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

Anna Levine Taub* Anton Kolotilin† Robert S. Gibbons‡
Ernst R. Berndt§

This Version: November, 2012

Abstract

Physicians prescribing drugs for patients with schizophrenia and related conditions are remarkably concentrated in their choice among antipsychotic drugs. We construct a model of physician learning-by-doing that generates several hypotheses amenable to empirical analyses. Using 2007 annual antipsychotic prescribing data from IMS Health on 15,037 physicians, we examine these predictions empirically. While prescribing behavior is generally concentrated, we find that, consistent with our model, prescribers having greater prescription volumes have less concentrated prescribing. Our model outperforms a competing theory concerning detailing by pharmaceutical representatives, and we provide a new correction for the mechanical bias present in other estimators used in the literature.

JEL Classification: I10; I11; D80; D83

Keywords: Antipsychotic, Pharmaceutical, Concentration, Learning, Prescription, Physician

*Cornerstone Research

†University of New South Wales, School of Economics

‡Sloan Distinguished Professor of Management, MIT Sloan School of Management, and NBER

§Louis E. Seley Professor in Applied Economics, MIT Sloan School of Management, and NBER

1 Introduction

1.1 Motivation and overview

Consider a physician seeing a patient with a confirmed diagnosis for which several alternative pharmaceutical treatments are available. Suppose that, given the clinical evidence, patient response to a given treatment is idiosyncratic and unpredictable in terms of both efficacy and side effects. What treatment algorithms might the physician employ to learn about the efficacy and tolerability of the alternative drug therapies for this and future similar patients?

One possibility is for the physician to concentrate her prescribing behavior—in the extreme, on just one drug. By observing this and future patients' responses to that drug, the physician can learn by doing, thereafter exploiting her accumulated knowledge about this drug. For example, the physician will learn how to counsel patients on the efficacy and side-effect responses they might experience, possible interactions with other drugs, and the best time of day to take the drug; in addition, she will learn how to adjust the dosage depending on patients' factors such as smoking behavior, thereby improving patient outcomes and engaging the patient in adherence and symptom remission.

Alternatively, the physician might diversify her prescriptions across several drugs, hoping to find the best match between different drugs and current and future similar patients. Specifically, based on information from a patient's history, familiarity with 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. Physicians continually face this tradeoff as they treat patients and invest in learning about available treatments. In this paper, we develop and test a model of physician learning by doing that addresses these issues.

Our theory predicts how different physicians locate along this concentration-diversification continuum. We also analyze whether physicians with concentrated prescriptions will converge (exhibiting near unanimity on the choice of a favorite drug) or diverge (with different physicians concentrating on different drugs). Our model predicts that path-dependence in learning by doing is a strong force towards the latter. In addition, our model predicts how

different young physicians will utilize older ("off-label") drugs. Finally, we use our model to guide our econometric specification.

We confront our model with data on a particular therapeutic class of drugs known as antipsychotics. Later in this Introduction, we provide a brief background on the history of antipsychotic drugs and the illnesses they treat. We also report preliminary evidence of heterogeneous concentration in prescribing behavior: a typical physician focuses disproportionately on one drug, but there is substantial heterogeneity across prescribers concerning their most-used drug.

These initial findings on heterogeneous concentration are consistent with our theoretical framework (emphasizing path dependence in learning by doing), from which we advance several novel hypotheses. We then discuss the data and econometric framework, including a new correction for the mechanical bias present in other estimators used in the literature, and present a substantial set of empirical findings that broadly accord with our model. We conclude by explaining why our model outperforms a competing theory (emphasizing detailing by pharmaceutical representatives), relating our findings to the geographical-variation literature, and suggesting directions for future research.

The issues in this paper are important: understanding factors affecting physicians' choices along the concentration-diversification continuum has significant commercial and public-health implications, particularly in the current context of promoting both the evidence-based and "personalized" practice of medicine. Perhaps not surprisingly, therefore, some of the issues we explore have been discussed by others. For example, Coscelli (2000), Coscelli and Shum (2004), and Frank and Zeckhauser (2007) considered concentrated prescribing behavior. Coscelli does not use a formal model, Coscelli and Shum use a learning model that would be inconsistent with several of our findings, and Frank and Zeckhauser offer a very different model that again does not fit with some of our results.[1] Turning from physicians to patients, Crawford and Shum (2005) and Dickstein (2012) have studied a problem complementary to ours: how a given patient's treatment regime evolves over time. In short, our model studies learning across patients, whereas these latter models study learning within patients. We can imagine interesting and testable implications from combining the two, and we hope that future work will pursue such possibilities.

Finally, turning from theory to evidence, many papers have analyzed whether unmeasured patient heterogeneity is responsible for physician-level findings in empirical analyses like ours. The overwhelming finding from this literature, with contributions both by health economists (e.g., Hellerstein (1998) and Zhang, Baicker, and Newhouse (2010)) and aca-

demic clinicians (e.g., Solomon et. al (2003) and Schneeweis et. al. (2005)), is that the estimated role of physicians in influencing treatment regimes is largely unaffected by incorporating patient-specific data. For example, the 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: the empirical results they obtained are largely quantitatively unaffected with alternative specifications incorporating patient-specific data. As Coscelli (2000: 354) summarized his early work with patient-level data: "These patterns demonstrate clearly that the probability of receiving a new treatment is significantly influenced by the doctor's identity, and that doctors differ in their choice among ... drugs for the same patient." Thus, similar to our hope that future theory will combine learning across patients and learning within patients, our hope is that future empirical work will combine longitudinal data on both physicians and patients, but the existing empirical literature gives us confidence that our results from physician-level data will persist.

1.2 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."[2] 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.[3]

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.[4] In the US, Medicaid is the largest payer of medical and drug benefits to people with schizophrenia.[5]

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 re-

sulted 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 (decrease in 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.[6]

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."[7]

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.[8]

Figure 1 about here

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.[9] As a proportion of all antipsychotic prescriptions, the atypical percentage 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.

1.3 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 and interactions 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 percentage 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. 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). Below we explore both these characteristics of prescribing behavior, both theoretically and empirically.

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 choice of their favorite drug. Thus, national market shares do not reflect homogeneous physicians each prescribing drugs in proportions approximating national shares, but rather portray heterogeneous physicians many of whom are highly concentrated on particular drugs. In comparison to the distribution of choices of highly concentrated prescribers given above, in our 2007 sample the national market percentages 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 concentration raise an intriguing possibility. 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 and persistent between-region differences in treatments and costs.[10] Could it be that our initial finding of heterogeneous concentration is driven by correspondingly large between-region variability in antipsychotic prescribing behavior? Alternatively, is most variability physician-specific, with regions relatively similar to each other? We address this issue in the concluding section. For now, we simply report the result that the large majority of variation is at the physician level.

This preliminary evidence leads us to focus on individual prescribers 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 theoretical issues we now turn our attention.

2 Towards a theory of prescriber learning and treatment behavior

2.1 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 genres: *perception*, *motivation*, *administration*, and *inspiration*, which we now briefly summarize.[11]

2.1.1 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.

2.1.2 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 variability among physicians' prescribing decisions. 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 and persistent brand allegiances among physicians.[12]

2.1.3 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 (and Medicare Part D required that any private prescription drug plan offer all but one of the atypical antipsychotic drugs on its formulary); many other private insurers had similar open formulary provisions.[13]

2.1.4 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 as they learn.

