug Diffusion Curves, by Drug, U.S. Physicians, January 2000—December 2010
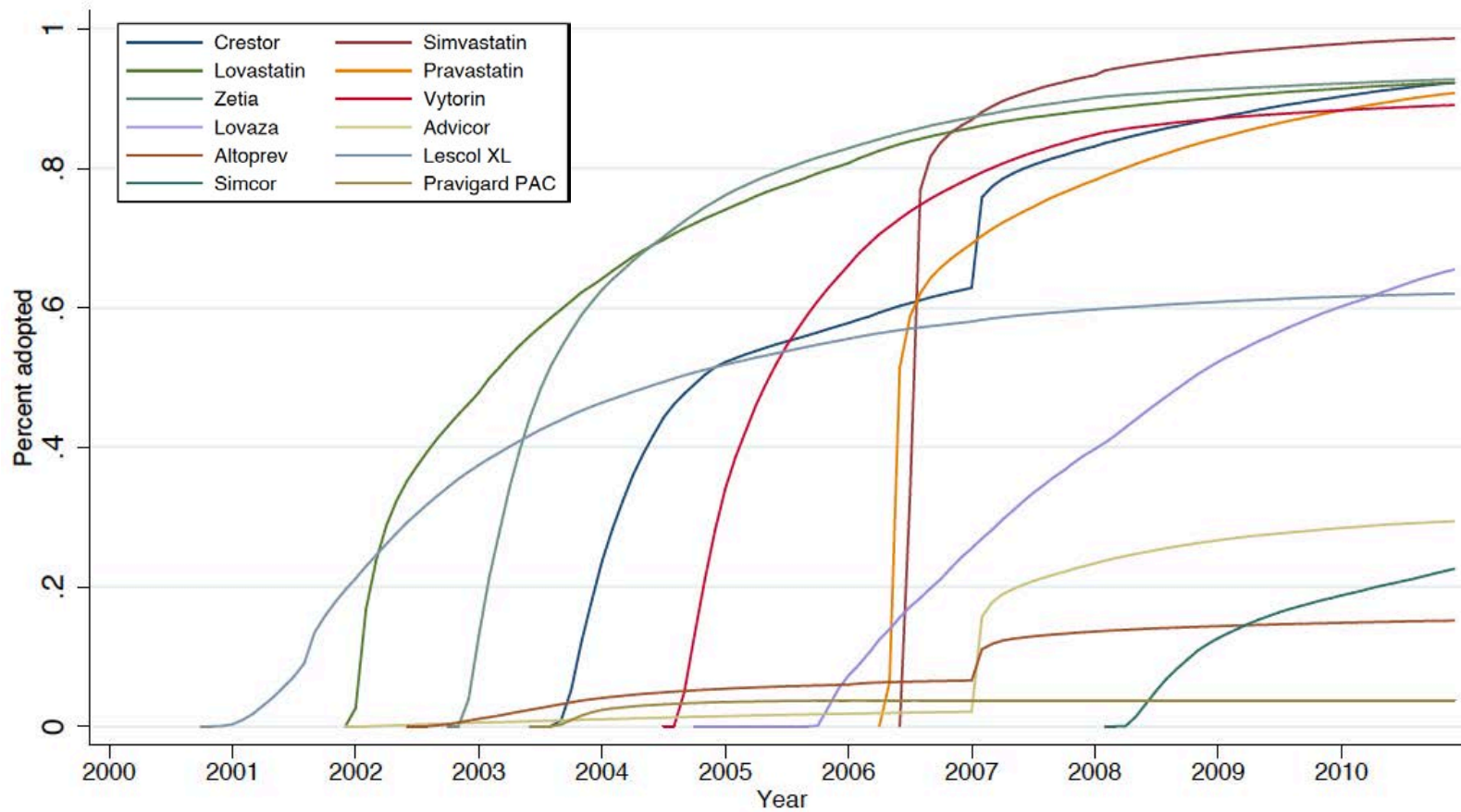
Notes: This figure plots the fraction of the 128,043 sample U.S. physicians that are associated with at least one prescription of the new drug indicated by the date marked on the horizontal axis, and shows the extent to which adoption rates differ across products. Market approval dates by drug are listed in Table 1. The prescription data cover January 2000 through December 2010 at a monthly frequency and are from IMS Health.

Figure 3: Heterogeneity in the Initial Use of a New Medical Technology, by U.S. Zipcode

