

## Information, learning, and drug diffusion: The case of Cox-2 inhibitors

Pradeep K. Chintagunta · Renna Jiang ·  
Ginger Z. Jin

Received: 2 July 2008 / Accepted: 14 July 2009 / Published online: 11 August 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** The recent withdrawal of Cox-2 Inhibitors has generated debate on the role of information in drug diffusion: can the market learn the efficacy of new drugs, or does it depend solely on manufacturer advertising and FDA updates? In this study, we use a novel data set to study the role of learning in the diffusion of three Cox-2 Inhibitors—Celebrex, Vioxx and Bextra—before the Vioxx withdrawal. Our study has two unique features: first, we observe each patient’s reported satisfaction after consuming a drug. This patient level data set, together with market level data on FDA updates, media coverage, academic articles, and pharmaceutical advertising, allows us to model individual prescription decisions. Second, we distinguish across-patient learning of a drug’s general efficacy from the within-patient learning of the match between a drug and a patient. Our results suggest that prescription choice is sensitive to many sources of information. At the beginning of 2001 and upon Bextra entry in January 2002, doctors held a strong prior belief about the efficacy of Celebrex, Vioxx, and Bextra. As a result, the learning from patient satisfaction is gradual and more concentrated on drug-patient

---

P. K. Chintagunta  
Booth School of Business, University of Chicago, Chicago, IL, USA  
e-mail: pradeep.chintagunta@ChicagoBooth.edu

R. Jiang (✉)  
Graduate School of Management, University of California at Davis, Davis, CA, USA  
e-mail: rejiang@ucdavis.edu

G. Z. Jin  
Department of Economics, University of Maryland, College Park, MD, USA  
e-mail: jin@econ.umd.edu

G. Z. Jin  
NBER, Cambridge, MA, USA

match than on across-patient spillovers. News articles are weakly beneficial for Cox-2 drug sales, but academic articles appear to be detrimental. The impact of FDA updates is close to zero once we control for academic articles, which suggests that FDA updates follow academic articles and therefore deliver little new information to doctors. Two counterfactual experiments are carried out to quantify the influence of information on market shares.

**Keywords** Learning · Drug diffusion · Prescription choice · Patient satisfaction

**JEL Classifications** D8 · I1 · M3 · C5

## 1 Introduction

Once a prescription drug is approved by the Food and Drug Administration (FDA), the market will generate thousands or even millions of instances of that drug's use in a short period of time. Can the market learn the efficacy and safety of new drugs, or does it solely depend on manufacturer advertising and FDA updates? This question is not only important for the everyday practice of prescribing, it also determines the optimal design of the FDA's post-marketing surveillance system, which is under intensive debate given the recent withdrawal of blockbuster drugs. In this paper, we use a novel data set to quantify how a doctor learns from patient experience, while controlling for other information channels including manufacturer advertising, FDA updates, and published articles.

In particular, we distinguish two types of physician learning: when a patient reports her drug use experience to a doctor, the reported information may reflect the drug's average quality or the patient's idiosyncratic match with the drug. The main task of the doctor is to decipher these two components as the former is applicable to all patients but the latter is only useful for the reporting patient. Throughout the paper, we label the learning of a drug's average quality as across-patient learning and the learning of a patient-drug match as within-patient learning.

Existing studies have focused on either across-patient learning (Ching 2005; Coscelli and Shum 2004; Narayanan et al. 2005) or within-patient learning (Crawford and Shum 2005) but not both. We believe the two types of learning are linked: doctors are not only uncertain about the average quality of a drug, they also have imperfect information on the specific match between a drug and a patient. Both uncertainties are embodied in one single report of patient satisfaction, hence ignoring any one of them is likely to introduce estimation bias in a model that accounts only for one of the two effects. In the model specified below, we show how across- and within-patient learning are mathematically linked in a Bayesian updating process.

Our empirical analysis focuses on Cox-2 Inhibitors, a new class of pain killers that underwent dramatic changes in a period of 6 years. Cox-2 Inhibitors

belong to a broader class of drugs called non-steroidal anti-inflammatory drugs (NSAIDs). Prior to the introduction of Cox-2 Inhibitors, patients typically were treated with traditional NSAIDs medication. Between 1998 and 2001, the FDA approved three Cyclooxygenase-2 (Cox-2) Inhibitors: Celebrex (Dec. 1998), Vioxx (May. 1999), and Bextra (Nov. 2001). All of them were heavily advertised as safer alternatives to then existing pain killers. By September 2004, the class had more than 10 million patients, annual sales had reached \$6 billion in 2003, and total advertising dollars spent in 2003 were as high as \$400 million. After a clinical trial associated Vioxx with severe cardiovascular (CV) risks, Merck withdrew the blockbuster drug in September 2004. CV risks and enhanced concerns on skin irritation led to the withdrawal of Bextra in April 2005. As of today, Celebrex is the only Cox-2 Inhibitor remaining on the market, with warnings added in April 2005.

Because the Vioxx withdrawal is likely to raise concerns about the other Cox-2 inhibitors,<sup>1</sup> we believe the role of information has changed dramatically before and after the Vioxx withdrawal. To better characterize the learning of new drugs, this paper focuses on the prescription decisions made before the end of 2003. The nine-month lag between the end of 2003 and the Vioxx withdrawal should be long enough to avoid any contamination from the withdrawal decision.

To empirically distinguish across- and within-patient learning, we use a unique data set obtained from a marketing research company, IPSOS. IPSOS tracked a nationally representative sample of patients from 1999 to 2005. Not only did IPSOS report every NSAIDs prescription received by the sampled patients (including traditional NSAIDs and Cox-2s), it started to keep a longitudinal record of patient satisfaction with these prescriptions from January 2001. These satisfaction measures, together with FDA updates, media coverage, academic articles, and manufacturer advertising, allow us to associate individual prescriptions with various sources of information.

Note that information content may differ across sources: for example, heart attack is rare and often urgent when it occurs. Patients that suffered from such an adverse event may not have time and opportunity to report this in the next doctor's visit. However, these events may be reviewed in an article published later on in the mass media or in academic journals. The accumulation of such events may also lead to some FDA actions. In comparison, minor side effects such as stomach upset and skin rash are noticeable to individual patients and are more likely incorporated in their satisfaction report. These potential differences motivate us to treat each information source differently.

Compared with the existing literature, our data are better-suited to modeling the across- and within-patient learning because we observe patients' satisfaction signals. Equipped with the patient satisfaction data, we assume doctors held a prior belief about Cox-2 inhibitors at the end of 2000, which summarizes all the information up to 2000. Starting Jan. 2001, doctors received

---

<sup>1</sup>The first official claim of CV risks being a class effect was documented by FDA in April 2005.

patient satisfaction information on a daily basis and used them to form posterior beliefs in a Bayesian fashion. To our knowledge, all the existing studies on drug learning have no direct panel data on patient feedback signals. Instead, authors assume that the unobserved signals conform to a given, i.e., assumed, statistical distribution. They then model prescription choice as a result of random draws from that distribution. Since we observe the realization of feedback signals, we can (a) impose fewer identification restrictions, (b) eliminate the computational burden of using simulation to integrate out the unobserved signals; and (c) make the model more parsimonious by eliminating the need to estimate the true drug qualities.

Despite the benefits associated with our data, they are still imperfect for integrating across- and within-patient learning because we do not observe physician identities. Thus, we need to make assumptions on the mechanism by which information is shared across patients. In particular, we assume that doctors in the same geographic area (in our case, census division) exchange opinions and learn from each others' patients' experiences. Since we do not observe physician identities, one can also think of this as assuming that patients directly share information and learn from each other's satisfaction within a geographic area. To mitigate the effect of arbitrary assumptions regarding the geographic area of information exchange, we investigate the scope of information pooling by changing the definition of geographic area and assessing model fit.

Our second contribution to the literature lies in collecting factors other than patient satisfaction that could potentially influence a doctor's prescription decision. Specifically, we allow FDA updates, manufacturer advertising,<sup>2</sup> news reports and academic articles to enter the utility function directly and therefore influence doctors' relative preference across drugs.<sup>3</sup> These data allow us to distinguish the impact of patient satisfaction from other factors.

Our results suggest that prescription choice is sensitive to many factors. At the beginning of 2001 and upon the Bextra entry in January 2002, doctors held a strong prior belief about the efficacy of Celebrex, Vioxx, and Bextra. As a result, the learning from patient satisfaction is gradual and more concentrated on drug-patient match than on across-patient spillovers. We also find that advertising and news articles are positively correlated with drug sales but academic articles appear to be detrimental. The impact of FDA updates is close to zero once we control for academic articles, which suggests that FDA updates follow academic articles and therefore deliver little news to doctors.

To better understand the relative importance of patient satisfaction versus academic articles, we conduct two counterfactual experiments: one expands the sharing of patient satisfaction from census division to nationwide, and the

---

<sup>2</sup>Which includes detailing, direct-to-consumer advertising, and journal advertising.

<sup>3</sup>These factors are too lumpy (i.e., only observed at the aggregate market level) relative to the patient level prescription data to accurately identify learning from these sources in a Bayesian updating framework. For more details please see Section 4.

other doubles the counts of academic publications. Both experiments attempt to capture the spirit of existing proposals for the reform of the FDA's post-marketing surveillance.

For example, Slater (2005) has urged the FDA to set up a nationwide database (via computer-assisted prescribing of bar-coded medications) to establish a rapid link between an event and prescription. In fact, private efforts such as [Sermo.com](#) has already facilitated nationwide sharing of patient experience among doctors. In a more comprehensive proposal, Ray and Stein (2006) suggest setting up the Center for Drug Information, which "coordinates the communication of accurate, unbiased information to practitioners and patients that promotes the use of drugs in accordance with the best available data." They argue that third parties such as academia have much less conflict of interest (than drug manufacturers) in marketing and could improve prescription practice by academic detailing. Indeed, the medical literature has proven the effectiveness of academic detailing which involves the face-to-face education of prescribers by pharmacists, physicians and nurses who are not compensated by the pharmaceutical company. It also involves mailing doctors a series of "unadvertisements" based on academic findings (Avorn and Soumerai 1983). Though doubling the count of academic publications is less realistic than the existing methods of academic detailing, it helps us quantify the effect of academic detailing and compare it directly to nationwide sharing of patient satisfaction.

The counterfactual predictions suggest that setting up a nationwide database of patient feedback encourages doctors to switch from traditional NSAIDs to Cox-2s, but increasing academic publications about Cox-2s steers market share away from Cox-2s. This suggests that patient feedback and academic articles may reflect different dimensions of drug quality, and hence do not substitute for each other.

The rest of the paper is organized as follows. Section 2 provides detailed information on the background of Cox-2 Inhibitors. Section 3 describes and summarizes the data. Section 4 presents the econometric model. In Section 5, we report empirical estimates, discuss robustness checks, and perform counterfactual predictions. Conclusions are offered in Section 6.

## 2 Background

Cox-2 Inhibitors were initially introduced to reduce the gastrointestinal (GI) risks of conventional non-steroidal anti-inflammatory drugs (NSAIDs) while maintaining the same efficacy in pain relief. Traditional NSAIDs, such as Aspirin, ibuprofen (Motrin) and naproxen (Naprosyn), block Cox-1 and Cox-2 enzymes and therefore impede the production of the chemical messengers (prostaglandins) that cause inflammation. However, since some Cox-1 enzyme exists in the stomach and its production of chemical messengers protects the inner stomach, blocking Cox-1 enzymes tends to reduce the mucus lining of the stomach, causing GI problems such as stomach upset, ulceration,

and bleeding. In comparison, the Cox-2 enzyme is located specifically in the areas that cause inflammation and not in the stomach. By selectively blocking the Cox-2 enzyme, Cox-2 inhibitors have the potential to reduce GI risks.<sup>4</sup>

Before FDA approval, clinical trials presented evidence that all three Cox-2s (Celebrex, Vioxx and Bextra) reduce the incidence of GI ulcers visualized at endoscopy compared to certain non-selective NSAIDs. But up to April 2005, only Vioxx demonstrated a reduced risk for serious GI bleeding in comparison with naproxen (FDA 2005). After FDA approval, all three Cox-2s were heavily marketed as being equally effective as traditional NSAIDs but with less adverse effects on the GI system.

The diffusion of Cox-2 inhibitors was very fast: according to the National Ambulatory Medical Care Survey (NAMCS) and the National Hospital Ambulatory Medicare Care Survey (NHAMCS), in 1999 (the first year of Cox-2 introduction), the number of ambulatory visits resulting in Cox-2 prescriptions were 15 million, slightly more than half of the visits that resulted in the prescriptions of traditional NSAIDs. By the end of 2000, the number of Cox-2 visits had exceeded those for traditional NSAIDs, reaching an estimate of 31.5 million. This growth continued in 2001, but at a much lower rate (Dai et al. 2005, Table 2).

In terms of prescriptions, according to the New Product Spectra (NPS),<sup>5</sup> the total number of new Cox2 prescriptions grew sharply from 61,066 in January 1999 to 2 million in December 2000, but leveled off after January 2001. The number of all Cox2 prescriptions (including new and old) demonstrated a similar pattern. Since Bextra was not approved until November 2001, its introduction was mainly market stealing (from Celebrex and Vioxx) rather than market expanding.

As NPS does not track drugs beyond 5 years of the launch, it does not cover Celebrex after 2003 and does not tell us the prescription trends for traditional pain-relievers. To develop a rough understanding of these trends, we plotted the monthly count of prescriptions observed in the individual-level IPSOS data for each Cox-2 as well as for traditional pain relievers as a whole by aggregating over individual prescriptions in each month. Although the number of individuals included in IPSOS is much smaller than those in the NPS, the diffusion patterns of Cox-2s between 1999 and 2003 obtained were very similar to those obtained from the NPS above. The aggregate IPSOS data also suggest that Cox-2s initially stole some market share from traditional pain killers, but the whole market expanded considerably between 2000 and 2003 before returning to the 1999 level at the end of 2005. The most obvious decline started in 2004 and accelerated with the withdrawal of Vioxx and Bextra.

---

<sup>4</sup>For a complete layperson description of Cox-2 inhibitors, readers can refer to [www.medicinenet.com](http://www.medicinenet.com).

<sup>5</sup>NPS is a database provided by IMS Health that tracks monthly number of prescriptions (new and refills) dispensed by pharmacists and monthly advertising activities of pharmaceutical manufacturers up to 60 months after initial launch.