2.2 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. Later we 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 $t \in \mathbb{N} = \{1, 2, ...\}$. 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. Specifically, 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 $S = \{s_1, \ldots, s_S\}$ 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 $D = \{d_1, \ldots, d_D\}$. 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 derived 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 action a that the physician undertakes, such as adjusting the dosage of the drug (a process known as titrating, perhaps because the patient is a heavy smoker), 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 (e.g., take once-a-day sedating drugs at night).[14] 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

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

where a denotes the complementary action the physician undertakes, and

$$x_{dt} = \theta_d + \varepsilon_{dt}.$$

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 ideal 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_d^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 $x_{dt} = \theta_d + \varepsilon_{dt}$, 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 specific 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 distinct drugs. However, the speed of learning the complementary action θ_d for each drug d depends only on how often the physician prescribes drug d, not on the drug or patient identity.

2.3 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 (1) and (2) 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 $\sigma_\varepsilon^2 = 0$), 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_d^2 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.[15]

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 var-

ious 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 – diverse concentration or deviation.

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.[16]

2.4 Analysis of the model and preliminary comparative statics

The optimal prescribing behavior of the physician can be characterized in a simple manner because our 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 $h_t = \times_{\tau=1}^{t-1} (s_\tau, d_\tau, a_\tau, x_{d_\tau \tau})$. The physician's policy decision is to choose a drug d and complementary action 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:

$$\begin{aligned} a(h_t) &= \mathbb{E}[\theta_d | h_t], \\ \mathbb{E}[b_{sdt} | h_t] &= B_{sd} - Var(\theta_d | h_t) - \sigma_\varepsilon^2, \end{aligned} \tag{1}$$

where $\mathbb{E}[\theta_d | h_t]$ and $Var(\theta_d | h_t)$ 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:

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

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, $\mathbb{E}[b_{sdt}|h_t]$ in (1), depends on d in two ways: the maximum benefit B_{sd}, which is already known, and the expected loss from imperfect complementary actions, $Var(\theta_d|h_t) + \sigma_\varepsilon^2$, 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 $Var(\theta_d|h_t)$. That is, the physician's prescribing behavior can be summarized by D state variables identified with posterior variances $Var(\theta_d|h_t)$ for $d \in D$. 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

$$B_{s'd'} - Var(\theta_{d'}|h_t) - \sigma_\varepsilon^2,$$
$$B_{s'd} - Var(\theta_d|h_t) - \sigma_\varepsilon^2.$$

Therefore, the myopic physician is trading off the difference between $B_{s'd'}$ and $B_{s'd}$ against the difference between $Var(\theta_{d'}|h_t)$ and $Var(\theta_d|h_t)$. 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

$$Var(\theta_d|h_t) < Var(\theta_{d'}|h_t) - (B_{s'd'} - B_{s'd}),$$

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 0 (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 0) will eventually prescribe $d^*(s)$ for every s and so have a common distribution of prescriptions, regardless of their initial history of patients.

To exposit all these ideas in a simple setting, in Appendix A we solve 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. As noted above, 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. Proposition 1 describes the solution to this example, and Corollaries 1 and 2 then show, respectively, that expected concentration and expected deviation are decreasing with volume.

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 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., $B_{s'd'} > B_{s'd}$).

To summarize the possible effects 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.

To illustrate the effects of new drugs and new physicians, we return to the example in Appendix A. We now enrich the example by assuming that 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 enriched version of our example is central to our discussion in Section 4.A of a competing hypothesis—namely, "detailing" by sales representatives from pharmaceutical firms, rather than our model of learning-by-doing: our model predicts that the propensity of young doctors to prescribe old drugs (i.e., drugs that stopped being detailed before the doctor began prescribing) is increasing the the doctor's prescription volume. As we describe in Section 4.1, we find empirical support for this prediction, which is contrary to the detailing hypothesis.

2.5 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 do high-volume physicians. In addition, we expect the set of drugs in the steady-state prescription set will 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. Note that in our framework training and experience are alternative sources of learning about a drug, i.e., they may substitute for one another.

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 waiting time between patients 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.

3 Data, methods and findings

3.1 Prescriptions data

Our data on prescribers' behavior are taken from the IMS Xponent$^{\text{TM}}$ 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.[17] We designate all other prescribers as other ("OTH").

To mitigate the possible impact of very low-volume prescribers, for the remainder 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). As seen in Table 1, 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 1 about here

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 0.33 (0.37 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 0.62 (0.67 for atypicals, slightly less than the 0.70 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^{TM} 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.[18]

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.[19]

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.[20]

3.2 Empirical framework and econometric methods

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

$$Y_i = \beta\left(\frac{1}{V_i}\right) + \varphi X_i + \varepsilon_i \tag{3}$$

where Y_i is one of two dependent variables (either C_i, a measure of the concentration of a physician's prescriptions, or D_i, a measure of the deviation of a physician's prescriptions from a specified average), V_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.[21] In some regressions we specify interaction variables, particularly among measures of inverse volume and physician specialty.

As a simple example, one measure of a prescriber's concentration C_i is the HHI of the physician's prescriptions. Since HHI will be bounded below and above by 0 and 1, we take account of this by employing appropriate econometric estimation methods.

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_{id}. Let the overall market share of drug d in region r be m_{dr}, where both s_{id} and m_{dr} are between zero and one. As a measure of the deviation of physician i's prescribing behavior from that of the regional market share, we calculate

$$D_i = \sum\nolimits_d (s_{id} - m_{dr})^2 = \text{HHI}_i + \text{HHI}_r - 2\sum\nolimits_d s_{id} m_{dr} \tag{4}$$

If every physician i in region r had the same prescribing share, D_i would equal zero. As physician prescribing behavior heterogeneity within region r increases, D_i increases.

Ellison and Glaeser (1997) have noted that one can expect a mechanical numerical reduction in the deviation measure D_i as volume increases at small volumes. To correct for this small-numbers volume issue in the deviation measure, they revise the raw deviation measure (4) as follows:

$$\widehat{D}_i = \frac{V_i}{V_i - 1}\left(D_i - (1 - \text{HHI}_r)\frac{1}{V_i}\right). \tag{5}$$

Hereafter we refer to this revised measure of deviation as *corrected deviation*. Analogous considerations imply that one can expect a mechanical numerical reduction in the Herfindahl-Hirschman measure of concentration as volume increases at small volumes. To correct for this small-numbers volume issue in the concentration measure, we amend the HHI index as follows, which yields an unbiased effect of volume on concentration:[22]

$$\widehat{C}_i = \frac{V_i}{V_i - 1}\left(C_i - \frac{1}{V_i}\right). \tag{6}$$

Hereafter we refer to this revised measure of concentration as *corrected concentration*.[23]

Regarding covariates, we take the age of the prescribing physician 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 but < 51, between 51 but less than 59, 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.[24]

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.