Panel A — One Month After Release
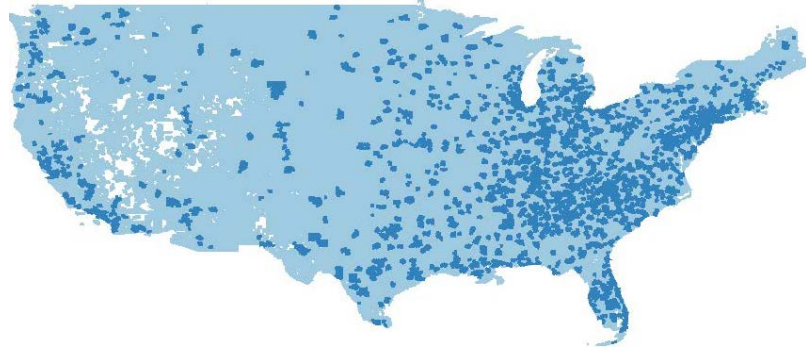
Panel B — Three Months After Release
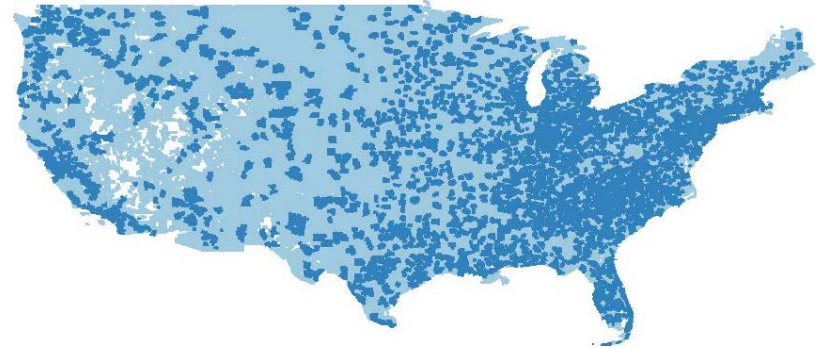
Panel C — Six Months After Release
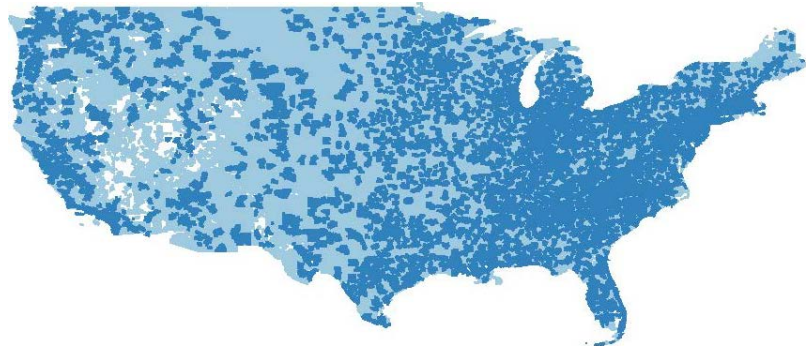
Panel D — Thirty-Six Months After Release
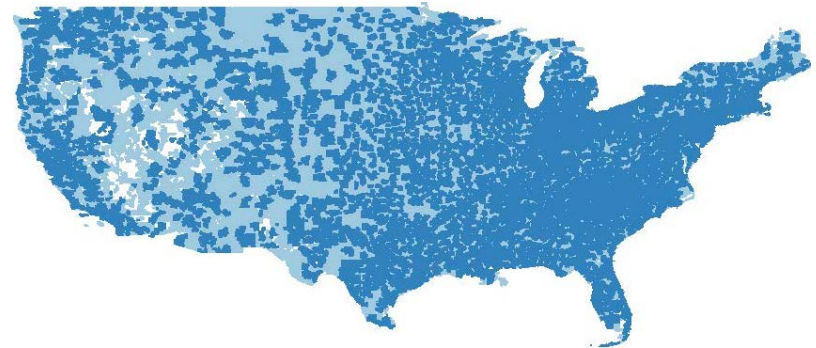
Notes: This figure illustrates the gradual diffusion of a new pharmaceutical drug, the statin Crestor, across zipcodes within the continental United States. Dark shades indicate zipcodes in which at least one prescription of Crestor has been written and filled, light shades indicate zipcodes in which Crestor has not yet been prescribed; areas shaded white contain no data. The four panels correspond to four points in time following the initial market introduction of Crestor in August 2003. These four points are September 2003 (Panel A), November 2003 (Panel B), February 2004 (Panel C), and August 2006 (Panel D). Prescription data are from IMS Health.

Appendix

A.1 Medical Innovation

Innovation in *hypercholesterolemia* and *dyslipidemia* therapy: Information about the evolving set of pharmaceutical therapies available for prescription was obtained from the U.S. Food and Drug Administration (FDA) for the period January 2000 through December 2010. Twelve new statin or lipid-lowering drugs, including new formulations, combinations, and versions, introduced during this period and are described below. These include three new molecular entities Crestor, Lovaza, and Zetia; three generic versions lovastatin (Mevacor), pravastatin (Pravachol), and simvastatin (Zocor); two new formulations Altoprev (extended-release Mevacor) and Lescol XL (extended-release Lescol); and four new drug combinations Advicor (extended-release niacin and Mevacor), Pravigard PAC (aspirin and Pravachol), Vytorin (Zetia and Zocor), Simcor (extended-release niacin and Zocor). A description of each drug innovation appears below based on publicly available data including approval letters and administrative, medical, and pharmacological review. Baycol was withdrawn early in the sample period in August 2001 and is thus omitted.

