The Heterogeneity of Concentrated Prescribing Behavior: Theory and Evidence from Antipsychotics*

by Anna Levine Taub[1], Anton Kolotilin[2], Robert S. Gibbons[3], and Ernst R. Berndt[4]

Abstract:

Physicians prescribing drugs for patients with schizophrenia and related conditions are remarkably concentrated in their choice among ten older typical and six newer atypical antipsychotic drugs. In 2007 the single antipsychotic drug most prescribed by an average physician accounted for 59% of all antipsychotic prescriptions written by that physician. Moreover, among physicians who concentrate their prescriptions on one or a few drugs, different physicians concentrate on distinct drugs. We construct a model of physician learning-by-doing that generates several hypotheses amenable to empirical analyses. Using 2007 annual antipsychotic prescribing data on 15,037 physicians from IMS Health, we examine these predictions empirically. While prescribing behavior is generally quite concentrated (varying somewhat depending on the concentration measure being used), prescribers having greater prescription volumes, those with training in psychiatry, male prescribers, and those not approaching retirement age tend to have less concentrated prescribing patterns, particularly among the newer generation of antipsychotics.

[1]Visting Assistant Professor, Northeastern University
[2]Ph.D. candidate, MIT Department of Economics
[3]Sloan Distinguished Professor of Management, MIT Sloan School of Management, and National Bureau of Economic Research
[4]Louis E. Seley Professor in Applied Economics, MIT Sloan School of Management, and National Bureau of Economic Research

*This research has benefited enormously from the IMS Health Services Research Network that has provided data and data assistance. Special thanks are due to Stu Feldman, Randolph Frankel, Cindy Halas, Robert Hunkler and Linda Matusiak at IMS Health. We have also benefited from feedback by seminar participants at Wharton, Northeastern University, Boston University School of Public Health and from the comments of Joseph Doyle, Ulrike Malmendier, Jonathan Skinner, Douglas Staiger and Richard Zeckhauser. The statements, findings, conclusions, views and opinions contained and expressed in this manuscript are based in part on 1996-2008 data obtained under license from IMS Health Incorporated: National Prescription Audit™, Xponent™ and American Medical Association Physician Masterfile™. All rights reserved. Such statements, findings, conclusions, views and opinions are not necessarily those of IMS Health Incorporated or any of its affiliated or subsidiary entities. This research has not been sponsored.
Document Name: Heterogeneity V44.doc Date: November 10, 2011

I. INTRODUCTION

A. *MOTIVATION AND BACKGROUND*

Consider a physician seeing a patient with a confirmed diagnosis for which a number of alternative pharmaceutical treatments are available. Patient response to the various treatments is idiosyncratic and unpredictable in terms of both efficacy and side effects, and there is little clinical evidentiary literature upon which the physician can base an *ex ante* choice among the alternative drug treatments. What treatment algorithms might the physician employ to learn about the efficacy and tolerability of the variety of possible drug therapies for the patient?

Given this uncertainty, one possibility is for the physician to concentrate prescribing behavior on a single drug. In this way the physician engages in learning by doing, observes responses to that drug, adjusts dosage strength as warranted, and thereby exploits her accumulating knowledge and experience. Simultaneously the physician could counsel the patient on the efficacy and side effect responses he/she might experience, possible interactions with other drugs, and best time of the day to take the drug, thereby improving patient adherence and outcomes, and engaging the patient to help maintain symptom remission.

Alternatively, the physician might diversify prescribing across several drugs in an attempt to personalize the treatment and find the best match between future patients and drugs. Specifically, based on information from the patient's history, limited experience with several other drugs, the existing scientific and clinical literature, conversations with fellow medical professionals in the local and larger geographical community, and perhaps interactions with pharmaceutical sales representatives, the physician might select the therapy that *a priori* appears to be the best match with the particular patient's characteristics (even if the physician is less able to counsel the patient on the side effects, interactions, and other aspects of the drug).

In short, the physician can learn from exploiting or exploring, concentrating or diversifying, or prescribing "ready-to-wear" vs. "custom-made" treatments.[i] Physicians continually face this tradeoff as they treat patients and invest in learning about available treatments. Can a theory of physician learning help us understand how the choice along this diversification-concentration continuum varies by the physician's specialty, volume of patients treated, and training? Also, if physicians concentrate their prescribing, does theory suggest a convergence and relative unanimity on their choice of a favorite drug, or will different physicians concentrate on distinct drugs? Finally, are there persistent geographic differences in physician prescribing behavior, analogous to the small-area variation phenomenon, or do physicians' drug utilization shares mimic regional or national shares?[ii] These are the issues we address in this research, with an application of our theoretical framework to a particular therapeutic class of drugs known as antipsychotics. The issues are important, for understanding factors affecting physicians' choices along the diversification-concentration continuum has significant commercial and public-health implications.

That physician prescribing behavior is relatively concentrated has been documented by, among others, Frank and Zeckhauser [2007].[iii] They report that among 1372 primary care physicians surveyed in 2004, the fraction of prescriptions accounted for by the most prescribed medication used by the physician was generally quite high across four acute and five chronic conditions (averaging 60%), but was about 13 percentage points lower for chronic than acute conditions. More generally, while in some cases patient demographic factors affected physician treatment variability, patient clinical factors played a startlingly minor role. Physician "ready-to-wear" treatment norms in some cases could be perceived as "sensible response to complex decision-making," but in other cases could be viewed as "disturbing" and "based on idiosyncratic physician-specific preferences or severe biases in the application of heuristics."[iv] Our theoretical framework extends their analysis in several ways.[v] We focus on antipsychotics (medicines treating primarily chronic conditions) and examine prescriber behavior across

a much larger number and variety of specialties. We also develop a theoretical framework that enables us to derive and examine several additional hypotheses empirically. However, unlike Frank and Zeckhauser, we observe only physicians, not the patients they treat.

We proceed as follows. First, we provide a brief background on several generations of antipsychotic drugs and the illnesses they treat. Next, we report preliminary evidence on concentrated vs. diversified prescribing behavior, and we utilize our initial findings on "heterogeneous concentration" to develop a theoretical framework (emphasizing learning-by-doing) that enables us to construct several hypotheses. We then discuss the data and econometric framework, and we present a substantial set of empirical findings. We conclude by relating our findings to the geographical-variation literature and suggesting directions for future research.

B. *ANTIPSYCHOTICS FOR THE TREATMENT OF SCHIZOPHRENIA AND RELATED CONDITIONS*

Schizophrenia is an incurable mental illness. It is characterized by "gross distortions of reality, disturbances of language and communications, withdrawal from social interaction, and disorganization and fragmentation of thought, perception and emotional reaction."[vi] Symptoms are both positive (hallucinations, delusions, voices) and negative (depression, lack of emotion). The prevalence of schizophrenia is 1-2%, with genetic factors at play but otherwise unknown etiology. The illness tends to strike males in late teens and early twenties, and females five or so years later. As the illness continues, persons with schizophrenia frequently experience unemployment, lose contact with their family, and become homeless; a substantial proportion experience periods of incarceration.[vii]

Because schizophrenia is a chronic illness affecting virtually all aspects of life of affected persons, the goals of treatment are to reduce or eliminate symptoms, maximize quality of life and adaptive functioning, and promote and maintain recovery from the adverse effects of illness to the maximum extent possible.[viii] In the US, Medicaid is the largest payer of medical and drug benefits to people with schizophrenia.[ix]

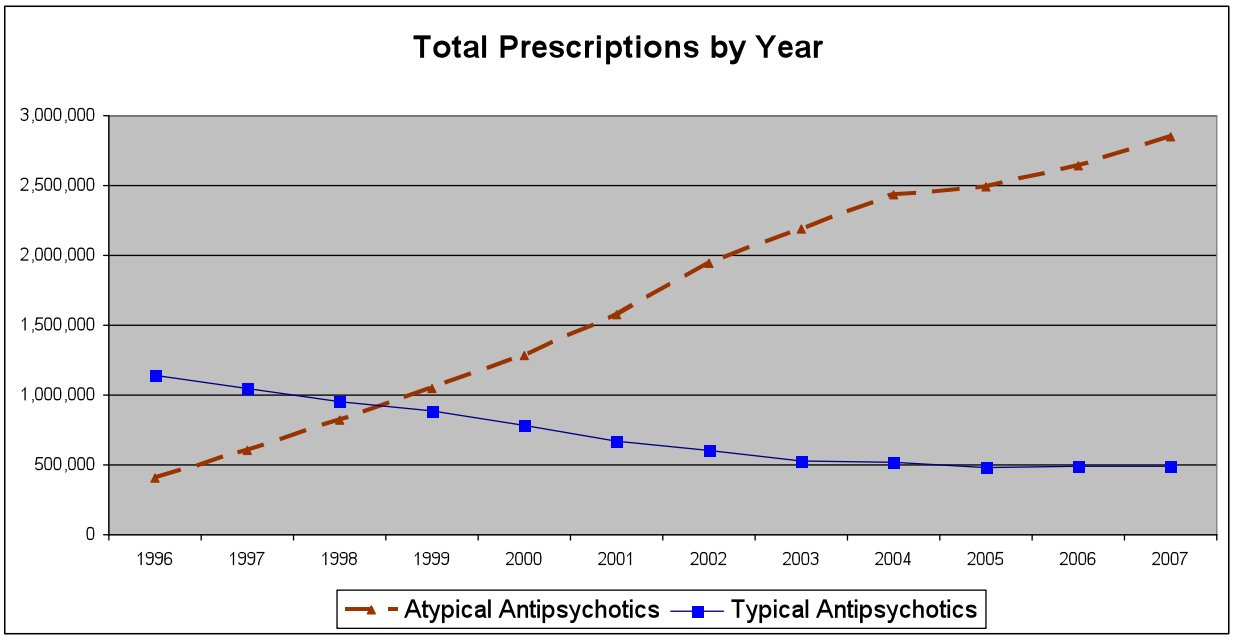
From 1955 up through the early 1990s, the mainstays of pharmacological treatment of schizophrenia were *conventional* or *typical* antipsychotic (also called *neuroleptic*) drugs that were more effective in treating the positive than the negative symptoms, but frequently resulted in extrapyramidal side effects (such as tardive dyskinesia—an involuntary movement disorder most often characterized by puckering of the lips and tongue, or writhing of the arms or legs) that may persist even after the drug is discontinued, and for which currently there is no effective treatment. In 1989, Clozaril (generic name clozapine) was approved by the U.S. Food and Drug Administration as the first in a new therapeutic class of drugs called *atypical* antipsychotics; this drug has also been dubbed a first-generation atypical (FGA). Although judged by many still to be the most effective among all antipsychotic drugs, for 1-2% of individuals taking clozapine a potentially fatal condition called agranulocytosis occurs (drop in the white blood cell count, leaving the immune system potentially fatally compromised). Patients taking clozapine must therefore have their white blood cell count measured by a laboratory test on a regular basis, and satisfactory laboratory test results must be communicated to the pharmacist before a prescription can be dispensed. For these and other reasons, currently clozapine is generally used only for individuals who do not respond to other antipsychotic treatments.[x]

Between 1993 and 2002, five so-called second-generation atypical (hereafter, SGA) antipsychotic molecules were approved by the FDA and launched in the US, including Risperdal (risperidone, 1993), Zyprexa (olanzapine, 1996), Seroquel (quetiapine, 1997), Geodon (ziprasidone, 2001) and Abilify (aripiprazole, 2002). Guidelines from the American Psychiatric Association state that although each of these five second-generation atypicals is approved for the treatment of schizophrenia (some later also received FDA approval for treatment of bipolar disease and major depressive disorder, as well as various pediatric/adolescent patient subpopulation approvals), they also note that "In addition to having therapeutic effects, both first- and second-generation antipsychotic agents can cause

a broad spectrum of side effects. Side effects are a crucial aspect of treatment because they often determine medication choice and are a primary reason for medication discontinuation."[xi]

Initially these SGAs were perceived as having similar efficacy for positive symptoms and superior efficacy for negative symptoms relative to typicals, but without the typicals' extrapyramidal and agranulocytosis side effects. However, beginning in about 2001-2002 and continuing to the present, a literature has developed regarding the association of SGAs with weight gain and the onset of diabetes, along with related metabolic syndrome side effects, particularly associated with the use of Zyprexa and clozapine and less so for Risperdal. Various professional treatment guidelines have counseled close scrutiny of individuals prescribed Zyprexa, clozapine and Risperdal. The FDA has ordered manufacturers to add bolded and boxed warnings to the product labels, initially for all atypicals, and later, to both typical and atypical antipsychotic labels. The labels have been augmented further with warnings regarding antipsychotic treatment of elderly patients with dementia, since this subpopulation appears to be at greater risk for stroke and death.[xii]

Figure 1: Number of Typical and Atypical Prescriptions, annually 1996-2007.



Source: Authors' calculations based on IMS Health Incorporated Xponent™ 1996-2007 data.

Despite this controversy, as seen in Figure 1, based on a 10% random sample of all antipsychotic prescribers in the U.S. (additional data details below), the number of atypical antipsychotic prescriptions dispensed between 1996 and 2007 increased about sevenfold from about 400,000 in 1996 to 2,800,000 in 2007, while the number of conventional or typical antipsychotic prescriptions fell 45% from 1,100,000 in 1996 to about 500,000 in 2003, and has stabilized at that level since then.[xiii] As a share of all antipsychotic prescriptions, the atypical share more than tripled from about 27% in 1996 to 85% in 2007. It is also noteworthy that, despite all the concerns about the safety and efficacy of antipsychotics, the total number of antipsychotic prescriptions dispensed in this 10% random sample – typical plus atypical – more than doubled between 1996 and 2007, from about 1,500,000 to about 3,300,000.

C. PRELIMINARY EVIDENCE ON CONCENTRATED VS. DIVERSIFIED PRESCRIBING BEHAVIOR

Although manufacturers received approval to market reformulated versions of several SGAs during the five years leading up to our 2007 sample period, no new major antipsychotic products were launched in the US during these years. Over the previous fifteen years, controversy regarding relative efficacy and tolerability of the six atypicals persisted, but prescribers learned about these drugs by observing how their patients responded, reading the clinical literature, and interacting with other professionals. These accumulated experiences have enabled prescribers to select a location along the diversification-concentration prescribing continuum.