**Table 1** Regulatory history of Cox-2s

Brand	Date	FDA decision
Celebrex	Dec. 31, 1998	Approval for rheumatoid and osteoarthritis
	Dec. 23, 1999	Approval for reducing the number of intestinal polyps in patients with familial adenomatous polyposis (FAP)
	Jun. 7, 2002	Labeling change because the Celecoxib Long-term Arthritis Safety Study (CLASS) does not associate Celebrex with significantly less GI risk as compared to traditional NSAIDS (ibuprofen or diclofenac)
	Dec. 23, 2004	Issuing a Public Health Advisory on an increased cardiovascular risk in association with Cox-2s and traditional NSAIDS based on long-term clinical trials
Vioxx	Apr. 7, 2005	New labeling that highlights cardiovascular risk
	May 20, 1999	Approved for osteoarthritis and pain
	Apr. 11, 2002	Approved for rheumatoid arthritis New warnings concerning reduced GI risk and increased cardiovascular risk based on the Vioxx Gastrointestinal Outcomes Research (VIGOR)
Bextra	Sept. 30, 2004	Withdrawal (voluntary by Merck)
	Nov. 16, 2001	Approved for osteoarthritis and rheumatoid arthritis
	Nov. 15, 2002	New warnings on life-threatening skin irritations
	Dec. 9, 2004	More warnings on skin irritations and cardiovascular risk
	Dec. 23, 2004	Issuing a Public Health Advisory on an increased cardiovascular risk in association with Cox-2s and traditional NSAIDS based on long-term clinical trials
	Apr. 7, 2005	Withdrawal (by Pfizer)

After a 3-year placebo-controlled clinical trial<sup>6</sup> showed that taking Vioxx 25 mg once daily doubles the risk of serious adverse cardiovascular (CV) events, Merck withdrew Vioxx on September 30, 2004. In April 2005, FDA's Arthritis and Drug Safety and Risk Management Advisory Committees reviewed the available data and concluded that (1) the increased CV risk is a class effect applying to all the Cox-2s and traditional NSAIDS; (2) Aside from the CV risk, Bextra is associated with an increased rate of serious and potentially life-threatening skin reactions and should be withdrawn from the market; (3) the overall benefits of Celebrex exceeded its potential risks, which allowed Celebrex to remain on the market but the label had to be revised to carry explicit warnings on potential CV and GI risks (FDA 2005). The FDA did not rank the three Cox-2s by their CV risks, but the evidence underlying the withdrawal requests suggests that the overall quality of Celebrex was better than the other two, with Vioxx being better than Bextra since only the latter was associated with skin irritations.

The adverse information about Cox-2 did not come all at once. Before the final withdrawal of Vioxx and Bextra, the FDA had taken several decisions regarding the side effects of each Cox-2 brand. As shown in Table 1, FDA initiated a label change for Celebrex in June 2002 because a long term clinical trial could not distinguish the amount of GI risk between Celebrex

<sup>6</sup>The clinical trial was called the Adenomatous Polyp Prevention on Vioxx (APPROVe).

and traditional NSAIDs (ibuprofen or diclofenac). This reverses the original understanding that Celebrex is safer because of lower GI risks. In comparison, Vioxx received new warnings about increased cardiovascular risk as early as April 2002. The first FDA warning of skin irritations applied to Bextra on Nov. 2002, and more Bextra warnings came in Dec. 2004 for both skin irritations and cardiovascular risk. One task of our study is to detect whether these FDA updates have any impact on the prescription decisions made by doctors before the Vioxx withdrawal.

### 3 Data summary

This section describes our data sources, summarizes the raw data, and presents simple data patterns that suggest across- and within-patient learning.

#### 3.1 Data description

We combine four data sources: (1) patient-level prescription and satisfaction data from the IPSOS patient diary database (IPSOS-PD), (2) monthly advertising expenditures obtained from the New Product Spectra (NPS) database, (3) the number of news articles covering Cox-2s derived from Lexis-Nexis for the period 1999 to 2005, and (4) the number of academic articles covering Cox-2s from Medline from 1999 to 2005.

In 1997, IPSOS created a national representative sample of 16,000 households and tracked their drug purchasing month by month.<sup>7</sup> The patient diary covers all the individuals within the sampled household. Each individual, if observed in the data, is viewed as one patient. Each record in the patient-level IPSOS data corresponds to one purchase of ethical drugs, including prescription and over-the-counter medications. The data used in this paper include all the individual records that IPSOS collected on traditional NSAIDs as well as on Cox-2s from January 1999 to December 2005.

Each record provides information on the patient's prescription date, age, sex, race, household income, education, copay, insurance status, and residential location defined by nine Census divisions and more than 200 DMAs (Designated Market Areas). Since over 80% of patients have health insurance and the self-reported copays are noisy and sometimes inconsistent with the reported drug insurance, we ignore price/copay information but include insurance status in the empirical analysis.

Specifically, IPSOS collects information on three types of insurance variables: i) a simple indicator of whether the patient has health insurance or not at the time of prescription (referred to as HEALTHINS); ii) an indicator of whether the patient has an insurance plan outside of Medicare or Medicaid

---

<sup>7</sup>Detailed description is available at <http://www.ipsos.ca/product.cfm?id=66&name=Healthcare&fn=health&fl=reid> and Bowman et al. (2004).



(referred to as INSPLAN); and iii) an indicator of whether the patient has any coverage for drug insurance (referred to as DRUGINS). One puzzling aspect of the data is that the correlations among the three insurance variables are between 0.12 to 0.24, which is not as high as expected. However, as we see later, they do seem to have some power explaining prescription behavior. We include all three variables in the model but only as controls. Our conversations with drug companies and insurers suggest that a majority of insurers excluded all Cox-2s from preferred formulary tiers.<sup>8</sup> If this applies to every insurer, the lack of formulary information should not undermine our estimation results, although it may explain why drug insurance makes a difference in the prescription choice between Cox-2s and non-Cox2s.

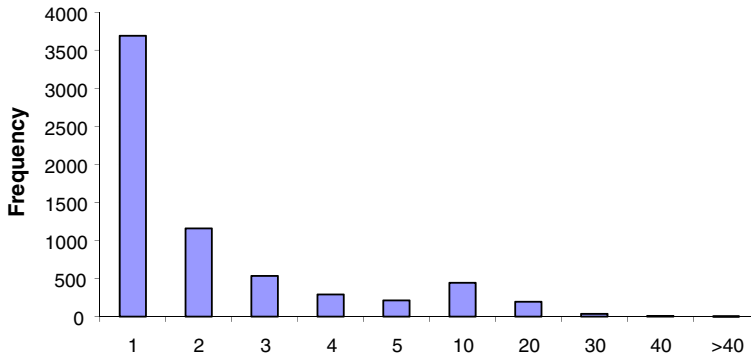
Starting from January 2001, the data also provide five satisfaction measures, reflecting patients' self reports on the effectiveness of the prescribed drug, its side effects, whether the drug works quickly, how long it lasts, and whether it is easy to take. Each satisfaction measure is obtained on a scale from 1 to 5, with 1 denoting extremely satisfied and 5 denoting extremely dissatisfied. Answers to these questions are likely to reflect the effects that are easily observable to patients (such as pain relief, stomach upset or skin irritation) but not heart attack or other life-threatening events. In this sense, the patient satisfaction data do not necessarily capture all the patient experience information conveyed to the doctor and our learning analysis is subject to this limitation.

The 1999–2005 IPSOS sample involves 28,601 patients and 136,950 observations of traditional NSAIDs and Cox-2s. Since many traditional NSAIDs (say Motrin) are available over the counter, we focus on prescriptions only. Out of the 57,942 filled prescriptions, 20.3% are for Celebrex, 13.6% for Vioxx, 3.9% for Bextra and the rest 62.2% for traditional NSAIDs. To ensure that this sample is indeed nationally representative, we calculate the number of COX-2 prescriptions and drug-specific market shares from the sample and compare their trends with those reported in the NPS. They are similar. We also regress the number of new COX-2 patients in our sample and the number of new COX-2 prescriptions in the NPS on various advertising variables, the regression coefficients and significance are comparable. These results reassured us about proceeding with the IPSOS data.

The sample is further reduced to 8,077 patients and 27,326 prescriptions after we (1) focus on the records with non-missing values in all five satisfaction questions, (2) delete observations that have missing Census division indicators, and (3) restrict the sample to 2001–2003 when advertising data are available from NPS. The reduction is largely due to the fact that IPSOS did not collect

---

<sup>8</sup>According to <http://www.fingertipformulary.com/>, which allows us to search for a drug's current formulary status by drug name and state in a hundred of medical plans, many traditional NSAIDs including Naproxen, Indomethacin, Diclofenac and Etodolac are listed in the most favorable formulary tier (Tier 1) while Celebrex is listed in the least favorable tier (Tier 3). Because the website only reports formulary as of the search time (January 4, 2008 in our case), we don't have data on the historical formulary status of Celebrex. Nor do we have data on Vioxx or Bextra because they have been withdrawn from market.



**Fig. 1** Number of Rx's per patient. Source: IPSOS patient diary data on NSAIDS prescriptions. Total 6,577 patients and 17,329 prescriptions

satisfaction data until 2001. Between 2001 and 2003, the reporting rate for satisfaction measures is 94.8%.<sup>9</sup> To best fit a model of how doctors learn from patient satisfaction, we focus on new patients that first appear in the data set on or after January 1, 2001. The main reason for discarding old patients is because doctors may have formed patient-specific priors based on their satisfaction before 2001, on which we have no information. However, the experiences of older patients may have contributed to doctor beliefs about average drug quality as of January 1, 2001, which will be captured in the model since we estimate the prior as of January 1, 2001.<sup>10</sup> Fortunately, there are not too many old patients: 6,577 out of the 8,077 patients (with non-missing satisfaction scores) are new since 2001, and these new patients account for 17,329 prescriptions.

We define a “run” as a sequence of one or more prescriptions of a single drug. For example, if a patient receives a prescription sequence A,A,A,B,C, we say that he has three runs, the length of each being 3, 1, 1. By this definition, the final sample of 17,329 prescriptions are classified into 7,998 runs. An average run consists of 2.17 prescriptions, and an average patient has 1.22 runs in our data.<sup>11</sup> By definition, new patients are likely to have fewer runs and fewer prescriptions per run, which explains why the number of prescriptions declined by 36.6% when we exclude old patients but the number of patients only goes down by 18.6%.

Conditional on the final sample of 6,577 new patients and 17,329 prescriptions, Fig. 1 shows that 56% of the patients were involved with prescription NSAIDs only once, and the vast majority (96%) occurred no more than 10 times. Table 2 presents the number of prescription switches between tradi-

<sup>9</sup>From 2001 to 2003, there are 28,866 NSAIDs prescriptions, of which 27,359 report all five satisfaction scores.

<sup>10</sup>What is ignored is how the old patients' satisfaction reported after 2001 contributes to the across-patient learning after 2001. We leave this issue for future research.

<sup>11</sup>The corresponding numbers are 2.37 and 1.23 in Crawford and Shum (2005).

**Table 2** Switching matrix

Frequency	Celebrex	Vioxx	Bextra	AllothRx	Switching rate
Celebrex	1949	100	41	237	0.16
Vioxx	96	1598	47	229	0.19
Bextra	25	22	389	71	0.23
AllothRx	228	228	97	5395	0.09

Note: Switches are from row to column. Conditional on patients who started after Jan. 1, 2001 and had at least two NSAIDS prescriptions before the end of 2003

tional NSAIDs and the three Cox-2s. By definition, a switch does not occur unless a patient has at least two prescriptions. On average, the switching rate of traditional NSAIDs (9%) is lower than that of Celebrex (16%), Vioxx (19%) and Bextra (23%). This is partly because we aggregate different brands of traditional NSAIDs into one category.

Table 3 summarizes satisfaction scores by drug and the five satisfaction questions. On average (across all five questions which we denote as  $satisf_{12345}$ ), patients are more satisfied with all three Cox-2s than they are with traditional NSAIDs, although the specific satisfaction for effectiveness is the lowest for Bextra. Within Cox-2s, Celebrex is the best in all five questions, with Vioxx being the worst in side effects and Bextra the worst in the other four. These patterns are hardly significant at conventional levels, but they are consistent with the fact that FDA kept Celebrex on the market but requested the withdrawal of Vioxx based on cardiovascular risk and the withdrawal of Bextra based on both cardiovascular risk and severe skin irritation. Another possible interpretation of why Bextra has the worst satisfaction score is that those who got Bextra are those who are more resistant to other Cox-2s and doctors prescribed Bextra to them as the last resort.

The five satisfaction measures are highly correlated (the correlations range from 0.87 to 0.97), so we will use their average  $satisf_{12345}$  in the final models. Averaging across the five satisfaction measures also allows us to smooth the discreteness in a single measure and therefore to get closer to the distributional assumption we need to make in the Bayesian model. We will revisit this issue when we present the structural results.

**Table 3** Mean and standard deviation of satisfaction scores (total: 6,577 patients 17,329 observations)

	allothRx	Celebrex	Vioxx	Bextra
Effectiveness	1.90 (1.11)	1.81 (1.03)	1.83 (1.03)	1.94 (1.13)
Side effects	1.98 (1.14)	1.81 (1.06)	1.89 (1.12)	1.82 (1.06)
Works quickly	2.03 (1.10)	1.94 (1.04)	1.99 (1.06)	2.00 (1.09)
How long does it last?	2.04 (1.11)	1.93 (1.04)	1.96 (1.06)	1.98 (1.06)
Easy to take	1.51 (1.01)	1.38 (0.87)	1.38 (0.89)	1.40 (0.92)
Average effectiveness ( $satisf_{134}$ )	1.99 (1.04)	1.89 (0.97)	1.93 (0.98)	1.97 (1.04)
Average across five ( $satisf_{12345}$ )	1.90 (0.94)	1.77 (0.87)	1.81 (0.88)	1.83 (0.91)

(1=extremely satisfied, 5=extremely dissatisfied)

One may argue that self-reported patient satisfaction does not necessarily reflect the true experience because patients may adjust their report according to other information available about the drug even if their real experience does not vary over time. To address this possibility, we compare the average satisfaction of all Cox-2 prescriptions before and after June 2002. Since FDA changed the label of Celebrex in June 2002 and issued a Vioxx warning in April 2002, June 2002 roughly captures the first official occurrence of adverse information for Cox-2 inhibitors. We find that the average satisfaction before and after June 2002 is quite similar (1.8530 vs. 1.8553) and the conclusion remains the same if we restrict the calculation to the 809 patients that appeared both before and after June 2002 (1.8018 vs. 1.8072).

All three Cox-2s were heavily marketed. The average monthly advertising expenditures (pooling detailing, journal advertising, and direct-to-consumer advertising) were 20.3M, 21.4M, and 10.5M dollars during the time period of 2001 to 2003 for Celebrex, Vioxx and Bextra, respectively. Although not reported here, the flow of advertising expenditure was comparable across drugs and even over time. Also, the trend of total advertising is quite similar to the trend of total prescriptions described previously. Since traditional NSAIDs involve a large number of brands and most of them had been on the market for a long time, we do not obtain advertising data for traditional NSAIDs. This is equivalent to assuming traditional NSAIDs have zero advertising since the start of our sample period.

