Online Appendix

The Impact of Information Technology on the Diffusion of New Pharmaceuticals

Kenneth J. Arrow, L. Kamran Bilir, Alan Sorensen

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 Baycol was withdrawn early in the sample period in August 2001 and is thus omitted.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 lipoproteins 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.

A.3 Endogenous Database Adoption: Instrumental Variables Results

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 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

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

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 $t-1$.[49] 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 the first-stage specification above. 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.[50]

Whether the bias in our baseline estimates is upward or downward hinges on the rela-

[49] 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.

[50] 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.

tionship 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.[51] This would tend to result in $\text{Cov}(Z_{it}, \varepsilon_{it}) \leq 0$ in (6), placing downward pressure on the estimate of β.

The estimates in Table A.6 are largely in line with the latter interpretation. Based on our preferred second-stage estimates in column 1, database users are 8.0 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. Columns 2 and 3 similarly suggest that the true impact of database adoption on diversity is an order of magnitude larger than in Table 5. Column 4 likewise suggests that column 3 of Table 5 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).[52] 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 Table A.6 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 and 5); third, we find that the results in Table A.6 are essentially unchanged when defining the instrument at a broader geographic unit (three digit zipcodes).

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

[52] First-stage estimates for Table A.6 appear in Tables A.7 and A.8.

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

		Within-Zipcode Variation, Zipcode Level			
	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 Diversity and Propensity, All U.S. Physicians, 2000—2010

Dependent Variable:	Number of Unique Drugs	Prescription HHI	1{(prescriptions of drug j by i at t) > 0}
	All Physicians		
	(1)	(2)	(3)
Database $_{it}$	0.0426***	-0.0037***	
	0.0081	0.0009	
Database$_{it}$			
x New $_{jt}$ x Generic $_{j}$			0.0266***
			0.0016
x New $_{jt}$ x Branded $_{j}$			-0.0089***
			0.0011
x Old $_{jt}$ x Generic $_{j}$			0.0351***
			0.0017
x Old $_{jt}$ x Branded $_{j}$			0.0019***
			0.0007
Prescription Volume $_{it-1}$	0.0192***	-0.0008***	-0.0007***
	0.0001	0.0000	0.0000
Physician FE	Y	Y	Y
Zipcode-Month FE	Y	Y	Y
Physician x t trends	Y	Y	Y
Observations	15510386	15510386	36238793
R^2	0.8785	0.7114	0.5277

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01 and [a] p < 0.11. This table provides least-squares estimates of a) equation (6) in columns 1-2, and b) equation (7) in column 3, for cholesterol drug prescriptions by U.S. physicians during January 2000 through December 2010 for all physicians. The dependent variable in column 1 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 column 2 is the prescription Herfindahl-Hirschman index for physician i in month t. The dependent variable in column 3 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. All regressions include the cholesterol drug prescription volume for physician i in month t-1, physician-specific time trends, and physician and zipcode-month fixed effects; column 3 also includes drug-month fixed effects. Standard errors clustered by zipcode appear below each point estimate.

Table A.3: Event Study Estimates, U.S. Physicians, Eventual Users, 2000—2010

Dependent Variable:	Number of Unique Drugs	Prescription HHI	Generic Share
	Eventual users		
	(1)	(2)	(3)
3 Years Before Adoption $_{it}$	0.0097	-0.0012	0.0014
	0.0112	0.0012	0.0010
2 Years Before Adoption $_{it}$	0.0174	-0.0010	0.0020
	0.0154	0.0017	0.0014
1 Year Before Adoption $_{it}$	0.0227	-0.0008	0.0023
	0.0188	0.0020	0.0017
Database Adoption Year $_{it}$	0.0437**	-0.0024	0.0030
	0.0210	0.0023	0.0019
1 Year After Adoption $_{it}$	0.0692***	-0.0055**	0.0049*
	0.0232	0.0025	0.0022
2 Years After Adoption $_{it}$	0.0851***	-0.0068**	0.0061**
	0.0258	0.0028	0.0024
3 Or More Years After Adoption $_{it}$	0.0818***	-0.0082***	0.0069**
	0.0291	0.0031	0.0027
Physician FE	Y	Y	Y
Zipcode-Month FE	Y	Y	Y
Prescription Volume $_{it-1}$	Y	Y	Y
Physician x t trends	Y	Y	Y
Observations	3013241	3013241	3013241
R^2	0.8941	0.7484	0.7484

