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

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

This Version: August, 2014

Abstract

We present two new findings based on annual antipsychotic US prescribing data from IMS Health on 2,867 psychiatrists who wrote 50 or more prescriptions in 2007. First, many of these psychiatrists have prescription patterns that are statistically significantly different than random draws from national market shares for prescriptions by psychiatrists. For example, many have prescription patterns that are significantly more concentrated than such draws. Second, among psychiatrists who are the most concentrated, different prescribers often concentrate on distinct drugs. Motivated by these two findings, we then construct a model of physician learning-by-doing that fits these facts and generates two further predictions: both concentration (on one or a few drugs) and deviation (from the prescription patterns of others) should be smaller for high-volume physicians. We find empirical support for these predictions. Furthermore, our model outperforms an alternative theory concerning detailing by pharmaceutical representatives.

JEL Classification: I10; I11; D80; D83

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

*Louis E. Seley Professor in Applied Economics, MIT Sloan School of Management, and NBER
†Sloan Distinguished Professor of Management, MIT Sloan School of Management, and NBER
‡University of New South Wales, School of Economics
§Cornerstone Research

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 deliver 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. In addition, physicians with concentrated prescriptions may converge (exhibiting near unanimity on the choice of a favorite drug) or diverge (with different physicians concentrating on distinct drugs). We explore these issues using 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.

In the body of the paper, we begin by describing our data on the antipsychotic prescriptions written by psychiatrists who wrote 50 or more prescriptions in 2007. We then present two new findings. First, many of these psychiatrists have prescription patterns that

are statistically significantly different than random draws from national market shares for prescriptions by psychiatrists. For example, many have prescription patterns that are significantly more concentrated than such draws. (As a stark illustration, in the data we describe and analyze below, the average share of antipsychotic prescriptions written for the psychiatrist's most-prescribed drug is 41%.) Second, but less formally, among psychiatrists who are the most concentrated, different prescribers concentrate on distinct drugs—the "heterogeneous concentration" in our title. In short, even in our sample of psychiatrists with high prescription volumes, national market shares for this population do not reflect homogeneous physicians each prescribing drugs in proportions approximating national shares, but rather the aggregate of heterogeneous physicians many of whom are highly concentrated, albeit on different drugs.

Motivated by these two findings, we construct a model of physician learning-by-doing. Our model predicts how different physicians locate along the concentration-diversification continuum. In particular, path-dependence in learning by doing is a strong force towards the heterogeneous concentration we observe. On the other hand, the model also predicts that the effects of this path-dependence should be smaller for physicians with larger prescription volumes. More specifically, both concentration (on one or a few drugs) and deviation (from the prescriptions patterns of others) should be smaller for high-volume physicians.

We find support for both of these predictions. In addition, we can distinguish our learning model from a competing hypothesis emphasizing detailing by pharmaceutical representatives: our model predicts that high-volume *young* physicians should write a higher share of their prescriptions for *old* drugs than low-volume young physicians should. We find support for this additional prediction, even when we define "old drugs" to be those that had ceased to be detailed before the young physicians' careers began—a finding the detailing alternative cannot explain.

The issues in this paper are important: Advances in the practice of evidence-based medicine are likely to be constrained considerably if physicians limit the evolution of their evidentiary platform to their own learning-by-doing experiences, down-weighting accumulating evidence reported by other prescribers. Concentrated, "one size fits all" prescribing behavior could fail to exploit opportunities to successfully tailor or "personalize" medical treatments to patients' idiosyncratic genetic, environmental and behavioral characteristics. In contemplating responses to these problems, however, it is of course important to understand the forces that drive concentrated prescribing. In the learning-by-doing model that we propose, physicians concentrate on a single drug because learning how to properly

manage the dosing and side effects associated with a variety of different drugs would worsen expected health outcomes for their patients while such learning took place. In this setting, policies that either diverted patients to more specialized physicians or lowered the costs associated with learning about how to manage the side effects and dosing of drugs would lead to more personalized medical treatments for patients.

Given the importance of this topic, it is not surprising that the issues we explore have been discussed by others. For example, Coscelli (2000), Coscelli and Shum (2004), and Frank and Zeckhauser (2007) also study concentrated prescribing behavior—empirically and, in some cases, theoretically. In contrast to existing work on concentrated prescribing behavior, our learning-by-doing model offers a new theory of "how physician beliefs form, and (if necessary) how they can be shaped"—issues that are "a key challenge for future research" (Cutler et al., 2013: 28).[1]

In complementary work, focusing on patients instead of physicians, Crawford and Shum (2005) and Dickstein (2012) analyze how a given patient's treatment regime evolves over time. More specifically, whereas we study what a physician can learn from one patient's experience with a given drug to forecast a future patient's experience with that drug, these studies of how a given patient's treatment evolves assume that there is no useful information from one patient's experience to improve the physician's forecast of a future patient's experience. We envision interesting and testable implications from combining our focus on learning across patients with this focus on learning within patients, and we hope that future work will pursue such possibilities.

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; Zhang et al., 2010; Cutler et al., 2013) and by academic clinicians (e.g., Solomon et al., 2003; Schneeweis et al., 2005)—is that the estimated role of physicians in influencing treatment regimes is largely unaffected by incorporating patient-specific data. We find three examples particularly striking. First, 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." Second, 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. Third, Cutler et al. (2013) find that patient demand is "relatively unimportant in explaining variations" in Medicare expenditures across Hospital Referral Regions, whereas "the single most important factor is physician beliefs about treatment." In short, 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 suggests that our results from physician-level data will very likely persist.