Table 2 about here

In Table 2 below we provide summary statistical information for both the dependent and explanatory variables employed in our analyses. The mean number of different antipsychotics prescribed is 4.41, of which 3.21 are atypicals; the mean percentage of atypical prescriptions is 88.05%. The average raw HHI over all antipsychotics is 0.48, while that for atypicals only is larger at 0.55; the mean raw deviation from national shares is 0.22. When corrected for small-volume biases, the corresponding corrected measures of concentration and deviation are 0.47 and 0.20, respectively. The average physician age is 50.60 years, with 27% of them being female.

3.3 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.

3.3.1 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 predicts greater deviation from national market shares for low- than for high-volume prescribers, other things equal, as well as for others whose present value of benefits from learning regarding variety is lower. Results from OLS estimation with deviation and corrected deviation (from national market shares) as the dependent variable are presented in Table 3.

Table 3 about here

Relative to our theoretical framework, two results stand out in Table 3. First, and more important, for each of the four interactions between specialty and inverse volume, deviation from national shares increases with inverse volume, and significantly so (with the coefficient estimates for the corrected deviation measure being slightly smaller than with raw deviation, but still significant). The finding that volume has a negative impact on deviation for all specialties is consistent with our learning-by-doing theoretical framework.[25]

Second, turning to the dummy variables for specialty (i.e., conditional on inverse volume by specialty), the raw and corrected deviations in prescribing behavior are smallest for PSY, followed by PCP, then the reference group, OTH and NEU (but the latter two are insignificantly different from each other). Variations across specialties (in both main effects and interactions) can be interpreted in terms of our framework if one imagines that training in different specialties conveys different information about complementary actions for various drugs.[26]

Moving away from our theoretical framework, various other results in Table 3 may be of interest. Relative to the youngest age quartile (under age 43), deviation increases monotonically and significantly with age quartile. Third, other things equal, female prescribers exhibit significantly more deviation. Fourth, physicians in solo practice exhibit slightly more deviation. Physicians practicing in the largest counties (>1,000,000 population) and small to medium population counties exhibit less deviation than those in smallest (under 150,000) counties, but hospital-based physicians, DOs, and opt-out prescribers do not differ from the reference group.

3.3.2 Concentration of antipsychotic prescribing: Physician prescribing antipsychotic HHI

Next we examine which physicians use a wider variety of drug molecules, as measured by the raw and corrected concentration measures in 2007 (see equation (6)). 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.

We estimate both a Tobit model (because HHI is censored above 1) and by ordinary least squares where the dependent variable is alternatively the raw and corrected overall antipsychotic HHI. Results are presented in Table 4; estimated magnitudes are the coefficient estimates. Tobit and OLS estimates on both raw and corrected HHI are qualitatively generally consistent across estimation methods, particularly in terms of relative magnitudes and statistical significance. Consistent with our theory, each of the four interactions between specialty and inverse volume is positive and statistically significant.[27] Turning to the specialty dummies we see that, conditional on volume, the concentration is lowest for PSY, followed by PCP, and then the excluded OTH reference group, with concentration for NEU being the greatest.

Table 4 about here

Moving away from our theoretical framework, among antipsychotics overall, only the oldest quartile of physicians have a statistically significant concentration difference (slightly greater than all other prescribers), but solo practitioners also prescribe in a slightly more concentrated manner than do those in group practice. There is no significant relationship involving population size, whether hospital-based, MD vs. DO, or physician opt-out, although female prescribers exhibit a significantly more concentrated prescribing behavior than do their male counterparts.

3.3.3 Robustness

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, we repeated the analysis using the logarithm of volume interacted with specialty as well as linear volume interacting with specialty. The estimate of the overall effect of volume on physician prescribing across specialties was essentially unchanged in both sets of

analyses. We also repeated our analysis limiting the sample just to the 3,431 psychiatrists. In addition, we repeated our analysis limiting the sample to those physicians having written at least 50 antipsychotic prescriptions in 2007 (and then we further limited the sample to those with at least 100 prescriptions). In these models, our main results on the effect of volume of prescribing remain unchanged. We also ran various regressions with number of distinct antipsychotic molecules overall, number of distinct atypical antipsychotic molecules, and number of distinct typical molecules as the dependent variable. Qualitative findings concerning effects of relative marginal volumes across specialty, gender and age were similar, although age patterns involving number of distinct atypical and typical molecules differed somewhat.

4 Discussion and conclusions

Before 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.

4.1 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" genre 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.[28]

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.[29] 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 heterogeneous 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 ever to have 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 are very unlikely to have 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.

To examine the competing detailing hypothesis, we 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. (See the Appendix.) For old physicians, however, the share of the new atypicals could increase or decrease with volume.[30] 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 these drugs 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 who 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 we employ in the regression reported below are the same as those specified in previous analyses. 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 5.[31]

Table 5 about here

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 effect of inverse volume is positive 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 learning 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 impact of inverse volume is much smaller in absolute value (at 215.403-136.807=78.596). 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: high-volume young physicians prescribe old drugs more often than do low-volume young physicians.

4.2 Relationship to existing literature and directions for future research

As noted in the Introduction, our evidence on heterogeneous concentration raises an intriguing possibility: is this variation due to between-region differences in treatments, or is most variability physician-specific and are regions relatively similar to each other?

To address this question, we note there are various ways one could measure the concentration behavior of prescriber i, which we denote C_i. A well-known measure of industry concentration is the Herfindahl-Hirschman Index (HHI).[32] 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 1. Then square the shares and sum over the j firms, yielding HHI= $\sum_j s_j^2$. 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).[33]

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 have played a prominent role in the Dartmouth regional variation and related literatures. Results are given in Table 6.

Table 6 about here

At the individual-prescriber level, prescribing behavior is very concentrated (HHI is almost 0.5), but there is also substantial variability among physicians' HHIs, 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. Specifically, 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 procedures 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 at the HRR and state levels of aggregation there is relatively little between-region variability.

The findings in Table 6 differ qualitatively from those reported by the Dartmouth Atlas project: 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.[34] 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 regional heterogeneity findings reported in the Dartmouth Atlas small-area variations literature.

While our theoretical framework can help explain persistent heterogeneity in concentrated prescribing behaviors, we ignore 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.[35]

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.[36] 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 percentage 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 declined most dramatically from 34% to 12%. Who were the prescribers who switched most rapidly – low or high volume, what specialties, gender, age group, solo vs. group practice – 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? More generally, 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.[37]

Our findings suggest that a significant proportion of the heterogeneity in the treatments patients receive depends upon physician preferences in treatment regime. It would also be informative and useful to identify specific patterns in physician decision-making that appear to indicate general differences in "practice style" across physicians, perhaps related to location of medical residency training, analogous to recent investigations characterizing "management style".[38]

4.3 Summary

We have developed and implemented empirically 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 ex ante 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 expected 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 optimally included 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 deviate more from national market shares than do the prescriptions of high-volume physicians. In short, concentration and deviation decrease with volume.

We confront the model with cross-sectional data on antipsychotic prescriptions, regressing concentration and deviation on the inverse volume and other characteristics of a prescriber. Standard measures of concentration and deviation have a finite-sample bias, so we construct an unbiased measure of concentration and use an unbiased measure of deviation constructed by Ellison and Glaeser (1997). As predicted by our model, we observe that higher-volume physicians have lower concentration and deviation in their prescriptions. To compare our learning model to a model of detailing by sales representatives to physicians, we regress the share of prescriptions written for new drugs on physician's age dummy, inverse volume, and interaction of the two. Consistent with our learning model but at odds with the detailing model, we find that the highest volume physicians in the youngest quartile prescribe a larger share of old drugs.