To address the possibility that news and scientific evidence may affect drug sales (Azoulay 2002; Venkataraman and Stremersch 2007), we count the number of news and journal articles related to Cox-2s from 1999 to 2005. Specifically, news articles are obtained from the Lexis-Nexis search of keywords Cox 2, Cox-2, Cox2, celebrex, vioxx, bextra, Cyclooxygenase-2, Cyclooxygenase2, and Cyclooxygenase 2 across all the U.S. newspapers and magazines. For each relevant article, we record title, publication date, publication region, and the news source. To focus on Cox-2 inhibitors, we delete articles that talk about Cox-1 and Cox2 enzyme but not inhibitors. Lexis-Nexis classifies articles into four regions: Midwest, Northeast, Southeast and Western. They are matched with the nine Census divisions (used in the IPSOS data) by the standard Census definition.<sup>12</sup> To account for the fact that some newspapers and magazines are read more often than others, we obtain the total circulation from the Audit Bureau of Circulations. Whenever applicable, we distinguish circulation on weekdays, Saturday and Sunday, and use the one that matches best with the publication date of the article. Articles that do not specify source or do not have circulation data for the specified source are excluded.

---

<sup>12</sup>Except for Southeast, the names of Lexis-Nexis regions match perfectly with those of Census regions. We interpret "Southeast" in Lexis-Nexis as "South" in Census. The crosswalk between the four Census regions and the nine Census divisions is available at [www.census.gov](http://www.census.gov).

From article titles, we define dummy variables indicating whether the article sounds negative, positive or neutral. For example, “Cox-2s increase the risk of ..” is counted negative but “Celebrex is easier on stomach” is counted positive. If the title includes both positive and negative words (or neither), it is counted neutral. The article title also tells us whether the article focuses on a particular Cox-2 brand or not. If yes, the article is only matched with the specific brand. If no, the article is presumably applicable to all the Cox-2s available on the market. In total, the Lexis-Nexis search results in 973 articles with valid circulation information, which includes 92 positive, 122 neutral and 756 negative articles.

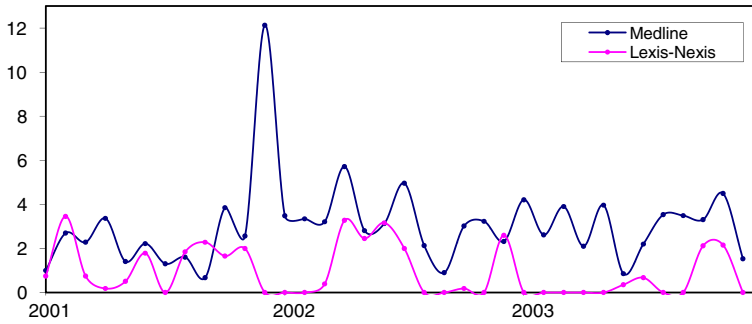
Academic articles about Cox-2 are gathered from Medline search of the same keywords, covering all the domestic and international journals in Medline. For each search result, we record title, abstract, publication date, and the name of the publishing journal. To focus on human subjects, we rule out articles that examine Cox-2 effects on animals only. Since most Medline journals are monthly or bi-monthly, we take the first day of the first issue month as the publication date. For example, both “April” and “April–June” issues are coded as published on April 1. Medline offers no regional distinction and more than 80% of articles do not focus on a specific brand name, so we assume all the non-specific articles applies to all the Cox-2s available on the market. The brand-specific articles are applied to the mentioned brand only.

Medline journals also differ greatly in terms of impact. To address this, we weigh each journal with the 2002 Science Gateway Impact Factor.<sup>13</sup> In total, we collect 1064 medical articles between 1999 and 2005, 950 of which have a valid impact factor. Missing impact factor is imputed by the mode of all the non-missing impacts. Like in Lexis-Nexis, we use title and abstract to classify Medline articles into negatives (13.44%), positive (28.19%) and neutral (58.36%). Note that the percent of negative titles is much lower for Medline articles than for news reports (78%). This suggests that the main effect of Medline articles is likely to come from the non-negatives. To simplify estimation, we pool positive and neutral as non-negatives but distinguish negatives and non-negatives for both types of articles.

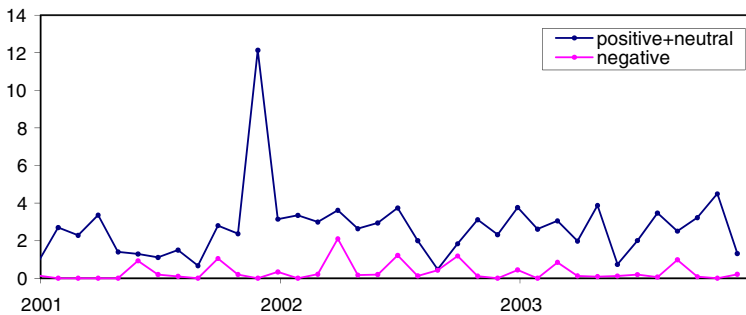
As a robustness check, we also record whether article authors are affiliated with a pharmaceutical company, a university, or other institutions, and whether the article talks about efficacy, side effects, or both. These variables are highly correlated with each other: for example articles affiliated with pharmaceutical companies are more likely to be non-negative and focus on efficacy. The high correlation prevents us from identifying the impact of each variable separately. Instead, we focus on negatives and non-negatives in the main specification, but discuss the effects of the other variables via a robustness check.

Figure 2 plots the weighted monthly counts where the weight is circulation for news articles and journal impact factor for Medline articles. Figures 3 and 4 decompose article counts into negative and non-negatives. One pattern

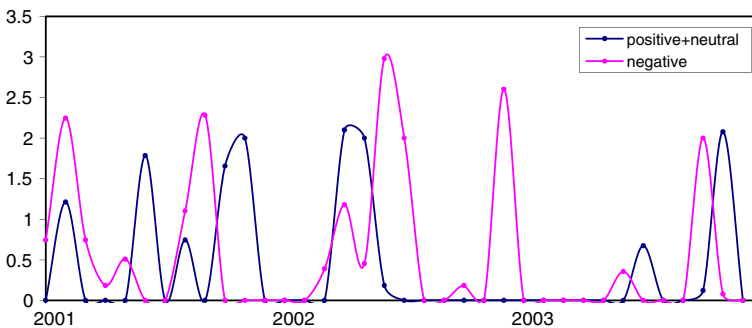
<sup>13</sup><http://www.sciencegateway.org/impact/if02a.html>



**Fig. 2** Total articles weighted (1 Lexis-Nexis=one Wall Street Journal article, 1 Medline=one JAMA article)



**Fig. 3** Medline articles weighted by impact factor (1=one JAMA article)



**Fig. 4** Lexis-Nexis articles weighted by circulation (1=one Wall Street Journal article). Source of Figs. 2–4: Lexis-Nexis 1999–2005 for news articles. Medline 1999–2005 for journal articles. News articles are weighted by newspaper circulations reported by the Audit Bureau of Circulations ([www.accessabs.com](http://www.accessabs.com)). Journal articles are weighted by the 2002 impact factor from Science Gateway (<http://www.sciencegateway.org/impact/if02a.html>). Positive, neutral and negative are defined by authors' reading of article title and abstract

that stands out most is the dramatic difference before and after 2004. Before the Vioxx withdrawal, the 1999–2003 period was characterized by occasional news and journal articles, in stable flow, and at most times non-negative in nature. In 2004 and 2005, huge spikes of negative news appear around the Vioxx withdrawal, the first lawsuit against Vioxx, and the withdrawal of Bextra. Medline articles also show a negative spike at the beginning of 2005, which we interpret as a lag effect of the Vioxx withdrawal in Sept. 2004. Based on these figures, we suspect the learning process may have changed substantially after the Vioxx withdrawal. In this paper, we focus on the pre-withdrawal period (2001–2003), while leaving the post period (2004–2005) for future research.

Finally, on the basis of Table 1, we create three dummy variables to indicate the FDA updates that occurred in our analysis period (2001–2003). Namely, new warnings added on Apr. 11, 2002 for Vioxx, new warnings added on Nov. 15, 2002 for Bextra, and label change as of Jun. 7, 2002 for Celebrex.

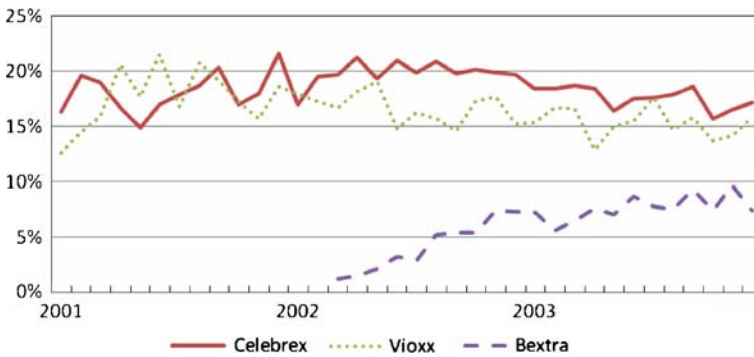
So far we have documented five sources of information: patient satisfaction, manufacturer advertising, news articles, Medline articles, and FDA updates. The time-series correlation across the five categories is no more than 0.3.<sup>14</sup> Such low correlation suggests that different sources may contain different types of information and it is possible to identify their impacts separately in a single model.

### 3.2 Basic evidence of learning

Since patient satisfaction is unique to our data, it is important to demonstrate its link with prescription decisions. In particular, if doctors learn anything from patient satisfaction, market shares should become more stable over time and satisfaction scores should correlate with drug market shares and drug switches within patient. To confirm this intuition, Fig. 5 presents the market share evolution in our sample period (January 2001 to December 2003) for Celebrex, Vioxx, Bextra, and all other NSAIDs prescriptions separately. The market share of traditional NSAIDs has dropped from over 70% to roughly 60% in March 2002 and remains stable afterwards. Similarly, Celebrex and Vioxx fluctuates a little bit less over time while Bextra market share picks up from 0% upon introduction to slightly below 10% at the end of 2003. These evolutionary patterns are consistent with learning. Figure 6 plots the evolution of average patient satisfaction, which shows no obvious up- or downward trends during 2001–2003. Consistent with the lack of change in the average satisfaction score before and after June 2002 (as reported in Section 3.1), this suggests that there is little evidence of patients adjusting their satisfaction report based on the sales or FDA updates of Cox-2 inhibitors. It is also comforting to note that, by

---

<sup>14</sup>One complication in calculating the correlations is that all these measures are in different time units. We choose to aggregate up to a monthly level, and then calculate correlations among the monthly level quantities.



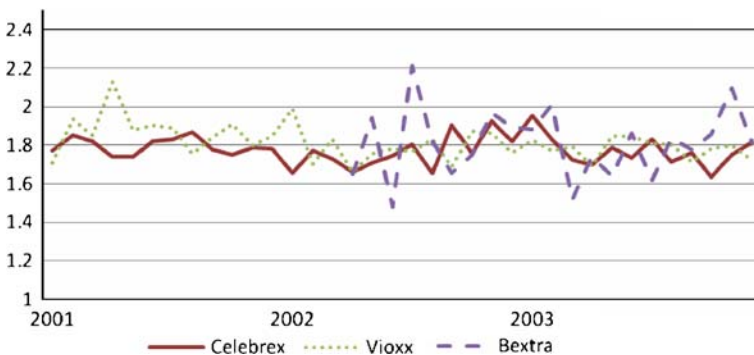
**Fig. 5** Cox-2 shares (6,577 patients, 17,329 Rx's)

the end of 2003, the order of average satisfaction score (as shown in Table 3) is consistent with the order of market shares within the three Cox-2 inhibitors (Celebrex>Vioxx>Bextra).

We then run a logit regression on whether the drug prescribed to patient  $p$  in time  $t$  is different from  $p$ 's last prescription (changes within the non-Cox2 NSAIDs are counted as non-switch). The key independent variable is the satisfaction scores patient  $p$  reported for the drug taken on the last prescription. Since this regression focuses on drug switch, we exclude first prescription (per patient) from our cleaned data, which leaves 2,887 patients and 13,637 prescriptions in the logit sample.

As shown in Table 4 Column (1), the more satisfied a patient is with the current prescription (i.e. the lower score of  $satisf_{12345}$ ), the less likely she switches to other brands. Decomposing satisfaction into different dimensions, Column (2) shows that the key effect of satisfaction is driven by drug efficacy ( $satisf_{134}$ ) instead of "side effects" ( $satisf_2$ ) or "easy to take" ( $satisf_5$ ).

Table 4 Column (3) adds other sources of information into the switch regression. Since advertising may potentially have an s-shape impact on drug diffusion, we use the inverse of the cumulative total advertising expenditure



**Fig. 6** Satisfaction (1=extremely satisfied)



**Table 4** Logit model on brand switching

	(1) Coefficients [ <i>t-stat</i> ]	(2) Coefficients [ <i>t-stat</i> ]	(3) Coefficients [ <i>t-stat</i> ]
Overall satisfaction (satisf12345)	0.1269*** [4.1958]		0.1406*** [4.5999]
Drug effectiveness (satisf134)		0.1318*** [3.0934]	
Side effects (satisf2)		0.0399 [1.0477]	
Ease to take (satisf5)		-0.0666* [-1.7272]	
Inverse of advertising			0.8509* [1.8357]
Medline article (neg)			-0.6724** [-2.4953]
Medline article (non-neg)			1.0848** [2.5531]
Lexis article (neg)			0.037 [0.7903]
Lexis article (non-neg)			-0.0852* [-1.7889]
After FDA updates			0.0088 [0.0572]
Intercept	-2.1172*** [-33.1160]	-2.1241*** [-32.4455]	-5.3796*** [-3.4811]
Log likelihood	-4189.9	-4186.9	-4092.7
# of patients	2887	2887	2887
# of prescriptions	13639	13639	13639

Dependent variable: switch=1 if switch brand from the last period (taking all traditional NSAIDS as one brand)

Independent variables: satisfaction scores (1=extremely satisfied, 5=extremely dissatisfied) reported in the last period, advertising (measured as the inverse of cumulative sum of total advertising expenditures up to the previous month), article counts (measured as the log of cumulative sum of weighted articles up to the previous day), and FDA update dummy

Source: IPSOS patient diary data 2001–2003

T-statistics in brackets. \*\*\* $p < 0.01$ , \*\* $p < 0.05$ , \* $p < 0.10$

since FDA approval (i.e. detailing + journal advertising + DTC advertising). This mimics the reciprocal model of advertising in the marketing literature (Lilien et al. 1992). Results are qualitatively similar if we use the total advertising in linear form. Aside from advertising, we also include Medline and Lexis-Nexis article counts up to  $t$ , and whether  $t$  is after the FDA update for the drug of  $p$ 's last prescription. The coefficient of  $satisf_{12345}$  is comparable to that in Column (1). As we expect, advertising and non-negative news articles deter switch but the other coefficients are either insignificant (the FDA update dummy and negative news articles) or counterintuitive (the negative and non-negative Medline articles). Note that this regression focuses on the information related to the last prescription taken by the same patient but ignores information of other available brands. This shortcoming will be corrected in our full model.

Another unique feature of our study is the distinction between across- and within-patient learning. Does the raw data contain evidence for both types of

learning? The simplest way to demonstrate across-patient learning is tracking nationwide market shares by drug-month. If across-patient learning exists, the market shares should stabilize over time. To quantify the stabilization, we compute the standard deviation of the monthly market share within 2001, 2002, and 2003 separately for each drug. Although not shown, we find that the standard deviation of monthly share declines year by year for all drugs, suggesting that the market shares become more stable over time.