Existing therapies available in January 2000:
1. Lescol (fluvastatin) is a statin marketed by Novartis since its FDA approval as a new molecular entity on December 31, 1993; its patent protection expired in 2012. Like other statins, its mechanism of action is to limit a specific enzyme in the liver, preventing cholesterol synthesis.
2. Lipitor (atorvastatin) is a statin marketed by Pfizer. Its mechanism of action is similar to that of fluvastatin, but unlike other statins, atorvastatin is a synthetic compound. The therapy was approved by the FDA as a new molecular entity on December 17, 1996. Between 1996 and 2012, Lipitor was the best-selling drug globally; its patent expired in November 2011.
3. Mevacor (lovastatin) is the first statin to receive FDA approval. The drug was approved as a new molecular entity on August 31, 1987 for sale in the United States by Merck. The therapy was protected by patents through June 2001.
4. Niaspan (extended-release niacin) is vitamin B_3, or nicotinic acid, and is marketed by Abbott Laboratories. Extended-release niacin was approved for sale in the United States on July 28, 1997.
5. Pravachol (pravastatin) is a statin marketed by Bristol Myers Squibb since its FDA approval on October 31, 1991. In addition to inhibiting cholesterol synthesis, Pravachol also inhibits low-density lipoprotein synthesis. Two clinical trials, each completed in November 2003, suggest Pravachol is outperformed by both Zocor and Lipitor. Patent protection expired in June 2006.
6. Zocor (simvastatin) is a statin marketed by Merck since its FDA approval as a new molecular entity on December 23, 1991. Zocor outperformed Pravachol in its prevention of cholesterol synthesis in a clinical trial completed in November 2003. Patent protection expired in April 2006.

New chemical entities, January 2000–December 2010:
1. Crestor (rosuvastatin calcium) is a new molecular entity approved by the FDA for sale in the United States by Astra Zeneca Pharmaceuticals on August 12, 2003. The molecule acts by reducing intestinal absorption of cholesterol and related phytosterols, and is thereby distinct relative to other statin therapies. The drug was approved for use in treating primary hypercholesterolemia and mixed dyslipidemia (by reducing total-C, LDL-C, and Apo B), and as an adjunct to other lipid-lowering treatments. It was thus approved for use alone or with other statins. A 2008 clinical trial revealed additional evidence supporting the superior performance of Crestor compared with a placebo treatment. Patent protection expires in January 2016.
2. Lovaza (omega-3-acid ethyl esters) is a new molecular entity introduced by Abbott labs and approved by the FDA on November 10, 2004. It was initially introduced under the trade name Omacor. Unlike statins, Lovaza is aimed at reducing tricylerides rather than low-density lipopro-

teins and may thus be combined with a statin as an adjunct therapy. Patent protection expired in September 2012.

3. Zetia (ezetimibe) is a new molecular entity introduced by Schering and approved by the FDA on October 25, 2002 for sale in the United States. The molecule acts by reducing intestinal absorption of cholesterol and related phytosterols, and is thus distinct from statins. The drug was initially approved for use in treating hypercholesterolemia for use alone or with other statins. In January 2008, a clinical trial found Zetia performed poorly compared with other therapies, and it was at that time recommended that Zetia not be prescribed except in cases for which all other cholesterol drugs had previously failed. Patent protection expires in April 2017.

New generic versions, January 2000–December 2010:
1. Lovastatin is the generic equivalent of Mevacor, and was initially approved by the FDA for sale in the United States by Geneva Pharmaceuticals applied on December 17, 2001.
2. Pravastatin is the generic equivalent of Pravachol, and was initially approved by the FDA for sale in the United States by Teva Pharmaceuticals on April 24, 2006.
3. Simvastatin is the generic equivalent of Zocor, and was initially approved by the FDA for sale in the United States by Teva Pharmaceuticals on June 23, 2006.

New dosage forms, January 2000–December 2010:
1. Altoprev (extended-release lovastatin) is a new dosage form and was approved by the FDA on June 26, 2002 for sale in the United States, following a new drug application by Aura Pharmaceuticals, Inc. of March 30, 2001. The approval is for use of Altoprev for lowering cholesterol and LDL-C to target levels along with diet and exercise, to slow the progression of atherosclerosis in patients with coronary heart disease, and to reduce total-C, LDL-C, Apo B, and triclycerides and to increase HDL-C in patients with dyslipoproteinemia. The drug was found to outperform Mevacor (lovastatin). Altoprev is protected by patents though at least December 2017.
2. Lescol XL (extended-release Lescol) is a new dosage form and was approved by the FDA for sale in the United States by Novartis on October 6, 2000. Patent protection expired in 2012.

New drug combinations, January 2000–December 2010:
1. Advicor (Mevacor and extended-release Niacin) is a new drug combination approved by the FDA on December 17, 2001 for sale in the United States by Kos Pharmaceuticals. Advicor was approved for use in treating primary hypercholesterolemia and mixed dyslipidemia in two types of patients: a) those treated with lovastatin who require further triglyceride lowering or HDL raising who may benefit from adding niacin to their regimen, and b) patients previously treated with niacin who require further LDL lowering and may benefit from having lovastatin added to their regimen. Thus, Advicor was not approved as an initial therapy for lowering LDL levels. Moreover, in clinical trials, Advicor was found to perform no better than Mevacor as a first-line agent.
2. Pravigard PAC (Pravachol and aspirin) is a new drug combination approved by the FDA on June 24, 2003 for sale in the United States by Bristol Myers Squibb.
3. Vytorin (Zetia and Zocor) is a new drug combination approved by the FDA for use, along with diet or with other lipid-lowering treatments to reduce total C, LDL-C and raise HDL-C, on July 23, 2004 by MSP Singapore company, LLC. The drug combination was more effective at lowering lipids, but was also associated with more adverse events (both serious and leading to discontinuation) than either monotherapy. In January 2008, a completed clinical trial revealed Zetia, a component of Vytorin, performed poorly relative to other therapies.
4. Simcor (simvastatin and extended-release niacin) is a new drug combination approved by the FDA on February 15, 2008 for sale in the United States by Abbott Laboratories. Like Advicor, Simcor is approved only as a second-line treatment for cases in which the monotherapy is considered to be inadequate.