By 2007, five years after the launch of the last SGA, how concentrated or diversified was physicians' prescribing behavior? We have two striking initial findings. First, concentration appears to be the dominant behavior: among prescribers who wrote at least twelve antipsychotic prescriptions in 2007, the average share of antipsychotic prescriptions written for the prescriber's favorite antipsychotic was 59%. Second, rather than exhibiting herd behavior (e.g., Banerjee, 1992), concentrated prescribers are quite heterogeneous in what they concentrate on, choosing different favorite drugs. For example, if

we (temporarily) limit the sample to very highly concentrated prescribers—those for whom in 2007 at least 75% of the atypical prescriptions written were for one drug (n=5,328)—we find substantial heterogeneity: 54.3% chose Seroquel as their favorite drug, 28.3% concentrated on Risperdal, 13% focused on Zyprexa, 2.5% on Abilify, 1.5% on Geodon, and 0.4% on clozapine.[xiv] We refer to the first phenomenon, where individual prescribers focus on only a few drugs, as *concentration* and the second, where a group of prescribers are dispersed around an average prescription pattern, as *deviation* (from, say, the national market shares), and we explore both below.

We conclude from this initial data examination that relatively concentrated prescribing behavior (a preference for one therapy for almost all patients) is the norm for prescribers of atypical antipsychotics, but that there is substantial heterogeneity across prescribers concerning what their favorite drug is. Thus, national market shares do not reflect homogeneous physicians each prescribing drugs in proportions that approximate the national shares, but rather portray heterogeneous physicians many of whom are highly concentrated on particular drugs (thereby generating the national shares largely via the number of concentrated prescribers who concentrate on a particular drug). In comparison to the choices of highly concentrated prescribers given above, in our 2007 sample the national market shares of the six atypicals were Seroquel 36.2%, Risperdal 27.2%, Abilify 13.8%, Zyprexa 13.1%, Geodon 7.3%, and clozapine 2.4%.

These initial findings of heterogeneous but concentrated prescribing raise an intriguing secondary issue. The highly publicized regional variation literature documents that within-region treatment variations for selected conditions experienced by Medicare patients are relatively small compared to much larger between-region differences in treatments and costs.[xv] Is there correspondingly large between-region variability in antipsychotic prescribing behavior, or is most variability physician-specific and are regions relatively similar to each other? Put differently, how does prescribing concentration vary with geographical aggregation?

To address this question, we note there are various ways one could measure the concentration behavior of prescriber i, C_i. A well-known measure of industry concentration is the Herfindahl-Hirschman Index (HHI). For a given industry or market, first rank the j = 1,…,J firms by some measure of size (e.g., revenues, employment, profits) with the first being the largest firm and the last the smallest. For each firm compute industry share s_j as its size measure divided by the total industry size measure, where the s_j share is between 0 and 100. Then square the shares and sum over the j firms, yielding HHI = $\sum_j s_j^2$. Note that the HHI ranges from zero to 10,000, with higher HHIs indicating greater concentration.[xvi] In the current context of an individual prescriber's behavior, we compute shares as the number of prescriptions written for a particular drug molecule divided by the total number of antipsychotic prescriptions written by that prescriber in 2007, and we then construct HHIs. Therefore a high HHI means that the individual prescriber is using one or at most several drugs predominately, while a low HHI implies she prescribes in a more varied manner (say, mimicking the national market shares).[xvii]

To mitigate the possible impact of very low-volume prescribers, for the rest of the paper we limit the sample to the 16,413 prescribers who in 2007 wrote at least 12 prescriptions for an antipsychotic (at least one a month). To analyze regional variation we restrict our sample to the 15,037 physician prescribers as we do not have geographic information for the non-physician prescribers in our sample. We compute mean HHIs and their variability (both standard deviations and coefficients of variation) at alternative levels of regional aggregation. While most geographical aggregates are obvious, we note that hospital referral regions (HRRs) represent 306 regional health care markets that play a prominent role in the Dartmouth regional variation and related literatures. Results are given in Table 1.

Table 1: Means, Standard Deviations and Coefficients of Variation for Antipsychotic HHIs for Alternative Geographical Aggregates, 2007

Geographic Aggregate	Mean HHI	Std. Dev	Coef. Of Variation	N
Individual Prescriber	4,841	2,323	0.48	15,037
County	3,469	1,800	0.52	1,860
Hospital Referral Region	2,208	409	0.19	306
State (plus District of Columbia)	2,060	167	0.08	51
Nation	2,018	na	na	1

IMS Health Incorporated Xponent™ 2007 data general prescriber sample data. Includes all physician prescribers who wrote at least 12 prescriptions for antipsychotics in 2007.

At the individual-prescriber level, prescribing behavior is very concentrated (HHI is almost 5000), but there is also substantial variability in the level of concentration, with the coefficient of variation being almost 0.5. However, as one aggregates into larger regions, not only is less concentrated prescribing observed, but so too is less variability in the level of concentration, particularly as one moves from the county to the HRR geographical aggregate. In particular, 93% of the difference in mean HHI between individual-prescriber and national-level shares disappears at the HRR level, and 99% disappears at the state level. Phelps [1992, pp. 25-26] has categorized coefficients of variation for surgical procedure in the 0.1 to 0.2 range as revealing "low variability," while those at 0.4 and greater are termed "high variability" procedures. Within that classification scheme, the concentration of antipsychotic prescribing behavior exhibits high variability at the individual-prescriber and county levels, but low variability at the HRR and larger regional aggregates. We conclude that while prescribing behavior is relatively concentrated at the level of the individual prescriber, and is considerably less concentrated at the county level, at both the individual prescriber and county level, antipsychotic prescribing behavior is highly variable and heterogeneous. In short, at the HRR and state levels of aggregation, there is relatively little between-region variability.

This preliminary evidence leads us to focus on individual prescribers rather than HRR- or state-level variability, and to inquire what theory of individual prescriber learning and treatment behavior can help us understand the two initial facts presented above: concentration, where individual prescribers focus on only a few drugs, and deviation, where a group of prescribers are dispersed around an average prescription pattern. We also ask whether the theory is able to generate additional predictions that can be assessed empirically. To those issues we now turn our attention.

II. TOWARDS A THEORY OF PRESCRIBER LEARNING AND TREATMENT BEHAVIOR

A. FOUR CLASSES OF EXPLANATIONS FOR HETEROGENEOUSLY CONCENTRATED PRESCRIBING

The economics and strategy literatures offer many explanations for different actors persistently responding in heterogeneous ways when faced with similar situations. Many of these explanations fall into one of the following four groups: perception, motivation, administration, and inspiration, which we now briefly summarize.[xviii]

1. Perception: We don't know we are behaving differently.

Physicians may disagree (without knowing it) about the best treatment for a particular patient. For example, suppose two medical studies arrived at different conclusions. One physician reads only one study, while the other physician reads only the other. In this case, both physicians are choosing what they believe is the best treatment for their patients and yet still choose to treat them in different ways. Physicians may persist in choosing different treatment regimes as long as they do not observe the treatment chosen by the other physician, the outcomes of the other physician's patients, or the article read by the other physician.

2. Motivation: We know we are behaving differently, but we don't want to change.

If physicians instead agreed on the most appropriate treatment but do not have the motivation to prescribe the optimal treatments for their patients, one may also observe very different prescribing decisions for each physician. If there is weak competition among physicians for patients, if knowledge

concerning which physicians are obtaining the most successful outcomes is difficult for patients to obtain, and/or if physicians' prescribing behaviors are reinforced by contacts with pharmaceutical sales representatives, then to the extent that physician-sales representative alliances are heterogeneous, we would expect to observe strong brand allegiances among physicians.[xix]

3. *Administration: We know we are behaving differently and we want to change, but we can't make the desired change happen.*

Alternatively, it could be that physicians have reached a consensus regarding what is the best treatment regime for a patient, and they may also want to give their patients the best care possible, but physicians face administrative or financial constraints preventing them from giving their patients the best treatment. For example, if the best treatment is drug A but only drug B is covered by a particular health plan's formulary, one may observe physicians using drug A whenever they can and drug B in all other cases. In this context one would observe very different prescribing behavior across physicians if their patients have different insurance coverage. In the context of antipsychotic drugs, however, Medicaid (the dominant payer for patients with schizophrenia), placed few if any restrictions on choice among the atypicals during our 2007 sample period (e.g., Medicare Part D required that any private prescription drug plan offer all but one of the atypical antipsychotic drugs on its formulary) and many other private insurers had similar formulary provisions.[xx]

4. *Inspiration: We know we're behaving differently, but we're doing the best we know how.*

Two other alternatives are that physicians may know there is a better treatment for their patients, but either they don't know which treatment is better or they need to learn more about the superior treatment in order for their patients to experience better outcomes. Roughly speaking, these two possibilities describe a bandit model and our learning-by-doing model, respectively. We say more about this distinction (and about why we chose our approach) below. For now, we simply note that, in either context, as physicians treat more patients they may learn from patients' responses to each

treatment. Given our preliminary empirical findings on concentrated prescribing behaviors documented above, the key question for any theoretical framework then becomes whether this learning causes physicians' behaviors to become more or less heterogeneous.

B. A MODEL OF PRESCRIBER LEARNING-BY-DOING

Although we do not *a priori* rule out the first three explanations underlying heterogeneously concentrated prescribing behavior (or the bandit version of the "inspiration" hypothesis), we now outline a model that formalizes the learning-by-doing hypothesis and motivates detailed empirical analyses. We later also consider a variant of the "motivation" hypothesis.

We assume that patients arrive sequentially to be seen by a physician (say, a female) and are indexed by periods in which they arrive $t \in \boldsymbol{N} = \{1, 2, \ldots\}$. That is, there are infinitely many patients and one physician. A new patient arrives at a physician's office at the beginning of each time interval *w.* That is, patient *t* arrives at the physician's office at the point in time *tw*, *w* later than patient *t-1* who arrived at *(t-1)w*. Let the continuous time discount rate be given by *r.* The physician observes that patient *t* has symptom *s* randomly drawn from the set of all possible symptoms $\boldsymbol{S} = \{1,\ldots,S\}$ with the corresponding probabilities $p_1,\ldots,p_S$. Symptoms are drawn independently across patients. The set of available drugs that treat these symptoms consists of $\boldsymbol{D} = \{1 \ldots D\}$. The maximum possible benefit of drug d for symptom *s* is B_{sd}. The ideal drug treatment for a given symptom *s* is indicated by *d*(s)*, meaning that $B_{sd^*(s)} > B_{sd}$ for all $d \neq d^*(s)$. The physician knows B_{sd} for all combinations of *s* in ***S*** and *d* in ***D***. That is, the learning in our model is not about the maximum possible benefit from drug d for a patient with symptom s; that ideal benefit is already known by the physician.

The therapy for a patient includes not only the drug, *d*, that the physician prescribes, but also any complementary actions *a* that the physician undertakes, such as adjusting the dosage of the drug (a process known as titrating), or any actions that affect the patient's adherence and outcomes, such as communicating information on possible side effects and their duration, possible adverse interactions

with other drugs, and/or the best time of the day to take the drug. In order to achieve the maximum potential benefit from a drug, the physician must undertake the ideal complementary action. It is this ideal complementary action that the physician learns about in our model. In particular, the realized effectiveness of drug *d* prescribed for patient *t* with symptom *s* is

$$b_{sdt} = B_{sd} - (a - x_{dt})^2 , \quad (1)$$

where a denotes the complementary action the physician undertakes, and

$$x_{dt} = \vartheta_d + \varepsilon_{dt} . \quad (2)$$

Thus, to achieve the maximum possible benefit ($b_{sdt} = B_{sd}$) from drug d for patient t with symptom s, the physician must choose the correct complementary actions for drug d and patient t ($a = x_{dt}$), where these actions depend on both the drug (ϑ_d) and the patient (ε_{dt}). As $|a - x_{dt}|$ increases, the realized benefit from drug *d* decreases at an increasing rate; as a result, even drug *d*(s)* can yield very poor outcomes if $|a - x_{dt}|$ is large. We assume ϑ_d and ε_{dt} are independent normally distributed random variables for all *d* and *t*, with mean zero and variances σ_ϑ^2 and σ_ε^2, respectively.

To simplify our analysis, we make a seemingly strong (but ultimately inconsequential) assumption: after prescribing drug *d* to patient *t* and undertaking complementary actions *a*, the physician observes x_{dt}. That is, the physician observes the complementary action that would have been optimal for the patient just treated, given the drug that was prescribed for that patient. Note that the physician does not observe $x_{d't}$ for $d' \neq d$ (i.e., the ideal actions had that patient been given another drug) or $x_{dt'}$ for $t' \neq t$ (i.e., the ideal actions for another patient given that drug). Note also that, because $x_{dt} = \vartheta_d + \varepsilon_{dt}$, we are not assuming that the physician observes what she would really like to know: ϑ_d. In short, our assumption gives the physician unrealistically much information about the patient just treated, but even this information still leaves the physician with much to learn about how to treat future patients.

Recall that the physician knows the maximum potential benefit from each drug B_{sd} as well as the distribution from which ϑ_d and ε_{dt} are drawn. Therefore the only uncertainty the physician faces is what complementary actions will work best for a particular drug and a particular patient.

It is useful to discuss the intuition underlying our model. Here the physician learns about ϑ_d by prescribing drug *d* and subsequently observing the ideal complementary action x_{dt} for patient *t*. Because the physician does not observe ϑ_d, she typically cannot learn everything she needs to know about a drug from treating one patient with this drug. Note that for simplicity we assume that the best action that the physician can potentially learn to make, ϑ_d, depends only on the drug prescribed but not on the symptom. A symptom in turn determines which drug has the highest potential for giving a patient the best outcomes, d*(s). We have also assumed that the variance of ϑ_d may depend on drug *d*, but the variance of ε_{dt} depends neither on drug *d* nor on patient *t*. Therefore, initially the physician may have different uncertainties associated with different drugs. However, the speed of learning the complementary action ϑ_d for each drug *d* depends on only how often the physician prescribes drug *d*, not on the drug or patient identity.

C. DISCUSSION OF THE MODEL

Our model builds on Jovanovic and Nyarko (1996), in which a decision maker also knows all parameters of the environment except the optimal complementary action. Their model also assumes a quadratic objective function and normally distributed random variables. The novel aspect of our model is random symptoms, which implies that the long-run prescribing behavior of the physician depends on the initial history of idiosyncratic patients' symptoms presented to her.