Because we do not observe the identity of the doctor, we have to assume that the across-patient information is shared within a specific geographic area. In the IPSOS data, the most detailed geographic area that yields a sufficient number of prescriptions for information sharing is census division. If information sharing is restricted to within each of the nine census divisions, we should observe significant heterogeneity of market shares across regions. In contrast, if information sharing is nationwide, market shares should be homogenous across regions. To test for these two extremes, we regress the number of prescription at a month-drug-division level on a full set of drug dummies and a full set of division dummies. The joint test of all division dummies having the same coefficient is rejected with a p-value less than  $1e-4$ . A more detailed look at the division coefficients suggest that each division is different from another, which motivates us to model across-patient learning by census division.

A careful reader may still wonder whether the observed heterogeneity of market share reflects demographic heterogeneity across division rather than distinctive learning within each division. Unfortunately, the regression reported above is conducted at the month-drug-division level, which makes it difficult to control for patient heterogeneity. However, as shown below, our full model examines the degree of learning *after* controlling for individual demographics including gender, age, education, income, and three measures of insurance status. Under that structure, we find that the model estimated with division-wide learning has a significantly better fit to the data than the model of nationwide learning. We should have found the opposite if the market share heterogeneity across divisions were solely attributed to the difference in observable demographics.

To better detect within-patient learning, we examine the number of switches in different phases of treatment. Taking each patient as the unit of analysis, we find that the number of switches in the first half of a patient's treatment regimen is always greater than the number of switches in the second half. This suggests that significant learning has taken place within each patient.

## 4 Econometric model and identification

### 4.1 Model

Consider a situation in which doctor  $d$  has concluded that patient  $p$  needs a pain relieving prescription of a fixed length starting from time  $t$ , but has not

determined which drug is the best choice. More specifically, the choice set includes traditional NSAIDs and whatever Cox-2s that are available at  $t$ . In making such choice, the doctor maximizes the patient’s expected utility for this *single* prescription.

Here we make three assumptions: in reality the doctor-patient relationship involves a number of information and incentive issues, and the doctor may not act as a perfect agent for the patient. We ignore such imperfections because we have no data on individual doctors. Second, we consider all the traditional NSAIDs as one drug and do not distinguish brands within this group. The main reason is that traditional NSAIDs involve dozens of brands and we do not have advertising and article reports for each specific brand. Treating traditional NSAIDs as one outside good helps us focus on the tradeoff between traditional NSAIDs and the three brands of Cox-2 Inhibitors. Third, we assume that each doctor is myopic and focuses on the current prescription. As detailed below, we assume that a doctor considers all the drug information available to her up to  $t$ , but she does not consider how experience learned from the current prescription would affect her future prescription choice on the same or other patients. For more discussion on forward-looking behavior, see the robustness checks section (and Crawford and Shum 2005).

We assume that patient  $p$ ’s CARA utility from a prescription of drug  $j$  can be written as:<sup>15</sup>

$$\tilde{V}_{pjt} = \left[ 1 - e^{-\gamma(\tilde{Q}_{pj} + \beta_{xj}X_{pt} + \beta_z Z_{jt} + \epsilon_{pj})} \right] / \gamma$$

where

- $\tilde{Q}_{pj}$  = doctor’s belief about drug  $j$ ’s quality for patient  $p$  at time  $t$ ;
- $\gamma$  = risk aversion parameter, non-negative. A zero  $\gamma$  implies risk neutrality;
- $X_{pt}$  = patient  $p$ ’s characteristics at time  $t$ ;
- $Z_{jt}$  = drug  $j$ ’s characteristics at time  $t$ ;
- $\epsilon_{pj}$  = extreme value error.

The information process is modeled as follows. Doctors are uncertain about  $\tilde{Q}_{pj}$ , which can be decomposed into two parts: the general quality of drug  $j$  that applies to every patient (referred to as  $Q_j$ ); and the specific match value between drug  $j$  and patient  $p$  (referred to as  $q_{pj}$ ). The true effect of drug  $j$  on patient  $p$  is therefore

$$Q_{pj} = Q_j + q_{pj}.$$

This term is fixed but unknown to the doctor or the researcher. Over the entire population, we assume  $q_{pj}$  is independent and identically distributed according to a normal distribution  $N(0, \sigma_{q_0}^2)$ .

<sup>15</sup>As  $\gamma$  goes to zero, the utility function becomes linear.

When drug  $j$  is first introduced to the market (or at the beginning of our data set), all doctors share two priors: for the general quality of drug  $j$ , the prior is

$$\tilde{Q}_{j0} \sim N\left(\bar{Q}_{j0}, \sigma_{\bar{Q}_{j0}}^2\right).$$

The prior for the patient-drug match ( $q_{pj}$ ) is mean independent of  $Q_{j0}$  and can be written as:

$$\tilde{q}_{pj0} \sim N\left(0, \sigma_{q_0}^2\right).$$

Together, the prior for the specific quality of drug  $j$  on patient  $p$  is

$$\tilde{Q}_{pj0} = \tilde{Q}_{j0} + \tilde{q}_{pj0} \sim N\left(\bar{Q}_{j0}, \sigma_{\bar{Q}_{j0}}^2 + \sigma_{q_0}^2\right).$$

We allow both  $\bar{Q}_{j0}$  and  $\sigma_{\bar{Q}_{j0}}$  to be drug-specific. This reflects the fact that the initial information about the average drug quality, whether it is from FDA guidelines, medical research, or patient experience, may differ across drugs. For example, the prior on Celebrex and Vioxx is defined as of January 1, 2001 and the prior on Bextra is defined as of March 1, 2002 (the first date that Bextra appears in our data set). Since doctors may have learned about Celebrex and Vioxx before 2001, the prior should be less dispersed for them than for Bextra. Since we put no restrictions on  $\sigma_{\bar{Q}_{j0}}$ , we can test this conjecture in the data. For simplicity, we assume the amount of patient heterogeneity (captured by  $\sigma_{q_0}$ ) is the same across all three drugs. We assume that doctors prior belief on the distribution of patient heterogeneity coincides with the actual distribution.

We assume doctors located in the same geographic area (say a Census region, a Census division, or a DMA) share information immediately and extensively. Assuming each prescription generates one signal, patient  $p$ 's satisfaction with drug  $j$  at time  $t$ , denoted as  $R_{pjt}$ , is a noisy but unbiased indicator of the true quality:<sup>16</sup>

$$R_{pjt} = \alpha_0 + \alpha_R \cdot (Q_j + q_{pj}) + v_{pjt}$$

where  $\alpha_0$  and  $\alpha_R$  equalize the scales of  $R$  and  $Q$ , and the signal noise  $v$  conforms to  $N(0, \sigma_v^2)$ .

Let  $n_{pjt}^R$  denote the number of satisfaction reports from patient  $p$  on drug  $j$  up to time  $t$ , and  $\bar{R}_{pjt}$  denote the average satisfaction across these  $n_{pjt}^R$  reports. At time  $t$ , doctors in the same area will use all the  $n_{pjt}^R$  signals across all local patients to update their beliefs on the average drug quality  $Q_j$ . However, because patients are independent from each other, the experience of patients other than  $p$  does not contain any information about  $q_{pj}$ .

With all the patient satisfaction information up to  $t$ , doctor's posterior on the effect of drug  $j$  on patient  $p$  can be decomposed into two parts: (1) doctor's

<sup>16</sup>Chan et al. (2007) do not observe the realization of signals. They use patients' stated switching reasons to create boundary conditions for those signals.

posterior about the general quality of drug  $j$ , and (2) doctor’s posterior about the specific match between drug  $j$  and patient  $p$ . That is:

$$\tilde{Q}_{pj_t} = \tilde{Q}_{j_t} + \tilde{q}_{pj_t}.$$

According to the Bayes rule (DeGroot 1970):

$$\begin{pmatrix} \tilde{Q}_{j_t} \\ \tilde{q}_{1j_t} \\ \vdots \\ \tilde{q}_{P_{j_t}j_t} \end{pmatrix} \sim N \left( \begin{pmatrix} \bar{Q}_{j_t} \\ \bar{q}_{1j_t} \\ \vdots \\ \bar{q}_{P_{j_t}j_t} \end{pmatrix}, \Sigma_{j_t} \right)$$

where

$$\begin{aligned} \bar{Q}_{j_t} &= \frac{\sum_p \frac{n_{pj_t}^R \cdot \alpha_R \cdot (\bar{R}_{pj_t} - \alpha_0)}{\sigma_v^2 + n_{pj_t}^R \cdot \alpha_R^2 \cdot \sigma_{q_0}^2} + \frac{\bar{Q}_{j_0}}{\sigma_{\bar{Q}_{j_0}}^2}}{\sum_p \frac{n_{pj_t}^R \cdot \alpha_R^2}{\sigma_v^2 + n_{pj_t}^R \cdot \alpha_R^2 \cdot \sigma_{q_0}^2} + \frac{1}{\sigma_{\bar{Q}_{j_0}}^2}} \\ \bar{q}_{pj_t} &= \frac{\sigma_{q_0}^2 \cdot n_{pj_t}^R \cdot \alpha_R \cdot (\bar{R}_{pj_t} - \alpha_0 - \alpha_R \cdot \bar{Q}_{j_t})}{\sigma_v^2 + n_{pj_t}^R \cdot \alpha_R^2 \cdot \sigma_{q_0}^2} \\ \Sigma_{j_t}^{-1} &= \begin{pmatrix} s & a_1 & \cdots & \cdots & a_{P_{j_t}} \\ a_1 & m_1 & & & \\ \vdots & & \ddots & & 0 \\ \vdots & 0 & & \ddots & \\ a_{P_{j_t}} & & & & m_{P_{j_t}} \end{pmatrix} \\ s &= \sum_p \frac{n_{pj_t}^R \cdot \alpha_R^2}{\sigma_v^2} + \frac{1}{\sigma_{\bar{Q}_{j_0}}^2} \\ a_p &= \frac{n_{pj_t}^R \cdot \alpha_R^2}{\sigma_v^2} \\ m_p &= \frac{n_{pj_t}^R \cdot \alpha_R^2}{\sigma_v^2} + \frac{1}{\sigma_{q_0}^2} \end{aligned}$$

Note that the two posterior beliefs,  $\tilde{Q}_{j_t}$  and  $\tilde{q}_{pj_t}$ , are correlated because both make use of the satisfaction information from patient  $p$ . Also note that as more patients become involved with the drug over time (i.e.,  $P_{j_t}$  increases over time), the length of the quality vector increases over time. That is, the size of  $\Sigma_{j_t}^{-1}$  increases over time. We give the formula for  $\Sigma_{j_t}^{-1}$  instead of  $\Sigma_{j_t}$  because the former is the natural product in deriving the posterior density. But we can show that the across-patient terms in  $\Sigma_{j_t}^{-1}$  are all zero, and we exploit this special structure to analytically invert it to get  $\Sigma_{j_t}$ . Inverting  $\Sigma_{j_t}^{-1}$  results in a matrix

without zero elements (see [Appendix](#)). This implies that the posterior of  $\tilde{q}_{pjt}$  is no longer independent across patients. This is because all the updates of  $q_{pj}$  rely on the update of  $Q_j$ , which in turn relies on satisfaction reports from all patients.

Equipped with the posterior updates, the expected utility is given by:

$$\begin{aligned} E[\tilde{V}_{pjt}] &= \left[ 1 - e^{-\gamma(\beta_{xj}X_{pt} + \beta_z Z_{jt} + \epsilon_{pjt})} E\left[ e^{-\gamma \tilde{Q}_{pjt}} \right] \right] / \gamma \\ &= \left[ 1 - e^{-\gamma(\beta_{xj}X_{pt} + \beta_z Z_{jt} + \epsilon_{pjt})} e^{-\gamma \bar{Q}_{pjt} + \frac{1}{2} \gamma^2 \sigma_{\bar{Q}_{pjt}}^2} \right] / \gamma \\ &= \left[ 1 - e^{-\gamma \left( \bar{Q}_{pjt} - \frac{1}{2} \gamma \sigma_{\bar{Q}_{pjt}}^2 + \beta_{xj}X_{pt} + \beta_z Z_{jt} + \epsilon_{pjt} \right)} \right] / \gamma \end{aligned}$$

Thus, maximizing the expected utility is equivalent to maximizing the following,

$$U_{pjt} = \bar{U}_{pjt} + \epsilon_{pjt} = \bar{Q}_{pjt} - \frac{1}{2} \gamma \sigma_{\bar{Q}_{pjt}}^2 + \beta_{xj}X_{pt} + \beta_z Z_{jt} + \epsilon_{pjt}$$

where  $\sigma_{\bar{Q}_{pjt}}^2$  denotes the posterior variance of  $\bar{Q}_{pjt}$ . In the [Appendix](#) we show that we can obtain  $\sigma_{\bar{Q}_{pjt}}^2$  directly from the elements in matrix  $\Sigma_j^{-1}$ . The standard logit probability (McFadden 1973) for patient  $p$  getting drug  $j$  at time  $t$  is:

$$PR_{pjt} = \frac{\exp(\bar{U}_{pjt})}{\sum_{j=1}^J \exp(\bar{U}_{pjt})}$$

From the prescribing probabilities, we can estimate parameters by maximizing the log likelihood function:

$$\ln(L) = \sum_{p,j,t} 1_{data=pjt} \cdot \ln(PR_{pjt})$$

The intuition behind the learning model can be summarized as following: before seeing patient  $p$ , doctor has a specific prior about the average quality of drug  $j$  ( $Q_j$ ) and the specific match between  $p$  and  $j$  ( $q_{pj}$ ). The true values of  $Q_j$  and  $q_{pj}$  are constant over time but the doctor is uncertain about them. When  $p$  reports a signal of satisfaction ( $R_{pjt}$ ), doctor recognizes it as a mixture of the true  $Q_j$ , the true  $q_{pj}$ , plus random noise. Note that every patient’s signal reflects  $Q_j$  but only patient  $p$ ’s signal reflects  $q_{pj}$ . This implies that doctor can use the average of every newly-reported signal to gather new information about  $Q_j$ . The formula for the posterior mean of  $Q_j$  reflects this simple updating process.

In comparison, the update on  $q_{pj}$  is much trickier: although in theory the satisfaction of other patients on drug  $j$  (labeled as  $R_{p'jt}$ ) does not reflect the idiosyncratic match value of  $q_{pj}$ , doctor will use part of  $R_{p'jt}$  to update her belief on  $Q_j$  and then employ the updated belief to better understand which

part of patient  $p$ 's signal reflects  $Q_j$  and which part reflects  $q_{pj}$ . Because of this link,  $R_{p'jt}$  will enter the posterior mean of  $q_{pj}$  indirectly and therefore the posterior means of  $Q_j$  and  $q_{pj}$  are interdependent. Similarly, although the true values of drug-patient match are independent across patients (by assumption), the posterior means of  $q_{pj}$  and  $q_{p'j}$  are not statistically independent. This complication highlights the fundamental connection between across- and within-patient learning and demonstrates why they must be modeled jointly.