Acknowledgements

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, the NBER, the University of Chicago, and the University of California – Los Angeles, and from the comments of Joseph Doyle, Marcela Horvitz-Lennon, Ulrike Malmendier, David Molitor, 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^{TM}, Xponent^{TM} and American Medical Association Physician Masterfile^{TM}. 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. Prof. Berndt has received an unrestricted gift from Pfizer, Inc., to perform health policy research administered through the MIT Sloan School of Management.

References

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

[2] 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.

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

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

[5] 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.

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

[7] 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.

[8] 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.

[9] 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.

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

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

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

[13] Crilly, John [2007], "The History of Clozapine and its Emergence in the US Market: A Review and Analysis", *History of Psychiatry* 18(1):39-60.

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

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

[16] Domino, Marisa Elena, Richard G. Frank and Ernst R. Berndt [2012], "The Diffusion of Antipsychotic Medications: Does the Close-Knittedness of Provider Networks Matter?", paper presented at 2009 Annual Meetings of the American Economic Association, San Francisco, California, January 4, 2009. Revised March 2012.

[17] 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.

[18] 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.

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

[20] 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.

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

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] 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.

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

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

[30] 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.

[31] 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.

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

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

[34] 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.

[35] 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.

[36] 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.

[37] 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.

[38] 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.

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

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

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

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] 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.

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

[48] Simpson, E. H. [1949], "Measurement of Diversity", *Nature* 163:688.

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

[50] 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.

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

Appendix A: A 2x2 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 $e^{-rw} = \delta$, $S = \{s_1, s_2\}$, $D = \{d_1, d_2\}$, $\Pr(s_2) = p_2 > 1/2$, $\sigma_1^2 = \sigma_2^2 = c > 1$, $\sigma_\varepsilon^2 = 0$, $B_{12} = B_{21} = 0$, $B_{11} = B_{22} = 1$.

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 θ_d about the ideal complementary action for each drug d (i.e., $\sigma_1^2 = \sigma_2^2 > 0$). However, the physician learns the ideal complementary action precisely after one prescription (i.e., $\sigma_\varepsilon^2 = 0$). As discussed in Section 2.3 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 $d^*(s_1) = d_1$ and $d^*(s_2) = d_2$ (i.e., $B_{11} > B_{12}$ and $B_{22} > B_{21}$). 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., $c > 1$).

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 online appendix.

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 exists 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.*

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 A1. 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_i for symptom s_i), 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.

Proposition 1 immediately implies that under reasonable restrictions on model parameters, concentration and deviation decrease with volume. For the concentration result, we just need to assume that parameters are such that the left panel of Figure A1 applies. For the deviation result, we also need to assume that the market shares are not extreme. In particular, we assume that the market share of the more popular drug is higher than the frequency of the symptom for which this drug is ideal. This assumption automatically holds if the economy is populated with physicians who may differ in volume but otherwise are identical. Further, we assume that the share of physicians who prescribe only the more popular drug is less than a half.

Corollary 1 *Suppose that Assumption 1 holds and $c > 2$. Then the expected concentration of a physician decreases with volume.*

Corollary 2 *Suppose that Assumption 1 holds, $c > 2$, and the market share m_2 of drug d_2 satisfies $m_2 \in [p_2, (1 + p_2)/2]$. Then the expected deviation of a physician decreases with volume.*

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 $T = 1, 2, 3$: 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 0.5) in 2007, we assume that $\Pr(s_2) = p_2 > 1/2$, again in keeping with Assumption 1. For example, let us set $p_2 = 6/7$. If we then proceed upwards in Figure A1 along 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 $T = 2$. Table A1 then represents prescription behaviors as a function of the colored and shaded regions in Figure A1.

Table A1 about here

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).

Appendix B: Proofs (not for publication)

Proof of Proposition 1

We need to consider 8 cases corresponding to all possible combinations of symptom $s \in \{s_1, s_2\}$ and posterior variances $Var(\theta_1|h_t) \in \{0, c\}$, $Var(\theta_2|h_t) \in \{0, c\}$. The optimal complimentary action given drug d is simply $a(h_t) = \mathbb{E}[\theta_d|h_t]$. We now turn to the optimal choice of drug. The optimal choice of drug for symptom s at history h_t depends on a triple $(s, Var(\theta_1|h_t), Var(\theta_2|h_t))$. The proof of Proposition 1 relies on the One–Shot Deviation Principle (hereafter abbreviated as OSDP) and consists of considering all 8 cases.

1. Consider cases with $(Var(\theta_1|h_t), Var(\theta_2|h_t)) = (0, 0)$.

   Lemma 1 *$d(s_1, 0, 0) = d_1$ and $d(s_2, 0, 0) = d_2$.*

   Proof. By OSDP, $d(s_1, 0, 0) = d_1$ and $d(s_2, 0, 0) = d_2$ iff $1 \geq 0$. ■

2. Consider cases with $(Var(\theta_1|h_t), Var(\theta_2|h_t)) \in \{(0, c), (c, 0)\}$.

   Lemma 2 *$d(s_1, 0, c) = d_1$ and $d(s_2, c, 0) = d_2$.*

   Proof. We prove that $d(s_1, 0, c) = d_1$ for $p_1 \in (0, 1)$. By symmetry, $d(s_2, c, 0) = d_2$ will hold. Suppose, to get a contradiction, that $d(s_1, 0, c) = d_2$. There are two cases to consider $d(s_2, 0, c) = d_2$ and $d(s_2, 0, c) = d_1$:

   - Case: $d(s_2, 0, c) = d_2$. By OSDP, the supposition $d(s_1, 0, c) = d_2$ gives

     $$-c + \delta \geq 1 + \delta(1 - p_1) - \delta c,$$

     which is equivalent to $\delta \geq \frac{1+c}{p_1+c}$. This inequality cannot be satisfied.

   - Case: $d(s_2, 0, c) = d_1$. By OSDP, the supposition $d(s_1, 0, c) = d_2$ gives

     $$-\frac{c\delta p_1}{1 - \delta p_2} + \left(\frac{\delta}{1 - \delta} - \frac{\delta p_1}{1 - \delta p_2}\right) \geq (1 - c) + \frac{\delta}{1 - \delta}.$$

     By OSDP, the supposition $d(s_1, 0, c) = d_2$ gives

     $$-c + \frac{\delta}{1 - \delta} \geq 1 - \frac{c\delta p_1}{1 - \delta p_2} + \left(\frac{\delta}{1 - \delta} - \frac{\delta p_1}{1 - \delta p_2}\right).$$

     Summing up these inequalities, we obtain $0 \geq 2$, which is false.