A.2 Data

U.S. Prescriptions for *Hypercholesterolemia* and *Dyslipidemia* Therapies: Prescription data for U.S. medical practitioners and each of the products described above were obtained from the IMS Health Xponent database. IMS Health draws its prescription data from a large but non-random sample of over 70 percent of U.S. pharmacies. As of the time our data were assembled, Xponent included direct information from over 38,000 retail stores, including approximately 119 mail-service pharmacies and 820 long-term care facilities; this compares with a universe of approximately 57,000 retail pharmacies, 327 mail-service outlets, and 3,000 long-term care facilities. In addition to observing directly dispensed prescription volumes (or "sell-out") for each sample pharmacy, IMS Health observes prescription drug purchase volumes (or "sell-in") for the universe of U.S. pharmacies and drugs—that is, including both sample and non-sample stores. To correct for sampling error and to ensure the data are representative, IMS Health has applied a proprietary procedure that a) combines sell-in and sell-out data for sample pharmacies to determine what ratio of purchased product is actually dispensed for each drug and store, b) uses this ratio (or "projection factor"), appropriately weighted by store type and proximity, to estimate the dispensed volume by drug for any store reporting sell-in but not sell-out volumes. Importantly, this projection and weighting procedure applies only to strictly positive prescription levels, but does not apply to zeros, enabling us to accurately track the initial adoption of new products over time for each physician.

The data IMS Health provided include prescriptions by 280,622 unique U.S. physicians for each product in each month during January 2000 through December 2010. To avoid studying physicians specialized outside cardiovascular care, we restrict analysis to physicians that prescribe at least some cholesterol products. Specifically, for a physician to be included in the dataset, he or she needs to have written at least ten filled prescriptions for cholesterol therapies during the calendar year 2010. The data provide precise identifying information for each prescribing physician, including the unique, 11-digit American Medical Association Medical Education Number, the first name, last name, and middle name, and the five-digit zipcode corresponding to the medical practice of the physician. From January 2006 through December 2010, the data provide additional detail regarding prescriptions: for each drug, a separate prescription count is observed for each of four payment methods, including Medicare Part D, Fee-for-Service Medicaid, cash, and commercial insurance. In the data, approximately half of dispensed prescriptions for cholesterol drugs correspond to individuals with commercial insurance; 34 percent obtain products through Medicare Part D, ten percent purchase medications with cash, and the remaining six percent are covered by Medicaid.

To prepare the data for analysis, we reshaped the files provided so that each row corresponds to a doctor-drug-month triplet. With guidance from IMS Health, zeros were explicitly introduced in this step for missing observations corresponding to existing products not associated with positive prescriptions in the IMS data. Starting in 2006, we aggregated prescriptions across methods of payment to arrive at a single number of prescriptions written by physician, drug, and month. We combined prescriptions for "Pravastatin" and "Pravastatin SOD", which are the same product, and did likewise for "Lovaza" and "Omacor", which are the same product. We dropped Baycol from the dataset. For some years, due to the projection calculation described above, the prescription variable was not a whole number; with guidance from IMS Health, we rounded the number of prescriptions to the nearest whole number. To abstract from physician entry during the sample period, we impose a sample restriction in addition to that described above: specifically, each physician included must prescribe at least ten cholesterol drugs during the calendar year 2000. Finally, we used information from the U.S. FDA to determine the approval date for each therapy. The first month after this date was determined to be the first month of a drug's market life in the United States. We created indicator variables for drugs that are new corresponding to the first six months of the drug's market

life in the United States, and separately, to the first 24 months of the drug's market life in the United States. We created indicator variables for generic products lovastatin, pravastatin, and simvastatin.

Electronic Database Use for *Hypercholesterolemia* and *Dyslipidemia* Therapies, by U.S. Physicians: We obtained data on individual physicians' information access from the leading U.S. point-of-care medical applications firm. For each physician, we observe the corresponding initial database registration date; this is used to construct the indicator variable Z_{it} that takes on a value of one for registered users, and that is otherwise zero. For each physician-product-month triplet, we also observe a proxy for the number of lookups completed. We use this proxy to construct a physician-specific indicator for database use that is equal to 1 if the doctor records at least one cholesterol drug lookup during the sample period, and that is otherwise zero. During January 2000 through December 2010, the share of sample physicians registered as database users rose from 0.003 to 0.451. Our analysis is thus based on a sample combining a) physicians that first registered during or before the sample period, and b) physicians that registered before the sample period, and c) physicians that never registered. Each physician is identified in the data by a unique, 11-digit American Medical Association Medical Education Number, first name, last name, middle name, and five-digit zipcode. These characteristics form the basis for a merge with the prescription information described above.