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 to December 2010, and for the subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs. The dependent variable in column 1 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 column 2 is the prescription Herfindahl-Hirschman index for physician i in month t. The dependent variable in column 3 is the generic share in prescriptions by physician i in month t. The specification replaces the Drug Database and Use indicator variable described in Table 2 with dummies for three years before adoption, two years before adoption, one year before adoption, the adoption year, one year after adoption, two years after adoption, and three or more years after adoption. 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. Standard errors clustered by zipcode appear below each point estimate.

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

Dependent Variable:	1{(prescriptions of drug *j* by *i* in drug *j*'s initial year) > 0}	Number of Unique Drugs	Prescription HHI
	Eventual adopters		
	(1)	(2)	(3)
Database x Low Usage $_i$	0.0025	0.0246***	0.0019***
	0.0025	0.0082	0.0009
Database x Medium Usage $_i$	-0.0044	0.0404***	-0.0040***
	0.0033	0.0110	0.0012
Database x Intense Usage $_i$	0.0038	0.0414***	-0.0023
	0.0036	0.0129	0.0014
Database x Low $_i$ x Generic $_j$	-0.0069		
	0.0043		
Database x Medium $_i$ x Generic $_j$	0.0074		
	0.0046		
Database x Intense $_i$ x Generic $_j$	0.0236***		
	0.0047		
Physician FE	Y	Y	Y
Zipcode-Drug FE	Y	N	N
Zipcode-Month FE	N	Y	Y
Prescription Volume $_{it(i)-1}$	Y	Y	Y
Physician x *t* trends	N	Y	Y
Observations	461653	6727828	6727828
Panel R^2	0.6571	0.8859	0.7283

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides least-squares estimates of a variant of a) equation (5) for U.S. physicians' prescription of twelve cholesterol drugs first approved for U.S. sale during January 2000—December 2008 (Table 1) in column 1, and b) equation (6) for cholesterol drug prescriptions by U.S. physicians during January 2000 through December 2010 in columns 2-3. Doctors are included if they eventually adopt the electronic reference; they need not have used it to search for information about cholesterol drugs. The dependent variable in column 1 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. The dependent variable in column 2 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 column 3 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 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. All regressions include physician fixed effects the cholesterol drug prescription volume for physician *i* in the month prior to drug *j*'s introduction; column 1 includes zipcode-drug fixed effects and columns 2-3 include zipcode-month fixed effects and doctor-specific time trends. Standard errors clustered by zipcode appear

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}
	Eventual adopters
	(1)
Database $_{it}$ x New $_{jt}$ x Generic $_{j}$	
x Low Usage $_{i}$	0.0038*
	0.0022
x Medium Usage $_{i}$	0.0118***
	0.0026
x Intense Usage $_{i}$	0.0392***
	0.0029
Database $_{it}$ x New $_{jt}$ x Branded $_{j}$	
x Low Usage $_{i}$	0.0036***
	0.0013
x Medium Usage $_{i}$	-0.0026
	0.0017
x Intense Usage $_{i}$	-0.0178***
	0.0020
Database $_{it}$ x Old $_{jt}$ x Generic $_{j}$	
x Low Usage $_{i}$	0.0070***
	0.0023
x Medium Usage $_{i}$	0.0109***
	0.0027
x Intense Usage $_{i}$	0.0613***
	0.0032
Database$_{it}$ x Old $_{jt}$ x Branded $_{j}$	
x Low Usage $_{i}$	0.0012
	0.0007
x Medium Usage $_{i}$	0.0016
	0.0011
x Intense Usage $_{i}$	-0.0015
	0.0015
Physician FE, Zipcode x Month FE, and Drug x Month FE	Y
Prescription Volume $_{it-1}$	Y
Physician x *t* trends	Y
Observations	12623365
R^2	0.5350

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides least-squares estimates for a variant of (7) for the subset of physicians that eventually adopt the drug reference; the specification is as described in Table 5, column 3, but considers differences in the intensity of database use. 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.