Our model has the same reduced form as another class of models called "learning" models, namely models of "learning curves" or "learning by doing," where benefits for each drug increase deterministically with the number of times the drug is prescribed. In particular, equations (3) and (4)

below imply that in our model the expected benefits from prescribing drug *d* for symptom *s* are equal to $B_{sd} - \frac{\sigma_\varepsilon^2 \sigma_d^2}{\sigma_\varepsilon^2 + \sigma_d^2 \# d} - \sigma_\varepsilon^2$, where *#d* is the number of times the physician prescribed drug *d*.

Moreover, if there is full learning about each drug after one prescription of the drug (i.e., if $\sigma^2_\varepsilon = 0$), then our model is equivalent to the following conceptually different model. There are benefits B_{sd} that the physician obtains if she prescribes drug d for symptom s. The physician incurs a fixed cost of σ^2_d when she prescribes drug *d* for the first time, and thereafter she incurs no cost when she prescribes drug *d*. This fixed cost can represent either the physical cost of reading instructions on how to use a new drug or the cognitive costs of switching from a customary drug to a new drug.

Our model also differs from the multi-armed bandit models (see e.g., Bergemann and Valimaki, 2006). In the multi-armed bandit analog of our model, the effectiveness of each drug B_{sd} would be unknown and there would be no complementary actions. That is, patients' experiences would be noisy signals for the true quality of a drug. Then, similarly to our model, in some cases physicians' prescribing choices would diverge even if initially they had the same beliefs about the efficacy of each drug. Crawford and Shum (2005), Ferreyra and Kosenok (2010), and Dickstein (2011) estimate models in this spirit, but they do not focus on either concentration or deviation in prescriptions by physicians.[xxi]

We now explain why we analyze and implement empirically our model rather than a multi-armed bandit model. A physician can observe the national market shares of the drugs, which provide that physician information about what other physicians prescribed (and, implicitly, something about what other physicians learned about the efficacy of various drugs). In a two-armed bandit model, if players observe each others' decisions, then eventually all players settle on the same decision with probability one (see Aoyagi, 1998). This prediction is in contradiction to one of our main preliminary empirical findings. More generally, in a multi-armed bandit model, if physicians observe nation-wide

market shares of all drugs, it is not clear that either form of heterogeneous concentration in physicians' prescribing behavior will arise.

In contrast, in our learning-by-doing model, the physician's prescribing behavior does not depend on whether the physician observes national market shares, because the underlying efficacy of each drug is already known by each physician. There is no spillover learning in our model because a physician must learn *how* to use a drug, and no amount of being told that other physicians have learned how to use it can teach the physician. That is, from the prescriber's perspective, each drug is an experience good rather than a search good.

D. ANALYSIS OF THE MODEL AND PRELIMINARY COMPARATIVE STATICS

The optimal prescribing behavior of the physician can be characterized in a simple manner because the model is stationary and the realized effectiveness has a quadratic structure with normally distributed uncertainty components. Denote the physician's history through patient t by $h_t = \times_{\tau=1}^{t-1}\left(s_\tau, d_\tau, a_\tau, x_{d_\tau \tau}\right)$. The physician's policy is to choose a drug d and complementary actions a, for each patient t with symptom s and at each history h_t.

Because complementary action a does not affect learning about ϑ_d, the optimal complementary action a and physician's expected instantaneous benefit from prescribing drug d for patient t are given by:

$a(h_t) = E[\vartheta_d | h_t]$, and

$$E[b_{sdt} | h_t] = B_{sd} - Var(\vartheta_d | h_t) - \sigma_\varepsilon^2, \qquad (3)$$

where $E[\vartheta_d | h_t]$ and $Var(\vartheta_d | h_t)$ denote the conditional expectation and variance of ϑ_d at history h_t. Moreover, the standard formula for Bayesian updating with normally distributed random variables yields:

$$\frac{1}{Var(\theta_d \mid h_t)} = \frac{1}{\sigma_d^2} + \frac{\#d(h_t)}{\sigma_\varepsilon^2}, \qquad (4)$$

where $\#d(h_t)$ denotes the number of patients to whom the physician prescribed drug *d* during history h_t. From these equations, we see that the more times a physician has prescribed drug *d*, the closer she will expect to be to achieving the second-best benefits of the drug *d* for a patient with symptom *s*, namely $B_{sd} - \sigma_\varepsilon^2$.

The optimized expected benefit from prescribing drug d to patient t with symptom *s* , $E[b_{sdt}| h_t]$ in (3), depends on *d* in two ways: the maximum benefit B_{sd}, which is already known, and the expected loss from imperfect complementary actions, $Var(\vartheta_d| h_t) + \sigma_\varepsilon^2$, which depends on the history h_t. Thus, the physician's optimal choice of drug for patient *t* depends on history h_t only through posterior variances $Var(\vartheta_d| h_t)$. That is, the physician's prescribing behavior can be summarized by *D* state variables identified with posterior variances $Var(\vartheta_d| h_t)$ for *d = 1, … D.* Therefore, to compare prescribing behavior of physicians with different histories, we need to compare only their posterior variances of ϑ_d.

We now discuss comparative-static results of the learning-by-doing model with respect to *w*, the waiting time between patients. Suppose first that *w* is large (i.e., the physician is a low-volume prescriber). In this case, the physician will eventually concentrate on a subset of drugs, in the sense that all future prescriptions will be from this subset, and each drug in this subset will be prescribed for some symptom. Moreover, this subset of drugs will depend on the initial history of patients' symptoms randomly presented to the physician. The intuition behind this is as follows. If the physician observes a sequence of patients with a given symptom *s*, then she chooses an appropriate drug, say *d*, for them. The physician will learn a great deal about this drug *d* and will be unwilling to switch to another drug *d'* when she sees a patient with symptom *s'* (even if *d'* would be more appropriate for *s'* if the physician had the same knowledge about drugs *d* and *d'*).

More formally, consider a physician's choice for a patient with symptom *s'* between two drugs *d'* and *d*. If the physician is myopic then the expected benefits to the patient from using drugs *d'* and *d* are given by

$$B_{s'd'} - Var(\vartheta_{d'} | h_t) - \sigma_\varepsilon^2 \quad \text{and} \qquad (5)$$

$$B_{s'd} - Var(\vartheta_d | h_t) - \sigma_\varepsilon^2. \qquad (6)$$

Therefore, the myopic physician is trading off the difference between $B_{s'd'}$ and $B_{s'd}$ against the difference between $Var(\vartheta_{d'} | h_t)$ and $Var(\vartheta_d | h_t)$. If the maximum potential benefit from drug *d'*, $B_{s'd'}$, is greater than that from drug *d*, $B_{s'd}$, but the physician has prescribed drug *d* more often than drug *d'* in the past so that $Var(\vartheta_d | h_t) < Var(\vartheta_{d'} | h_t) - (B_{s'd'} - B_{s'd})$, then she will choose drug *d*.

As *w* is decreased (i.e., the volume of patients seen by the physician increases), the model implies that physicians have a larger incentive to invest in learning how to use new or different drugs effectively. The set of drugs a physician eventually uses will still depend on the initial history of symptoms the physician has seen, but this dependence becomes weaker as patient volume increases. Therefore we would expect to see less concentrated prescribing with increases in patient volume, all else equal.

Finally, as *w* decreases to zero (i.e., the physician sees patients almost continuously), the set of drugs that the physician will prescribe will cease to depend on the symptoms of the initial patients that the physician randomly sees. More formally, if we assume that there are sufficiently many different symptoms such that each drug *d* in ***D*** is optimal for some symptoms *s* in ***S*** (i.e., for each *d* there exists *s* such that $d^*(s)=d$), then a very high-volume physician will eventually learn a great deal about optimal complementary actions ϑ_d for each drug *d* in ***D*** and prescribe $d^*(s)$ for every *s*.

As noted in the Introduction, our initial examination of the data revealed two striking facts: not only concentration, as we have just discussed, but also deviation (say, from national market shares). The above intuition about concentration applies to deviation as well: because the long-run prescriptions of

physicians with low volume are influenced by the random initial history of patients the physician treats, we expect low-volume physicians to be not only concentrated in their prescriptions but also different from each other and hence from national shares, whereas physicians with very high volumes (i.e., *w* approaching zero) will eventually prescribe *d*(s)* for every *s* and so have a common distribution of prescriptions, regardless of their initial history of patients.

To conclude this description of our theoretical framework, we now address two features of our data that are outside the abstract model developed thus far: new drugs and new physicians. New drugs that appear during a given physician's career are straightforward to add to our model, as follows. Suppose that after the history $h_t = \times_{\tau=1}^{t-1}\left(s_\tau, d_\tau, a_\tau, x_{d_\tau \tau}\right)$ in which each prescribed drug d_τ was necessarily chosen from the original set of available drugs D, a new drug *d′* becomes available. For simplicity, suppose that (a) the introduction of drug *d′* is a complete surprise to the physician and (b) the physician believes that no other drugs will be introduced during the remainder of her career. In this case, our model effectively starts over when the new drug *d′* is introduced, with the proviso that if drug *d* in *D* was prescribed during history h_t then the physician's uncertainty about complementary actions for drug *d* is now lower than it was when she started seeing patients. As a result of this reduction in uncertainty, it can be optimal for the physician (and her patients) to prescribe a drug *d* from *D* for both symptoms *s* and *s′*, even if drug *d′* would be preferred for symptom *s′* in the absence of such uncertainty (i.e., $B_{sd'} > B_{sd}$).

To summarize the possible effect of a new drug, recall that in our original model, if a physician's volume is not too high, then her early random exposure to particular symptoms and drugs can cause her steady-state prescriptions to be concentrated on a subset of drugs. A similar logic holds here, but it can apply also to higher-volume physicians who had prescribed every drug *d* in *D* before the new drug *d′* appeared.

In addition to new drugs appearing over time, our data also include new physicians appearing over time. For a given physician, who starts seeing patients at a given date, the set of drugs available at that date is the set *D* in our model, and for this physician any new drugs that appear subsequently can be handled as just described. New issues arise, however, if one attempts to compare prescription behaviors between physicians who started seeing patients at different dates, and when different drugs are available at these different dates. In brief, we have just argued that, for a given physician, the effect of volume on steady-state prescription concentration can differ before versus after a new drug appears, so it is likewise the case that the effect of volume on concentration can differ between physicians who start seeing patients at different dates, with more drugs available at the later date.

To exposit all these ideas in a simple setting, in Appendix A we develop an example of our model. To accelerate physicians' progress towards steady-state prescription behaviors, we assume that $\sigma_\varepsilon^2 = 0$, so that a physician learns everything about a drug's complementary actions after prescribing the drug just once. The original uncertainty about the drug's complementary actions, σ_d^2, can then be viewed as a one-time cost of learning about the drug, in the sense that the expected benefit from prescribing drug *d* for symptom *s* is now $B_{sd} - \sigma_d^2$ the first time the drug is prescribed and B_{sd} thereafter. In this example, only drug d_1 is available in the first period, but both drug d_2 and a new cohort of physicians appear in the second period. This structure of the example ensures that the steady state is reached in the third period. We then analyze how steady-state prescription rates vary across drugs and physicians. This example will be especially relevant for our discussion in Section IV.A of a competing hypothesis (namely, "detailing" by sales representatives from pharmaceutical firms, rather than our model of learning-by-doing).

E. FROM THEORY TOWARDS EVIDENCE

Our main theoretical framework (before the introduction of new drugs or new physicians) suggests that low-volume physicians may concentrate on a smaller subset of steady-state drugs than will high-volume physicians, since low-volume physicians have a smaller incentive to invest in learning how to use different drugs effectively than high-volume physicians do. In addition, we expect which drugs are in the steady-state prescription set to vary more among low- than high-volume physicians, because the eventual treatment decisions of low-volume physicians depend more on their random patient history than do those of high-volume physicians.

We also expect that differences in physicians' specialties can influence steady-state prescription decisions. In particular, training in different specialties may include more or less information about complementary actions for different drugs, so σ_d^2 may differ across specialties, and training may also influence a physician's ability to learn from observing x_{dt}, in the sense that σ_ε^2 may differ across specialties. Like higher volume, lower values of these two variances lead to less concentrated steady-state prescription patterns.

Finally, we expect older physicians to experiment with new drugs less than do younger physicians, for two reasons. First, as suggested above, older physicians will have prescribed more old drugs than younger physicians. Second (but not yet in our model), older physicians approaching retirement have shorter planning horizons than do younger physicians. To capture the latter somewhat loosely in our model, we can imagine that physicians closer to retirement have a higher discount rate *r* when a new drug arrives. Similarly to differences in patient arrival rate, *w*, physicians with higher discount rates, *r*, are less likely to experiment with new drugs.

We now describe the data utilized in our analysis, the econometric methods we implement, and our findings concerning the extent to which the predictions of this model are consistent with prescribing behavior observed in our data.

III. DATA, METHODS AND FINDINGS

A. PRESCRIPTIONS DATA

Our data on prescribers' behavior are taken from the IMS Xponent data source that tracks prescribing behavior by linking individual retail and mail-order dispensed pharmacy prescriptions to the prescriber identification number. A 10% random sample of all prescribers who wrote at least one antipsychotic prescription in 1996 was drawn, and these prescribers are followed on a monthly basis from January 1996 through September 2008. Each year after 1996 the sample is refreshed by adding a 10% sample of new antipsychotic prescribers. These prescribers are "new" in the sense that they are new to the sample; they may have been prescribing antipsychotics for many years. For each physician prescriber, we have matched geographical, training and office-practice data from the registry at the American Medical Association. Our data are a cross-section of prescribers in 2007, five years after the market introduction of the last branded atypical antipsychotic medication (and ten or more years after four of the six atypicals were introduced). To mitigate the possible impact of very low-volume prescribers we limit the sample to the 16,413 prescribers who in 2007 wrote at least 12 prescriptions for an antipsychotic (at least one a month).

We aggregate various specialties into five groups. Primary care physicians ("PCPs") include internal medicine, family medicine and practice, pediatrics, and general practice prescribers. Another group of prescribers is psychiatrists ("PSY"), which includes not only general psychiatry but also child - adolescent and geriatric psychiatry. The neurologist group ("NEU") includes those in general neurology, as well as geriatric and child neurologists. A fourth group of prescribers encompasses non-physicians ("NPs"), primarily nurse practitioners and physician assistants.[xxii] We designate all other prescribers as other ("OTH").