Table A.1: Descriptive Statistics, Within-Zipcode Prescribing Variation Across Physicians

	Product:	lovastatin	pravastatin	simvastatin	Generic
Variable		(1)	(2)	(3)	(4)

Panel A		St Dev of Share in Total Rx, by Zipcode			
Final month, December 2010					
Mean		0.079	0.103	0.196	0.191
St Dev		0.077	0.082	0.094	0.096
5th Percentile		0.003	0.011	0.046	0.043
25th Percentile		0.027	0.049	0.133	0.122
Median		0.057	0.085	0.196	0.191
75th Percentile		0.107	0.134	0.248	0.249
95th Percentile		0.225	0.249	0.349	0.349
Panel B		St Dev of Generic Rx Share, by Zipcode			
Six months after generic release					
Molecule-specfic branded drug		Mevacor	Pravachol	Zocor	
Mean		0.276	0.225	0.142	
St Dev		0.186	0.157	0.115	
5th Percentile		0	0	0	
25th Percentile		0.112	0.115	0.059	
Median		0.315	0.219	0.118	
75th Percentile		0.417	0.326	0.202	
95th Percentile		0.535	0.500	0.355	
Panel C		St Dev of Generic Rx Share, by Zipcode			
Final month, December 2010					
Molecule-specfic branded drug		Mevacor	Pravachol	Zocor	
Mean		0.001	0.014	0.006	
St Dev		0.014	0.051	0.022	
5th Percentile		0	0	0	
25th Percentile		0	0	0	
Median		0	0	0	
75th Percentile		0	0	0.004	
95th Percentile		0	0.260	0.029	

Notes: This table describes the distribution across U.S. zipcodes of within-zipcode prescribing variation across local physicians. As in Table 3, Panel A describes within-zipcode prescribing variation in December 2010; Panels B and C describe the local variation in physicians' within-molecule substitution toward generics for lovastatin (column 1), pravastatin (column 2), and simvastatin (column 3). Panel B describes this variation in substitution six months after the generic release in question, while Panel C describes variation in prescribing in the final sample period, December 2010. The upper-left number in Panel A (mean, lovastatin, 0.079) is the average, across zipcodes, of the standard deviation across local physicians in the fraction of their total December 2010 prescriptions that are accounted for by generic lovastatin; the upper-left number in Panel B (mean, lovastatin, 0.276) is the average, across zipcodes, of the standard deviation across local physicians in the fraction of their total Mevacor plus generic lovastatin prescriptions that are accounted for by generic lovastatin in October 2002, six months after expiration of the Mevacor patent; the upper-left number in Panel C is the analogous statistic for December 2010. Generic approval dates are from the U.S. Food and Drug Administration; all other variables are from IMS Health.

Table A.2: Prescription Propensity, Two-Stage Least Squares, First-Stage Estimates, U.S. Physicians, 2000—2010

	All physicians	Eventual users	All physicians	Eventual users
	(1)	(2)	(3)	(4)
Panel A	First stage for Database $_{it}$ x New $_{jt}$ x Generic $_j$		First stage for Database $_{it}$ x Old $_{jt}$ x Generic $_j$	
Adoption Share in Zipcode $_{it-1}$ x New $_{jt}$ x Generic $_j$	0.9578***	0.9800***	-0.0044***	0.0297***
	0.0005	0.0012	0.0009	0.0015
Adoption Share in Zipcode $_{it-1}$ x New $_{jt}$ x Branded $_j$	-0.0032***	0.0309***	-0.0035***	0.0278***
	0.0004	0.0008	0.0006	0.0010
Adoption Share in Zipcode $_{it-1}$ x Old $_{jt}$ x Generic $_j$	-0.0034***	0.0385***	0.9613***	0.5601***
	0.0003	0.0008	0.0006	0.0010
Adoption Share in Zipcode $_{it-1}$ x Old $_{jt}$ x Branded $_j$	-0.0033***	0.0358***	-0.0039***	0.0364***
	0.0002	0.0005	0.0004	0.0006
Panel B	First stage for Database $_{it}$ x New $_{jt}$ x Branded $_j$		First stage for Database $_{it}$ x Old $_{jt}$ x Branded $_j$	
Adoption Share in Zipcode $_{it-1}$ x New $_{jt}$ x Generic $_j$	-0.0112***	0.1483***	-0.0577***	0.4793***
	0.0008	0.0021	0.0015	0.0037
Adoption Share in Zipcode $_{it-1}$ x New $_{jt}$ x Branded $_j$	0.9454***	1.2834***	-0.0520***	0.4466***
	0.0006	0.0014	0.0011	0.0025
Adoption Share in Zipcode $_{it-1}$ x Old $_{jt}$ x Generic $_j$	-0.0106***	0.1640***	-0.0541***	0.5915***
	0.0005	0.0014	0.0010	0.0024
Adoption Share in Zipcode $_{it-1}$ x Old $_{jt}$ x Branded $_j$	-0.0106***	0.1559***	0.9081***	1.2694***
	0.0004	0.0009	0.0006	0.0015
Physician FE	Y	Y	Y	Y
Drug x Month FE	Y	Y	Y	Y
Prescription Volume $_{it-1}$	Y	Y	Y	Y
Observations	36238793	7674288	36238793	7674288
First-Stage F statistic	7.9e+05	95954	7.9e+05	95954

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides first-stage estimates corresponding to equation (7) and the second-stage estimates in Table 9. All regressions include physician and drug-month fixed effects, as well as the cholesterol drug prescription volume for physician i in month t-1. Standard errors appear below each point estimate.