#### 4.2 Estimation issues

The model presented above focuses on one type of signal, patient satisfaction. In reality, there are many types of signals. FDA updates, media reports, academic articles and manufacturer advertising could all be viewed as noisy signals of the average drug efficacy that affects doctor's Bayesian update. However, estimating the Bayesian role of these signals requires each one of them have enough variation over time and across patients. In a Bayesian world, lack of variation adds to the difficulty in estimating the precision of a signal. When we allow both advertising and patient satisfaction to enter the Bayesian updating process, the model estimation has trouble converging. When it does converge, the variance term corresponding to advertising is extremely large, suggesting that the monthly advertising data may not provide enough variation to identify the variance. Given that FDA updates and article data have even less variation than advertising, it is difficult to model all of them in the framework of Bayesian learning.

To address this computation problem, we model patient satisfaction as a signal that contributes to the Bayesian learning but treat all the other factors as drug attributes ( $Z_{jt}$ ) that directly enter the utility function. This implies that all the true effects of advertising, if they exist, are captured in the coefficient of advertising. Because drug manufacturers may adjust advertising intensity by historical or predicted sales and we do not address the potential endogeneity problem, we treat advertising as pure control.

Specifically, the model described above circumvents the estimation difficulty but still allows all types of factors to play a role in prescription choice. The disadvantage is that we can no longer rely on the Bayesian structure to describe how historical information in FDA updates, advertising, news report and Medline articles affect a patient's expected utility. Rather, we define  $Z_{jt}$  as a vector, where each non-advertising element corresponds to the log of the cumulative sum of one factor. To better capture a potential s-shape impact of advertising, we use the inverse of cumulative total advertising (detailing + journal advertising + DTC) instead of advertising itself (Lilien et al. 1992).

Since the model treats patient satisfaction and other sources of information differently, the magnitudes of their structural coefficients are not directly comparable. As shown below, we evaluate their relative importance by (1) comparing models with and without certain information, and (2) using our preferred model to predict drug diffusion in (hypothetical) scenarios that vary by information structure.

Another estimation issue is whether we should treat traditional NSAIDs, Celebrex, Vioxx and Bextra as four branches in a simple logit, or assume a nested logit structure where a doctor first chooses between traditional NSAIDs and Cox-2, and then decides which brand is the best within the nest of Cox-2. We have estimated both, results are almost identical (in both likelihood value and coefficient magnitude). The parameter that describes the substitutability of the two nests is estimated at 0.99, which implies that the nested logit is analytically the same as the simple logit. In light of this finding, we only report the results based on the simple logit model.

### 4.3 Identification

Overall, the econometric model includes four sets of parameters:  $[\beta_{xj}, \beta_z]$  capture the effects of individual demographics and drug attributes,  $[\bar{Q}_{j0}, \sigma_{Q_{j0}}, \sigma_{q_0}]$  capture doctor's prior,  $[\alpha_0, \alpha_R, \sigma_v]$  capture the importance of patient satisfaction, and  $\gamma$  captures doctor's risk preference. As discussed above, FDA updates, inverse of manufacturer advertising, news reports, and Medline articles are treated as drug attributes, and their impact on patient utility are captured in  $\beta_z$ .

The identification of  $\beta_{xj}$  comes from the time-invariant prescription pattern across patients. For example, if Cox-2 prescriptions tend to be concentrated in the elderly, it translates into a significant and positive coefficient corresponding to the interaction of Cox-2 and age. Similarly,  $\beta_z$  is identified from the co-movements of drug market shares and various drug information. In principle, causality could go either way for advertising: on the one hand, advertising may trigger sales; on the other hand, historical or predicted sales patterns may motivate changes in advertising intensity. This implies that the coefficient for advertising is better interpreted as the correlation between advertising and sales rather than a causal effect.

The prior means of drug quality,  $\bar{Q}_{j0}$ , are identified from initial market shares. Because we include traditional NSAIDs as the outside good whose efficacy is well-known to doctors, we normalize its  $Q$  as zero. The prior of the three Cox-2s are all identified relative to the traditional NSAIDs. However, patient satisfaction  $R$  is reported in absolute terms. Apparently, the noise in  $R$ , denoted by  $\sigma_v$ , is determined by the heterogeneity in  $R$ . Since we assume  $R$  equals a linear function of true quality  $Q_{pj}$  plus noise, we can derive  $\sigma_v$  by regressing  $R_{pjt}$  on a full set of patient-drug dummies and calculating the standard deviation of the residuals. This procedure does not require any prescription data, so we estimate  $\sigma_v$  and fix it when estimating the full model.

Parameters,  $\alpha_0$  and  $\alpha_R$ , describe the scale difference between satisfaction  $R$  and true quality  $Q_{pj}$ . However, since we do not know  $Q_{pj}$ , they must be proxied by the posteriors, which are in turn reflected in evolving market shares. If the diffusion path is flat for each drug, the lack of updating implies that patient satisfaction has little impact, which amounts to  $\alpha_R = 0$ . If drug  $j$ 's diffusion path is positively related to drug  $j$ 's average satisfaction over time,



it implies a significant, positive  $\alpha_R$ . The other term,  $\alpha_0$ , is simply an intercept that is derived from the relative scale of  $R$  and  $Q$ .

The dispersion on the prior of the average quality of drug  $j$ , namely  $\sigma_{Q_{j0}}$ , is identified by the speed of diffusion. According to the Bayesian formula, the mean of the posterior,  $\bar{Q}_{jt}$ , is essentially a weighted average between  $R$  and the prior mean  $\bar{Q}_{j0}$ , while the weights are inversely related to the amount of noise in the two terms. Since we already identify the noise of  $R$ , a relatively small (large)  $\sigma_{Q_{j0}}$  implies that doctors believe the prior is relatively precise (noisy) and therefore put less (more) weight on patient satisfaction, which results in slow (fast) learning.

Similarly, the dispersion on the prior of patient-drug match, namely  $\sigma_{q_0}$ , is identified by how fast doctors update their patient-specific beliefs. Small (large)  $\sigma_{q_0}$  implies that patient  $p$ 's doctor is reluctant (eager) to revise her prior after she receives  $p$ 's satisfaction report, because she thinks the report is relatively noisy (precise).

The risk aversion parameter,  $\gamma$ , is identified by a functional form restriction. As noted in Coscelli and Shum (2004), the data only tell us about the term  $\bar{Q}_{pjt} - \frac{1}{2}\gamma\sigma_{Q_{pjt}}^2$ . The fact that we assumed a CARA utility function leads to a linear decomposition into the mean and variance terms.

## 5 Results

As described in Section 3, we focus on the patients that first appear in the data on or after January 1, 2001. The analysis sample ends at December 31, 2003 and is conditional on the prescriptions that come with valid answers for all five satisfaction questions. The final sample involves 6,577 patients and 17,329 prescriptions.

### 5.1 Benchmark model without learning

Before estimating the structural model, we check two benchmark models. These benchmarks utilize a discrete choice framework but do not incorporate a learning structure. Comparing them with our structural model will help us understand the importance of the learning structure. Specifically, Benchmark I estimates the prescription choice within traditional NSAIDs and the three brands of Cox-2s, assuming that the utility of patient  $p$  using drug  $j$  is:

$$U_{pjt} = \beta_{j0} + \beta_s \overline{satisf}_{jt} + \beta_{xj} X_{pt} + \beta_z Z_{jt} + \epsilon_{pjt}.$$

Here  $\overline{satisf}_{jt}$  denotes the average satisfaction reported for drug  $j$  up to time  $t$ . To capture the fundamental difference across drugs, we also include a set of drug dummies, whose impacts on utility are captured by coefficients  $\beta_{j0}$ .

Benchmark II omits patient satisfaction in the utility function so that a comparison of the two benchmark models would highlight the role of patient satisfaction. Specifically, the utility function for Benchmark II is:

$$U_{pjt} = \beta_{j0} + \beta_{xj}X_{pt} + \beta_z Z_{jt} + \epsilon_{pjt}.$$

Assuming logit errors, we can write out the probability of patient  $p$  choosing drug  $j$  and maximize the overall likelihood. We normalize the satisfaction measure as  $6 - \text{satisf}_{12345}$  so that a positive coefficient on patient satisfaction implies that the more satisfied patients are, the better the drug choice is. Since the benchmark models do not incorporate the learning structure, in order to capture all the information available up to the study period, we compute the satisfaction variable as the average of all satisfaction reports up to one month before the prescription month.

To be consistent with the structural model, we use the inverse of total advertising cumulated from the day of drug entry up to one month before the prescription month. We have tried other definitions, including the cumulative sum itself (with or without log), the advertising flow (instead of cumulative sum), and the monthly average of the cumulative sum. Results are qualitatively similar.

To estimate the extent to which doctors prescribe based on observable patient demographics, we allow the coefficient of patient demographics ( $\beta_{xj}$ ) to vary by whether drug  $j$  is a traditional NSAIDs or a Cox-2. In other words, these coefficients capture doctors' preferences between traditional NSAIDs and Cox-2s, but not within Cox-2s. Allowing  $\beta_{xj}$  to vary by Cox-2 brand does not change the results.<sup>17</sup>

As shown in Table 5, when we include patient satisfaction and other sources of information in Benchmark I, patient satisfaction has a positive and significant impact for all three Cox-2s. The satisfaction coefficient is larger for Bextra, probably because Bextra is newer than the other two drugs. In terms of other information, the coefficient of inverse advertising is negative as expected but indistinguishable from zero at the 95% confidence level. The coefficients for Lexis-Nexis articles are significantly positive (and more prominent in the non-negative ones), but both coefficients for negative and non-negative Medline articles are insignificant. In contrast, the coefficient of FDA updates is positive (and marginally significant), which is surprising given the fact that most FDA updates have negative content. The three intercepts suggest that Celebrex and Vioxx are viewed better than Bextra, everything else being equal. This reflects the fact that Bextra has the smallest market share among the three Cox-2s. In demographics, older, high-income males with private health insurance are more likely to receive Cox-2 prescriptions.

Omitting patient satisfaction leads to a worse fit in Benchmark II. In comparison with Benchmark I, advertising appears to be much more important

<sup>17</sup>Another way to address patient demographics is including them as concomitant variables, as suggested in Dayton and Macready (1988).

**Table 5** Benchmark models—discrete choice model without learning structure

	(1) Coefficients [ <i>t</i> -stat]	(2) Coefficients [ <i>t</i> -stat]
Intercept of Celebrex	-4.6079*** [-5.3824]	-2.1321*** [-2.8039]
Intercept of Vioxx	-3.5984*** [-4.1337]	-2.2336*** [-2.8906]
Intercept of Bextra	-9.8366*** [-9.2781]	-2.8402*** [-3.6176]
(6-Satisf12345) for Celebrex	0.5446*** [7.8586]	
(6-Satisf12345) for Vioxx	0.2870*** [4.4635]	
(6-Satisf12345) for Bextra	1.6105*** [10.519]	
Inverse of advertising	-0.2749 [-1.0472]	-2.0548*** [-5.6204]
Cox2 * age	0.0333*** [30.273]	0.0337*** [30.6364]
Cox2 * sex	-0.0967*** [-2.8441]	-0.1086*** [-3.157]
Cox2 * low income	-0.1319*** [-3.961]	-0.1225*** [-3.5714]
Cox2 * low education	-0.0089 [-0.2618]	-0.0136 [-0.3988]
Cox2 * HEALTHINS	0.0606 [0.8301]	0.0698 [0.9790]
Cox2 * INSPLAN	0.6539*** [11.911]	0.6653*** [11.9443]
Cox2 * DRUGINS	-0.1759*** [-5.0986]	-0.1729*** [-4.9119]
Medline article (neg)	-0.2032 [-1.623]	-0.1867 [-1.6179]
Medline article (non-neg)	0.1910 [0.8534]	0.1457 [0.6912]
Lexis article (neg)	0.0626*** [3.3122]	0.0606*** [3.1563]
Lexis article (non-neg)	0.0903*** [4.8548]	0.0889*** [4.7287]
After FDA update	0.0969* [1.7211]	0.1091* [1.8746]
Log likelihood	-17226	-17315
# of patients	6577	6577
# of prescriptions	17329	17329

Notes: Throughout Tables 5 to 9, the default drug is traditional NSAIDS. T-statistics in brackets. Satisfaction is measured by 6-satisf12345, computed as the average of all patient satisfaction up to the month before prescription. Advertising variable is measured as the inverse of cumulative sum of advertising expenditures up to the previous month. Articles are measured as the log of cumulative sum of weighted articles up to the previous day \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

in this case. Further, the coefficient of the Bextra dummy is no longer worse than those of Celebrex and Vioxx. As we see below, these results suggest that a discrete choice model without patient satisfaction is subject to omitted variable bias.

### 5.2 Model with learning

The results on the two benchmark models encourage us to think more systematically about patient satisfaction. Accordingly, the structural model

adds a Bayesian learning structure on top of the classical discrete choice framework.

Recall that each individual satisfaction measure is discrete but the five satisfaction measures are very closely correlated (with correlation coefficient ranging between 0.87 and 0.97). These high correlations motivate us to use  $satisf_{12345}$  as a continuous measure of  $R_{pj}$ . As discussed in Section 4.3, we estimate the structural model in two steps: first, we regress  $R_{pj}$  on a full set of patient-drug ( $pj$ ) dummies, and compute the residuals' standard deviation. According to our model, this standard deviation gives us an unbiased estimate of  $\sigma_v$ . With R-square 0.697, the regression produces  $\sigma_v = 0.496$ . Ideally, we need the residual to be normally distributed so that the model can yield close solutions to the posterior belief. Although not shown here, a plot of the histogram of these residuals shows that the distribution is symmetric and close to the bell shape. Having said that, we acknowledge the potential approximation error that could be caused by treating the discrete satisfaction scores as continuous signals. In the second step, we set  $\sigma_v$  at 0.496 and search for the best parameters that maximize the overall log likelihood.

A potential concern is that treating  $v$  as a normally distributed variable may make the signal  $R$  go beyond the range of 1 to 5. Though the probability of this is positive, we argue it is reasonably small and will not generate severe estimation bias. Specifically, the average  $satisf_{12345}$  throughout our whole sample is 1.794 and the estimated standard deviation of  $v$  is 0.496. Given the normal assumption, this implies that the probability of a signal below 1 is less than 5.5% and the probability of it being above 5 is less than 0.01%.<sup>18</sup>

Results reported below assume that doctors talk to each other within a census division. As discussed in Section 3, we observe significant heterogeneity of market shares across divisions, which suggests that information is not fully shared across divisions. As a confirmation, we also run the structural model assuming nationwide information pooling and find that it generates a significantly worse fit to the data.

Table 6 presents three sets of structural results: Column (1) presents a *BASIC* model that incorporates all sources of information. To gauge the relative importance of within-patient and across-patient learning, Column (2) ignores within-patient learning (by setting  $\sigma_{q_0} = 0$ ) and Column (3) ignores across-patient learning (by setting  $\sigma_{Q_{j_0}} = 0$ ).