1.2 Antipsychotics for the treatment of schizophrenia and related conditions

Schizophrenia is an incurable mental illness 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" (Mosby's Medical, Nursing, & Allied Health Dictionary, 1998: 1456). 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 undergo periods of incarceration (Domino et al., 2004).

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 (American Psychiatric Association, 2004: 9). In the US, Medicaid is the largest payer of medical and drug benefits to people with schizophrenia (Duggan, 2005).

From 1955 through the early 1990s, the mainstays of pharmacological treatment of schizophrenia were *conventional* or *typical* antipsychotic (also called *neuroleptic*) drugs that were more effective in treating the positive than the negative symptoms, but frequently resulted in extrapyramidal side effects (such as tardive dyskinesia—an involuntary movement disorder 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.[2] In 1989, Clozaril (generic name clozapine) was approved by

the U.S. Food and Drug Administration (FDA) as the first in a new 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 (Frank et al., 2004).[3]

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.*" (American Psychiatric Association, 2004: 66, italics added). Learning about such side effects is central to our theoretical model.

Initially these SGAs were perceived as having similar efficacy for positive symptoms and superior efficacy for negative symptoms relative to typicals, but without the older drugs' extrapyramidal and agranulocytosis side effects. However, beginning in about 2001-2002 and continuing to the present, a literature has developed associating 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 all typical and all atypical antipsychotic labels. The labels have been augmented further with warnings regarding antipsychotic treatment of elderly patients with dementia, since evidence suggests

this subpopulation is at greater risk for stroke and death.[4]

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.[5] In comparison, the number of conventional or typical antipsychotic prescriptions fell 55% from 1,100,000 in 1996 to about 500,000 in 2003 and has stabilized at that level since then. 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.

2 Data and initial findings

2.1 Prescription data

Our data on prescribers' behaviors are taken from the IMS Xponent^{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. 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, medical training and office-practice data from the registry at the American Medical Association.

Our data are a cross-section of prescribers in 2007. Although manufacturers received approval to market reformulated versions of several SGAs during the five years leading up to our 2007 sample, no new major antipsychotic products were launched in the US during these years, and 2007 is ten or more years after four of the six atypicals were introduced.

We link the prescriber identifiers in the IMS Xponent^{TM} data base to the American Medical Association ("AMA") Masterfile Directory, which provides education, training,

specialty certification and demographic data on most physicians and type of practice as of 2008. In addition, each prescriber in our sample is assigned a geographical location based on their 2007 location. The resulting dataset includes physicians in various specialties, such as psychiatry (general, child-adolescent, and geriatric), neurology (general, child, and geriatric), primary care (internal medicine, family medicine, pediatrics, and general practice), and others.

Our theoretical approach below applies to a group of physicians each treating patients with symptoms drawn randomly from a fixed distribution, so we hereafter restrict attention to psychiatrists (although our main empirical findings are similar when we analyze the full dataset and interact the main regressors with the aggregate specialty groups above). In addition, to mitigate the possible impact of very low-volume prescribers, for the remainder of the paper we limit the sample to the 2,867 psychiatrists who in 2007 wrote at least 50 prescriptions for an antipsychotic. Again, our main results are qualitatively similar for samples with lower prescription thresholds (e.g., 12 per year) and higher (e.g., 100 per year). In fact, our sample shrinks by only 16% when we shift from the lower prescription threshold of 12 per year to our main sample with threshold 50.[6]

Several final features of the physician data set are also worth noting. First, we have data on only physicians and their prescribing behavior, not on the patients they see. Second, IMS keeps track of prescribers who 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.[7]

2.2 Initial findings on heterogeneously concentraded prescribing behavior

To describe our initial findings, we must first correct for the mechanical bias present in other estimators due to sample sizes (i.e., prescriptions per physician) too small to invoke the law of large numbers. We begin by exploring the deviation of a physician's prescriptions—say, from national market shares for the population of physicians in question. Consider physician i prescribing drug d, and denote the share of prescriptions written by this physician for drug d as s_{id}. Let the national market share of drug d be m_d, where both s_{id} and m_d are between zero and one. As a measure of the deviation of physician i's prescribing behavior

from national market shares, we calculate

$$D_i = \sum\nolimits_d \left(s_{id} - m_d\right)^2 . \tag{1}$$

If every physician had the same prescription behavior, D_i would equal zero. As physician prescribing behavior heterogeneity increases, D_i increases.

Ellison and Glaeser (1997) note that, at small volumes, there will be a mechanical reduction in the deviation measure D_i as volume increases. To correct for this small-volume issue in the deviation measure, they revise the raw deviation measure (1) as follows:

$$\widehat{D}_i = \frac{V_i}{V_i - 1}\left(D_i - \left(1 - \sum\nolimits_d m_d^2\right)\frac{1}{V_i}\right). \tag{2}$$

where V_i is the volume of prescriptions written by physician i. Hereafter we refer to this revised measure of deviation as "corrected deviation."

We document below not only that many physicians have prescription patterns that are statistically significantly different than would be predicted by random draws from national market shares, but also (as an example) that many physicians have prescription patterns that are statistically significantly more concentrated than would be predicted by such random draws. One measure of a prescriber's concentration C_i is the HHI of the physician's prescriptions, $C_i = \sum_d s_{id}^2$. As with deviation, however, at small volumes, there will be a mechanical reduction in the HHI measure of concentration as volume increases. To correct for this small-numbers volume issue in the concentration measure, we amend the raw HHI index as follows:

$$\widehat{C}_i = \frac{V_i}{V_i - 1}\left(C_i - \frac{1}{V_i}\right). \tag{3}$$

Hereafter we refer to this revised measure of concentration as "corrected concentration."[8]

Using these corrected measures of deviation and concentration, we now present two striking initial findings. First, even in our sample of psychiatrists with high prescription volumes, many physicians have prescription patterns that are statistically significantly different than random draws from national market shares.