Table A.3: Time to First Prescription of a New Drug, Intensity of Use, U.S. Physicians, 2000—2010

Dependent Variable:	Indicator for prescription within first year of drug introduction					
	All physicians				Eventual adopters	
	(1)	(2)	(3)	(4)	(5)	(6)
Database $_{ij}$ x Low Usage $_i$	0.0032**	0.0046**	0.0022	0.0049**	-0.0128***	0.0022
	0.0016	0.0019	0.0019	0.0022	0.0021	0.0023
Database $_{ij}$ x Medium Usage $_i$	0.0067***	0.0042	0.0018	-0.0011	-0.0125***	-0.0044
	0.0018	0.0027	0.0020	0.0029	0.0022	0.0032
Database $_{ij}$ x Intense Usage $_i$	0.0341***	0.0154***	0.0275***	0.0056*	0.0136***	0.0026
	0.0020	0.0030	0.0023	0.0033	0.0024	0.0033
Database $_{ij}$ x Low Usage $_i$ x Generic $_j$			0.0039	0.0005	0.0019	-0.0054
			0.0026	0.0026	0.0036	0.0037
Database $_{ij}$ x Medium Usage $_i$ x Generic $_j$			0.0172***	0.0156***	0.0149***	0.0096**
			0.0031	0.0032	0.0041	0.0042
Database $_{ij}$ x Intense Usage $_i$ x Generic $_j$			0.0232***	0.0282***	0.0205***	0.0222***
			0.0034	0.0034	0.0042	0.0044
Physician FE	N	Y	N	Y	N	Y
Zipcode-Drug FE	Y	Y	Y	Y	Y	Y
Prescription Volume $_{it(j)-1}$	Y	Y	Y	Y	Y	Y
Observations	1513408	1513408	1513408	1513408	480869	480869
R^2	0.4557	0.5690	0.4557	0.5690	0.4759	0.5845

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides least-squares estimates for a variant of (5). The binary dependent variable captures the time lapse between FDA approval of drug j and physician i's initial prescription of it, taking a value of 1 if initial prescription occurs within a year of FDA approval; specification are included for the full sample of physicians (columns 1-4) and for the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (columns 5-6). Database is the Drug Database indicator variable described in Table 2, and takes a value of 1 for a physician user with database access at the time drug j receives FDA approval. Generic indicates the products pravastatin, lovastatin, and simvastatin. Low, medium, and intense usage denote non-overlapping categories of physicians who, conditional on adoption, use the database to look up cholesterol drugs to differing extents; the usage proxy is zero for low-intensity users, between zero and 14 for medium-intensity users, and above 14 for high-intensity users. Regressions include zipcode-drug (columns 1-6) and physician (columns 2, 4, 6) fixed effects as well as the cholesterol drug prescription volume for physician i in the month prior to drug j's introduction. Standard errors clustered by zipcode appear below each point estimate; results are qualitatively identical when errors are clustered by physician.

Table A.4: Prescription Diversity, Intensity of Use, U.S. Physicians, 2000—2010

Dependent Variable:	Number of Unique Drugs	Prescription HHI	Number of Unique Drugs	Prescription HHI
	All physicians		Eventual adopters	
	(1)	(2)	(3)	(4)
Database $_{it}$ x Low Usage $_i$	0.0605***	-0.0032***	-0.0082***	0.0007***
	0.0027	0.0003	0.0030	0.0003
Database $_{it}$ x Medium Usage $_i$	0.0618***	-0.0034***	-0.0092***	-0.0000
	0.0037	0.0004	0.0039	0.0005
Database $_{it}$ x Intense Usage $_i$	0.2283***	-0.0104***	0.1543***	-0.0060***
	0.0043	0.0005	0.0046	0.0005
Physician FE	Y	Y	Y	Y
Zipcode-Month FE	Y	Y	Y	Y
Prescription Volume $_{it-1}$	Y	Y	Y	Y
Observations	15510386	15510386	6727828	6727828
R^2	0.8328	0.6317	0.8441	0.6523

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides least-squares estimates for a variant of (6) for cholesterol drug prescriptions by U.S. physicians during January 2000 through December 2010, including all sample physicians (columns 1-2) and the subset of physicians that eventually adopt the electronic reference, but need not use it to search for information about cholesterol drugs (columns 3-4). The dependent variable in columns 1 and 3 captures the prescription diversity of physician i as the number of unique drugs j that are prescribed by i during month t. The dependent variable in columns 2 and 4 is the prescription Herfindahl-Hirschman index for physician i in month t. Database is the Drug Database indicator variable described in Table 2, and takes a value of 1 for a physician user with database access in month t; it is otherwise zero. Low, medium, and intense usage denote non-overlapping categories of physicians who, conditional on adoption, use the database to look up cholesterol drugs to differing extents; the usage proxy is zero for low-intensity users, between zero and 14 for medium-intensity users, and above 14 for high-intensity users. All regressions include physician and zipcode-month fixed effects, as well as the cholesterol drug prescription volume for physician i in month t-1. Standard errors clustered by zipcode-month appear below each point estimate.