As seen in Table 2, although PCPs comprise about 50% of our sample of 16,413 prescribers, in 2007 they and the relatively populous OTH group of prescribers wrote relatively few antipsychotic and atypical prescriptions, averaging less than 70 annually. In contrast, PSYs averaged more than 600 antipsychotic (554 atypical) prescriptions annually, several times the second leading prescribers – NPs, with about 200 antipsychotic (185 atypical) prescriptions annually. NEU prescribers write on average almost 100 antipsychotic (87 atypical) prescriptions annually.

Table 2: Mean Values of Characteristics of 2007 Prescriber Sample, by Prescriber Specialty

Specialty Group	Number of Prescribers	Antipsychotic Annual Rx	Atypical Annual Rx	No. Distinct Antipsychotics	No. Distinct Atypicals	Antipsychotic HHI	Atypical HHI	% Antipsychotic Rxs for Atypicals
PSY	3,431	611.03	554.45	7.26	4.71	3,261	3,659	91.37
NEU	688	97.53	86.57	3.23	2.39	6,050	6,956	85.30
PCP	8,536	66.49	59.02	3.78	2.90	5,002	5,720	86.85
OTH	2,382	54.42	49.27	2.95	2.39	6,194	6,679	88.35
NP	1,376	200.11	185.38	4.34	3.30	4,974	5,364	92.19

Notes: NEU – general, geriatric and child neurologists; PCP – primary care physicians, internal medicine, family medicine and practice, pediatrics, and general practice; PSY – general, child-adolescent and geriatric psychiatry; NP – non-physician prescribers, nurse practitioners and physician assistants; OTH – all other prescribers.

All values calculated using IMS Health Incorporated Xponent™ general prescriber sample 2007 data for prescribers writing at least 12 antipsychotic prescriptions.

Even in these raw data, one begins to see patterns in the concentration of prescribing behavior. For example, PSYs, the highest-volume prescribers, prescribe on average the largest distinct number of antipsychotics (7.26) and atypicals (4.71), and they exhibit the least concentrated antipsychotic prescribing behavior, having on average an HHI of 3,261 (3,659 for atypicals). In contrast, OTH physicians, the lowest-volume prescribers, use the smallest number of distinct antipsychotic (2.95) and atypical (2.39) molecules, and they are the most concentrated prescribers, having an HHI of 6,194 (6,679 for atypicals, slightly less than the 6,956 atypical HHI for NEU prescribers). While NPs are second only to PSYs in terms of annual volume, in terms of both the variety of drugs they use and their concentration, their behavior is quite similar to that of the relatively low-volume PCPs.

We link the prescriber identifiers in the IMS Xponent data base to the American Medical Association ("AMA") directory of physicians. Notably, while the AMA Masterfile Directory has education, training, specialty certification and demographic data on most physicians and type of practice as of 2008, there is no comparable data available on NP nurse practitioners or physician assistants and therefore for our subsequent empirical analyses we exclude all NPs.[xxiii]

Finally, each prescriber in our sample is assigned a geographical location based on their 2007 location. In addition to the obvious country, state and national aggregates, we also examine hospital referral regions (HRRs) that have played a prominent role in analyses by the Dartmouth Atlas Project researchers.[xxiv]

Several features of the physician data set are worth noting. First, we have data on only physicians/NPs and their prescribing behavior, not on the patients they see. Second, IMS keeps track of prescribers that are deceased or retire, using look-back windows with no prescribing activity for one year forward and one year backward. Third, because the sample starts with prescribers who wrote at least one antipsychotic prescription in 1996 (who are then followed through September 2008, unless they die or retire), the set of prescribers in the sample is likely older than would be observed in an entirely new random sample drawn in, say, 2007.[xxv]

B. EMPIRICAL FRAMEWORK AND ECONOMETRIC METHODS

The cross-sectional regression specification we take to the 2007 data is of the following general form:

$$Y_i = \beta Volume_i + \varphi X_i + \varepsilon_i \qquad (7)$$

where Y_i is one of two dependent variables (either D_i, a measure of the deviation of a physician's prescriptions from a specified average, or C_i, a measure of the concentration of a physician's prescriptions), $Volume_i$ is the number of prescriptions from prescriber i, and X_i is a vector of covariates,

all of which are described in more detail below. As simple examples, two of our measures of concentration C_i are: (a) the number of different molecules a physician prescribes and (b) the HHI of the physician's prescriptions. Since HHI will be between 0 and 10,000, we take account of this in our analysis by employing appropriate econometric estimation methods. In some regressions we specify interaction variables, particularly among measures of volume and physician specialty.

Regarding covariates, the age of the prescribing physician is taken from the AMA Masterfile Directory. In our empirical analysis we use age quartiles as indicator-variable regressors instead of merely the raw age of the physician. This allows us to evaluate effects that may be nonlinear in age. The age quartiles are less than 43, between 43 and 50, between 51 and 58, and age 59 and greater.

While we do not have any information about patients, several practice-setting variables help us partially control for the patient mix seen by a given physician. In particular, we observe the specialty of the physician as well as whether the physician is hospital or office-based, and the county/region in which the practice is located. We expect, as Table 2 reports, that specialty is also correlated with antipsychotic prescribing volume.

In terms of differential learning costs (σ_d^2 and σ_ε^2 in our model), we might expect the learning costs for physicians to vary depending on their training and/or current practice environment. In particular, we control for whether the physician practices in a group or has a solo practice, the population of the county in which the physician practices, and whether the physician has an MD or DO degree.[xxvi]

Finally, women and men might use this technology in different ways, although our theory has nothing to say about this. Therefore, we control for the gender of the physician. In addition, some physicians ask that their prescribing data not be shared with pharmaceutical or other for-profit organizations. We will examine whether these "opt-out" physicians appear to differ from other physicians in their prescribing behavior.

The concentration of a physician's prescriptions can be quantified alternatively by the number of different antipsychotics prescribed or the Herfindahl-Hirschman Index described above. The deviation of a physician's prescriptions (say, from regional market shares) can be quantified as follows. Consider physician i prescribing drug d in geographical region r, and denote the share of prescriptions written by this physician for drug d as s_{idr}. Let the overall market share of drug d in region r be m_{dr}, where both s_{idr} and m_{dr} are between zero and 100. As a measure of the deviation of physician i's prescribing behavior from that of the regional market share, we calculate

$$D_{ir} = \sum_d (s_{idr} - m_{dr})^2 = HHI_i + HHI_r - 2\sum_d s_{idr}m_{dr} \quad (8)$$

If every physician in region r had the same prescribing share, D_{ir} would equal zero. As physician prescribing behavior heterogeneity within region r increases, D_{ir} increases.[xxvii]

In Table 3 below we provide summary statistical information for both the dependent and explanatory variables employed in our analyses. The mean number of different antipsychotics

Table 3: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Minimum	Maximum
Number of Different Antipsychotics Prescribed	15,037	4.41	2.64	1	15
HHI of Individual Physician's Antipsychotic Prescribing	15,037	4,841	2,323	1,196	10,000
Deviation of Physician's Antipsychotic prescribing from National Market Shares	15,037	2,184	1,901	10	10,100
Number of Different Atypicals Prescribed	15,037	3.21	1.48	0	6
HHI of Individual Physician's Atypical Prescribing	14,865	5,452	2,381	1,701	10,000
% of Prescriptions for Antipsychotics that were for Atypicals	15,037	88.05	20.01	0	100
Total Yearly Antipsychotic Prescriptions	15,037	190	464	12	7,186
Total Yearly Atypical Antipsychotic Prescriptions	15,037	172	417	0	6,780
Prescriber Age	15,037	50.60	10.89	26	92
PCP	15,037	0.57	0.50	0	1
PSY	15,037	0.23	0.42	0	1
NEU	15,037	0.05	0.21	0	1
OTH	15,037	0.16	0.42	0	1
Solo Practice	15,037	0.20	0.40	0	1
Population (county)	15,037	1,022,341	1,752,971	1,299	9,734,701
Female	15,037	0.27	0.44	0	1
Hospital Based Physician	15,037	0.08	0.27	0	1
DO Flag	15,037	0.09	0.28	0	1
Physician Opt Out	15,037	0.04	0.19	0	1

All values calculated using IMS Health Incorporated Xponent™ general prescriber sample 2007 data. Sample includes all physician prescribers that wrote at least 12 prescriptions for antipsychotics in 2007.

prescribed is 4.41, of which 3.21 are atypicals; the mean share of atypical prescriptions is 88.05%. The average HHI over all antipsychotics is 4,841, while that for atypicals only is larger at 5,452; the mean deviation from national shares is 2,184. The average physician age is 50.60 years, with 27% of them being female.

C. RESULTS

The reference group in all our regressions is a young (under age 43) male physician, practicing in a county with less than 150,000 residents, who has an MD degree, is not hospital-based, did not request that his prescribing information be withheld for companies interested in it for marketing purposes, and whose specialty is one that typically does not prescribe many antipsychotics (OTH). All coefficient estimates therefore compare how the prescribing behavior of a particular physician having different characteristics compares to physicians in the excluded reference group.

1. Deviation in Prescribing Behavior

We begin our empirical analysis by examining the deviation of any individual physician's prescribing behavior from national market shares. Recall that because of possibly varying random initial experience with an antipsychotic drug about which a physician attempted to learn more, our theoretical framework hypothesizes greater deviation for low-volume than high-volume prescribers, other things equal, as well as for others whose present value of benefits from learning regarding variety is lower. Initial visual inspection of the deviation data suggested a lognormal distribution. Results from OLS estimation with log of deviation as the dependent variable are presented in Table 4.

Table 4: Deviation of Physician's Antipsychotic Prescribing Shares from National Market Shares

	Coefficient
OTH*Total Yearly Antipsychotic Prescriptions	-0.00169***
	(0.000157)
PCP*Total Yearly Antipsychotic Prescriptions	-0.00200***
	(0.000087)
PSY*Total Yearly Antipsychotic Prescriptions	-.000511***
	(0.000018)
NEU*Total Yearly Antipsychotic Prescriptions	-.000751***
	(0.000171)
Age Quartile 43-50^	0.0132
	(0.0196)
Age Quartile 51-58^	0.0458**
	(0.0195)
Age Quartile 59+^	0.134***
	(0.0206)
PCP^	-0.300***
	(0.0223)
PSY^	-1.079***
	(0.0267)
NEU^	-0.0647
	(0.0411)
Female^	0.0905***
	(0.0161)
Population 150,000-500,000 (county)^	-0.0194
	(0.0192)
Population 500,000-1,000,000 (county)^	0.0608***
	(0.0203)
Population more than 1,000,000 (county)^	0.0252
	(0.0197)
Solo Practice^	0.0468***
	(0.0173)
Hospital Based Physician^	-0.0349
	(0.0259)
DO Flag^	-0.0379
	(0.0247)
Physician Opt Out^	-0.0325
	(0.0370)
Cons	7.77***
	(0.0270)
Number of Observations	*15,037*
R^2	*0.28*
Mean of dependent variable	*7.276*

* , **, ***, indicate significance at the 10%, 5%, and 1% respectively. ^ dy/dx for a dummy variable represents effect of a discrete change from 0 to 1. All values calculated using IMS Health Incorporated Xponent™ general prescriber sample 2007 data, and population estimates from the US Census Bureau. Sample includes all physician prescribers that wrote at least 12 prescriptions for antipsychotics in 2007.

Several results stand out. First, for each of the four specialty-volume interaction variables, deviation from national shares decreases with volume, and significantly so; the negative volume effect is smallest for NEU prescribers, followed by PSY, PCP, and largest for OTH prescribers. Second, conditional on volume, deviation prescribing behavior is smallest for PSY, then PCP, NEU and the reference group, OTH (the latter two are insignificantly different from each other). Relative to the youngest age quartile (under age 43), deviation increases with age, significantly for those age 51-58 and particularly for the oldest greater than age 58 quartile. Third, other things equal, female prescribers exhibit significantly more deviation. Fourth, physicians in solo practice exhibit more deviation. Physicians practicing in medium to large counties (500,000 to 1,000,000 population) exhibit more deviation than those in small (under 500,000) counties, but hospital-based physicians, DOs, and opt-out prescribers do not differ from the reference group. These results, particularly those involving volume and prescriber specialty, are consistent with predictions based on our learning-by-doing theoretical framework.[xxviii]

2. Concentration of Antipsychotic Prescribing: Number of Distinct Antipsychotic Drug Molecules Prescribed and Physician Prescribing Antipsychotic HHI

Next we examine which physicians use a wider variety of drug molecules. We employ two measures of variety. The first is the number of distinct antipsychotic drug molecules prescribed in 2007, while the second is the HHI of antipsychotic prescriptions in 2007. Since our theoretical framework suggests relationships among molecule variety, prescriber volume, specialty, and their interactions, we first inquire whether our results are consistent with the learning-by-doing model, and then discuss results involving other covariates about which our theory has no suggested relationships.

Results from estimation of a Poisson specification relating the number of distinct antipsychotic drugs a physician prescribes to a host of explanatory variables are presented in the first column of Table 5; entries in the table are estimated marginal effects evaluated at variable means. For each of the four specialty-interaction terms, the number of distinct molecules prescribed increases significantly with

volume; the positive volume effects are largest for OTH, followed by PCP and NEU, and are smallest for PSY prescribers; the volume impact on variety is about four times larger for OTH than for PSY prescribers. Conditional on volume, the three specialty dummy variables show that NEU, PCP and PSY specialists prescribe an ever greater variety of antipsychotics relative to OTH physicians. These results are consistent with the notion that high volume and PSY specialty training are alternative ways of learning (i.e., reducing σ_d^2 in our model).

The effect of age on overall antipsychotic prescribing variety is not monotonic; while marginal effects of the two middle-age quartiles are significantly positive and greater than the reference youngest quartile, the marginal effect for the oldest quartile is not different from that of the youngest. Female physicians on average prescribe fewer (about 9%) distinct antipsychotics than males, and physicians residing in the more populous counties prescribe fewer distinct antipsychotics than those in the least populous, although the marginal effect is not monotonic. While antipsychotic variety prescribing does not differ between solo and group practices, and between physicians opting or not opting out of data sharing agreements, both hospital-based and DO physicians prescribe antipsychotics with greater variety than do the non-hospital based or MD physicians, respectively.