Table A.6: Two-Stage Least Squares, Prescription Outcomes, U.S. Physicians, 2000—2010

Dependent Variable:	1{Prescription in 1st Year}	Number of Unique Drugs	Prescription HHI	Prescription Indicator
	Eventual users			
	(1)	(2)	(3)	(4)
Database $_{ij}$	-0.0110			
	0.0121			
Database $_{ij}$ x Generic $_{j}$	0.0804***			
	0.0196			
Database $_{it}$		0.2848***	-0.0376***	
		0.0603	0.0071	
x New $_{jt}$ x Generic $_{j}$				0.0516***
				0.0069
x New $_{jt}$ x Branded $_{j}$				-0.0343***
				0.0036
x Old $_{jt}$ x Generic $_{j}$				0.2401***
				0.0133
x Old $_{jt}$ x Branded $_{j}$				-0.0165***
				0.0019
Prescription Volume $_{it(j)-1}$	Y	Y	Y	Y
Physician FE	Y	Y	Y	Y
Drug FE	Y	N	N	N
Month FE	N	Y	Y	N
Drug x Month FE	N	N	N	Y
Physician x *t* trends	N	Y	Y	Y
Observations	290898	3013241	3013241	7674288
R^2	0.5697	0.8610	0.6747	0.5357
First-Stage *F* Statistic	577	2716	2716	810

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides two-stage least squares of a) equation (5) for U.S. physicians' prescription of cholesterol drugs first approved for U.S. sale during January 2000—December 2008 (Table 1) in column 1, b) equation (6) in columns 2-3, and c) equation (7) in column 4. Dependent and independent variables are as described in Table 4 (column 1) and Table 5 (columns 2-4). 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. Generic indicates the products pravastatin, lovastatin, and simvastatin. All regressions include physician fixed effects and the cholesterol drug prescription volume for physician *i* in the month prior to drug *j*'s introduction; other controls are as indicated above. First-stage estimates for columns 1-3 are in Table A.7 and for column 4 are in Table A.8. Standard errors clustered by zipcode appear below each point estimate.

Table A.7: First-Stage Estimates, U.S. Physicians, 2000—2010

Dependent Variable:	1{Prescription in 1st Year}	Number of Unique Drugs	Prescription HHI
	Eventual users		
	(1)	(2)	(3)
Panel A	First stage for Database $_{ij}$		
Adoption Share in Zipcode $_{it\text{-}1}$	1.0621***		
	0.0091		
Adoption Share in Zipcode $_{it(j)\text{-}1}$ x Generic $_j$	-0.1249***		
	0.0095		
Panel B	First stage for Database $_{ij}$ x Generic $_j$		
Adoption Share in Zipcode $_{it(j)\text{-}1}$	-0.0641***		
	0.0074		
Adoption Share in Zipcode $_{it(j)\text{-}1}$ x Generic $_j$	0.7308***		
	0.0077		
Panel C	First stage for Database $_{it}$		
Adoption Share in Zipcode $_{it\text{-}1}$		0.1294***	0.1294***
		0.0007	0.0007
Drug FE	Y	N	N
Physician FE	Y	Y	Y
Month FE	N	Y	Y
Physician x t trends	N	Y	Y
Prescription Volume $_{it(j)\text{-}1}$	Y	Y	Y
Observations	290898	3013241	3013241
First-Stage F Statistic	577	2716	2716

Notes: * p < 0.10, ** p < 0.05, *** p < 0.01. This table provides first-stage estimates corresponding to equation (5) (column 1) and equation (6) (columns 2-3); second-stage estimates appear in Table A.6, columns 1-3. All regressions include physician fixed effects as well as the cholesterol drug prescription volum for physician i in month t-1. Standard errors appear below each point estimate.