To assess the statistical significance of such findings, we follow Ellison and Glaeser (1997) by computing the statistic

$$t_i = \frac{D_i - \left(1 - \sum_d m_d^2\right)/V_i}{\sqrt{Var\left(D_i\right)}} \tag{4}$$

for each physician. Given our sample of psychiatrists with high prescription volumes, we can apply the Central Limit Theorem to show that if $|t_i| > 2$ then the null hypothesis

is rejected for psychiatrist i at the 5% level. In our sample, the 5th percentile in the distribution of t_i is 4.85, meaning that for over 95% of the psychiatrists in our sample, we reject the null at the 5% level.

Complementing this finding about the statistical significance of deviation from national market shares, we also show that many physicians have prescription patterns that are significantly more concentrated than would be predicted by random draws from national market shares. To do so, we compute an analogous statistic

$$T_i = \frac{C_i - \left(\frac{V_i - 1}{V_i} \sum_d m_d^2 + \frac{1}{V_i}\right)}{\sqrt{Var\left(C_i\right)}} \tag{5}$$

for each physician. Applying the Central Limit Theorem again gives that if $|T_i| > 2$ then the null hypothesis is rejected for psychiatrist i at the 5% level. In our sample, the 25th percentile in the distribution of T_i is 2.83, meaning that for over 75% of the psychiatrists in our sample, we reject the null at the 5% level.

In addition to these statistical tests, we also argue (although less formally) that, among psychiatrists who deviate from psychiatrist national market shares by a similar amount, there is substantial heterogeneity in prescription patterns. For example, if we (temporarily) limit the sample to psychiatrists with the most concentrated prescribing, i.e. those for whom $|T_i| > 2$ (n=2,348), 54% chose Seroquel as their favorite drug, 28% Risperdal, 10% Abilify, 5% Zyprexa, 1% Geodon, and 1% clozapine, with 1% for typical antipsychotics.[9]

To recap, even in our sample of psychiatrists with high prescription volumes, national market shares for this population do not reflect homogeneous physicians each prescribing drugs in proportions approximating national shares, but rather the aggregate of heterogeneous physicians many of whom are highly concentrated, albeit on different drugs.

3 Towards a theory of prescriber learning and treatment behavior

3.1 A model of prescriber learning-by-doing

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. The time between the arrivals of successive patients is w; so patient t arrives at the physician's office at the point in time tw. The continuous time discount rate is 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 $d^*(s)$, 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).[10] 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.

To formalize the process of learning about complementary actions, we assume that the realized effectiveness of drug d prescribed for patient t with symptom s is

$$b_{sdt} = B_{sd} - (a - x_{dt})^2, \tag{6}$$

where a denotes the complementary action the physician undertakes, and

$$x_{dt} = \theta_d + \varepsilon_{dt}. \tag{7}$$

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 the optimal drug for the patient's symptom, $d^*(s)$, 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.

Recall that the physician knows the maximum potential benefit from each drug B_{sd}. The only uncertainty the physician faces is what complementary actions will work best for

a specific drug and a particular patient. From (7), the ideal complementary actions for a given drug have a patient-specific component ε_{dt} and a general component θ_d. Because the patient-specific components ε_{dt} are independent across patients, there is nothing to learn from, say, ε_{d1} about ε_{dt} for $t > 1$. Rather, the physician uses experience with this drug from past patients to learn the general component θ_d, which is relevant for future patients receiving this drug.

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 the complementary action that would have been optimal (i.e., x_{dt}). Note that the physician does not observe the ideal actions had that patient been given another drug (i.e., $x_{d't}$ for $d' \neq d$) or the ideal actions for another patient given that drug (i.e., $x_{dt'}$ for $t' \neq t$). In short, our assumption gives the physician an unrealistically large amount of information about the patient just treated, but even all this information still leaves the physician with much to learn about the average ideal complementary actions for a given drug (and an average patient), θ_d.

The intuition underlying our model is simple. 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 cannot learn everything she needs to know about a drug from treating one patient with this drug. We have 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 on only how often the physician prescribes drug d, not on the drug or patient identity.

3.2 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 symptoms presented to her.

Our model has the same reduced form as another class of models, also 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 (8) and (9) 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.

In addition, 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 may diverge even if initially they had the same beliefs about the efficacy of each drug. Crawford and Shum (2005), Ferreyra and Kosenok (2011), and Dickstein (2012) estimate models in this spirit, but they do not focus on either concentration or deviation in prescriptions by physicians (not to mention the effects of volume on concentration and deviation).[11]

While bandit models have important applications when physicians are trying to learn the true qualities of drugs, these models are less useful in our setting where physicians need to learn the complementary actions for a drug. In particular, in a two-armed bandit model, if players observe each others' decisions, then eventually all players settle on the same decision with probability one (Aoyagi, 1998). And in our setting, a physician can observe the national market shares of the drugs, which provide that physician information about what other physicians prescribed (and, implicitly, some information about what other physicians learned about the efficacy of various drugs) and still make different prescribing decisions. Thus, a two-armed bandit model would contradict one of our main empirical findings—heterogeneous concentration. More generally, in a multi-armed bandit model, if physicians observe national market shares of all drugs, it is not clear that any of our empirical findings would arise—concentration, deviation, variation in both concentration and deviation with volume, and variation in young physicians' use of old drugs with volume.