Table A.5: Prescription Propensity, Intensity of Use, U.S. Physicians, 2000—2010

	Dependent Variable:	1{(prescriptions of drug j by i at t) > 0}	
		All physicians	Eventual adopters
		(1)	(2)
Database $_{it}$ x New$_{jt}$ x Generic$_j$			
	x Low Usage $_i$	0.0092***	0.0059***
		0.0014	0.0018
	x Medium Usage $_i$	0.0153***	0.0122***
		0.0018	0.0021
	x High Usage $_i$	0.0395***	0.0356***
		0.0020	0.0023
Database $_{it}$ x New$_{jt}$ x Branded$_j$			
	x Low Usage $_i$	0.0046***	0.0059***
		0.0009	0.0010
	x Medium Usage $_i$	-0.0021***	-0.0009
		0.0012	0.0012
	x High Usage $_i$	-0.0198***	-0.0197***
		0.0015	0.0016
Database $_{it}$ x Old$_{jt}$ x Generic$_j$			
	x Low Usage $_i$	0.0119***	0.0091***
		0.0013	0.0017
	x Medium Usage $_i$	0.0127***	0.0101***
		0.0016	0.0020
	x High Usage $_i$	0.0592***	0.0555***
		0.0018	0.0022
Database $_{it}$ x Old$_{jt}$ x Branded$_j$			
	x Low Usage $_i$	0.0021***	0.0035***
		0.0006	0.0006
	x Medium Usage $_i$	0.0002	0.0015*
		0.0008	0.0009
	x High Usage $_i$	-0.0034***	-0.0030***
		0.0011	0.0011
Physician FE		Y	Y
Zipcode x Month FE		Y	Y
Drug x Month FE		Y	Y
Prescription Volume $_{it-1}$		Y	Y
Observations		28103453	12623365
R^2		0.5187	0.5271

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides least-squares estimates for a variant of (7) for all sample physicians (column 1) and the subset of physicians that eventually adopt the drug reference (column 2). The dependent variable is an indicator for whether the doctor i prescribes drug j during month t. Database is the Drug Database indicator variable described in Table 2. New denotes a drug j that is within 24 months of its initial FDA approval in month t; drug j is otherwise Old. Generic indicates the products pravastatin, lovastatin, and simvastatin; other drugs are Branded. Low, medium, and intense usage denote non-overlapping categories of physicians who, conditional on adoption, use the database to look up cholesterol drugs to differing extents; the usage proxy is zero for low-intensity users, between zero and 14 for medium-intensity users, and above 14 for high-intensity users. All regressions include physician, zipcode-month, and drug-month fixed effects, as well as the cholesterol drug prescription volume for physician i in month t-1. For computational ease, the estimates above rely only on observations in January 2000 and every subsequent June and December. Standard errors clustered by zipcode-month appear below each point estimate.

Table A.6: Time to First Prescription, Mandatory Substitution Laws, U.S. Physicians, 2000—2010

	Dependent Variable:	Indicator for prescription within first year of drug introduction					
		All physicians			Eventual users		
		Mandatory Substitution Law			Mandatory Substitution Law		
		Yes	No	All	Yes	No	All
		(1)	(2)	(3)	(4)	(5)	(6)
Database $_{ij}$		0.0022	-0.0018	-0.0005	-0.0016	-0.0062	-0.0042
		0.0036	0.0025	0.0025	0.0041	0.0029	0.0028
Database $_{ij}$ x Mandatory Substitution $_i$				0.0001			-0.0014
				0.0043			0.0043
Database $_{ij}$ x Generic $_j$		0.0264***	0.0251***	0.0236***	0.0257***	0.0180***	0.0172***
		0.0038	0.0026	0.0026	0.0060	0.0041	0.0040
Database $_{ij}$ x Generic $_j$ x Mandatory $_i$				0.0058			0.0100
				0.0045			0.0068
Generic $_j$ x Mandatory $_i$				0.0175***			0.0129**
				0.0017			0.0054
Drug FE		Y	Y	Y	Y	Y	Y
Physician FE		Y	Y	Y	Y	Y	Y
Prescription Volume $_{it(j)-1}$		Y	Y	Y	Y	Y	Y
Observations		500158	1013050	1513261	99631	212685	313646
R^2		0.5566	0.5525	0.5535	0.5713	0.5656	0.5673

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides least-squares estimates for (5) and a variant thereof. The binary dependent variable captures the time lapse between FDA approval of drug j and physician i's initial prescription of it, taking a value of 1 if initial prescription occurs within a year of FDA approval; specification are included for the full sample of physicians (columns 1-4) and for the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (columns 5-6). Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access at the time drug j receives FDA approval. Generic indicates the products pravastatin, lovastatin, and simvastatin. Estimates are presented for two subsamples: physicians located in states with active mandatory substitution laws (columns 1,4) in the initial period and those without such laws (columns 2,5); the full-sample results in columns 3 and 6 include interactions with an indicator for whether physician i's state has an active mandatory substitution law. Regressions include drug (columns 1-6) and physician (columns 2, 4, 6) fixed effects, as well as the cholesterol drug prescription volume for doctor i in the month prior to drug j's introduction. Standard errors clustered by physician appear below each point estimate.