All three models set the risk parameter as zero (which implies risk-neutrality). When we estimate the full model with risk preference, the risk parameter is extremely close to zero ( $\hat{\gamma} = 1.2e - 23$  with t-stat less than 0.01).

<sup>18</sup>These probability estimates are obtained using the cumulative density function of a normal distribution with mean 1.794 and standard deviation 0.496.

**Table 6** Models with different learning structure

	(1) Basic model	(2) Basic model with across-learning only	(3) Basic model with within-learning only
$\alpha_0$	-16.7348*** [-4.4589]	-78.2229 [-0.4911]	-5.6142** [-2.1568]
$\alpha_R$	2.5693*** [7.6483]	103.279 [0.4823]	2.3425*** [8.1898]
$\alpha_v$	0.4960	0.4960	0.4960
$\gamma$ (absolute risk aversion)	Set at 0	Set at 0	Set at 0
Q0_celebrex	2.6971*** [2.8217]	-0.1770 [-1.0656]	-0.2610 [-0.2746]
Q0_vioxx	2.3237** [2.3099]	-0.0496 [-0.2453]	-0.3616 [-0.3718]
Q0_bextra	2.3090** [2.2647]	-1.3623* [-5.9807]	-0.8229 [-0.8335]
$\sigma_{q0}$	0.3066*** [7.4460]		0.2927*** [8.1967]
$\sigma_{Q0}$ celebrex	0.0177*** [7.0597]	0.0006 [0.4727]	
$\sigma_{Q0}$ vioxx	0.0199*** [7.1608]	0.0004 [0.4748]	
$\sigma_{Q0}$ bextra	0.0294*** [6.7797]	0.0009 [0.4759]	
Inverse of advertising	-0.5673*** [-2.7105]	-0.6465* [-2.3424]	-0.9521*** [-4.3958]
Cox2 * age	0.0177*** [13.3565]	0.0337* [28.6727]	0.0174*** [12.0504]
Cox2 * sex	-0.1219*** [-2.9104]	-0.1052* [-3.001]	-0.1291*** [-3.1486]
Cox2 * low income	-0.2007*** [-4.8697]	-0.1208* [-3.5537]	-0.2033*** [-5.0342]
Cox2 * low education	-0.0684 [-1.6274]	-0.0174 [-0.5031]	-0.0739 * [-1.7888]
Cox2 * HEALTHINS	-0.0015 [-0.0174]	0.0741 [1.0262]	-0.0069 [-0.0821]
Cox2 * INSPLAN	0.4330*** [6.5639]	0.674* [12.2724]	0.4171*** [6.3665]
Cox2 * DRUGINS	-0.2277*** [-5.3704]	-0.1833* [-5.246]	-0.1923*** [-4.6508]
Medline article (neg)	-0.7520*** [-4.4846]	-0.0555 [-0.5389]	-0.4343*** [-2.8955]
Medline article (non-neg)	-0.9671*** [-3.511]	-0.4621* [-7.4174]	-0.2654 [-0.9925]
Lexis article (neg)	0.0911*** [3.7882]	0.0797* [4.1067]	0.1099*** [4.6372]
Lexis article (non-neg)	0.0173 [0.7274]	0.0897* [4.7318]	0.0086 [0.362]
After FDA update	-0.0803 [-1.0502]	-0.0689 [-1.1513]	0.0161 [0.2325]
Log likelihood	-11376	-17259	-11565
# of patients	6577	6577	6577
# of prescriptions	17329	17329	17329

This implies that prescription choice has little to do with risk preference: a patient stays on the old prescription not because her doctor is afraid of trying a new brand. Rather, it is probably because the patient is satisfied with the old prescription, or because the other sources of information do not produce any significant news against the old brand. Since including the risk parameter prolongs estimation a great deal and all the other parameters do not change much when we set  $\gamma = 0$ , we only report results that assume risk neutrality.

Three findings stand out in Table 6. First, there is significant learning from patient satisfaction. On the one hand, the positive, significant estimate of  $\alpha_R$  suggests that doctors believe the satisfaction reports from patients are correlated with drug efficacy and therefore use them to update the prior. On the other hand, the magnitudes of  $\sigma_{Q_{j0}}$  are much smaller than both the noise in the satisfaction report (i.e.  $\sigma_v$ ) and the dispersion of patient-drug match (i.e.  $\sigma_{q_0}$ ). This suggests that doctors hold strong priors about the average efficacy of the three drugs. As a result, although they value the satisfaction reports, the updating on the general drug quality is slow. In comparison, the learning on the specific match between a drug and a patient is faster, because the magnitude of  $\sigma_{q_0}$  is much closer to that of  $\sigma_v$ .

This interpretation is consistent with the comparison across Columns (1), (2) and (3). The overall likelihood in Column (1) ( $-11376$ ) is significantly better than that in Columns (2) and (3) ( $-17259$ ,  $-11565$ ), suggesting that both across- and within-patient learning are important in our data. However, the likelihood (and point estimates) in Column (3) is much closer to Column (1). This implies that a larger part of the data variation is driven by within-patient learning, the same conclusion as we have inferred from the relative magnitudes of  $\sigma_{q_0}$ ,  $\sigma_{Q_{j0}}$ , and  $\sigma_v$ . Along the same lines, we note that structural models including within-patient learning (Columns (1) and (3)) fit the data much better than the benchmark models in Table 5, but  $\alpha_R$  becomes insignificant when we ignore within-patient learning in Column (2).

Coefficients corresponding to other sources of information are mixed. As we expect, inverse of advertising is significantly negative. However, since drug manufacturers may change advertising intensity according to predicted sales change in the near future, this coefficient may capture some demand factor that manufacturers observe but we do not. The concern of endogeneity prompts us to treat advertising as a pure control and not as having any causal effect.

News articles have a positive influence on prescriptions, no matter whether these titles sound negative or non-negative. This result is puzzling: it seems to suggest that news articles play a greater role in informing doctors/patients of the existence of Cox-2s rather than revealing the quality of Cox-2s. One possible explanation is that most news are picked up by patients; when they inquire about the drug in a doctor's office, the doctor relies on his own experience with the drug or his reading of professional articles, but not the content of the news article. However this does not explain why negative news have a larger coefficient that is more statistically significant than that of positive news. We suspect it is either due to the measurement error in our raw data or to the fact that many news articles in our data are negative.

In contrast, a medical article about Cox-2s has a significant negative impact on prescription sales, even if its title and abstract are non-negative. Note that most of the non-negative articles are neutral, which mentions both positive and negative effects of Cox-2s. Our findings suggest that doctors lay more emphasis on the negative contents of Medline articles, or tend to interpret Medline publication as a negative signal against Cox-2s. The coefficient of FDA update is negative as we expect, but statistically indistinguishable from zero. One possible explanation is the FDA updates lag behind Medline articles and therefore deliver little new information to doctors.

To better understand the relative importance of information, Table 7 re-estimates the BASIC model by excluding news reports (Column (2)) or medical articles (Column (3)). Comparing Column (3) with the BASIC model (results repeated in Table 7 Column (1)), we find that excluding Medline articles does not affect the qualitative role of patient satisfaction, but it makes the coefficient of the FDA updates much more negative than in the BASIC model ( $-0.6988$ , with  $t$ -stat  $-14.97$ ) versus  $-0.0803$  (with  $t$ -stat  $-1.05$ ). The coefficient magnitude for advertising also increases substantially. In comparison, excluding news reports alone (Column (2)) produces more similar results to the BASIC model. FDA updates seem to be a redundant follow-up from the medical literature: once we control for Medline articles, the coefficient of FDA updates is close to zero. But negative news articles continue to have a positive impact on drug prescription, with or without the control of Medline articles. This suggests that news articles (even if with negative titles) probably inform patients about the availability of Cox-2s. Patients then bring this information to the doctor's notice, and this informative role is not closely correlated with professional opinion about Cox-2s.

Comparing estimates within the three Cox-2s, we find the prior mean ( $Q_0$ ) of Bextra is always smaller than that of Vioxx and Celebrex. This is consistent with the small market share of Bextra. In all specifications, the prior dispersion ( $\sigma_{Q_0}$ ) is greater for Bextra than for Celebrex and Vioxx. This finding reflects the late entry of Bextra.

Some sensitivity occurs in the absolute magnitude of  $Q_0$ : the three  $Q_0$ s are positive in the BASIC model; but when we exclude Medline articles, they all turn negative (Table 7 Column (3)). This seemingly sensitive result is indeed sensible: because the BASIC model controls for the number of Medline articles in the utility function,  $Q_0$  should be interpreted as the prior mean of a Cox-2 conditional on *non-zero* Medline articles. When we omit Medline articles, the estimated  $Q_0$  represents the prior mean of a Cox-2 conditional on its *average* count of Medline articles. Since most Medline articles have a negative effect on the probability of choosing Cox-2, this explains why  $Q_0$  turns negative if we exclude Medline articles.

The coefficients of demographics are stable across specifications. Results suggest that older, better-income, and better educated males have a greater tendency of receiving Cox-2. Different insurance variables have different signs: being privately insured is associated with a greater likelihood of receiving Cox-2, but drug insurance is negatively correlated with Cox-2 prescription.

**Table 7** Learning models with and without medical and news articles

	(1) Basic model	(2) Basic model with Medline only	(3) Basic model with Lexis only
$\alpha_0$	-16.7348*** [-4.4589]	-17.7292*** [-6.0502]	-2.2200*** [-2.7177]
$\alpha_R$	2.5693*** [7.6483]	2.6205*** [8.7214]	2.5611*** [8.0841]
$\alpha_v$	0.4960	0.4960	0.4960
$\gamma$ (absolute risk aversion)	Set at 0	Set at 0	Set at 0
Q0_celebrex	2.6971*** [2.8217]	2.8960*** [2.8864]	-2.3770*** [-21.0303]
Q0_vioxx	2.3237** [2.3099]	2.5121** [2.3804]	-2.3183*** [-20.3324]
Q0_bextra	2.3090** [2.2647]	2.4614** [2.2928]	-2.5121*** [-20.5037]
$\sigma_{q0}$	0.3066*** [7.4460]	0.2979*** [8.4998]	0.2963*** [8.1301]
$\sigma_{Q0}$ celebrex	0.0177*** [7.0597]	0.0175*** [7.7912]	0.0122*** [7.4432]
$\sigma_{Q0}$ vioxx	0.0199*** [7.1608]	0.0195*** [8.1070]	0.0120*** [7.4906]
$\sigma_{Q0}$ bextra	0.0294*** [6.7797]	0.0276*** [7.5103]	0.0003 [1.5751]
Inverse of Advertising	-0.5673*** [-2.7105]	-0.5594** [-2.5702]	-1.1450*** [-4.552]
Cox2 * age	0.0177*** [13.3565]	0.0176*** [13.1596]	0.0181*** [13.9286]
Cox2 * sex	-0.1219*** [-2.9104]	-0.1219*** [-2.9018]	-0.1234*** [-3.0968]
Cox2 * low income	-0.2007*** [-4.8697]	-0.1995*** [-4.8751]	-0.1974*** [-4.8776]
Cox2 * low education	-0.0684 [-1.6274]	-0.0641 [-1.5144]	-0.0682* [-1.6595]
Cox2 * HEALTHINS	-0.0015 [-0.0174]	-0.0032 [-0.0377]	0.0285 [0.3349]
Cox2 * INSPLAN	0.4330*** [6.5639]	0.4323*** [6.6111]	0.4291*** [6.6021]
Cox2 * DRUGINS	-0.2277*** [-5.3704]	-0.2252*** [-5.2828]	-0.2643*** [-6.5615]
Medline article (neg)	-0.7520*** [-4.4846]	-0.6739*** [-3.9215]	
Medline article (non-neg)	-0.9671*** [-3.511]	-1.0345*** [-3.6034]	
Lexis article (neg)	0.0911*** [3.7882]		0.0601** [2.5662]
Lexis article (non-neg)	0.0173 [0.7274]		-0.0112 [-0.4759]
After FDA update	-0.0803 [-1.0502]	-0.068 [-0.8983]	-0.6988*** [-14.9698]
Log likelihood	-11376	-11387	-11539
# of patients	6577	6577	6577
# of prescriptions	17329	17329	17329



The latter may be explained by the non-favorable formulary status of Cox-2 relative to traditional NSAIDs. However, the potential for measurement errors in these insurance variables suggest that we regard these variables as pure controls rather than ascribe any specific economic meaning. All these findings are similar to what we have seen in the benchmark models without learning (Table 5).

Overall, results suggest that patient satisfaction, advertising, news reports and the medical literature are all important in prescription choice. Specifically, at the beginning of 2001 and upon the Bextra entry in January 2002, doctors held a strong prior belief about the efficacy of Celebrex, Vioxx, and Bextra, and learned gradually from patient satisfaction. We find evidence for both across- and within-patient learning, but within-patient learning explains much more variation in the data. Other sources of information are important as well: news articles and advertising are positively correlated with prescription, but Medline articles appear to be detrimental for drug sales. The impact of FDA updates is close to zero once we control for Medline articles. This suggests that the contents of FDA updates have already been included in Medline articles and therefore deliver little new information to doctors.

### 5.3 Model with learning and unobserved heterogeneity

One may argue that a doctor observes more patient-specific information than just her satisfaction before writing any prescription. Such information, including the patient's medical history and the nature of her demand for pain relief, may inform the doctor about whether the patient is suitable for a specific drug. Because we as researchers do not observe such information, we might mis-attribute some unobserved heterogeneity to learning.

To address this issue, we add patient-drug random effects  $\theta_{pj}$  to the utility function:

$$E[\tilde{V}_{pj}] = -e^{-\gamma(\theta_{pj} + \beta_{xj}X_{pj} + \beta_z Z_{jt} + \epsilon_{pj})} E\left[e^{-\gamma\tilde{Q}_{pj}}\right].$$

We estimate three models with random effects, the first two assume  $\theta_{pj}$  conforms to a discrete distribution that includes two or three “types” of patients, while the third model assumes  $\theta_{pj}$  is normal ( $N(0, \sigma_{\theta_j})$ , i.i.d. across patients).<sup>19</sup>

As shown in Table 8 Columns (2) and (3), allowing two or three distinct patient types improves the model fit a great deal (log L changes from  $-11376$  to  $-10181$  and  $-10086$ ) but the main results remain stable. Similar to the BASIC

<sup>19</sup>For the third model, we use simulated MLE with 20 draws.