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. As a result, observing other physicians' prescription decisions conveys no useful information: 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.[12]

3.3 Analysis of the model and preliminary comparative statics

We assume that the physician seeks to maximize the expected present value of the sum of the realized effectiveness (b_{sdt}) of the drugs she prescribes to the sequence of patients she treats ($t = 1, 2, \ldots$). 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:

$$\begin{aligned} a(h_t) &= \mathbb{E}[\theta_d | h_t], \\ \mathbb{E}[b_{sdt} | h_t] &= B_{sd} - Var(\theta_d | h_t) - \sigma_\varepsilon^2, \end{aligned} \tag{8}$$

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:

$$\frac{1}{Var(\theta_d | h_t)} = \frac{1}{\sigma_d^2} + \frac{\#d(h_t)}{\sigma_\varepsilon^2}, \tag{9}$$

where $\#d(h_t)$ denotes the number of patients to whom the physician prescribed drug d during history h_t. From (8) and (9), 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 – that is, $\mathbb{E}[b_{sdt} | h_t]$ in (8) – depends on d in two ways: the maximum benefit B_{sd}, which is

already known, and the expected loss from imperfect complementary actions, $Var\left(\theta_d|h_t\right)+\sigma_\varepsilon^2$, which depends on the history h_t only through posterior variances $Var\left(\theta_d|h_t\right)$. That is, the physician's prescribing behavior can be summarized by D state variables identified with posterior variances $Var\left(\theta_d|h_t\right)$ 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.

To prepare for the empirical work below, we now discuss comparative-static results of the learning-by-doing model with respect to w, the time between the arrivals of successive 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 an initial sequence of patients each of whom has a given symptom s, then she will choose 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

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

Therefore, the myopic physician compares the difference between $B_{s'd'}$ and $B_{s'd}$ to the difference between $Var\left(\theta_{d'}|h_t\right)$ and $Var\left(\theta_d|h_t\right)$. 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

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

then she will choose drug d.

Now consider physicians who are not myopic. For physicians with extremely high values of w (i.e., very low patient volumes), the prescription behavior that optimizes the expected present value of the realized effectiveness (b_{sdt}) is essentially the myopic behavior just

described. For physicians with lower values of w (i.e., higher volumes of patients), however, optimal prescription behavior now accounts for the fact that learning more about a new drug today will improve effectiveness for future patients given this drug, who may arrive soon. In this sense, high-volume prescribers 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 expect to see lower concentration and lower deviation with increases in patient volume, all else equal.

Finally, as w approaches zero (i.e., the physician sees patients almost continuously), the set of drugs that the physician eventually prescribes ceases 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 physician with very high patient volume 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.

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. 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.

4 Heterogeneous concentration, deviation and prescription volumes

4.1 Empirical framework and econometric methods

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

$$Y_i = \beta \ln V_i + \varphi X_i + \varepsilon_i \tag{10}$$

where Y_i is either the corrected measure of deviation from national psychiatrist market shares in (2) or the corrected measure of concentration in (3), V_i is the number of antipsychotic prescriptions written by psychiatrist i in 2007, and X_i is a vector of covariates described below. We take the age of the psychiatrist from the AMA Masterfile Directory.

As a flexible specification, we use age quartiles (age < 45, $45 <$ age < 54, $54 <$ age < 62, and age > 62) as indicator-variable regressors instead of the raw age metric.

Although our theory has nothing to say about gender, female and male psychiatrists may use this technology in different ways. Therefore we control for the gender of the psychiatrist. In addition, some psychiatrists ask that their prescribing data not be shared with pharmaceutical or other for-profit organizations; thus we add an "opt-out" indicator variable to the specification.

The theory above emphasizes differential learning costs (σ_ε^2 and σ_d^2 in our model), and we might expect the learning costs for psychiatrists to depend on their training and/or current practice environment. In particular, we control for whether the psychiatrist practices in a group or has a solo practice, is hospital-based or not, the population of the county in which the psychiatrist practices (in thousands), and whether the psychiatrist has an MD or DO degree.[13]

Summary statistics for all the dependent and explanatory variables in our analyses are presented in Table 1. In part because we limit our sample to psychiatrists writing at least 50 prescriptions, the raw and corrected concentration and deviation are very similar.

Table 1 about here

4.2 Results: deviation and concentration models

Results from OLS estimation with corrected deviation from national psychiatrist market shares and corrected concentration as dependent variables are reported in Table 2 below. The omitted reference case is a male psychiatrist under age 45 in a group practice that is not hospital-based, who has not opted out of the AMA Masterfile registry and has an MD degree. We present parameter estimates on only the $\ln V_i$, age quartiles and female gender variables; estimates for the other covariates are available upon request.[14]

Table 2 about here

As seen in the first column of Table 2, in the corrected deviation model the coefficient on $\ln V_i$ is negative and statistically significant, consistent with our learning-by-doing theoretical framework. Psychiatrists in the two oldest age quartiles are more deviant than psychiatrists under age 45, although the relationship is not monotonic with age. While female psychiatrists are very slightly less deviant than males, this effect is not statistically significant.

Also consistent with our theoretical framework, corrected concentration declines with increases in prescribing volume, and significantly so. Although none of the age quartile coefficient estimates is statistically significant, female psychiatrists are slightly more concentrated in their antipsychotic prescribing behavior than are their male counterparts.

4.2.1 Robustness

We have undertaken a number of robustness checks. For example, we repeated the analysis allowing volume to enter in different ways. In particular we estimated the model using $1/V_i$ as well as just linear V_i. Although neither of those fit the data as well as $\ln V_i$, the negative sign of the estimated effect of volume on psychiatrist deviation and HHI concentration prescribing was robust.