■

Lemma 3 *$d(s_2, 0, c) = d_2$ iff $\delta \geq \frac{c-1}{c-p_1}$ (boundary between red and yellow areas), and $d(s_1, c, 0) = d_1$ iff $\delta \geq \frac{c-1}{c-p_2}$ (boundary between yellow and green areas).*

Proof. We prove the part with $d(s_2, 0, c) = d_2$, the part with $d(s_1, c, 0) = d_1$ is obtained by replacing p_1 with p_2. By OSDP, $d(s_2, 0, c) = d_2$ iff

$$\begin{aligned} (1-c) + \frac{\delta}{1-\delta} &\geq \frac{(1-c)\,\delta p_2}{1-\delta p_1} + \left(\frac{\delta}{1-\delta} - \frac{\delta p_2}{1-\delta p_1}\right), \\ \delta &\geq \frac{c-1}{c-p_1}. \end{aligned}$$

■

3. Consider cases with $(Var(\theta_1|h_t), Var(\theta_2|h_t)) = (c, c)$. There are 3 possible continuation subcases depending on whether $\delta < \frac{c-1}{c-p_1}$, $\frac{c-1}{c-p_1} \leq \delta < \frac{c-1}{c-p_2}$, or $\delta \geq \frac{c-1}{c-p_2}$. We consider each of these subcases in turn.

   (a) Consider subcase $\delta < \frac{c-1}{c-p_1}$ (red area).

   Lemma 4 *$d(s_2, c, c) = d_2$.*

   Proof. By OSDP, $d(s_2, c, c) = d_2$ iff

   $$(1-c) + \frac{\delta p_2}{1-\delta} \geq -c + \frac{\delta p_1}{1-\delta},$$

   which is trivially satisfied. ■

   Lemma 5 *$d(s_1, c, c) = 1$ iff $\delta \leq \frac{1}{2p_2}$ (boundary between light and dark red areas). $d(s_1, c, c) = d_1$ is satisfied for all δ and p_2 iff $c \leq 2$.*

   Proof. By OSDP, $d(s_1, c, c) = d_1$ iff

   $$\begin{aligned} (1-c) + \frac{\delta p_1}{1-\delta} &\geq -c + \frac{\delta p_2}{1-\delta}, \\ \delta &\leq \frac{1}{2p_2}. \end{aligned}$$

   Moreover, $d(s_1, c, c) = 1$ is satisfied for all δ and p_2 iff subcase $\delta < \frac{c-1}{c-p_1}$ implies $\delta \leq \frac{1}{2p_2}$, which is equivalent to

   $$\frac{1}{2p_2} - \frac{c-1}{c-(1-p_2)} \geq 0$$

for all $p_2 \geq \frac{1}{2}$. By simplifying the inequality, we obtain $c \leq \frac{3p_2-1}{2p_2-1}$ for all $p_2 \geq \frac{1}{2}$, which is equivalent to $c \leq 2$. ■

(b) Consider subcase $\frac{c-1}{c-p_1} \leq \delta < \frac{c-1}{c-p_2}$ (yellow area).

Lemma 6 $d(s_2, c, c) = d_2$.

Proof. By OSDP, $d(s_2, c, c) = d_2$ iff

$$(1-c) + \frac{\delta p_2}{1-\delta} \geq -c + \left(\frac{\delta}{1-\delta} - \frac{\delta p_2}{1-\delta p_1}\right) + \frac{(1-c)\,\delta p_2}{1-\delta p_1}.$$

In the subcase $\frac{c-1}{c-p_1} \leq \delta < \frac{c-1}{c-p_2}$, $d(1, c, 0) = 2$. Therefore, conditional on $d(s_2, c, c) = d_2$, the physician is better off by prescribing $d(s_1, c, 0) = d_2$ in the future:

$$(1-c) + \frac{\delta p_2}{1-\delta} \geq (1-c) + \left(\frac{\delta}{1-\delta} - \frac{\delta p_1}{1-\delta p_2}\right) + \frac{(1-c)\,\delta p_1}{1-\delta p_2}.$$

Therefore, the initial inequality is satisfied because $\frac{\delta p_2}{1-\delta p_1} > \frac{\delta p_1}{1-\delta p_2}$ for $p_2 > \frac{1}{2}$. ■

Lemma 7 $d(s_1, c, c) = d_1$ *iff* $\frac{1-\delta p_2}{1-\delta} \geq \frac{c\delta p_2}{1-\delta p_1}$ *(boundary between light and dark yellow areas).*

Proof. By OSDP, $d(s_1, c, c) = d_1$ iff

$$\begin{aligned} 1 - c + \frac{(1-c)\,\delta p_2}{1-\delta p_1} + \left(\frac{\delta}{1-\delta} - \frac{\delta p_2}{1-\delta p_1}\right) &\geq -c + \frac{\delta p_2}{1-\delta}, \\ \frac{1-\delta p_2}{1-\delta} &\geq \frac{c\delta p_2}{1-\delta p_1}. \end{aligned}$$

■

(c) Consider subcase $\delta \geq \frac{c-1}{c-p_2}$ (green area).

Lemma 8 $d(s_2, c, c) = d_2$.

Proof. By OSDP, $d(s_2, c, c) = d_2$ iff

$$\begin{aligned} (1-c) + \frac{(1-c)\,\delta p_1}{1-\delta p_2} + \left(\frac{\delta}{1-\delta} - \frac{\delta p_1}{1-\delta p_2}\right) &\geq -c + \frac{(1-c)\,\delta p_2}{1-\delta p_1} + \left(\frac{\delta}{1-\delta} - \frac{\delta p_2}{1-\delta p_1}\right), \\ 1 &\geq -c\left(\frac{\delta p_2}{1-\delta p_1} - \frac{\delta p_1}{1-\delta p_2}\right), \end{aligned}$$

which is satisfied because $\frac{\delta p_2}{1-\delta p_1} > \frac{\delta p_1}{1-\delta p_2}$ for $p_2 > \frac{1}{2}$. ■

Lemma 9 *$d(s_1, c, c) = d_1$ (no dark green area).*

Proof. By OSDP, $d(s_1, c, c) = d_1$ iff

$$\begin{aligned} 1 - c + \frac{(1-c)\,\delta p_2}{1-\delta p_1} + \left(\frac{\delta}{1-\delta} - \frac{\delta p_2}{1-\delta p_1}\right) &\geq -c + \frac{(1-c)\,\delta p_1}{1-\delta p_1} + \left(\frac{\delta}{1-\delta} - \frac{\delta p_1}{1-\delta p_2}\right), \\ 1 &\geq c\left[\frac{\delta p_1}{1-\delta p_2} - \frac{\delta p_2}{1-\delta p_1}\right], \end{aligned}$$

which is satisfied because $\frac{\delta p_2}{1-\delta p_1} > \frac{\delta p_1}{1-\delta p_2}$ for $p_2 > \frac{1}{2}$. ■

Proof of Corollary 1

We compute expected concentration for each area in turn.

In light red, dark red, and dark yellow areas, expected concentration is:

$$C_r = 1.$$

In light yellow area, expected concentration is:

$$C_y = p_1\left(p_1^2 + p_2^2\right) + p_2.$$

In light green area, expected concentration is:

$$C_g = p_1^2 + p_2^2.$$

Noting that $C_r > C_y > C_g$, the left panel of Figure A.1 immediately implies that expected concentration decreases with δ.

Proof of Corollary 2

We compute expected deviation for each area in turn under fixed market shares m_d of each drug d. We use equalities $p_1 + p_2 = 1$ and $m_1 + m_2 = 1$ to express all results in terms of p_2 and m_2 only.