Table A.8: First-Stage Estimates, U.S. Physicians, 2000—2010

	Eventual users	
	(1)	(2)
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 $_{j}$	0.9921***	0.0970***
	0.0013	0.0015
Adoption Share in Zipcode $_{it-1}$ x New $_{jt}$ x Branded $_{j}$	0.0390***	0.0820***
	0.0009	0.0011
Adoption Share in Zipcode $_{it-1}$ x Old $_{jt}$ x Generic $_{j}$	0.0583***	0.6574***
	0.0009	0.0011
Adoption Share in Zipcode $_{it-1}$ x Old $_{jt}$ x Branded $_{j}$	0.0519***	0.1199***
	0.0007	0.0008
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 $_{j}$	0.1214***	0.8576***
	0.0021	0.0036
Adoption Share in Zipcode $_{it-1}$ x New $_{jt}$ x Branded $_{j}$	1.2427***	0.6763***
	0.0015	0.0026
Adoption Share in Zipcode $_{it-1}$ x Old $_{jt}$ x Generic $_{j}$	0.14300***	1.2164***
	0.0016	0.0027
Adoption Share in Zipcode $_{it-1}$ x Old $_{jt}$ x Branded $_{j}$	0.1297***	1.7685***
	0.0011	0.0019
Physician FE	Y	Y
Physician x *t* trends	Y	Y
Drug x Month FE	Y	Y
Prescription Volume $_{it-1}$	Y	Y
Observations	7674288	7674288
First-Stage F statistic	810	810

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 A.6, column 4. 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.9: Time to First Prescription, Mandatory Substitution Laws, U.S. Physicians, 2000—2010

Dependent Variable:	Indicator for prescription within first year of drug introduction		
	Eventual users		
	Mandatory Substitution Law		
	Yes	No	All
	(1)	(2)	(3)
Database $_{ij}$	-0.0021	-0.0062**	-0.0045
	0.0044	0.0031	0.0030
Database $_{ij}$ x Mandatory Substitution $_{i}$			-0.0008
			0.0050
Database $_{ij}$ x Generic $_{j}$	0.0274***	0.0184***	0.0173***
	0.0065	0.0045	0.0044
Database $_{ij}$ x Generic $_{j}$ x Mandatory $_{i}$			0.0125*
			0.0075
Generic $_{j}$ x Mandatory $_{i}$			0.0084
			0.0067
Drug FE	Y	Y	Y
Physician FE	Y	Y	Y
Prescription Volume $_{it(j)-1}$	Y	Y	Y
Observations	93902	196957	290874
R^2	0.5730	0.5694	0.5703

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 subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs. 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 (column 1) in the initial period and those without such laws (column 2); the full-sample results in column 3 include interactions with an indicator for whether physician *i*'s state has an active mandatory substitution law. 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 zipcode appear below each point estimate.

Table A.10: Time to First Prescription, Pharmaceutical Innovation, U.S. Physicians, 2000—2010

Dependent Variable:	Indicator for prescription within first year of drug introduction			
	Eventual users			
	Medical Patents			
	High	Low	All	All
	(1)	(2)	(3)	(4)
Database $_{ij}$	-0.0087	-0.0092	-0.0050*	0.0010
	0.0107	0.0099	0.0027	0.0058
Database $_{ij}$ x High Patents $_{i}$			-0.0022	
			0.0109	
Database $_{ij}$ x Low Patents $_{i}$			0.0074	
			0.0108	
Database $_{ij}$ x Patents $_{i}$				-0.0012
				0.0011
Database $_{ij}$ x Generic $_{j}$	0.0252	0.0385**	0.0205***	0.0415***
	0.0159	0.0161	0.0039	0.0091
Database $_{ij}$ x Generic $_{j}$ x High Patents $_{i}$			-0.0147	
			0.0168	
Database $_{ij}$ x Generic $_{j}$ x Low Patents $_{i}$			0.0323**	
			0.0164	
Database $_{ij}$ x Generic $_{j}$ x Patents $_{i}$				-0.0042**
				0.0017
Generic $_{j}$ x High Patents $_{i}$			0.0284*	
			0.0156	
Generic $_{j}$ x Low Patents $_{i}$			-0.0268*	
			0.0138	
Generic $_{j}$ x Patents $_{i}$				0.0056***
				0.0015
Drug FE, Physician FE	Y	Y	Y	Y
Prescription Volume $_{it(j)-1}$	Y	Y	Y	Y
Observations	16017	16932	290898	290898
R^2	0.5772	0.5650	0.5703	0.5703

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 subset of physicians that eventually adopt and use the electronic reference to search for information about cholesterol drugs. 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 and 2 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 (column 1), Low includes those in the bottom five percent (column 2). Full-sample results include interactions with indicators for High and Low medical patenting (column 3) and the log number of medical patents (column 4). 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 zipcode appear below each point estimate.

Table A.11: 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.