We also experimented with using different antipsychotic prescribing frequency cutoffs when constructing our dataset we use in our regressions (at least 12, 75, and 100 in 2007). The sign of the estimated effect of volume on psychiatrist deviation and concentration remained unchanged.

We also explored whether similar results hold if we include physicians from all of specialties in our full dataset. In this regression specification we interacted the volume measure with physician-specialty dummies (psychiatrist, neurologist, primary care, and other) thereby allowing flexibility in how high and low volume physicians in different specialties differ from one another. Consistent with Table 2, we found that both psychiatrist deviation from national market shares and psychiatrist HHI concentration were negatively related to volume. In addition, we found negative estimated effects for the other specialties. In fact, in this full dataset (and with the lower prescription threshold of 12), the volume effects were even stronger and the R^2 was more than twice as high. The estimated volume effects were generally largest for "other" physicians, followed by primary care physicians, then neurologists, and smallest for psychiatrists. One plausible interpretation of this last finding is that greater residency training concerning antipsychotics to some extent substitutes for physician experience gained by prescribing a certain type of drug more frequently.

5 Exploring a competing hypothesis: detailing

There may be alternative frameworks that help explain the variations we observe in physician prescribing behavior. One plausible competing hypothesis involves pharmaceutical sales representatives (called "detailers") who may target their sales efforts at high-volume

prescribers. More specifically, suppose that, rather than high-volume prescribing behavior generating greater prescription heterogeneity through the logic of our learning-by-doing model, high-volume prescribers were instead 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). In this competing hypothesis it is the increase in detailing that leads to less concentrated prescribing by high-volume physicians—perhaps because some detailers provide persuasive information, or because writing a few prescriptions for each detailed drug is a reciprocal form of behavior providing some positive feedback from the prescriber to the various detailers.[15]

In evaluating this competing hypothesis, it is important 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 few incentives for its manufacturer to detail physicians: the brand would be unable to appropriate many benefits, which for the most part would instead accrue to the generics.[16]) An implication is that drugs having lost market exclusivity many years ago are unlikely to have been detailed to young doctors practicing in 2007.

In order to compare the predictions of the competing hypothesis (that physician detailing drives heterogeneous prescribing behavior) to the predictions of our learning-by-doing model, we distinguish "old" antipsychotic drugs approved and launched in the US before 1990 (i.e., Clozaril and all the typical antipsychotic drugs) from "new" antipsychotic drugs (i.e., all SGA atypical antipsychotics, the earliest of which was Risperdal, approved in 1993). Note that the ten typical drugs prescribed in our 2007 sample were approved for marketing by the FDA between 1957 and 1984, and Clozaril (also an 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. Thus, any effect that detailing of old drugs might have on prescriptions in 2007 would be a very long-run effect. More importantly, for our purposes, such detailing of old drugs could not have been targeted at physicians who were not practicing when this detailing occurred, before 1996.

To highlight prescriber-age issues, in this section we limit the sample to the oldest (age 62 and up in 2007) and youngest (age below 45 in 2007) age quartiles of psychiatrists writing at least 50 antipsychotic prescriptions in 2007 (1,410 psychiatrists). The youngest quartile of psychiatrists were between the ages of 17 and 33 in 1996, when the last old drug experienced generic entry. These youngest psychiatrists are therefore very unlikely ever to

have been detailed on an old drug. In contrast, the oldest quartile of psychiatrists were age 51 and older in 1996. These oldest psychiatrists therefore could have been detailed on an old drug.

If pharmaceutical detailing were the primary driver of psychiatrists' choice of which antipsychotic class of drugs to prescribe (old vs. new), then the youngest psychiatrists should prescribe very few of the older drugs. In addition, we would expect high-volume young psychiatrists (who are likely visited the most by pharmaceutical detailers promoting new drugs) to be the least likely to prescribe older drugs.

In contrast, under our learning-by-doing model, the share of old drugs prescribed by young psychiatrists should *increase* with volume for high enough volumes (see the Appendix): in our framework, high-volume young psychiatrists have an incentive to invest in learning the complementary actions for old drugs because these drugs deliver the highest benefits for some (albeit a small minority of) patients. On the other hand, young psychiatrists with low volumes typically have insufficient incentive to invest in learning the complementary actions for old drugs (because the first set of patients they encountered typically had symptoms best treated by the new drug, so they prescribed the new drugs and learned about their complementary actions).

As the dependent variable we employ the psychiatrist's share of total antipsychotic prescriptions written for the old drugs, where the share ranges from zero to 100. (Since in our data this share sometimes is 100, we employ Tobit rather than OLS estimation.) If the detailing hypothesis were the primary driver of prescriber choice, for young psychiatrists the old share would increase with volume. If instead our learning-by-doing model were the primary driver of prescriber choice, for young psychiatrists the old share would *increase* with volume. To allow for differential volume effects by age, we specify a model with an interaction between volume and age >62. The results are shown in Table 3 below.

Consistent with both the detailing and learning-by-doing hypotheses, Table 3 shows that the oldest psychiatrists prescribe larger shares of the old drugs, but that this share decreases with an older psychiatrist's prescription volume. More importantly, since the estimated effect of the $\ln V_i$ is positive and significant, the highest volume psychiatrists in the youngest quartile prescribe a larger share of old drugs. This is consistent with our learning-by-doing framework, but is at odds with the detailing hypothesis, because these youngest high-volume psychiatrists are likely to have been heavily detailed on new drugs, but are likely never to have been detailed on the old drugs.