Next we examine the concentration of physician prescribing as measured by the HHI of the physician's prescriptions among all antipsychotics. Initial visual data inspection suggested a log normal distribution of HHIs, censored from above at 10,000. We therefore estimate a Tobit model where the dependent variable is the log of overall antipsychotic HHI. Results are presented in the last column of Table 5; estimated magnitudes are the marginal effects evaluated at variable means. Each of the four specialty-volume interaction variables is negative and statistically significant, with the smallest negative volume effect on concentration being that for PSY, then for NEU and PCP, and largest for OTH

Table 5: Concentration of Antipsychotic Prescribing: Number of molecules, and HHI of Physician's Antipsychotic Prescribing (Marginal effects at Variable means)

	Total Number of Distinct Molecules	Log (Antipsychotic Prescription HHI)
OTH*Total Yearly Antipsychotic Prescriptions	0.00395***	-0.00102***
	(0.00021)	(0.00008)
PCP*Total Yearly Antipsychotic Prescriptions	0.00271***	-0.0009729***
	(0.00008)	(0.00004)
PSY*Total Yearly Antipsychotic Prescriptions	.00100***	-0.000195***
	(0.00002)	(0.00001)
NEU*Total Yearly Antipsychotic Prescriptions	0.00241***	-0.000532***
	(0.00022)	(0.00009)
Age Quartile 43-50^	0.111**	-0.00756
	(0.047)	(0.010)
Age Quartile 51-58^	0.163***	-0.0127
	(0.047)	(0.010)
Age Quartile 59+^	0.0290	0.0145
	(0.048)	(0.010)
PCP^	1.046***	-0.220***
	(0.056)	(0.011)
PSY^	4.252***	-0.632***
	(0.099)	(0.014)
NEU^	0.369***	-0.0294
	(0.116)	(0.021)
Female^	-0.385***	0.0786***
	(0.036)	(0.008)
Population 150,000-500,000 (county)^	-0.115**	0.0138
	(0.044)	(0.010)
Population 500,000-1,000,000 (county)^	-0.290***	0.0423***
	(0.046)	(0.010)
Population more than 1,000,000 (county)^	-0.242***	0.0281***
	(0.045)	(0.010)
Solo Practice^	-0.0526	0.0120
	(0.041)	(0.009)
Hospital Based Physician^	0.151***	-0.0144
	(0.058)	(0.013)
DO Flag^	0.110*	-0.0157
	(0.061)	(0.012)
Physician Opt Out^	0.0407	-0.0309*
	(0.086)	(0.019)
Pseudo R^2	*0.149*	*0.21*
Number of Observations	*15,037*	*15,037*
Mean of dependent variable	*4.41*	*8.37*
^ indicates dummy variable		*Rt censored obs = 1,146*

* , **, ***, indicate significance at the 10%, 5%, and 1% respectively. ^ dy/dx for a dummy variable represents effect of a discrete change from 0 to 1. All values calculated using IMS Health Incorporated Xponent™ general prescriber sample 2007 data, and population estimates from the US Census Bureau. Sample includes all physician prescribers that wrote at least 12 prescriptions for antipsychotics in 2007.

prescribers; the OTH negative volume effect is about five times larger than that for PSY physicians. Conditional on volume, PSY concentration is about 47% less ($e^{-0.632}$ = .53) than that for OTH, PCP concentration is 20% less, and NEU is 3% less (the last not significantly different from OTH). Among antipsychotics overall, there is no significant relationship between age quartile and concentration, nor is there any significant concentration impact of being in solo practice, hospital-based, or a DO rather than the reference group prescriber. Those practicing in counties having populations of at least 500,000 tended to be more concentrated overall than prescribers in the smallest counties (under 150,000), and those restricting data sharing were slightly but significantly less concentrated in their prescribing behavior.

3. Concentration Among Newer Drugs: Number of Distinct Atypical Drug Molecules Prescribed and Physician Atypical Antipsychotic HHI

As noted earlier, the first atypical antipsychotic was launched in the US in 1990, and between 1993 and 2007 five other new atypicals were approved by the FDA. These newer generation antipsychotics rapidly became dominant; from 1996 to 2007 the share of all antipsychotic prescriptions dispensed as atypicals in our prescriber sample increased from 27% to 85%. We now examine factors affecting the variety and concentration of prescribing behavior for the newer atypicals. Marginal effects (evaluated at sample means) on the number of distinct atypical molecules prescribed, based on Poisson model estimation, are given in the first set of columns in Table 6, while marginal effects on the log of atypical prescription HHIs, based on Tobit estimation, are presented in the final set of columns. Although the estimated marginal effects on atypical variety and concentration are in many cases quite similar to those on overall antipsychotic variety and concentration, several differences are worth noting.

Table 6: Concentration of Atypical Prescribing: Number of molecules, and HHI of Physician's Atypical prescribing (Marginal effects at Variable means)

	Total Number of Distinct Molecules	**Log (Antipsychotic Prescription HHI)**
OTH*Total Yearly Antipsychotic Prescriptions	0.00395***	-0.00102***
	(0.00021)	(0.00008)
PCP*Total Yearly Antipsychotic Prescriptions	0.00271***	-0.0009729***
	(0.00008)	(0.00004)
PSY*Total Yearly Antipsychotic Prescriptions	.00100***	-0.000195***
	(0.00002)	(0.00001)
NEU*Total Yearly Antipsychotic Prescriptions	0.00241***	-0.000532***
	(0.00022)	(0.00009)
Age Quartile 43-50^	0.111**	-0.00756
	(0.047)	(0.010)
Age Quartile 51-58^	0.163***	-0.0127
	(0.047)	(0.010)
Age Quartile 59+^	0.0290	0.0145
	(0.048)	(0.010)
PCP^	1.046***	-0.220***
	(0.056)	(0.011)
PSY^	4.252***	-0.632***
	(0.099)	(0.014)
NEU^	0.369***	-0.0294
	(0.116)	(0.021)
Female^	-0.385***	0.0786***
	(0.036)	(0.008)
Population 150,000-500,000 (county)^	-0.115**	0.0138
	(0.044)	(0.010)
Population 500,000-1,000,000 (county)^	-0.290***	0.0423***
	(0.046)	(0.010)
Population more than 1,000,000 (county)^	-0.242***	0.0281***
	(0.045)	(0.010)
Solo Practice^	-0.0526	0.0120
	(0.041)	(0.009)
Hospital Based Physician^	0.151***	-0.0144
	(0.058)	(0.013)
DO Flag^	0.110*	-0.0157
	(0.061)	(0.012)
Physician Opt Out^	0.0407	-0.0309*
	(0.086)	(0.019)
Pseudo R^2	*0.149*	*0.21*
Number of Observations	*15,037*	*15,037*
Mean of dependent variable	*4.41*	*8.37*
^ indicates dummy variable		*Rt censored obs = 1,146*

* , **, ***, indicate significance at the 10%, 5%, and 1% respectively. ^ dy/dx for a dummy variable represents effect of a discrete change from 0 to 1. All values calculated using IMS Health Incorporated Xponent™ general prescriber sample 2007 data, and population estimates from the US Census Bureau. Sample includes all physician prescribers that wrote at least 12 prescriptions for antipsychotics in 2007.

As was the case for antipsychotics overall, for the distinct number of atypicals prescribed the positive and significant volume effect is largest for OTH prescribers, followed by PCP and NEU physicians, with PSY prescribers experiencing the smallest volume impact; the volume effect is about seven times larger for OTH than for PSY prescribers. Conditional on antipsychotic volume, NEU physicians utilize a slighty but insignificantly smaller number of distinct atypicals, whereas PCPs and particularly PSYs employ a significantly greater atypical armamentarium variety. The effect of age quartile on variety differs for atypicals relative to antipsychotics overall. For atypicals, the only significant age quartile is the oldest, who, other things equal, prescribe a smaller number of distinct atypicals than do the youngest quartile; for antipsychotics overall, it was the two middle aged quartiles that prescribed a greater variety of distinct molecules. We return to consider this differential age result in greater detail later. As was the case for antipsychotics overall, other things equal, female physicians prescribe a smaller number of distinct atypicals, as do those in all but the smallest population counties (with the negative population effect again being non-monotonic); relative to the reference group, there is no significant difference in the number of distinct atypicals prescribed by those in solo practice, hospital-based practices, or by the opt-out physicians; DO physicians, however, utilize a larger number of distinct atypicals than do MDs.

As seen in the final columns of Table 6, the volume impact on atypical HHIs is significantly negative for all specialties, largest for OTHs, followed by PCPs and NEUs, and smallest for PSYs; this volume effect is about six times larger for OTH than for PSY prescribers. Conditional on volume, while PSY and then PCP physicians have the lowest atypical HHIs, NEU physicians have greater atypical concentration than do OTHs; for antipsychotics overall, the NEU impact was negative but insignificant. As they age, older physicians in the 51-58 become more concentrated in their atypical prescribing (borderline significance), but particularly the oldest over age 58 quartile that is approaching retirement evidence the most concentrated atypical prescribing behavior. For antipsychotics overall, none of the

age quartile impacts were significant. This difference in prescriber-age effects for the overall antipsychotic versus atypical only concentration suggests that the oldest physicians are disproportionate users of the older conventional antipsychotics. We discuss this phenomenon in more detail below. Although female and solo practice physicians have more concentrated atypical prescribing behavior, as do those practicing in counties with population over 500,000, hospital-based, DOs and opt-out physicians are significantly less concentrated in their atypical prescribing.

4. Other Results

We have undertaken a number of robustness checks, mostly involving the relationships between specialty and volume in the various deviation and concentration regressions. For example, to check whether our specialty-volume interaction term estimates in fact simply reflected a nonlinear, quadratic relationship in volume, we added quadratic volume interacted with specialty. While some of the additional terms were statistically significant, the estimate of the overall effect of volume on physician prescribing across specialties was essentially unchanged. We also limited the sample just to the 3,431 psychiatrists. In that model, while the volume term had a significant negative coefficient estimate, it again was very small; the significance of several estimates on other variables became insignificant, but our qualitative findings were essentially unchanged. Finally, our results are qualitatively similar if we use linear regressions instead of Poisson or Tobit regressions.

IV. **DISCUSSION AND CONCLUSIONS**

Before summarizing and concluding, we consider a competing model that attempts to explain physician prescribing behavior – that of detailing by sales representatives to physicians. We then relate our findings to various existing literatures.

A. EXPLORING A COMPETING HYPOTHESIS: PHYSICIAN DETAILING

There are several plausible competing hypotheses to ours concerning factors affecting physicians' prescribing behavior. One competing hypothesis consistent with the "motivation" rather than the "inspiration" class of explanations discussed earlier involves selection by pharmaceutical sales representatives ("detailers" who "detail" physicians) to high-volume prescribers. Suppose that, instead of high-volume prescribing generating greater physician prescribing heterogeneity through the logic of our learning-by-doing model, one hypothesized that high-volume prescribers are exposed to detailing by a greater number of different pharmaceutical manufacturers than are low-volume prescribers (because of the large returns potentially realized by pharmaceutical detailing when a high-volume prescriber is persuaded to prescribe a particular branded drug by a detailer). Either because some detailers provide persuasive information or because writing a few prescriptions for each detailed drug provided is a reciprocal form of behavior providing some positive feedback from the prescriber to the various detailers, in this competing hypothesis it is the increased detailing that leads to less concentrated prescribing by high-volume physicians, rather than the physician's learning-by-doing in response to larger patient volumes.[xxix]

In evaluating this plausible competing hypothesis, it is useful to note that drugs are detailed only when they are on patent or have market exclusivity for other reasons; after a branded drug faces generic competition, there are no incentives for its manufacturer to detail physicians, for the brand would be unable to appropriate many benefits, which for the most part would instead accrue to the generics.[xxx] An implication is that drugs having lost market exclusivity many years ago are unlikely to have been detailed to young doctors practicing in 2007, although older physicians in 2007 may have been detailed on them years ago, earlier in their career, or may have become familiar with them during their residency training when they were the only antipsychotics available on the market.

In order to compare the predictions of the competing hypothesis that physician detailing drives physician prescribing behavior to the predictions of our model, we separate antipsychotic drugs into

"old drugs" approved and launched in the US before 1990 (Clozaril and all the typical antipsychotics) and "new drugs" (all SGA atypical antipsychotics, the earliest of which was Risperdal, approved in 1993), and we compare the behavior of the oldest and youngest quartiles of physicians. The ten typical drugs prescribed by physicians in our 2007 sample were approved for marketing by the FDA between 1957 and 1984, while Clozaril, a FGA, was approved in 1989; they all experienced generic entry by 1996, many much earlier in the 1980s. An implication is that none of these old drugs was detailed after 1996.

The oldest quartile of physicians in our 2007 sample is comprised of physicians aged 59 and up, who in 1996 were age 48 and older. These physicians were almost surely all the way through their training and had been practicing for some time when the first SGA atypical, Risperdal, was approved in 1993 (when they were age 45 and older). In contrast, the youngest quartile of our 2007 sample is comprised of physicians from the age of 26 to 42, who in 1996 (when the last old drug experienced generic entry) were between the ages of 15 and 31; they are therefore unlikely to have ever been detailed on an old drug, and certainly would not have been detailed on them in 2007 or several years earlier. Moreover, they would have been between ages 12 and 28 when in 1993 the first SGA, Risperdal, was approved. Most of these youngest quartile physicians had either not yet enrolled in medical school or were still in their residencies after at least one of the SGAs was approved. While both the oldest and youngest quartile physicians may have been detailed on the new drugs in recent years, and both have not been detailed on old drugs in recent years, it is possible that the oldest physician cohort has some memories of being detailed on and/or actually prescribing the old drugs intensively earlier in their careers. We will compare how the oldest physicians and youngest physicians use of new drugs varies with their overall antipsychotic prescribing volume.

If pharmaceutical detailing influence were the primary driver of physicians' choice of which antipsychotic class of drugs to prescribe (old vs. new), then we would expect the youngest physicians to prescribe very few of the older drugs. In addition, we would expect high-volume young physicians (who

are likely visited the most by pharmaceutical detailers promoting new drugs) to be the least likely to prescribe older drugs. On the other hand, we would expect the oldest physicians to prescribe both new and old drugs, as these physicians were likely detailed on the older drugs before these drugs went off patent.