In light red area, expected deviation is:

$$\begin{aligned} D_{lr} &= p_1\left[(1-m_1)^2 + (0-m_2)^2\right] + p_2\left[(0-m_1)^2 + (1-m_2)^2\right] \\ &= 2\left[(1-p_2)\,m_2^2 + p_2\,(1-m_2)^2\right]. \end{aligned}$$

In dark red and dark yellow areas, expected deviation is:

$$\begin{aligned} D_{dry} &= (0-m_1)^2 + (1-m_2)^2 \\ &= 2(1-m_2)^2. \end{aligned}$$

In light yellow area, expected deviation is:

$$\begin{aligned} D_{ly} &= p_1\left[(p_1-m_1)^2 + (p_2-m_2)^2\right] + p_2\left[(0-m_1)^2 + (1-m_2)^2\right] \\ &= 2(1-p_2)(p_2-m_2)^2 + 2p_2(1-m_2)^2. \end{aligned}$$

In light green area, expected deviation is:

$$\begin{aligned} D_g &= (p_1-m_1)^2 + (p_2-m_2)^2 \\ &= 2(p_2-m_2)^2. \end{aligned}$$

It is easy to verify that $D_{lr} > D_{dry}$ if $m_2 \geq p_2$ and $D_{dry} \geq D_{ly} \geq D_g$ iff $m_2 \leq \frac{1+p_2}{2}$. Combining these observations, the left panel of Figure A.1 immediately implies that expected deviation decreases with δ if $m_2 \in \left[p_2, \frac{1+p_2}{2}\right]$. Finally we note that if the market share m_2 of drug 2 is generated by physicians who may have different δ but otherwise are identical, then $m_2 \geq p_2$ always holds and $m_2 \leq \frac{1+p_2}{2}$ holds if the share of physicians in the dark areas is less than a half.

Notes

[1]Coscelli and Shum analyze a two-armed bandit model of learning about the efficacy of one new drug. In this model, if prescribers could observe national market shares, then they would all make the same prescription for a given patient, whereas in our model, physician-specific learning by doing rationalizes heterogeneous concentration as optimal behavior even when physicians can observe national market shares. Frank and Zeckhauser informally discuss a "Sensible Use of Norms" hypothesis based on a multi-armed bandit model and a "My Way" hypothesis where "physicians regularly prescribe a therapy that is quite different from the choice that would be made by other physicians" (p. 1008). Because their bandit model ignores learning across patients, they interpret evidence of the My Way hypothesis as physicians "engaging in some highly suboptimal therapeutic practices" (p. 1125), whereas in our model such heterogeneous concentration by physicians is optimal. Finally, neither model makes our predictions about the effect of volume on concentration or the use of old drugs by new prescribers.

[2]Mosby's Medical, Nursing, & Allied Health Dictionary (1998), p. 1456.

[3]Domino, Norton, Morrissey and Thakur (2004).

[4]American Psychiatric Association (2004), p. 9.

[5]Duggan (2005).

[6]Frank, Berndt, Busch and Lehman (2004). For a history of clozapine and discussion of antitrust issues raised by the laboratory test results requirement, see Crilly (2007).

[7]American Psychiatric Association (2004), p. 66.

[8]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".

[9]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.

[10]See, for example, Skinner and Fisher (1997), Fisher, Wennberg, Stukel et al. (2003a,b) and Yasaitis, Fisher, Skinner et al. (2009).

[11]We thank Jan Rivkin for teaching us these "4 'tions," which we adapt here for our own purposes.

[12]An early discussion of these principal-agent issues is found in Pauly (1980), albeit in the context of hospital treatments, not pharmaceuticals.

[13]For discussion, see Frank and Glied (2006) and Huskamp (2003).

[14]We are indebted to Marcela Horvitz-Lennon, M.D., for discussion of physicians' common complementary actions when prescribing antipsychotic drugs to people with schizophrenia.

[15]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.

[16]For a model of antipsychotic and antidepressant prescribing behavior incorporating spillovers depending on the "close-knittedness" of prescribers, see Domino, Frank and Berndt (2012).

[17]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.

[18]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 physician prescribers.

[19]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 solely by migration patterns of patients. For further details, see Dartmouth Atlas Project, http://www.dartmouthatlas.org.

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

[21]In our model, the prescribing behavior depends on e^{-rw_i}, where $w_i = 1/V_i$ is physician i's waiting time between patients, so the prescribing behavior depends on e^{-r/V_i}, which is approximately $1 - r/V_i$.

[22]To see why we use this corrected measure of concentration, suppose that a physician i prescribes a drug d with probability p_d independently across periods and that the realized share of a drug d is s_{id}. Then the expectation of $\widehat{C}_i$ is $\sum_d p_d^2$. Specifically,

$$\mathbb{E}\left[\widehat{C}_i\right] = \frac{V_i}{V_i - 1}\left(\mathbb{E}\left[\sum\nolimits_d s_{id}^2\right] - \frac{1}{V_i}\right) = \sum\nolimits_d p_d^2$$

because

$$\mathbb{E}\left[\sum\nolimits_d s_{id}^2\right] = \sum\nolimits_d \left(Var\left(s_{id}\right) + p_d^2\right) = \sum\nolimits_d \left(\frac{p_d\left(1 - p_d\right)}{V_i} + p_d^2\right) = \frac{V_i}{V_i - 1}\sum\nolimits_d p_d^2 + \frac{1}{V_i}.$$

[23]For another measure attempting to correct HHI for small number volume issues, see Stern and Trajtenberg (1998).

[24]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."

[25]While the estimated coefficients on inverse volume speak to our theory, to ease interpretation one can also compute the marginal effect of volume on deviation as $-\beta/V^2$. For the specialties OTH, PCP, PSY and NEU, the marginal effects of volume on corrected deviation evaluated at the mean of specialty volume are -0.00087, -0.00063, -0.000012 and -0.00026. Thus, at mean specialty volumes, the negative effect of volume on corrected deviation is smallest for PSY prescribers, followed by NEU, then PCP, and largest for OTH prescribers.

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

[27]Marginal effects (again equal to $-\beta/V^2$ at mean specialty volume) based on Tobit estimates on corrected HHI are -0.0011 for OTH, -0.00081 for PCPs, -0.000016 for PSY prescribers, and -0.00032 for NEU. Hence, as with corrected deviation, at mean specialty volumes, the effect of volume on corrected concentration is smallest for PSY, then for NEU and PCP, and largest for OTH prescribers.

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

[29]For discussion and empirical evidence, see Berndt, Kyle and Ling (2003).

[30]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.

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

[32]Interestingly, the Herfindah-Hirschman formula was originally developed in ecology as a measure of biodiversity. See Simpson (1949), who defined group shares as summing to one, and hence produced a concentration measure with range [0,1]. While some later applications in industrial organization use percentage shares summing to 100, and hence produce a concentration measure with range [0,10,000], other economic applications use the share and concentration measures with range [0,1] (see, for example Stern and Trajtenberg (1998)), and we do likewise.

[33]For example, if a prescriber used only three of the atypicals with prescription shares of 0.65, 0.25 and 0.10, the HHI would be 0.495; if, however, all six were used equally (each 0.167), the HHI would be 0.167. Note that the U.S. Department of Justice horizontal merger guidelines measure shares as percentages on [0,100] and hence produce a concentration measure on [0,10,000]. The guidelines state that when a merger results in a change in the measure of more than 100 points and with the merged firm generating a post-merger industry measure 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.