Table 3 about here

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 psychiatrists: *High-volume young psychiatrists prescribe old drugs more often than do low-volume young psychiatrists.*

6 Discussion and conclusion

We conclude by (1) exploring the connection between our results and the literature on regional-variation, (2) discussing possibilities for future work, and (3) summarizing the paper.

6.1 Regional variation

Our 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 (e.g., Skinner and Fisher, 1997; Fisher et al., 2003a,b; Yasaitis et al., 2009). Could it be that our findings of heterogeneous concentration are driven by correspondingly large between-region variability in antipsychotic prescribing behavior? Alternatively, is most variability in antipsychotic prescribing behavior psychiatrist-specific, with regional patterns similar to each other?

To analyze regional variation, we return to our full sample of 2,867 psychiatrists writing at least 50 antipsychotic prescriptions in 2007. We compute mean HHIs and their variability (both standard deviations and coefficients of variation) at alternative levels of regional aggregation. While most of these regional aggregates are familiar, 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 4.

As expected from our results in Section 2.2, prescriber behavior is more concentrated at the individual level than are national market shares (mean corrected HHI falls from 0.28 at the individual level to 0.19 nationally), with monotonically decreasing concentration as one considers larger aggregate regions (from county to HRR to state to nation). Also consistent with Section 2.2, there is more variability in prescribing behavior across psychiatrists at the individual level than across larger aggregate regions: the coefficient of variation of corrected

HHI is 0.39 at the individual level but falls monotonically to 0.10 at the state level, and the mean corrected deviation (defined in Section 2.2) likewise falls monotonically from 0.077 at the individual level to 0.007 at the state level. Phelps (1992: 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 close to high variability at the individual-prescriber level, moderate variability at the county level, and low variability at the HRR and especially at state levels.

Table 4 about here

In short, in our data, practice heterogeneity is much greater at the individual level than at the HRR level: the coefficient of variation of corrected HHI is almost twice as high at the individual level as at the HRR level, almost reaching Phelps' "high" threshold for the former and barely exceeding his "low" ceiling for the latter. Cutler et al. (2013) offer related evidence that the strongest determinant of regional variation is differences in physician beliefs about the efficacy of particular therapies.

6.2 Future work

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 remain controversial issues, even after publication of a substantial number of articles over the last decade, including those based on randomized controlled clinical trials.[17] 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 from34% 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 by different prescribers 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 regulatory-related information dissemination strategies.[18]

In addition to the dynamics of prescribing behavior for individual physicians, it would be interesting to study the dynamics of such behaviors for groups of physicians. In this paper, our model and empirics ignored 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 geographical 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 national efforts to enhance the practice of evidence-based medicine.[19]

Finally, our findings (and those of others) suggest that a significant proportion of the heterogeneity in the treatments patients receive depends upon physician preferences concerning treatment regime. Our model and empirics focus on the roles of initial patients and future volume in determining a physician's prescription pattern. It would also be informative and useful to identify other physician experiences that generate differences in "practice style" across physicians, perhaps related to location of medical residency training, analogous to recent investigations characterizing "management style".[20]

6.3 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 (i.e., the benefit to the patient if the optimal complementary actions are taken), 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 to certain 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. 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. In short, both concentration and deviation decrease with volume.

We have confronted this model with cross-sectional data on antipsychotic prescriptions, regressing corrected deviation and corrected concentration on the volume and other characteristics of psychiatrists writing at least 50 antipsychotic prescriptions in 2007. As predicted by our model, we observe that higher-volume psychiatrists have lower deviation and concentration in their prescribing behavior.

To compare our learning model to a model of detailing by sales representatives to psychiatrists, we regress the share of prescriptions written for new drugs on the psychiatrist's age quartile, total volume of antipsychotic prescriptions written, and the interaction of the two. Consistent with our learning model but at odds with the detailing model, we find that the higher volume psychiatrists in the youngest age quartile prescribe a larger share of old drugs than do their lower volume counterparts.

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}. 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. All authors contributed equally to this research; the alphabetical ordering is by convention.

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

[4] 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.

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

[6] 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.

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

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

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

[10] 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.

[11] Cutler, David M., Jonathan Skinner, Ariel Dora Stern and David Wennberg [2013], "Physician Beliefs and Patient Preferences: A New Look At Regional Variation in Health Care Spending", Cambridge, MA: National Bureau of Economic Research, Working Paper No. 19320.

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

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

[14] 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.

[15] 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.

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

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

[22] 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.

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

[24] 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.

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

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

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

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

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

[34] 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.

[35] 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.

[36] 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.

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

[38] Solomon, Daniel H., Sebastian Schneeweis, Robert J. Glynn, Raise Levin and Jerry Avorn [2003], "Determinants of Selective Cyclooxygenase-2 Inhibitor Prescribing: Are

Patient or Physician Characteristics More Important?", *American Journal of Medicine*, 115:715-720.

[39] 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.

[40] 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.

[41] 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.

[42] 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.

[43] Zhang, Yuting, Katherine Baicker and Joseph P. Newhouse [2010], "Geographic Variation in Medicare Drug Spending", *New England Journal of Medicine*, 363(5):405-409.

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 time between the arrivals of successive 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 3.2 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 $T = 2$. Designating (x, y) to mean that a physician is prescribing fraction x of d_1 and fraction y of d_2, where $x + y = 1$, we then have the following prescription behaviors as a function of the colored and shaded regions in Figure A1.

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

For old physicians, concentration falls with volume, the number of atypicals increases with volume, and the share of atypicals increases with volume. For new physicians, concen-

tration 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).