**Table 8** Learning models with unobserved heterogeneity

	(1) Basic model	(2) Basic model plus 2-type random effects	(3) Basic model plus 3-type random effects	(4) Basic model plus normal random effects
$\alpha_0$	-16.7348*** [-4.4589]	-16.7665** [-2.3868]	-15.062*** [-3.5513]	-10.075*** [-2.9900]
$\alpha_R$	2.5693*** [7.6483]	2.8871*** [4.3816]	1.9062*** [7.3826]	1.7146*** [7.2924]
$\alpha_v$	0.4960	0.4960	0.4960	0.4960
$\gamma$ (absolute risk aversion)	Set at 0	Set at 0	Set at 0	Set at 0
Q0_celebrex	2.6971*** [2.8217]	1.5902 [1.0334]	4.1650** [2.4419]	3.5070** [2.1934]
Q0_vioxx	2.3237** [2.3099]	1.2475 [0.7738]	3.7237** [2.0994]	3.1889* [1.8939]
Q0_bextra	2.309** [2.2647]	1.2234 [0.7508]	3.8992** [2.1668]	3.1233* [1.8424]
$\sigma_{q0}$	0.3066*** [7.4460]	0.2748*** [4.2553]	0.3913*** [7.0522]	0.6977*** [5.4915]
$\sigma_{Q0}$ celebrex	0.0177*** [7.0597]	0.0214*** [4.1597]	0.0330*** [6.8166]	0.0557*** [5.3064]
$\sigma_{Q0}$ vioxx	0.0199*** [7.1608]	0.0243*** [4.1382]	0.0372*** [6.8097]	0.0620*** [5.3305]
$\sigma_{Q0}$ bextra	0.0294*** [6.7797]	0.0331*** [4.0741]	0.0491*** [6.4683]	0.0921*** [5.0416]
Inverse of advertising	-0.5673*** [-2.7105]	-0.5303** [-2.3151]	-0.5321** [-2.3092]	-0.7083*** [-2.6497]
Cox2* age	0.0177*** [13.3565]	0.0492*** [19.2723]	0.0591*** [17.8998]	0.0409*** [17.7789]
Cox2* sex	-0.1219*** [-2.9104]	-0.2491*** [-3.5346]	-0.3748*** [-4.2314]	-0.2620*** [-4.2175]
Cox2* low income	-0.2007*** [-4.8697]	-0.2969*** [-3.4475]	-0.4637*** [-3.9185]	-0.2380*** [-3.4795]
Cox2* low education	-0.0684 [-1.6274]	-0.1103 [-1.2236]	-0.1254 [-1.1796]	-0.0220 [-0.3115]
Cox2* HEALTHINS	-0.0015 [-0.0174]	0.1170 [0.6776]	0.1276 [0.7203]	0.0999 [0.7562]
Cox2* INSPLAN	0.4330*** [6.5639]	0.6913*** [6.1174]	0.8165*** [6.2004]	0.6453*** [6.5212]
Cox2* DRUGINS	-0.2277*** [-5.3704]	-0.4658*** [-5.4517]	-0.2257** [-2.1169]	-0.1558** [-2.2431]
Medline article (neg)	-0.7520*** [-4.4846]	-0.9028*** [-3.7983]	-1.1146*** [-4.4816]	-0.8561*** [-3.2479]
Medline article (non-neg)	-0.9671*** [-3.511]	-1.2490*** [-2.8952]	-1.4297*** [-3.0832]	-1.4646*** [-3.2011]
Lexis article (neg)	0.0911*** [3.7882]	0.0590 [1.5536]	0.0278 [0.6364]	0.1308*** [3.1384]
Lexis article (non-neg)	0.0173 [0.7274]	0.0177 [0.5646]	-0.0519 [-1.5106]	-0.0011 [-0.0267]
After FDA update	-0.0803 [-1.0502]	-0.0607 [-0.6458]	-0.0666 [-0.6900]	-0.1233 [-1.1698]
$\theta_{\text{celebrex}}$ of type 2		4.1771*** [44.343]	-2.7667*** [-13.048]	
$\theta_{\text{vioxx}}$ of type 2		3.9791*** [42.811]	-2.8848*** [-14.102]	
$\theta_{\text{bextra}}$ of type 2		4.3559*** [31.646]	-2.0819*** [-6.5969]	

**Table 8** (continued)

	(1) Basic model	(2) Basic model plus 2-type random effects	(3) Basic model plus 3-type random effects	(4) Basic model plus normal random effects
Probability of type 1		0.5846*** [48.141]	0.2731*** [12.079]	
$\theta_{\text{celebrex}}$ of type 3			3.3490*** [24.0066]	
$\theta_{\text{vioxx}}$ of type 3			3.0855*** [21.7766]	
$\theta_{\text{bextra}}$ of type 3			3.3047*** [18.506]	
Probability of type 2			0.4025*** [18.9100]	
$\sigma$ ind-celebrex RE				2.4923*** [25.0627]
$\sigma$ ind-vioxx RE				2.3108*** [25.7732]
$\sigma$ ind-bextra RE				2.0277*** [17.7305]
Log likelihood	-11376	-10181	-10086	-10577
# of patients	6577	6577	6577	6577
# of prescriptions	17329	17329	17329	17329
BIC	22967	20616	20465	21398

model, doctors learn from patient feedback and the learning is more within-patient than across-patients. Inverse advertising still has a negative coefficient, but news articles are no longer significant. In comparison, the coefficients of medical articles remain negative and highly significant. In fact, controlling for 3 patient types increases the magnitudes of the medical article coefficients by about 50% (as compared to the BASIC model), implying that ignoring unobserved heterogeneity may lead to biased estimates.

The model with normal random effects (Table 8, Column (4)) produces qualitatively similar parameter estimates and the log likelihood is worse than what we get with two patient types. Thus, the three-patient-type model captures most unobserved heterogeneity. In addition, the BIC criterion favors the 3-type model, too. Therefore, we denote the 3-type model as our preferred model and use it for counterfactual simulations at the end of this section.

### 5.4 Robustness checks

In this subsection we discuss several robustness checks on the BASIC model.

*Forward-looking behavior of physicians* In contrast to several other researches that have studied forward-looking behavior (Crawford and Shum 2005; Ching 2005; Erdem and Keane 1996), our model assumes that each

doctor focuses only on the current prescription situation. We do not model forward-looking not only because it simplifies the econometric model, but also because of the nature of the product category that we look at. In the data, a large proportion of patients have only one prescription and the potential risk of malpractice is likely to prevent doctors from experimenting.<sup>20</sup> In addition, we carried out the following simple test and did not find evidence supporting the forward-looking hypothesis.

Consider a risk neutral patient who is completely new to the Cox-2 category after all three Cox-2s become available. Since Bextra is the newest member in the category, it is by definition the least known alternative. If the patient's doctor is forward looking, the motivation to experiment would lead him to first prescribe Bextra to collect information. If on the other hand the prescription is driven by what the doctor has already learned about the drug quality, then he is more likely to prescribe either of the two older drugs that on average have greater posterior mean quality than Bextra. Indeed, among 1,255 such new patients, only 200 were given Bextra as their first prescription while the remaining majority were prescribed either Celebrex or Vioxx.

Therefore, we believe that although experimentation might be relevant for some product categories, it is unlikely to be a key issue for our study. We will leave the possibility of studying forward-looking behavior for future research.

*Sampling weights* While our data contain a nationally representative sample of households, we do not observe the whole population. In reality, doctors may use the experience of all patients to form beliefs about drug quality. Intuitively, ignoring part of the population tends to miss part of the across-patient learning and therefore mis-characterize the importance of across- and within-patient learning.

To address this issue, we make use of sampling weights that are available to us in the data.<sup>21</sup> If individual A has a sampling weight of 100, we assume doctors (in A's Census division) observe 100 patients whose demographics, prescription history, and satisfaction index are identical to A's. By this assumption, we inflate the individual records by sampling weights and then re-estimate the BASIC model. Statistically speaking, this is equivalent to asserting that, when doctors summarize patient feedback into the posterior belief, they assign more importance to the patients who represent more of the population in our original data.

---

<sup>20</sup>The fear of malpractice is not necessarily inconsistent with risk neutrality. A risk neutral doctor will try to avoid malpractice if it implies large compensation or end of his/her career.

<sup>21</sup>IPSOS has a sophisticated program that assigns weights to panelists. The weights change with time. So we obtained weights for the time period that is relevant for our study. These weights reflect the overall representativeness of each panelist during a specific time period.

Results incorporating sampling weight are presented in Table 9 Column (2).<sup>22</sup> Compared with the unweighted results (Table 9 Column (1)), adding sampling weights does not change qualitative conclusions:  $\alpha_R$  is still positive and highly significant, implying that doctors learn from patient feedback. Like before, estimated  $\sigma_{q_0}$  is much larger than the three  $\sigma_{Q_0}$ . This indicates that the prior of patient-drug match is more dispersed than the prior of average drug quality, hence doctors learn faster within a patient than across patients. In fact, adding sampling weights enlarges the difference between  $\sigma_{q_0}$  and  $\sigma_{Q_0}$ , which suggests that our unweighted results may even underestimate the importance of within-patient learning. This change is intuitive because across-patient learning is identified from prescription correlations across different patients. When we inflate the data by sampling weights, we attenuate the observed correlation among a greater population, which reduces the amount of learning obtained from each single patient. Parameters on demographics and the other information variables hardly change. Since the log likelihood (-11375) is extremely similar to what we get from the unweighted model (-11376), we are confident that our main results (unweighted) are robust to including sampling weights.

*Functional form of advertising* In the BASIC model, we use the inverse of total cumulative advertising, which entails three assumptions: first, drug diffusion follows a reciprocal model as dictated by the inverse of advertising; second, advertising does not depreciate over time; third, different forms of advertising are pooled together.

Strictly speaking, all three assumptions are subject to question. Since any functional form of advertising is arbitrary, we re-estimate the BASIC model with many alternative specifications: (1) using advertising or log advertising instead of the inverse; (2) using detailing and DTCA separately instead of the total of detailing, journal advertising and DTCA; (3) using flow of advertising instead of the cumulative sum; (4) estimating monthly depreciation rates for detailing and DTCA; and (5) lagging advertising by 3,6,9 and 12 months.

Across these specifications, the qualitative results on all the non-advertising variables are similar to what we had before, but the coefficient(s) on advertising is sensitive to specifications. As shown in Table 9 Column (3), when we include log(detailing) and log(DTCA) separately, both coefficients are significant but detailing is positive while DTCA is negative. We suspect the negative sign of DTCA is due to endogenous determination of DTCA or omitted variable bias. In theory, the same concern exists for any other type of advertising. Because we do not have valid instruments to control for such endogeneity,

---

<sup>22</sup>We are able to obtain sampling weights for only a subset of individuals in our data. We use two steps to impute the missing weights. First, focusing on the individuals that have sampling weights, we regress the reported weights on a polynomial function of demographic variables and whether the studied individual has returned the diary on a regular basis. In the second step, we use the predicted relationship to predict sampling weights for the rest of the sample. Results reported in Table 9 are based on the imputed weights.

**Table 9** Robustness check on sampling weights and advertising

	(1) Basic model	(2) Basic model plus sampling weights	(3) Basic model using alternative adv.
$\alpha_0$	-16.7348*** [-4.4589]	-16.8017*** [-7.4687]	-19.2098*** [-3.9187]
$\alpha_R$	2.5693*** [7.6483]	2.5797*** [11.9663]	2.6403*** [6.1506]
$\alpha_v$	0.4960	0.4960	0.4960
$\gamma$ (absolute risk aversion)	Set at 0	Set at 0	Set at 0
Q0_celebrex	2.6971*** [2.8217]	2.7038*** [4.7383]	3.3535*** [2.9303]
Q0_vioxx	2.3237** [2.3099]	2.3324*** [3.9314]	3.2387*** [2.7111]
Q0_bextra	2.309** [2.2647]	2.3078*** [3.7378]	2.4873*** [1.1891]
$\sigma_{Q0}$	0.3066*** [7.4460]	0.3063*** [13.1666]	0.2910*** [6.0368]
$\sigma_{Q0}$ celebrex	0.0177*** [7.0597]	0.0066*** [9.8280]	0.0170*** [5.8207]
$\sigma_{Q0}$ vioxx	0.0199*** [7.1608]	0.0075*** [10.0908]	0.0185*** [5.8224]
$\sigma_{Q0}$ bextra	0.0294*** [6.7797]	0.0111*** [9.5648]	0.0265*** [5.4660]
Inverse of advertising	-0.5673*** [-2.7105]	-0.5655*** [-2.8889]	
Log total detailing			0.2486** [2.1864]
Log total DTCA			-0.1704*** [-4.6389]
Cox2 * age	0.0177*** [13.3565]	0.0177*** [13.2926]	0.0177*** [13.2303]
Cox2 * sex	-0.1219*** [-2.9104]	-0.1217*** [-2.8713]	-0.1191*** [-2.8355]
Cox2 * low income	-0.2007*** [-4.8697]	-0.2007*** [-4.8256]	-0.2042*** [-4.9374]
Cox2 * low education	-0.0684 [-1.6274]	-0.0677 [-1.5916]	-0.0667 [-1.5834]
Cox2 * HEALTHINS	-0.0015 [-0.0174]	-0.0011 [-0.013]	0.0062 [0.0717]
Cox2 * INSPLAN	0.433*** [6.5639]	0.4331*** [6.5698]	0.4294*** [6.4765]
Cox2 * DRUGINS	-0.2277*** [-5.3704]	-0.2276*** [-5.272]	-0.2293*** [-5.3778]
Medline article (neg)	-0.752*** [-4.4846]	-0.7481*** [-6.0478]	-0.5583*** [-2.9170]
Medline article	-0.9671*** [-3.511]	-0.9683*** [-5.9626]	-1.3000*** [-4.0813]
Lexis article (neg)	0.0911*** [3.7882]	0.0903*** [3.7466]	0.1709*** [4.4297]
Lexis article (non-neg)	0.0173 [0.7274]	0.0178 [0.7513]	0.0277 [1.1587]
After FDA update	-0.0803 [-1.0502]	-0.0811 [-1.0566]	-0.1296* [-1.6721]
Log likelihood	-11376	-11375	-11367
# of prescriptions	17329	17329	17329

we treat advertising as a control and do not interpret its coefficient as having a causal effect. Fortunately, the effects of all the other variables are stable across specifications. Since these non-advertising variables are beyond the control of drug manufacturers, they are immune from reverse causality.

*Patient demographics* Strictly speaking, patient demographics may play two roles in prescription decisions: first, doctors may have a *fixed* view of drugs that match best with various demographic characteristics. To fully account for such practice, we should allow the coefficients of each patient's demographics ( $\beta_{xj}$ ) to vary by brand for each of the 4 alternatives, instead of Cox-2s versus traditional NSAIDs. Given the large number of demographics included in the basic model, we estimate brand-specific  $\beta_{xj}$  on the demographic variable that has the most predictive power in prescription decision—patient age. The re-estimated basic model does not show much improvement in the likelihood (from  $-11376$  to  $-11374$ ) and the magnitude of the age coefficient is similar across the three Cox-2 brands. At the same time, results on all the information variables remain unchanged.

Another channel for patient demographics to influence prescription decisions is through the learning structure. It is not difficult to see that doctors may be more likely to apply the experience of elderly male patients to other elderly males than to young females. However, it is extremely difficult to account for demographic-specific learning in the structural model, because some key demographic variables are continuous (say age) and any demographic grouping seems arbitrary. Keeping this caveat in mind, we emphasize that the learning estimates presented in this paper represent the average amount of learning across all demographic groups.