[34]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 R2 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 4 and 5 – while the fixed effects are jointly significant, they have little effects on the point estimates or statistical significance reported earlier.

[35]For an attempt to incorporate spillovers from "close knitted prescribers" in the context of antipsychotics, see Domino, Frank and Berndt (2012).

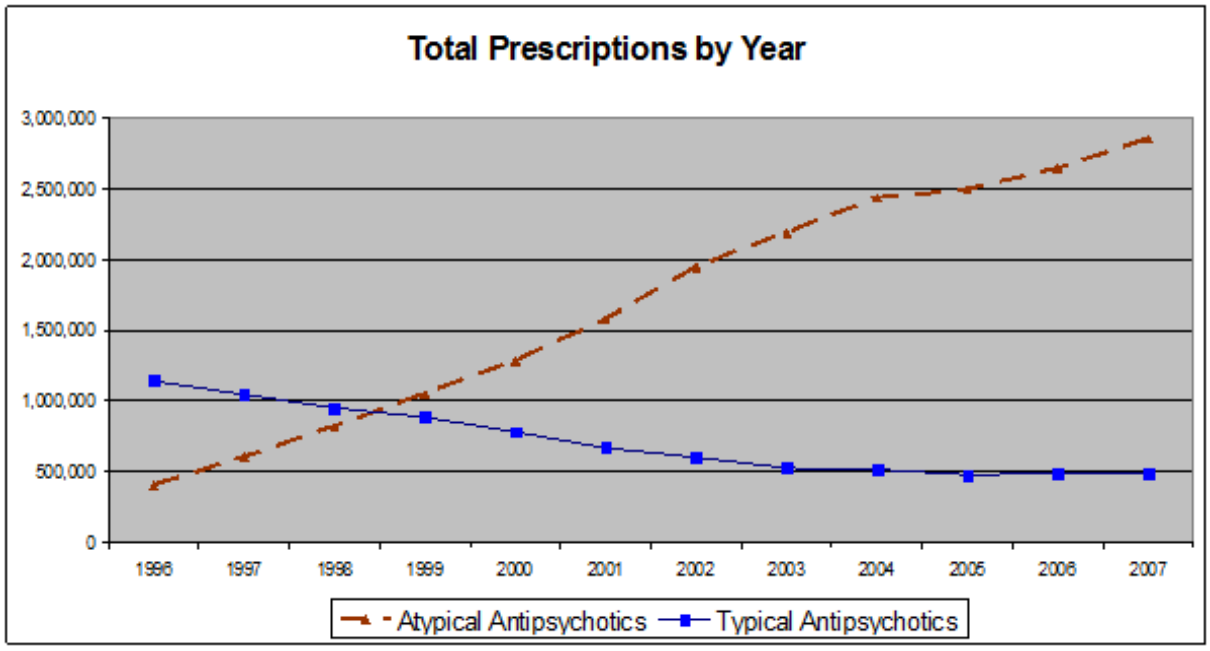
[36]An appendix to this paper available from the lead author, "Timelines Appendix" provides further details. Among the more notable publications are those based on the CATIE study; see, for example, Lieberman, Stroup, McEvoy et al. (2005), White (2006) and Kraemer, Glick and Klein (2009).

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

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

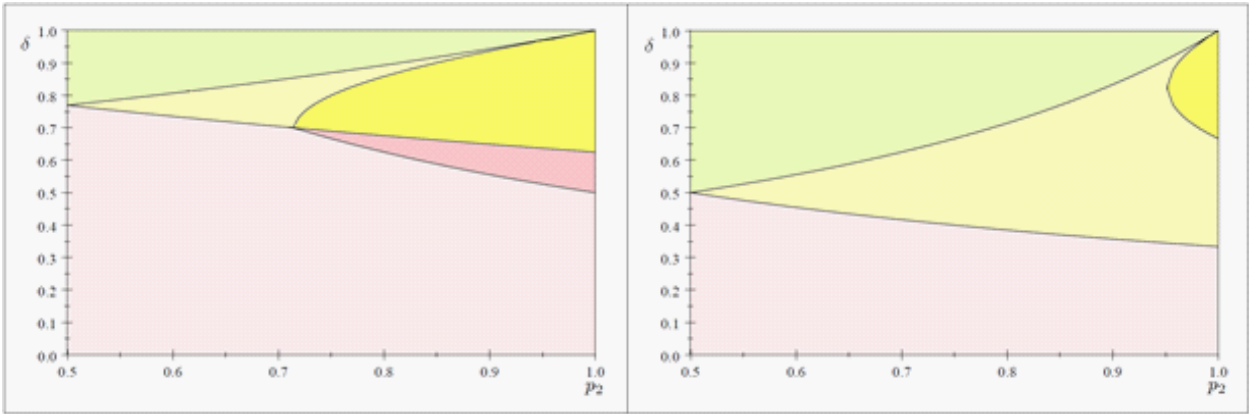
Figures

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



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

Figure A1: Physican's Prescribing Behavior in a 2x2 Example.



Left panel: $c = 8/3 > 2$. Right panel: $c = 3/2 < 2$.

Tables

Table 1: Mean Values of Characteristics of 2007 Prescriber Sample, by Prescriber Specialty

Specialty Group	Number of Prescribers	Antipsychotic Annual Rx	Atypical Annual Rx	# Distinct Antipsychotics	# Distinct Atypicals	Antipsychotic HHI	Atypical HHI	% Antipsychotic Rxs for Atypicals
PSY	3,431	611.03	554.45	7.26	4.71	0.33	0.37	91.37
NEU	688	97.53	86.57	3.23	2.39	0.61	0.70	85.30
PCP	8,536	66.49	59.02	3.78	2.90	0.50	0.57	86.85
OTH	2,382	54.42	49.27	2.95	2.39	0.62	0.67	88.35
NP	1,376	200.11	185.38	4.34	3.30	0.50	0.54	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^{TM} general prescriber sample 2007 data for prescribers writing at least 12 antipsychotic prescriptions.

Table 2: Summary Statistics

Variable	Obs	Mean	Std. Dev.	Minimum	Maximum
Number of Different Antipsychotics Prescribed	15,037	4.41	2.64	1	15
HHI for Antipsychotic Prescriptions	15,037	0.48	0.23	0.12	1
Corrected HHI for Antipsychotic Prescriptions	15,037	0.47	0.23	0.12	1
Deviation for Antipsychotic Prescriptions	15,037	0.22	0.19	0.00	1
Corrected Deviation for Antipsychotic Prescriptions	15,037	0.20	0.19	-0.05	1
Number of Different Atypicals Prescribed	15,037	3.21	1.48	0	6
HHI for Atypical Prescriptions	14,865	0.55	0.24	0.17	1
% of Atypical Prescriptions	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^{TM} general prescriber sample 2007 data.
Sample includes all physician prescribers that wrote at least 12 prescriptions for antipsychotics in 2007.