Notes

[1]Coscelli does not use a formal model. Coscelli and Shum use a two-armed bandit model. Frank and Zeckhauser informally discuss a "Sensible Use of Norms" hypothesis also based on a bandit model. See Section 3.2 for a comparison of bandit models to our approach. Frank and Zeckhauser also discuss 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). They interpret their My Way hypothesis as physicians "engaging in some highly suboptimal therapeutic practices" (p. 1125), whereas in our model such heterogeneous concentration by physicians may be optimal. Finally, none of these alternative models generates predictions about either (a) the effect of volume on concentration and deviation or (b) the prescriptions for old drugs by young physicians.

[2]In an earlier version of this manuscript (Taub et al., 2011), we included in our analyses among the typical antipsychotics an old drug named prochlorperazine (Compazine) that was FDA approved both for treatment of schizophrenia and 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. The drugs we therefore count as typical antipsychotics are fluphenazine, haloperidol, loxapine, molindone, pimozide, perphenazine, thioridazine, thiothixene, chlorpromazine, and trifluoperazine.

[3]For a history of clozapine and discussion of antitrust issues raised by the laboratory test results requirement, see Crilly (2007).

[4]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".

[5]We will use "prescribed", "written" and "dispensed" interchangeably, but the IMS Health Xponent data are based on dispensed prescriptions; for a variety of reasons, a physician may 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.

[6]Results for these alternative samples are available upon request.

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

[8]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

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

[9]In comparison, in our 2007 sample of high-volume psychiatrists, the national market percentages of the six atypicals were Seroquel 30%, Risperdal 24%, Abilify 15%, Zyprexa 11%, Geodon 8%, and clozapine 3%, with 9% for typical antipsychotics.

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

[11]More specifically, Crawford and Shum (2005) and Dickstein (2012) use patient-level data, so they can analyze a patient's learning but not a prescriber's concentration. In contrast, Ferreyra and Kosenok (2011) share our focus on prescriber learning and analyze prescriber data, but they focus on learning to prescribe a single new drug, rather than on the steady-state concentration or deviation of prescriptions.

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

[13]DO is doctor of osteopathy. Mosby's Medical, Nursing, & Allied Health Dictionary (1998: 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." Although the vast majority of psychiatrists in our sample have an MD degree, a number have DO training and degree.

[14]In the corrected-deviation regression, the only coefficient estimate that is significantly different from zero is that for the hospital-based psychiatrist: -0.010, standard error of 0.004. In the corrected-concentration regression, the only statistically significant estimates are those for the hospital-based psychiatrist (-0.012, standard error 0.006) and that for the opt-out psychiatrist (-0.038, standard error 0.010).

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

[16]For discussion and empirical evidence, see Berndt et al. (2003).

[17]Among the more notable publications are those based on the CATIE study; see, for example, Lieberman et al. (2005), White (2006) and Kraemer et al. (2009).

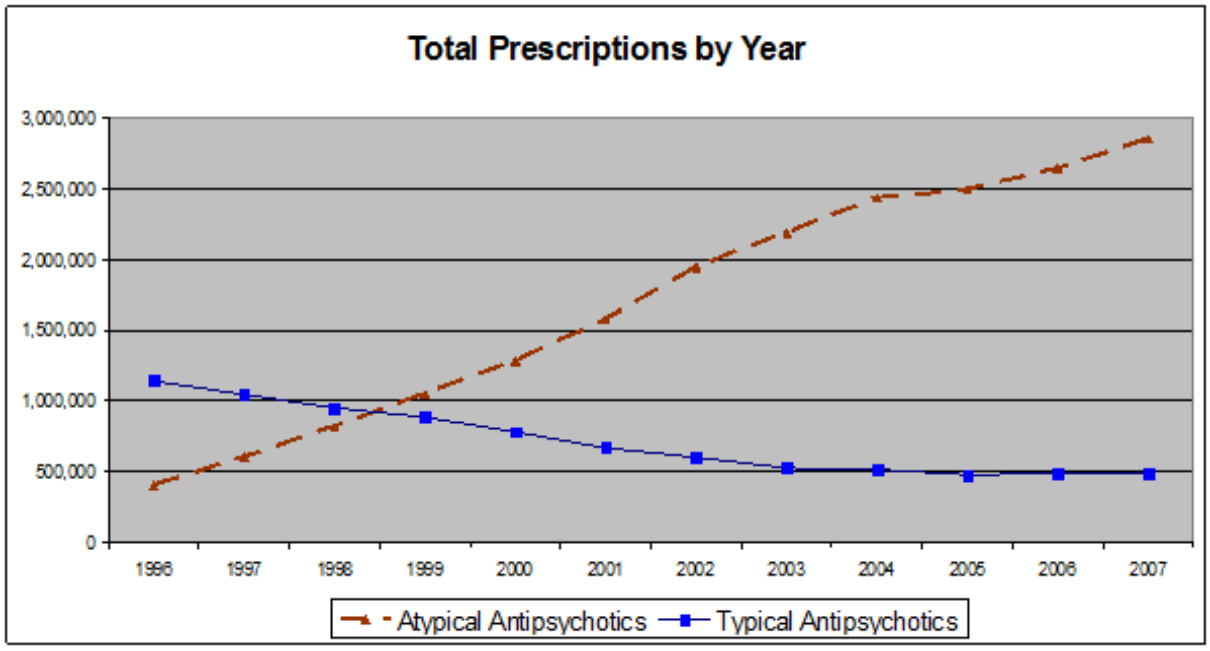
[18]The only research on this topic of which we are aware is that by Hoblyn et al. (2006).

[19]For an attempt to incorporate spillovers from"close knitted prescribers" in the context of antipsychotic prescribing, see Domino et al. (2012).

[20]See, for example, Bertrand and Schoar (2003) and Kaplan et al. (2008).