Returning to our learning-by-doing model, a prediction we obtain is that the share of new atypicals prescribed by young physicians should fall with volume for high enough volumes. For old physicians, however, the share of the new atypicals could increase or decrease with volume.[xxxi] The first of these results is the most important: in our framework, high-volume young physicians have an incentive to invest in learning the complementary actions for old drugs (the typical antipsychotics and Clozaril) because these drugs deliver the highest benefits for some (albeit a small minority of) patients. Moreover both old and young physicians with low volumes have insufficient incentive to invest in learning the complementary actions for some class of drugs, but for old physicians it is the new drug class about which they don't learn (because they learned about old drugs when they were the only ones available), whereas for new physicians it is most often the old drug class about which they don't learn (because their first set of patients had symptoms best treated by the new drug and so the physician prescribed the new drugs and learned about their complementary actions).

In order to evaluate these predictions empirically we examine the prescribing behavior of psychiatrists in our sample. Given their high volume, it is these physicians that are likely subject to the most visits by pharmaceutical sales representatives ("detailers") and hence are the physicians for whom we would most likely expect to observe the influence of detailing.

As the dependent variable we employ the psychiatrist's share of total antipsychotic prescriptions written for the new atypicals. If the detailing hypothesis were the primary driver of prescriber choice, for young physicians the new (old) share would be high (low) and would increase (decrease) with

volume, whereas for old physicians the same general pattern would be observed, except perhaps that as volume increased older physicians might be less inclined to increase their use of new drugs, given memories of their use and detailing of old drugs earlier in their careers, so the positive volume impact on new share would be smaller than for younger physicians. If instead the learning hypothesis were the primary driver of prescriber choice, for old physicians the share of new (old) drugs would be smaller (larger) but would increase (decrease) with volume. For young physicians, however, the share of new (old) drugs would be higher (lower), but would decline (increase) with volume.

The explanatory variables in the regression reported below are the same as those specified in previous analyses, while in order to maximize the age difference the 2007 sample is restricted to 1,844 psychiatrists in the oldest (age 59 and over) and youngest (age 26-42) physician age quartiles. Results from the regression are presented in Table 8.[xxxii]

Several findings are particularly notable. First, all else equal, older physicians prescribe a lower percentage of new drugs, consistent with both hypotheses. Second, however, since the estimated marginal effect on total yearly antipsychotic prescription volumes is negative and significant, the highest volume physicians in the youngest quartile prescribe a smaller share of new (larger share of old) drugs. This is consistent with our framework, but is at odds with the detailing hypothesis, for these youngest high volume prescribers are likely to have been heavily detailed on new drugs, but are likely never to have been detailed on the old drugs. Third, while higher volume physicians in the oldest age quartile also prescribe a smaller share of new drugs (larger share of old drugs), this marginal impact of volume is much smaller in absolute value (at -0.0049 + 0.0026 = -0.0023). As noted above, in our model this effect could have either sign. We conclude, therefore, that while the predictions of our learning-by-doing model are generally observed in the prescribing data, a crucial prediction of the detailing hypothesis is at odds with the prescribing behavior we observe among young physicians.

Table 8: Tobit Regression (Marginal Effects Estimated at Variable Means) on Percent of All Antipsychotic Prescriptions written for" New Drugs" in 2007

	dy/dx
Physician Age 59+^	-6.92***
Total Yearly Antipsychotic Prescriptions	-0.0049***
(Physician Age 59+)^(Total Yearly Antipsychotic Prescriptions)	0.0026*
Female^	3.23***
Population 150,000-500,000 (county)^	1.25
Population 500,000-1,000,000 (county)^	0.66
Population more than 1,000,000 (county)^	0.30
Solo Practice^	0.048
Hospital Based Physician^	-2.55*
DO Flag^	0.70
Physician Opt Out^	-5.71**
Number of Observations= 1,843	
Pseudo R^2= 0.0089	
Left Censored = 0 Right Censored = 440	
Mean of dependent variable	*88.29*
^ dy/dx is f or a discrete change of a dummy variable f rom 0 to 1	

* , **, ***, indicate significance at the 10%, 5%, and 1% respectively. ^ dy/dx for a dummy variable represents effect of a discrete change from 0 to 1. All values calculated using IMS Health Incorporated Xponent™ general prescriber sample 2007 data, and population estimates from the US Census Bureau. New drugs are defined as SGA atypicals. Sample is comprised of the Oldest (59 +) and youngest (26-42) quartile of psychiatrists.

B. RELATIONSHIP TO EXISTING LITERATURE

In the Introduction we noted that this research builds on the insights and empirical findings reported by Frank and Zeckhauser [2007] regarding primary care physician "ready-to-wear" vs. "custom-made" treatment of patients. Our empirical analyses have both a wider range of physician specialty prescribers yet also a more focused treatment choice – antipsychotic drugs as maintenance treatments for schizophrenia and related chronic conditions. These findings largely complement and extend those reported by Frank and Zeckhauser.

However, our empirical findings on regional disparities differ in large part from those reported by the Dartmouth Atlas project authors. As reported in Table 1, regional heterogeneity as measured by coefficients of variation considered to be high (above 0.4) occur in our antipsychotic concentration prescribing behavior only at the individual prescriber and county level of aggregation, but are low (less than 0.2) at the HRR and greater levels of geographical aggregation.[xxxiii] We conclude that the variability in antipsychotic behavior that we observe is at the level of the individual prescriber, and that this prescriber behavior is remarkably similar across HRRs and states, in contrast to much of the Dartmouth Atlas small variations literature.

We note that other researchers have recently reported that regional disparities are in some cases much less than the Medicare surgical and related procedures reported by the various Dartmouth Atlas collaborators. For example, Rettenmeier and Saving [2009] report that rankings of states on the basis of Medicare per enrollee health care spending differ substantially from those based on per capita health care spending on the non-Medicare/Medicaid population. Zhang, Baicker and Newhouse [2010] examined annual 2007 inpatient, outpatient and pharmaceutical spending for 533,170 beneficiaries simultaneously enrolled in Medicare Part A and B, and in stand-alone Part D Medicare plans. They find that across the 306 HRRs pharmaceutical spending varies less than medical spending.

C. SUMMARY AND DIRECTIONS FOR FUTURE RESEARCH

We develop a model in which a physician treats a sequence of patients with random symptoms. For each patient, the physician prescribes a drug and chooses a complementary action. The physician knows the maximum possible benefit from prescribing any drug for any symptom, but does not know the complementary actions that achieve this maximum benefit for any given drug. By prescribing a drug, choosing complementary actions, and observing the patient's response, the physician learns about the appropriate complementary actions for that drug. Thus, in our model, there is learning by doing, causing physicians to be more adept at choosing complementary actions for drugs they have prescribed

previously than for drugs they have not yet prescribed. On the other hand, knowing that some drugs are well suited for some symptoms, physicians may optimally prescribe an unfamiliar drug in response to a new symptom, especially if this and other symptoms that may be well addressed by this drug are likely to recur in future patients.

The main predictions of our model arise from considering differences in optimal prescribing behavior for physicians treating different volumes of patients. In particular, past volume influences the extent of learning by doing and hence a physician's ability to choose appropriate complementary actions for familiar drugs, whereas future volume influences the expected benefits to future patients from prescribing an unfamiliar drug for the current patient, so as to learn more about its appropriate complementary actions. High-volume physicians are thus more likely to prescribe a wide range of drugs and to prescribe them at rates closer to the national market shares. Low-volume physicians, in contrast, may optimally treat the patients they see by learning a great deal about appropriate complementary actions for a small subset of the available drugs and not prescribing drugs from outside this subset. Furthermore, the drugs in this subset depend on the random symptoms presented by the patients the physician treats early in her career. As a result, low-volume physicians may prescribe drugs at rates that differ greatly from national market shares. In short, concentration and deviation decrease with volume.

Our empirical analysis produced several findings. As predicted by our model, we observe that higher-volume physicians use a wider variety of drugs. This is true when we examine physicians' use of antipsychotics overall and when we focus on only their use of atypical antipsychotics. In addition, and also as predicted by our model, we find that higher-volume physicians prescribe drugs at rates that more closely mimic national market shares.

Moving from our model's main predictions about the effects of patient volume on prescriber behavior, we also considered the effects of differences in medical specialty training, which we model as differences in the initial uncertainty about a drug's appropriate complementary actions. We find that

volume matters most for those physicians who are primary care physicians (PCPs) or who are trained in specialties that do not typically prescribe antipsychotics in large volumes—i.e., the OTH prescribers or neurologists (NEU). In all analyses, volume matters also for PSYs, and in all regressions it matters considerably less for PSY physicians than for other prescribers. We interpret this smaller effect of volume for PSYs as medical specialty training being a substitute for experience gained from high-volume prescribing in developing knowledge concerning optimal complementary behavior for the various antipsychotic medications.[xxxiv] Another interpretation of this finding is that psychiatrists see a wider range of patients. We cannot differentiate between these two explanations without having data on patients, so our model takes the set of symptoms and the probabilities of their arrival as fixed, and our empirical work includes specialty dummies and interactions.

While this framework can help explain persistent heterogeneity in concentrated prescribing behaviors, and may thereby be consistent with the Frank and Zeckhauser "sensible use of norms" behavior, our theoretical framework entirely ignores learning from others, spillovers, and herding behavior. Chandra and Staiger [2007] have developed and estimated a model that focuses on productivity spillovers related to local specialization in heart attack care, whereby excellence in one clinical approach in a local market raises the average skill of other practitioners of that approach operating in the same market. This in turn leads to greater specialization and reduces both the absolute and relative productivity of practitioners using alternative approaches. Homogeneity in clinical approach within a geographic area, and substantial heterogeneity across areas, can reflect what may also be sensible and useful since they stem from positive spillover effects from local specialization. In future research, it would be useful to attempt to incorporate various types of spillover effects into physician prescribing behavior. This is particularly important, since learning from sources other than one's own prescribing behavior is a critical component in efforts to enhance the practice of evidence-based medicine.

A major limitation of this study is that we do not observe data on the patient populations treated by physicians. However, we note that the literature cited and results obtained by Frank and Zeckhauser [2007] suggest that, other than through demographics, variations in patient condition severity and clinical manifestations are remarkably unrelated to physician practice behavior, and that the results they obtained are largely quantitatively unaffected with alternative specifications incorporating patient-specific data. The dominant role of physicians over patients in influencing choice of medication has also been reported elsewhere, both by other health economists (e.g., Hellerstein [1997] and Zhang, Baicker and Newhouse [2010]) and by academic clinicians (e.g., Schneeweis, Glynn, Avorn and Solomon [2005]).

Several interesting future research projects have emerged from our study. As noted earlier, the relative efficacy, tolerability and cost-effectiveness of the various typical and atypical antipsychotics remains a controversial issue, even after publication of a substantial number of articles over the last decade, including those based on randomized clinical trials.[xxxv] What is less controversial is that this dispute has had a substantial impact on changing prescription shares of the various antipsychotics. Our IMS Health data reveal that between 2002 and 2008, the Seroquel prescription share increased from 21% to 37%, Abilify from 0% to 16%, Geodon from 4% to 7%, even as the Risperdal share declined from 35% to 26%, and that of Zyprexa from 34% to 12%. Who were the prescribers who switched most rapidly – low or high volume, what specialties, gender, age group – and who were those who changed relatively little? What were the relative responses to the FDA issuing bold boxed warnings, to professional associations revising treatment guidelines, to publication of major findings in medical journals? How well does our theoretical framework, implemented here in a cross-sectional context, predict dynamic behavior of physicians? Understanding which prescribers respond most and which the least would provide valuable information to guide future information dissemination strategies.[xxxvi]

Our findings suggest that a significant proportion of the heterogeneity in the treatments patients receive depends upon physician preferences in treatment regime. It would be informative and useful to identify specific patterns in physician decision-making that appear to indicate general differences in "style" across physician practices, perhaps related to location of medical residency training, analogous to recent investigations characterizing "management style".[xxxvii]

REFERENCES

American Psychiatric Association [2004], *Practice Guideline for the Treatment of Patients with Schizophrenia*, Second Edition, available online at http://www.psych.org.

Aoyagi, Masaki. [1998], "Mutual Observability and the Convergence of Actions in a Multi-Person Two-Armed Bandit Model", *Journal of Economic Theory* 82:405-24.

Banerjee, Abhijit V. [1992], "A Simple Model of Herd Behavior", *Quarterly Journal of Economics* 107(3):797-817.

Bergemann, Dirk and Juuso Valimaki. [2006], "Bandit Problems," in S. Durlauf, and L. Blume, eds.: The New Palgrave Dictionary of Economics (Macmillan Press, Basingstoke).

Berndt, Ernst R., Margaret K. Kyle and Davina C. Y Ling [2003], "The Long Shadow of Patent Expiration: Generic Entry and Rx to OTC Switches," ch. 8 in Robert C. Feenstra and Matthew D. Shapiro, eds., *Scanner Data and Price Indexes*, NBER Series on the Conference on Research in Income and Wealth, Vol. 61, Chicago: University of Chicago Press for the National Bureau of Economic Research, 229-267.

Bertrand, Marianne and Antoinette Schoar [2003], "Managing with Style: The Effect of Managers on Firm Policies", *Quarterly Journal of Economics* 118(4):1169-1208.

Bikhchandani, Sushil, David Hirschleifer and Ivo Welch [1992], "A Theory of Fads, Fashion, Custom, and Cultural Exchange as Information Cascades", *Journal of Political Economy* 100(5):992-1026.

Chandra, Amitabh and Douglas L. Staiger [2007], "Productivity Spillovers in Health Care: Evidence from the Treatment of Heart Attacks", *Journal of Political Economy* 115(1):103-41.

Cipher, Daisha J. and Roderick S. Hooker [2006], "Prescribing Trends by Nurse Practitioners and Physician Assistants in the United States", *Journal of the American Academy of Nurse Practitioners* 18:291-6.

Coscelli, Andrea [2000], "The Importance of Doctors' and Patients' Preferences in the Prescription Decision", *Journal of Industrial Economics* 48:349-69.

Coscelli, Andrea and Matthew Shum [2004], "An Empirical Model of Learning and Patient Spillovers in New Drug Entry", *Journal of Econometrics* 122:213-46.