Table A.7: Time to First Prescription, Pharmaceutical Innovation, U.S. Physicians, 2000—2010

Dependent Variable:	Indicator for prescription within first year of drug introduction							
	All physicians				Eventual users			
	Medical Patents		All	All	Medical Patents		All	All
	High	Low	All	All	High	Low	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Database $_{ij}$	-0.0000	0.0015	-0.0008	0.0074	-0.0095	-0.0139	-0.0048*	0.0029
	0.0096	0.0090	0.0022	0.0050	0.0209	0.0109	0.0025	0.0051
Database $_{ij}$ x High Patents $_i$			-0.0038				-0.0037	
			0.0096				0.0095	
Database $_{ij}$ x Low Patents $_i$			0.0116				0.0084	
			0.0090				0.0092	
Database $_{ij}$ x Patents $_i$				-0.0016*				-0.0015
				0.0009				0.0009
Database $_{ij}$ x Generic $_j$	0.0274***	0.0347***	0.0246***	0.0343***	0.0276*	0.0383**	0.0194***	0.0359***
	0.0096	0.0095	0.0022	0.0053	0.0152	0.0156	0.0035	0.0080
Database $_{ij}$ x Generic $_j$ x High Patents $_i$			-0.0009				-0.0089	
			0.0097				0.0145	
Database $_{ij}$ x Generic $_j$ x Low Patents $_i$			0.0139				0.0288**	
			0.0097				0.0144	
Database $_{ij}$ x Generic $_j$ x Patents $_i$				-0.0019*				-0.0032**
				0.0009				0.0015
Generic $_j$ x High Patents $_i$			0.0158***				0.0253**	
			0.0037				0.0116	
Generic $_j$ x Low Patents $_i$			-0.0119***				-0.0259**	
			0.0036				0.0113	
Generic $_j$ x Patents $_i$				0.0035***				0.0050***
				0.0004				0.0012
Drug FE, Physician FE	Y	Y	Y	Y	Y	Y	Y	Y
Prescription Volume $_{it(j)-1}$	Y	Y	Y	Y	Y	Y	Y	Y
Observations	74766	78995	1513408	1513408	15412	15090	313699	313699
R^2	0.5516	0.6167	0.5535	0.6220	0.5750	0.6350	0.5672	0.6359

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table provides least-squares estimates for (5) and a variant thereof. The binary dependent variable captures the time lapse between FDA approval of drug j and physician i's initial prescription of it, taking a value of 1 if initial prescription occurs within a year of FDA approval; specification are included for the full sample of physicians (columns 1-4) and for the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs (columns 5-6). Database is the Drug Database and Use indicator variable described in Table 2, and takes a value of 1 for a physician user with database access at the time drug j receives FDA approval. Generic indicates the products pravastatin, lovastatin, and simvastatin. Estimates in columns 1, 2, 5, and 6 are for subsamples: High includes all physicians in U.S. 4-digit zipcodes that are in the top five percent based on the number of USPTO medical patents granted (columns 1,5), Low includes those in the bottom five percent (columns 2,6). Full-sample results include interactions with indicators for High and Low medical patenting (columns 3 and 7) and the log number of medical patents (columns 4 and 8). All regressions include drug and physician fixed effects, as well as the cholesterol drug prescription volume for doctor i in the month prior to drug j's introduction. Standard errors clustered by physician appear below each point estimate.

Table A.8: Prescribing Heterogeneity, U.S. Physicians, January 2000 and December 2010

	Euclidean Distance Between $i's$ Prescriptions and the Average		
	Database $_i$ = 0	Database $_i$ = 1	All
	(1)	(2)	(3)
Panel A	December 2010		
Mean	0.1762	0.1522	
Estimated difference in means			-0.0236***
Standard error			0.0014
Panel B	January 2000		
Mean	0.2162	0.2037	
Estimated difference in means			-0.0093***
Standard error			0.0015
Panel C	December 2010 and January 2000		
Difference in means	-0.0400	-0.0515	
Estimated difference in differences			-0.0107***
Standard error			0.0017
Estimated average change			-0.0447***
Standard error			0.0011

Notes: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. This table summarizes prescription heterogeneity across U.S. physicians and over time. Columns 1 and 2 indicate the average Euclidean distance (norm) between a) the vector of physician-i prescription shares across drugs j and b) the vector of average prescription shares, in December 2010 (Panel A) and in January 2000 (Panel B) for physicians without access to the electronic database in December 2010 (column 1) and for physicians with access in December 2010 (column 2). Column 3 presents estimates from two cross-section regressions in which the mean Euclidean distance between physician i and his group average is the dependent variable, regressed on an indicator for database access in December 2010 and zipcode fixed effects. Panel C provides difference-in-differences estimates with two time periods (January 2000 and December 2010); the dependent variable is as in Panels A and B, and is regressed on an indicator for December 2010, its interaction with the indicator for database access, and physician fixed effects. Standard errors appear below each point estimate.