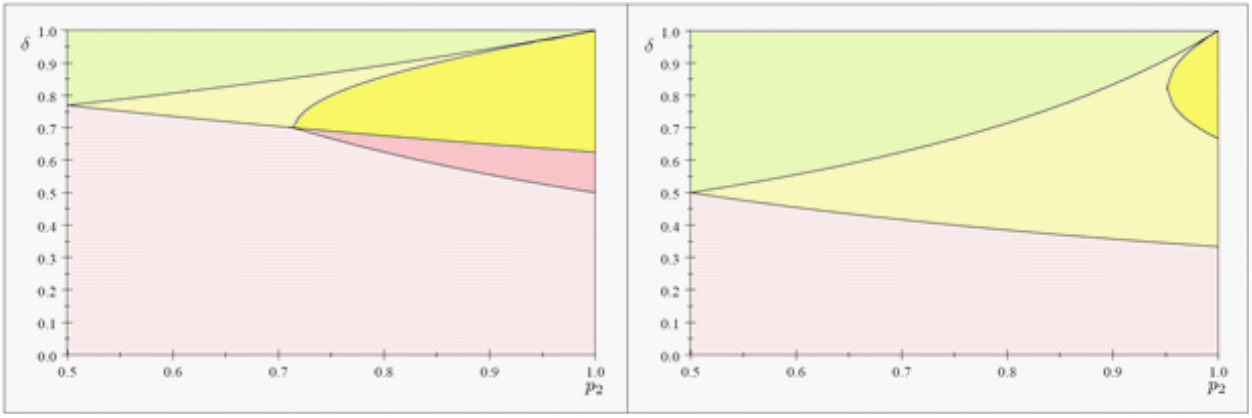
Figures

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: Summary Statistics for 2007 Prescriber Sample

Variable	Mean	Std. Dev.	Minimum	Maximum
HHI of Individual Physician's Antipsychotic Prescribing	0.29	0.11	0.12	0.95
Corrected HHI of Individual Physician's Antipsychotic Prescribing	0.28	0.11	0.12	0.95
Deviation of Physician's Antipsychotic prescribing from Nat. Mkt. Shares	0.08	0.08	0.002	0.86
Corrected Deviation of Physician's Antipsychotic prescribing from Nat. Mkt. Shares	0.08	0.08	-0.01	0.86
Total Yearly Antipsychotic Prescriptions	726	849	50	7186
Prescriber Age	53	12	28	92
Solo Practice	0.16	0.37	0	1
Population (county) in thousands	1,240	1,840	3	9,735
Female	0.34	0.47	0	1
Hospital Based Physician	0.14	0.34	0	1
DO Flag	0.04	0.18	0	1
Physician Opt Out	0.04	0.19	0	1
Number of Observations	2,867			

All values calculated using IMS Health Incorporated Xponent™ general prescriber sample 2007 data for psychiatrists writing at least 50 antipsychotic prescriptions.

Table 2: OLS Estimates of Corrected Deviation and Corrected Concentration

	Corrected Deviation	Corrected HHI
Log(Total Yearly Antipsychotic Prescriptions)	-0.025***	-0.034***
	[0.001]	[0.002]
Age Quartile 45-53^	0.006	0.004
	[0.004]	[0.005]
Age Quartile 54-61^	0.016***	0.007
	[0.004]	[0.005]
Age Quartile 62+^	0.011***	-0.006
	[0.004]	[0.006]
Female^	-0.001	0.010**
	[0.003]	[0.004]
Number of Observations	2,867	2,867
R^2	0.13	0.14
^ indicates dummy variable		

Dependent Variables: Corrected Deviation of Psychiatrist Antipsychotic Prescribing from National Psychiatrist Market Shares as in (2), Corrected Concentration of Psychiatrist's Antipsychotic Prescribing Shares as in (3). Standard errors in square brackets. *** p< 0.01, ** p<0.05, * p<0.10

Table 3: Tobit Estimates of Percentage of Prescriptions for Old Drugs

	% Rxs for New Drugs
Physician Age 62+^	13.278***
	[4.023]
Log(Total Yearly Antipsychotic Prescriptions)	2.252***
	[0.476]
(Physician Age 62+)^(Log(Total Yearly Antipsychotic Prescriptions))	-1.331**
	[0.659]
Female^	-2.157***
	[0.812]
Number of Observations= 1,410	
Pseudo R^2= 0.0123	
Left Censored = 0 Right Censored = 221	
Mean of dependent variable	*11.01*
^ indicates dummy variable	

* , **, ***, indicate significance at the 10%, 5%, and 1% respectively. All values calculated using IMS Health Incorporated Xponent™ general prescriber sample 2007 data, and population estimates from the US Census Bureau. New drugs are defined as SGA atypicals. Sample is comprised of the Oldest (62 +) and youngest (28-44) quartile of psychiatrists.

Table 4: Means, Standard Deviations and Coefficients of Variation for Antipsychotic HHIs: Alternative Geographical Aggregates, 2007

Geographic Aggregate	Mean Corrected HHI	HHI Std. Dev.	Coef. of Variation	Mean Corrected Deviation	Deviation Std. Dev.
Individual Prescriber	0.28	0.11	0.39	0.077	0.080
County	0.24	0.08	0.32	0.041	0.051
Hospital Refferal Region	0.22	0.05	0.22	0.023	0.036
State (plus District of Columbia	0.20	0.02	0.10	0.007	0.010
Nation	0.19	na	na	na	na

IMS Health Incorporated Xponent™ 2007 data general prescriber sample data. Includes all psychiatrists who wrote at least 50 prescriptions for antipsychotics in 2007.