Crawford, Gregory and Matthew Shum [2005], "Uncertainty and Learning in Pharmaceutical Demand", *Econometrica* 73:1137-73, July.

Dartmouth Atlas Project. *The Dartmouth Atlas of Health Care.* Available online at http://www.dartmouthatlas.org.

Dickstein, Michael J. [2011], "Efficient Provision of Experience Goods: Evidence from Antidepressant Choice", Harvard, unpublished.

Domino, Marisa Elena, Edward C. Norton, Joseph P. Morrissey and Neil Thakur [2004], "Cost Shifting to Jails after a Change to Managed Mental Health Care", *Health Services Research* 39(5):1379-1401.

Doyle, Joseph J., Steven Ewer and Todd Wagner [2008], "Returns to Physician Human Capital: Analyzing Patients Randomized to Physician Teams", Cambridge, MA: National Bureau of Economic Research, Working Paper No. 14174, July.

Duggan, Mark [2005], "Do New Prescription Drugs Pay for Themselves? The Case of Second Generation Antipsychotics", *Journal of Health Economics* 24:1-31.

Ellison, Glenn D. and Edward L. Glaeser [1997], "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy* 105(5):889-927.

Ferreyra, Maria Marta and Grigory Kosenok [2010], "Learning about New Products: An Empirical Study of Physicians' Behavior", *Economic Inquiry*, forthcoming.

Fisher, Elliot S., John E. Wennberg, Therese A. Stukel et al. [2003a], "The Implications of Regional Variations in Medicare Spending, Part 1: The Content, Quality, and Accessibility of Care", *Annals of Internal Medicine* 138(4):273-87.

Fisher, Elliot S., John E. Wennberg, Therese A. Stukel et al. [2003b], "The Implications of Regional Variations in Medicare Spending, Part 2: Health Outcomes and Satisfaction with Care", *Annals of Internal Medicine* 138(4):299-98.

Frank, Richard G., Ernst R. Berndt, Alisa B. Busch and Anthony F. Lehman [2004], "Quality-Constant Price Indexes for the Ongoing Treatment of Schizophrenia: An Exploratory Study", *Quarterly Review of Economics and Finance*, 44(3):390-409, July.

Frank, Richard G. and Sherry A. Glied [2003], *Better but Not Well: Mental Health Policy in the U.S. since 1950*. Baltimore: Johns Hopkins University Press.

Frank, Richard G. and Richard J. Zeckhauser [2007], "Custom-made versus ready-to-wear treatments: Behavioral propensities in physicians' choices", *Journal of Health Economics* 26(6):1101-227, December.

Gawande, Anul 2009], "The cost conundrum", *New Yorker* (serial on the Internet), June 1. Available online from http://www.newyorker.com/reporting/2009/06/01fa_fact_gawande.

Gibbons, Robert S. and Rebecca M. Henderson [2011], "Relational Contracts and Organizational Capabilities," forthcoming in *Organization Science*.

Hellerstein, Judith [1997], "The Importance of the Physician in the Generic versus Trade-Name Decision", *RAND Journal of Economics* 27:108-36.

Hoblyn, Jennifer, Art Noda, Jerome A. Yesavage et al. [2006], "Factors in Choosing Atypical Antipsychotics: Toward Understanding the Bases of Physicians' Prescribing Decisions", *Journal of Psychiatric Research* 40:160-6.

Hooker, Roderick S. and Daisha J. Cipher [2005], "Physician Assistant and Nurse Practitioner Prescribing: 1997-2002", *Journal of Rural Health* 21(4):355-60, Fall.

Huskamp,, Haiden A. [2003], "Managing Psychotropic Drug Costs: Will Formularies Work?, *Health Affairs* 22(5):84-96, September/October.

Jovanovic, Boyan and Yaw Nyarko [1996], "Learning by Doing and the Choice of Technology", *Econometrica* 64(6):1299-310.

Kaplan, Steven N., Mark M. Klebanov and Morten Sorensen [2008], "Which CEO Characteristics and Abilities Matter?", Chicago: University of Chicago Booth School of Management, unpublished, available from skaplan@uchicago.edu.

Kraemer, Helena Chmura, Ira D. Glick and Donald F. Klein [2009], "Clinical Trials Design Lessons From the CATIE Study", *American Journal of Psychiatry*, 166(11):1222-8, November.

Lieberman, Jeffrey A., T. Scott Stroup, Joseph P. McEvoy et al. [2005], "Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia: Primary Efficacy and Safety Outcomes of the Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) Schizophrenia Trial, *New England Journal of Medicine* 353:1209-23.

Malmendier, Ulrike and Klaus M. Schmidt [2011], "You Owe Me", unpublished working paper, Berkeley CA: University of California at Berkeley, Department of Economics, June 15. Available from ulrike@econ.berkeley.edu.

Morgan, Perri A. and Roderick S. Hooker [2010], "Choice of Specialties Among Physician Assistants in the United States", *Health Affairs* 29(5):887-92, May.

Mosby's Medical, Nursing, & Allied Health Dictionary [1998], ed. Kenneth N. Anderson, Fifth Edition, St. Louis: Mosby-Year Book, Inc.

Pauly, Mark V. [1980], *Doctors and Their Workshops*, Chicago: University of Chicago Press.

Phelps, Charles E. [2000], "Information Diffusion and Best Practice Adoption", ch. 5 in Anthony J. Culyer and Joseph P. Newhouse, eds., *Handbook of Health Economics*, Vol. 1A, Amsterdam: Elsevier Science B.V., 223-64.

Phelps, Charles E. [1992], "Diffusion of Information in Medical Care", *Journal of Economic Perspectives* 6(3):23-42, Summer.

Phelps, Charles E. and Cathleen Mooney [1993], "Variations in Medical Practice Use: Causes and Consequences", ch. 7 in Richard J. Arnould, Robert F. Rich and William D. White, eds., *Competitive Approaches to Health Care Reform*, Washington DC: The Urban Institute Press, pp. 139-78.

Pohl, Joanne M., Charlene Hanson, Jamesetta A. Newland and Linda Cronenwett [2010], "Unleashing Nurse Practitioners' Potential to Deliver Primary Care and Lead Teams", *Health Affairs* 29(5):900-5, May.

Rettenmeier, Andrew J. and Thomas R. Saving [2009], "Perspectives on the Geographic Variation in Health Care Spending", College Station, TX: Private Enterprise Research Center, Texas A&M University, Working Paper #0911, December.

Schneeweiss, Sebastian, Robert J. Glynn, Jerry Avorn and Daniel H. Solomon [2005], "A Medicare database review found that physician preferences increasingly outweighed patient characteristics as determinants of first-time prescriptions for COX-2 inhibitors", *Journal of Clinical Epidemiology*, Brief Reports, 58:98-102.

Shell, Renee C. [2001], "Antidepressant Prescribing Practices of Nurse Practitioners", *The Nurse Practitioner* 26(7):42-7, July.

Skinner, Jonathan and Elliott Fisher [1997], "Regional Disparities in Medicare Expenditures: An Opportunity for Reform", *National Tax Journal* 50(3), September.

Stern, Scott M. and Manuel Trajtenberg [1998], "Empirical Implications of Physician Authority in Pharmaceutical Decisionmaking", Cambridge, MA: National Bureau of Economic Research, Working Paper No. 6851, December.

Taub, Anna A. Levine, Anton Kolotilin, Robert S. Gibbons and Ernst R. Berndt [2011], "The Diversity of Concentrated Prescribing Behavior: An Application to Antipsychotics", Cambridge, MA: National Bureau of Economic Research, Working Paper No. 16823, February.

Wennberg, John E., Elliott S. Fisher and Jonathan S. Skinner [2002], "Geography and the Debate Over Medicare Reform", *Health Affairs Web Exclusive*, 13 February, W98-W114.

Wennberg, John E., Elliott S. Fisher, Jonathan S. Skinner and Kristen K. Bronner [2007], "Extending the P4P Agenda, Part 2: How Medicare Can Reduce Waste and Improve the Care of the Chronically Ill", *Health Affairs* 26(6):1575-85, November/December.

Wennberg, John E., J. L. Freeman and W. J. Culp [1987]. "Are Hospital Services Rationed in New Haven or Over-Utilised in Boston? *Lancet* 1:1185-8, May 23.

White, Randall F. [2006], "The Significance of CATIE: An Expert Interview with Jeffrey A. Lieberman, MD", *Medscape Psychiatry & Mental Health*, online posted March 15, 2006.

Yasaitis, Laura, Elliott S. Fisher, Jonathan S. Skinner and Amitabh Chandra [2009], "Hospital Quality and Intensity of Spending: Is There An Association?", *Health Affairs Web Exclusive*, 28(4):W566-W572, 21 May.

Zhang, Yuting, Katherine Baicker and Joseph P. Newhouse [2010], "Geographic Variation in Medicare Drug Spending", *New England Journal of Medicine* 363(5):405-9, July 29.

Appendix A: A *2 × 2 EXAMPLE*

To obtain more precise comparative-static results (and to illustrate the logic of the model more generally), consider a simple example that satisfies the following assumption:

Assumption 1: $e^{-rw}=\delta$, $\boldsymbol{S}=\boldsymbol{D}=\{1,2\}$, $Pr(s_2)=p_2>1/2$, $\sigma_1^2=\sigma_2^2=c>1$, $\sigma_\varepsilon^2=0$, $B_{12}=B_{21}=0$, $B_{11}=B_{22}=1$.

A verbal interpretation of Assumption 1 is the following. We define δ as $\delta=e^{-rw}$. Therefore, a higher value of δ corresponds to a physician who has a shorter waiting time between patients and hence sees a higher volume of patients. There are two drugs $d=1$ and $d=2$, and two symptoms $s=1$ and $s=2$. Symptoms $s=2$ and $s=1$ are realized with probabilities p_2 and $p_1=1-p_2$, respectively. Symptom $s=2$ occurs more often than symptom $s=1$ (i.e., $p_2>1/2$). Therefore, drug $d=2$ is more likely to be ideal for a randomly drawn symptom. In all other respects, drugs and symptoms are symmetric (i.e., $B_{11}=B_{22}$, $B_{12}=B_{21}$, and $\sigma_1^2=\sigma_2^2$).

Before seeing any patients, the physician has the same uncertainty about the ideal complementary action for each drug ϑ_d (i.e., $\sigma_1^2=\sigma_2^2=\sigma^2>0$). However, the physician learns the ideal complementary action precisely after one prescription (i.e., $\sigma_\varepsilon^2=0$). As discussed in Section D of the main text, this learning assumption implies that the physician incurs a fixed cost $c=\sigma^2$ when she prescribes drug d for the first time, and thereafter she incurs no cost when she prescribes drug d.

The ideal drugs for given symptoms are normalized in such a way that $d^*(1)=1$ and $d^*(2)=2$ (i.e., $B_{11},B_{22}>0$). Without loss of generality, we can normalize $B_{12}=B_{21}=0$ because only the relative benefits $B_{22}-B_{21}$ and $B_{11}-B_{12}$ matter for the physician's choice of drug d. Likewise, without loss of generality we can jointly rescale B_{11}, B_{22}, and σ^2 so that $B_{11}=B_{22}=1$. Finally, to make the analysis interesting, we assume that the myopic physician concentrates on the drug prescribed to the first patient (i.e., $\sigma^2>B_{11}-B_{12}=B_{22}-B_{21}=1$).

In Proposition 1, we fully characterize the physician's optimal prescribing behavior under Assumption 1. Figure A1 illustrates different cases that arise in the model depending on parameter values. The explicit formulas for the boundaries of different regions of Figure A1 are given in the proof of Proposition 1 in an appendix available from the lead author.

Proposition 1 *Let Assumption 1 hold. There are six different cases that can arise in the model that correspond to the combination of a color (green, yellow, red) and a shade (light, dark) shown in Figure A1. (The dark red area is empty iff c≥2.)*

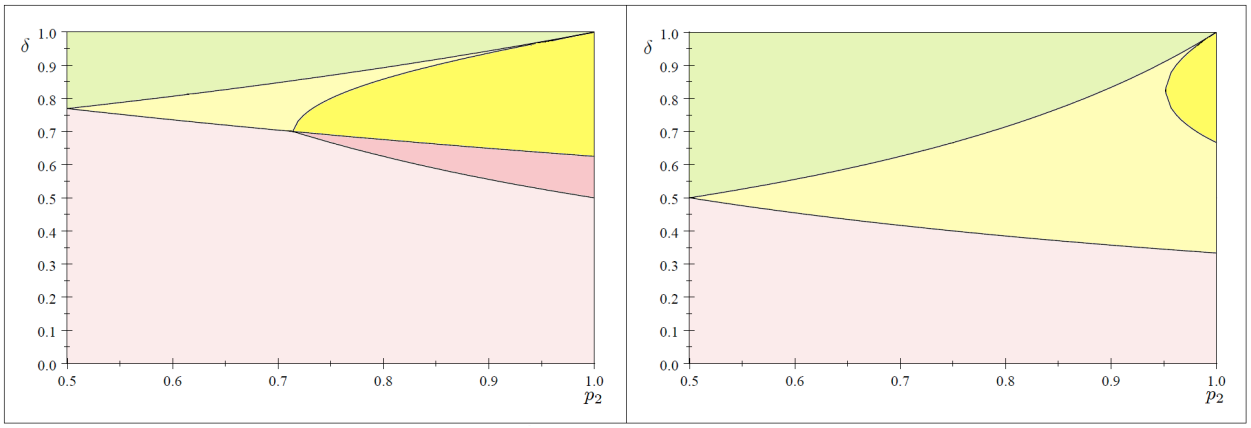
In the first period, the physician prescribes:

- *the ideal drug in the light color areas;*
- *the drug d=2 in the dark color areas.*

Starting from the second period the physician prescribes:

- *the ideal drug in the green area;*
- *the ideal drug or the drug d=2 depending on whether d=1 or d=2 was prescribed in the first period, respectively, in the yellow areas;*
- *the drug prescribed in the first period in the red areas.*

Figure A1. Left panel: *c=8/3>2*; Right panel: *c=3/2<2*.