Table 3: Deviation of Physician's Prescribing Shares from National Market Shares

	Deviation	Corrected Deviation
OTH*(1/Total Yearly Antipsychotic Prescriptions)	2.957***	2.573***
	[0.139]	[0.145]
PCP*(1/Total Yearly Antipsychotic Prescriptions)	3.263***	2.796***
	[0.081]	[0.084]
PSY*(1/Total Yearly Antipsychotic Prescriptions)	5.009***	4.534***
	[0.166]	[0.173]
NEU*(1/Total Yearly Antipsychotic Prescriptions)	2.872***	2.464***
	[0.270]	[0.282]
Age Quartile 43-50^	0.009**	0.009**
	[0.004]	[0.004]
Age Quartile 51-58^	0.019***	0.020***
	[0.004]	[0.004]
Age Quartile 59+^	0.032***	0.034***
	[0.004]	[0.004]
PCP^	-0.076***	-0.076***
	[0.007]	[0.008]
PSY^	-0.144***	-0.142***
	[0.007]	[0.008]
NEU^	0.005	0.006
	[0.012]	[0.012]
Female^	0.013***	0.013***
	[0.003]	[0.003]
Population 150,000-500,000 (county)^	-0.013***	-0.014***
	[0.004]	[0.004]
Population 500,000-1,000,000 (county)^	-0.007*	-0.008**
	[0.004]	[0.004]
Population more than 1,000,000 (county)^	-0.016***	-0.017***
	[0.004]	[0.004]
Solo Practice^	0.011***	0.011***
	[0.003]	[0.003]
Hospital Based Physician^	-0.005	-0.006
	[0.005]	[0.005]
DO Flag^	-0.006	-0.007
	[0.005]	[0.005]
Physician Opt Out^	-0.002	-0.002
	[0.007]	[0.007]
Cons	0.191***	0.188***
	[0.007]	[0.008]
Number of Observations	15,037	15,037
R^2	0.279	0.226

Linear regression is estimated. Standard errors are in brackets. ^indicates dummy variable. * , **, ***, indicate significance at the 10%, 5%, and 1% respectively. All values calculated using IMS Health Incorporated Xponent^{TM} 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. The corrected deviation measure is based on equation (5).

Table 4: Concentration of Physician's Antipsychotic Prescribing Shares

	HHI		Corrected HHI	
	Linear	Tobit	Linear	Tobit
OTH*(1/Total Yearly Antipsychotic Prescriptions)	3.620***	4.192***	3.365***	3.943***
	[0.164]	[0.179]	[0.170]	[0.186]
PCP*(1/Total Yearly Antipsychotic Prescriptions)	3.940***	4.218***	3.584***	3.866***
	[0.095]	[0.103]	[0.099]	[0.107]
PSY*(1/Total Yearly Antipsychotic Prescriptions)	6.236***	6.526***	5.832***	6.125***
	[0.195]	[0.213]	[0.203]	[0.221]
NEU*(1/Total Yearly Antipsychotic Prescriptions)	3.322***	3.780***	3.047***	3.512***
	[0.318]	[0.348]	[0.330]	[0.361]
Age Quartile 43-50^	0.004	0.005	0.005	0.006
	[0.004]	[0.005]	[0.005]	[0.005]
Age Quartile 51-58^	0.004	0.005	0.005	0.006
	[0.004]	[0.005]	[0.005]	[0.005]
Age Quartile 59+^	0.011**	0.014***	0.012**	0.015***
	[0.005]	[0.005]	[0.005]	[0.005]
PCP^	-0.095***	-0.094***	-0.094***	-0.093***
	[0.009]	[0.009]	[0.009]	[0.010]
PSY^	-0.216***	-0.213***	-0.213***	-0.210***
	[0.009]	[0.009]	[0.009]	[0.010]
NEU^	0.040***	0.044***	0.041***	0.045***
	[0.014]	[0.015]	[0.014]	[0.016]
Female^	0.030***	0.031***	0.030***	0.032***
	[0.004]	[0.004]	[0.004]	[0.004]
Population 150,000-500,000 (county)^	-0.004	-0.005	-0.004	-0.005
	[0.004]	[0.005]	[0.005]	[0.005]
Population 500,000-1,000,000 (county)^	0.001	-0.001	0.000	-0.002
	[0.005]	[0.005]	[0.005]	[0.005]
Population more than 1,000,000 (county)^	-0.005	-0.007	-0.006	-0.008*
	[0.004]	[0.005]	[0.005]	[0.005]
Solo Practice^	0.009**	0.009**	0.009**	0.009**
	[0.004]	[0.004]	[0.004]	[0.004]
Hospital Based Physician^	-0.009	-0.009	-0.009	-0.009
	[0.006]	[0.006]	[0.006]	[0.007]
DO Flag^	-0.008	-0.009	-0.008	-0.009
	[0.006]	[0.006]	[0.006]	[0.006]
Physician Opt Out^	-0.002	-0.003	-0.002	-0.002
	[0.008]	[0.009]	[0.009]	[0.009]
Cons	0.462***	0.457***	0.458***	0.453***
	[0.009]	[0.009]	[0.009]	[0.010]
Number of Observations	15,037	15,037	15,037	15,037
R^2	0.332		0.292	
Num Observations Rt. Censored		1,146		1146

Tobit and linear regressions are estimated. Standard errors are in brackets. ^indicates dummy variable. * , **, ***, indicate significance at the 10%, 5%, and 1% respectively. All values calculated using IMS Health Incorporated Xponent^{TM} 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. The corrected HHI measure is based on equation (6).

Table 5: Percent of All Antipsychotic Prescriptions written for *New Drugs* in 2007

	% Rxs for New Drugs
Physician Age 59+^	-4.552***
	[1.163]
1/Total Yearly Antipsychotic Prescriptions	215.403***
	[45.557]
(Physician Age 59+)^(1/Total Yearly Antipsychotic Prescriptions)	-136.807**
	[57.192]
Female^	3.730***
	[1.026]
Population 150,000-500,000 (county)^	1.637
	[1.550]
Population 500,000-1,000,000 (county)^	1.053
	[1.516]
Population more than 1,000,000 (county)^	0.771
	[1.453]
Solo Practice^	0.486
	[1.331]
Hospital Based Physician^	-2.439*
	[1.369]
DO Flag^	0.97
	[2.712]
Physician Opt Out^	-5.314**
	[2.293]
Constant	90.754***
	[1.521]
Number of Observations	1,843
Pseudo R^2	0.0089
Num Observations Rt. Censored	440
Mean of dependent variable	88.29

Tobit regression is estimated. Standard errors are in brackets. ^indicates dummy variable. * , **, ***, indicate significance at the 10%, 5%, and 1% respectively. All values calculated using IMS Health Incorporated Xponent^{TM} 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.

Table 6: Means, Standard Deviations and Coefficients of Variation for Antipsychotic HHIs

Geographic Aggregate	**Mean HHI**	**Std. Dev**	**Coef. Of Variation**	**N**
Individual Prescriber	0.48	0.23	0.48	15,037
County	0.35	0.18	0.52	1,860
Hospital Referral Region	0.22	0.04	0.19	306
State (plus District of Columbia)	0.21	0.02	0.08	51
Nation	0.20	na	na	1

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

Table A1: Prescribing Behavior in a 2x2 Example

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

(x,y) means that a physician is prescribing fraction x of d_1 and fraction y of d_2, where $x+y=1$.