*Medline articles* The negative coefficient on non-negative medline articles is counterintuitive. To better understand the statistical forces underlying this coefficient, we conduct a number of robustness checks.

First, we re-estimate the basic model by decomposing the non-negative medline articles into positive and neutral articles. Results suggest that the negative coefficient of non-negative articles is primarily driven by a negative response to positive articles. Once we control for positive articles separately, the response to neutral articles becomes positive but insignificant with 95% confidence level.

To address the suspicion that doctors may view a positive article from a pharmaceutical company employee as a negative signal, we conduct a second robustness check by including variables for author affiliation. Results suggest a strong negative response to company affiliation and including affiliation reduces the significance of the responses to positive/negative/neutral articles. In comparison, including variables describing whether an article focuses on efficacy or side effects generate very noisy results. Among a number of specifications we have tried, only in one case do we observe negative and significant response to side-effects articles. The efficacy indicator is never significant.

**Table 10** Model fit: % of correct prediction of the actual RX choice

Discrete choice model without learning structure	Basic model	Basic, within-patient learning only	Basic, across-patient learning only	Basic + 3 patient-types (preferred)
61.17%	78.96%	78.35%	60.58%	85.49%

(based on the estimation sample of 17,329 RXs)

We suspect many of the noisy results are driven by the high correlations across the different sets of variables: for example, company-affiliated articles are more likely to be positive and positive articles are more likely to focus on efficacy instead of side effects. Thus including all of them in one specification is likely to generate a collinearity problem. Given that all the other information variables do not change much when we try different specifications on the medline articles, we believe the basic model is a reasonable simplification.

### 5.5 Model fit and counterfactual predictions

This subsection examines the relative importance of different sources of information. Treating the BASIC model with 3-patient-type random effects (Table 8 Column (3)) as our preferred model, we predict the number of prescriptions for three scenarios and compare them with the actual data.

The first scenario is our preferred model, which takes all sources of information as given and reports the predicted prescription counts by drug-month. This scenario indicates a good fit to the data: As shown in Table 10, for each of the 17,329 prescriptions considered in our estimation sample, we are able to predict the actual prescription choice correctly 85.5% of the time. In comparison, the percentage of correct prediction is 61.2% for the logit model without learning structure, 79.0% for the basic model, 78.4% for the basic model with within-patient learning only, and 60.6% for the basic model with across-patient learning only. Another measure of model fit is the percent of market share deviations from the actual data. Taking month-drug as the unit of observation, our preferred model has an average absolute percentage deviation of 26.5% if we focus on the prediction of Cox-2s, or 20.7% if the calculation includes non-Cox2s.<sup>23</sup> This suggests that, on average, our prediction of a Cox-2's monthly market share deviates from its actual share by 26.5%.

The second scenario assumes that patient feedback is shared nationwide instead of within a census division. This scenario reflects a recent proposal of FDA setting up a nationwide database to share patient feedback among doctors (Slater 2005). Since the satisfaction signals are observed, this counterfactual experiment can be easily implemented by allowing every patient to

<sup>23</sup>For each month-drug, absolute percentage deviation is defined as  $\text{abs}(\text{predicted number of RX} - \text{actual number of RX}) / \text{actual number of RX}$ . We then average this index across all month-drugs for each Cox-2 (which yields 26.5%) or for all four drugs (which yields 20.7%).



learn from everyone else in the data set, as opposed to only from those in the same geographic area. The third scenario assumes double counts of medical articles. This may be achieved by, e.g., subsidizing journal subscription, mailing a summary of academic findings to doctor office, or encouraging pharmacists and other health care professionals to educate doctors in an office visit. Readers can also interpret the last scenario as a greater intensity of “academic detailing”, which has been proposed as a potential improvement in FDA’s post-marketing surveillance (Ray and Stein 2006). In addition to mimicking proposals for the FDA reform, these two scenario also allow us to compare the effects of learning from individual patient satisfaction versus reading academic publications. These two are not readily comparable in the reported coefficients because one is modeled as Bayesian updating but the other enters the utility function directly.

It is important to realize the limitation of our counterfactual experiments. As implied by the data collection process, our patient diary data are likely to capture the effects that are easily observable to patients (such as pain relief, stomach upset or skin irritation) but not life-threatening events like likely heart attack. For this reason, predictions generated from our second scenario do not fully capture the actual effects of pooling all patient feedbacks in a nationwide database. To the extent that some patient experience, especially heart attacks and other severe events, are reported in academic articles, nationwide pooling of information may be translated into a greater intensity of academic dissemination, which is partly captured in the third scenario. Lastly, all of our empirical model focuses on a specific drug class (NSAIDs) in a specific time period (2001–2003), and therefore conclusions drawn from the counterfactual experiments are not necessarily applicable to other drugs and other time.

Comparing the two hypothetical scenarios against the actual data, Table 11 reports the predicted percentage change in the market share of Celebrex, Vioxx, Bextra and non-Cox2s from January 2001 to December 2003. Expanding census division learning to the national level makes a big difference: because patients report higher satisfaction for Celebrex than for Vioxx, Bextra, or traditional NSAIDs (see summary in Table 3), a nationwide database encourages switching towards Cox-2 inhibitors. The percentage change of market share is the lowest for Celebrex because Celebrex has the largest sales among the three Cox-2 inhibitors. The effect of more Medline publications is opposite to pooling patient feedback: compared to the actual market shares, doubling

**Table 11** Counterfactuals

Counterfactual scenarios	% change in the market share of			
	Celebrex	Vioxx	Bextra	All others
#1: Nationwide sharing of patient feedbacks	15.01%	20.82%	21.05%	−10.74%
#2: Double academic articles	−30.44%	−25.36%	−27.11%	17.03%

Medline articles would increase the market share of traditional NSAIDs by 17.03%, while depressing the market share of Cox-2 inhibitors by 25–30%.

## 6 Conclusions

Acquiring information about drug efficacy is not only at the center of FDA regulations, but also the key element driving each prescription decision in doctor's office. Using a unique data set from patient diaries, we estimate how patient satisfaction and other factors affect the diffusion of Cox-2 inhibitors from 2001 to 2003. Our results suggest that prescription choice is sensitive to many sources of information, including patient satisfaction, Medline articles, news report and manufacturer advertising. In comparison, the impact of FDA updates is close to zero once we control for Medline articles. This suggests that the contents of FDA updates have already been included in Medline articles and therefore deliver no new information to doctors. This also confirms the view that FDA postmarketing surveillance lags behind the medical literature and has room to improve.

According to our counterfactual predictions, setting up a nationwide database of patient feedback encourages doctors to switch from traditional NSAIDs to Cox-2s, but increasing Medline publications about Cox-2s steals market share away from Cox-2s. This suggests that patient feedback and academic articles may reflect different dimensions of drug quality, and these two sources of information do not necessarily substitute for each other.

Despite our efforts devoted at gathering every piece of information about Cox-2, our results are subject to several limitations: first of all, the patient diary data do not contain doctor identities and only represents a sample of all the Cox-2 patients. Both tend to undermine our ability to precisely estimate how doctors learn across patients. Second, our patient satisfaction data are self-reported. This does not necessarily generate a specific bias as compared to the patients' real experiences, but it does put more weight on the symptoms that patients can observe easily and care to report to their doctors. Third, although our model of patient-drug match already incorporates heterogeneity reflected in the satisfaction data, it is possible that there remain some patient attributes observable to doctors but not to researchers. We use patient-drug random effects to control for such unobserved heterogeneity, but we might still be ignoring some sources of heterogeneity. Finally, manufacturers may advertise more in a period that they expect to have low sales, thus introducing an endogeneity problem. This suggests that the coefficients of advertising should be interpreted as the correlation between advertising and prescription choice, rather than as having a causal impact.

In summary, this is a first attempt at using actual consumer (or patient) feedback information in the context of a learning model. Future research can look at including other sources of information within the formal learning framework proposed here.

**Acknowledgements** We wish to thank the editor and two anonymous referees for constructive comments and suggestions. We also thank Simon Anderson, John Chao, Andrew Ching, Greg Crawford, Tülin Erdem, Phillip Leslie, Jackie Luan, Puneet Manchanda, Sean Nicholson, Peter Reiss, Jesse Shapiro, Vishal Singh, Russ Winer, Marta Wosinska, seminar attendees at the University of Pennsylvania, Stanford University, and New York University, as well as the participants at the 2006 American Society of Health Economists Conference, the 2007 Conference on Empirical IO, the 2007 International Industrial Organization Conference, and the 2007 Quantitative Marketing and Economics Conference. Research assistance from Yan Chen is greatly appreciated. All errors remain ours.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

**Appendix: Formula for the posterior variance of  $\tilde{Q}_{pjt}$**

The purpose of this appendix is to show that we can obtain  $var(\tilde{Q}_{pjt})$  directly from the elements in  $\Sigma^{-1}$ . Since  $\tilde{Q}_{pjt} = \tilde{Q}_{jt} + \tilde{q}_{pjt}$ , in order to compute the variance of  $\tilde{Q}_{pjt}$ , we need to first obtain the variances of  $\tilde{Q}_{jt}$  and  $\tilde{q}_{pjt}$  (diagonals of  $\Sigma$ ), as well as the covariance between them (first row of  $\Sigma$ ). As stated in the paper,  $\Sigma^{-1}$  is the natural product in deriving the Normal posterior density. We explore the special structure of  $\Sigma^{-1}$ , and provide the analytical formula for  $\Sigma$  and  $var(\tilde{Q}_{pjt})$  as follows. Denote

$$\Sigma^{-1} = \{a_{ij}\} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{12} & a_{22} & 0 & \cdots & 0 \\ a_{13} & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{1n} & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

$\Sigma^{-1}$  is symmetric with non-negative elements; the submatrix after crossing out the first row and first column is diagonal.

Step 1: Compute the determinant of  $\Sigma^{-1}$   
 $|\Sigma^{-1}| = a_{11} \cdot term1 \cdot (1 - term2)$ , where  $term1 = \prod_{k=2}^n a_{kk}$ ,  $term2 = \sum_{j=2}^n \frac{a_{1j}^2}{a_{11}a_{jj}}$

Step 2: Compute the elements in  $(\Sigma^{-1})^* = \{a_{ij}^*\}$   
 First,  $a_{11}^* = term1$ . Next, for  $i, j = 2, \dots, n$ ,

$$a_{1i}^* = a_{i1}^* = -term1 \cdot \frac{a_{1i}}{a_{ii}}$$

$$a_{ii}^* = term1 \cdot \left( 1 - term2 + \frac{a_{1i}^2}{a_{11}a_{ii}} \right) \cdot \frac{a_{11}}{a_{ii}}$$

$$a_{ij}^* = term1 \cdot \frac{a_{1i}a_{1j}}{a_{ii}a_{jj}}, \text{ for } i \neq j$$

Step 3: Obtain  $\Sigma = \frac{1}{|\Sigma^{-1}|} (\Sigma^{-1})^*$

Step 4: Compute  $\text{var}(\tilde{Q}_{pit})$  as:

$$\begin{aligned} \text{var}(\tilde{Q}_{pit}) &= \text{var}(\tilde{Q}_{jt}) + \text{var}(\tilde{q}_{pit}) + 2\text{cov}(\tilde{Q}_{jt}, \tilde{q}_{pit}) \\ &= \Sigma_{11} + \Sigma_{ii} + 2\Sigma_{1i} \\ &= \frac{1}{|\Sigma^{-1}|} (a_{11}^* + a_{ii}^* + 2a_{1i}^*) \\ &= \frac{1}{|\Sigma^{-1}|} \left[ \text{term1} + \text{term1} \cdot \left( 1 - \text{term2} + \frac{a_{1i}^2}{a_{11}a_{ii}} \right) \right. \\ &\quad \left. \cdot \frac{a_{11}}{a_{ii}} - 2\text{term1} \cdot \frac{a_{1i}}{a_{ii}} \right] \\ &= \frac{1}{a_{ii}} + \frac{1}{a_{11}(1 - \text{term2})} \left( 1 - \frac{a_{1i}}{a_{ii}} \right)^2 \end{aligned}$$

where  $i > 1$ . Thus, we can obtain  $\text{var}(\tilde{Q}_{pit})$  directly from the elements in  $\Sigma^{-1}$ .

## References

- Avorn, J., & Soumerai, S. B. (1983). Improving drug-therapy decisions through educational outreach: A randomized controlled trial of academically based 'detailing'. *New England Journal of Medicine*, 308(24), 1457–1463.
- Azoulay, P. (2002). Do pharmaceutical sales respond to scientific evidence? *Journal of Economics and Management Strategy*, 11(4), 551–594.
- Bowman, D., Heilman, C., & Seetharaman, S. (2004). Determinants of product-use compliance behavior. *Journal of Marketing Research*, 41, 324–338.
- Chan, T., Narasimhan, C., & Xie, Y. (2007). *Impact of treatment effectiveness and side effects on prescription choices*. Working paper, Washington University at St. Louis.
- Ching, A. (2005). *Consumer learning and heterogeneity: Dynamics of demand for prescription drugs after patent expiration*. Working paper, University of Toronto.
- Coscelli, A., & Shum, M. (2004). An empirical model of learning and patient spillovers in new drug entry. *Journal of Econometrics*, 122, 213–246.
- Crawford, G., & Shum, M. (2005). Uncertainty and learning in pharmaceutical demand. *Econometrica*, 73, 1137–1173.
- Dai, C., Stafford, R. S., & Alexander, C. G. (2005). National trends in cyclooxygenase-2 inhibitor use since market release: Nonselective diffusion of a selectively cost-effective innovation. *Achieves of Internal Medicine*, 165, 171–177.
- Dayton, M., & Macready, G. (1988). Concomitant-variable latent class models. *Journal of the American Statistical Association*, 83, 173–178.
- DeGroot, M. (1970). *Optimal statistical decisions*. New York: McGraw-Hill.
- Erdem, T., & Keane, M. (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods market. *Marketing Science*, 15, 1–20.
- Food and Drug Administration (2005). *Analysis and recommendations for agency action regarding non-steroidal anti-inflammatory drugs and cardiovascular risk*. FDA Memorandum as of April 6, 2005.
- Lilien, G. L., Kotler, P., & Moorthy, K. S. (1992). *Marketing models*. Englewood Cliffs: Prentice Hall.

- McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Eds.), *Frontiers in econometrics*. New York: Academic.
- Narayanan, S., Manchanda, P., & Chintagunta, P. K. (2005). Temporal differences in the role of marketing communication in new product categories. *Journal of Marketing Research*, *42*, 278–290.
- Ray, W. A., & Stein, C. M. (2006). Reform of drug regulation beyond an independent drug-safety board. *The New England Journal of Medicine*, *354*, 194–201.
- Slater, E. E. (2005). Today's FDA. *The New England Journal of Medicine*, *352*, 293–297.
- Venkataraman, S., & Stremersch, S. (2007). The debate on influencing doctors' decisions: Are drug characteristics the missing link? *Management Science*, *53*(11), 1688–1701.