To provide intuition for Proposition 1, we explain color and shade regions of Figure A1 in turn. We begin by explaining different colors in Figure A1. A low-volume physician (red area) never experiments. She always concentrates on the drugs prescribed in the past. An intermediate-volume physician (yellow area) is willing to experiment and prescribe a new drug only if this new drug is more likely to be the ideal drug than the drug she prescribed in the past. As the probability that the new drug is ideal increases, a physician has higher incentives to experiment with the new drug. This corresponds to the decreasing boundary between the red and yellow areas on Figure Aa1. A high-volume physician (green area) is always willing to experiment and prescribe a new drug. As the probability that the new drug is ideal decreases, a physician has lower incentives to experiment with the new drug. This corresponds to the increasing boundary between the yellow and green areas on Figure A1.

We now explain shades (light and dark areas) in Figure A1. Shades determine what drug a physician prescribes at the beginning of her career. In light areas, an inexperienced physician prescribes the ideal drug (drug $d=s$ for symptom s), whereas in dark areas she prescribes the more popular drug (drug $d=2$) regardless of symptoms. Note that in dark areas, the inexperienced physician prescribes the more popular drug even though this drug may be suboptimal for the patient. This occurs because the inexperienced physician expects the more popular drug to be optimal for most future patients, so she invests in learning how to use this drug at the beginning of her career. Note that in the dark yellow area the physician concentrates on the most popular drug her entire career. However, she would diversify and always prescribe the ideal drug in the long run if she were forced to prescribe the less popular drug at the beginning of her career.

Finally, we explain why a physician prescribes the more popular drug at the beginning of her career only if she sees an intermediate volume of patients and the more popular drug is very likely to be ideal (i.e., why the dark area occurs at intermediate values of δ and high values of p_2). A low-volume

physician prescribes the ideal drug because she is not willing to invest in learning any drug (e.g., as volume goes to zero, the physician becomes myopic and so does what is best for the current patient). In contrast, a high-volume physician prescribes the ideal drug because she is willing to invest in learning complementary actions for both drugs. Therefore, only an intermediate-volume physician can invest in learning only the more popular drug. The intermediate-volume physician invests in learning only about the more popular drug only if this more popular drug is very likely to be ideal in the future.

COMPARING COHORTS OF PHYSICIANS AND ERAS OF DRUGS

We now use this 2x2 example to build intuition for what our model predicts about the prescriptions of typical versus atypical antipsychotics by old versus young physicians. Specifically, consider the following sequence of eras denoted T = 1, 2, and 3: at T = 1, a cohort of "old" physicians is trained and has access to only typical antipsychotics; at T = 2, a cohort of "young" physicians is trained (and the "old" continue to practice) and all physicians have access to both typical and atypical drugs; finally, at T = 3, both cohorts are practicing and have access to both kinds of drugs. We will view T = 3 as 2007, the year of our data. We now explore what the 2x2 example predicts about prescriptions in T = 3.

In T = 1, there are two possible symptoms (s_1 and s_2), a cohort of physicians beginning their prescribing careers (hereafter, "old physicians"), and only one drug available (which we will interpret as a typical antipsychotic and label as d_1). For these old physicians during T = 1, all they can do is prescribe d_1, so they do so for all symptoms (s_1 and s_2). As a result, because Assumption 1 implies full learning after one prescription, these old physicians know perfectly how to take complementary actions for d_1 in the future.

In T = 2, another drug becomes available (which we will interpret as an atypical antipsychotic and label as d_2) and a new cohort of physicians begin their prescribing careers (hereafter, "new physicians"). Both old and new physicians know that drug d_i is the best prescription for symptom s_i, in the sense that

this prescription maximizes B_{sd}. The only difference between the new and old physicians is that the new physicians do not yet know how to take complementary actions for either drug (d_1 or d_2), whereas the old physicians do know how to do this for the typical (d_1) but not for the atypical (d_2).

Because the market share of atypicals relative to typicals is very large (much greater than 50%) in 2007, we assume that Prob(s_2) = $p_2 > ½$, again in keeping with Assumption 1. For example, if we set the market share of atypicals at about 6/7, then p_2 is 6/7. If we then proceed upwards in Figure A1along a vertical line at $p_2 = 6/7$, we are comparing physicians with different volumes.

Recall that old and new physicians have different histories at T = 3. For new physicians, T = 3 is their second period, so their prescription at T = 3 depends on their history at T = 2. For old physicians, T = 3 is their third period, so their prescription at T = 3 depends on their history at T = 1 and the fact that the new drug arrived at T = 2. Designating (x, y) to mean that a physician is prescribing fraction x of d_1 and fraction y of d_2, where x + y = 1, we then have the following prescription behaviors as a function of the colored and shaded regions in Figure A1.

	Old physicians	**New physicians**
Light red	all are (1, 0)	$1-p_2$ are (1, 0); p_2 are (0, 1)
Dark red	all are (1, 0)	all are (0, 1)
Dark yellow	all are ($1-p_2$, p_2)	all are (0, 1)
Light yellow	all are ($1-p_2$, p_2)	$1-p_2$ are ($1-p_2$, p_2); p_2 are (0, 1)
Light green	all are ($1-p_2$, p_2)	all are ($1-p_2$, p_2)

For old physicians, concentration falls with volume, the number of atypicals increases with volume, and the share of atypicals increases with volume. For new physicians, concentration falls with volume, the number of atypicals weakly increases with volume, and the share of atypicals falls with volume for sufficiently high volumes. The last of these results is the most important: high-volume young physicians have an incentive to invest in learning the complementary actions for old drugs (typical antipsychotics) because these drugs deliver the highest benefits for some (albeit a small minority) of patients. Alternatively, viewing the table from the opposite perspective, both old and young physicians with low volumes have insufficient incentive to invest in learning the complementary actions for a drug, but for old physicians it is the new drug about which they don't learn (because they learned about the old drug when it was the only one available), whereas for new physicians it is most often the old drug about which they don't learn (because their first patient had symptom s_2 and so the physician prescribed d_2 and learned about its complementary actions).

ENDNOTES

[i] These are the terms used by Frank and Zeckhauser [2007].
[ii] See, for example, Skinner and Fisher [1997], Fisher, Wennberg, Stukel et al. [2003a], and Yasaitis, Fisher, Skinner et al. [2009].
[iii] Also see Coscelli [2000] and Coscelli and Shum [2004].
[iv] Frank-Zeckhauser [2007], pp. 1125-6.
[v] It also builds on the framework outlined by Phelps and Mooney [1993] and Phelps [1992,2000].
[vi] Mosby's Medical, Nursing, & Allied Health Dictionary [1998], p. 1456.
[vii] Domino, Norton, Morrissey and Thakur [2004].
[viii] American Psychiatric Association [2004], p. 9.
[ix] Duggan [2005].
[x] Frank, Berndt, Busch and Lehman [2004].
[xi] American Psychiatric Association [2004], p. 66.
[xii] Additional controversy emerged when major studies, published in 2005 and 2006, raised issues regarding whether there were any significant efficacy and tolerability differences between the costly SGAs and the older off-patent conventional antipsychotics, as well as differences among the five SGAs. Important issues regarding the statistical power of these studies to detect differences, were they present, have also been raised, and currently whether there are any significant differences among and between the conventional and SGA antipsychotics remains controversial and unresolved. For further details and references, see the Appendix available from the lead author, "Timelines – U.S. Food and Drug Administration Approvals and Indications, and Significant Events Concerning Antipsychotic Drugs".
[xiii] Although at times we will use the words "prescribed", "written" and "dispensed" interchangeably, the IMS Health Xponent data are based on dispensed prescriptions; for a variety of reasons, a physician can prescribe a Product X but it may not be dispensed at all, or in fact after consulting with the prescriber the pharmacist may dispense product Y.
[xiv] The 75% cutoff is a stringent one, for the patient population seen by a prescriber is likely somewhat heterogeneous, with some patients having failed to respond to various medicines (perhaps including the physician's favorite one) and others having a history of effective response to another drug. For the latter, a physician may be reluctant to switch from an effective drug to the physician's favorite one, given the medical profession's adage "Don't shoot a singing bird."
[xv] See, for example, Skinner and Fisher [1997], Fisher, Wennberg, Stukel et al. [2003a,b] and Yasaitis, Fisher, Skinner et al. [2009].
[xvi] The U.S. Department of Justice horizontal merger guidelines state that when a merger results in a change in the HHI of more than 100 points and with the merged firm generating a post-merger industry HHI of > 1800, the merger will be presumed to create or enhance market power or facilitate its exercise, and will likely be very closely scrutinized by the Department of Justice, and perhaps even challenged. The merger guidelines can be accessed online at http://www.justice.gov/atr/public/guidelines/horiz_book/15.html.
[xvii] For example, if a prescriber only used three of the atypicals with prescription shares of 65%, 25% and 10%, the HHI would be 4950; if however, all six were used equally (each 16.67%), the HHI would be 1667.33.,
[xviii] We thank Jan Rivkin for teaching us these "4 'tions," which we adapt here for our own purposes.
[xix] An early discussion of these principal-agent issues is found in Pauly [1980], albeit in the context of hospital treatments, not pharmaceuticals.

[xx] For discussion, see Frank and Glied [2006] and Huskamp [2003].

[xxi] More specifically, Crawford and Shum (2005) and Dickstein (2011) use patient-level data, so they can analyze a patient's learning but not a prescriber's concentration. In contrast, Ferreyra and Kosenok (2009) share our focus on prescriber learning and analyze prescriber data, but they focus on learning to prescribe a single new drug, rather than on steady-state concentration of prescriptions.

[xxii] Many states have licensed nurse practitioners and certain physician assistants to write prescriptions, under varying physician supervision provisions. In the current context of antipsychotic drugs, it is worth noting that in one survey of nurse practitioners, almost one-third of patients they treated were seen for mental health problems. For further details, see, for example, Cipher and Hooker [2006], Hooker and Cipher [2005], Morgan and Hooker [2010], Pohl, Hanson, Newland and Cronenwett [2010] and Shell [2001]. Notably, in preliminary data analyses examining relative antipsychotic prescribing by specialty, nurse practitioners were the fourth largest specialty, comprising 20,872 of the 224,259 (9.3%) prescribers in the top eleven specialties.

[xxiii] In addition to excluding the 1,376 non-physician prescribers, we dropped 205 observations for which county codes were missing, three with missing gender information, and two observations for which age information was an unreasonable outlier. In an earlier version of this manuscript (Taub, Kolotilin, Gibbons and Berndt [2011], we included in our analyses among the typical antipsychotics an old drug named Prochlorperazine (Compazine), a drug that was FDA approved both for treatment of schizophrenia and for nausea. Since its primary use has been for nausea, and since the branded version has now been withdrawn from the US market, we exclude that drug from our set of antipsychotics. For a substantial number of primarily OTH prescribers, this was the only antipsychotic prescribed, and then in very small numbers. When this drug was excluded from the analyses, we were left with a total of 15,037 physicians.

[xxiv] HRRs represent regional health care markets for tertiary medical care that generally requires the services of a major referral center, primarily for major cardiovascular surgery procedures and neurosurgery; HRRs have been developed by and are maintained by the Dartmouth Atlas Project. HRRs may cross state and county borders because they are determined solelyby migration paterns of patients. For further details, see Dartmouth Atlas Project, http://www.dartmouthatlas.org.

[xxv] In a Physician Sample appendix, available from the lead author, we discuss this latter point in more detail.

[xxvi] DO is doctor of osteopathy. Mosby's Medical Dictionary [1998, p. 1169] defines osteopathy as "a therapeutic approach to the practice of medicine that uses all the usual forms of medical diagnosis and therapy, including drugs, surgery, and radiation, but that places greater emphasis on the influence of the relationship between the organs and the musculoskeletal system than traditional medicine does. Osteopathic physicians recognize and correct structural problems using manipulation."

[xxvii] Alternative measures of concentration that account explicitly for inherently "noisy" concentration from low volume prescribers have been developed by Ellison and Glaeser [1997] in the context of the geographic concentration of manufacturing industries, and adapted to the context of prescription pharmaceuticals by Stern and Trajtenberg [1998]. Our deviation measure is closely related to the Ellison-Glaeser concept.

[xxviii] Results are very similar when the prescriber deviation measure is relative to HRR or state market shares, rather than national market shares.

[xxix] For a model of reciprocal behavior in response to gift giving and experimental evidence, see Malmendier and Schmidt [2011].

[xxx] For discussion and empirical evidence, see Berndt, Kyle and Ling [2003].

[xxxi] For high-volume oldest physicians, there are two effects: because they have high future volume, they have incentives to learn new drugs as they are introduced, but because they have high past volume, they have substantial experience with existing drugs and so optimally continue to prescribe them.

[xxxii] The results are qualitatively similar if we use the full sample instead of just the PSY subsample.

[xxxiii] To examine regional disparities in greater detail, we estimated regressions with the prescriber's total 2007 volume of antipsychotic prescriptions (or its logarithm) as the dependent variable and the set of non-volume explanatory variables specified in the tabled results as explanatory variables; we then estimate this equation with and without HRR fixed effects added, and examine how much incremental explanatory power is provided by the HRR fixed effects. Although we find the 305 fixed effects are jointly highly significant, their incremental contribution to goodness of fit is *de minimus*. Specifically, for the volume levels regression, addition of 305 fixed effects increases the R^2 from 0.2622 to 0.2747 (a 4.8% proportional increase), while for the log volume specification the increase is from 0.3713 to 0.3855 (a 3.8% proportional increase). Similar findings emerged when we added fixed effects to the regression models whose results were reported in Tables 7 and 8 – while the fixed effects are jointly significant, they have little effects on the point estimates or statistical significance reported earlier.

[xxxiv] In a related finding, Doyle, Ewer and Wagner [2008] find that while treating randomized patients at the same hospital, medical residents from a lower-ranked medical school were able to substitute diagnostic tests and specialist consultation for the more rapid judgments made by residents from a higher-ranked medical school, achieving the same outcomes on average but at a higher cost.

[xxxv] An appendix to this paper available from the lead author, "Timelines Appendix" provides further details. Among the more notable publications are those based on the CATIE study; see, for example, Lieberman, Stroup, McEvoy et al. [2005], White [2006] and Kraemer, Glick and Klein [2009].

[xxxvi] The only research on this topic of which we are aware is that by Hoblyn, Noda, Yesavage et al. [2006].

[xxxvii] See, for example, Bertrand-Schoar [2003] and Kaplan, Klebanov and Sorensen [2008].