Heterogeneous Learning and the
Targeting of Marketing Communication for New Products

Sridhar Narayanan
Puneet Manchanda

August 2005

1 Sridhar Narayanan is Assistant Professor of Marketing at the Graduate School of Business, Stanford University and Puneet Manchanda is Associate Professor of Marketing at the Graduate School of Business, University of Chicago. The authors would like to thank Pradeep Chintagunta, Peter Rossi, Guenter Hitsch, seminar participants at Cornell, Dartmouth, Erasmus, Emory, Harvard, Hong Kong University of Science and Technology, Minnesota, New York University, Northwestern, Purdue, Stanford, SUNY-Buffalo, UCLA, University of Southern California, University of Texas at Dallas, Washington University at St. Louis and Wisconsin and participants at the Advanced Research Techniques Forum (2005), Marketing Dynamics Conference (2005) and Bayesian Methods in Marketing Conference at Linz (2005). Manchanda would also like to acknowledge support from the Kilts Center for Marketing and the Beatrice Foods Faculty Fund at the University of Chicago. All correspondence may be addressed to the first author at Stanford University, Graduate School of Business, 518 Memorial Way, Stanford, CA 94305 (Tel: 650-723-9675, Fax: 650-725-7979) or via email at sridhar.narayanan@stanford.edu.
Heterogeneous Learning and the Targeting of Marketing Communication for New Products

Abstract

New product launches are often accompanied by extensive marketing communication campaigns. Firms’ allocation decisions for these marketing communication expenditures have two dimensions – across consumers and over time. What makes this problem hard in the case of new products is that consumers are uncertain about the quality of new products and learn about them through marketing communication. Further, different consumers may have different rates of learning about product quality, i.e. there may be heterogeneous learning. Thus, consumer responsiveness to marketing communication would vary along two dimensions. For each consumer, this responsiveness would vary over time, as she learns about product quality. Across consumers, there would be differences in responsiveness in each time period. For optimal allocation of marketing communication across both consumers and time, firms would need estimates of how responsiveness to marketing communication varies across consumers and over time.

Past studies in this area have typically studied one of these two dimensions in which responsiveness varies. They have either looked at heterogeneity in responsiveness across agents or the variation in responsiveness over time. In the context of new products, past research has looked at how consumer learning about product quality causes responsiveness to vary over time. However, there is no study that we are aware of that allows for heterogeneous learning rates, i.e. heterogeneity in how consumers learn over time. In this study, we develop the methodology for estimating individual-level parameters of learning for consumers that differ on their learning processes and use a rich panel dataset that allows us to estimate these parameters of the model.

To obtain individual-level estimates of learning, we add a hierarchical Bayesian structure to the Bayesian learning model. We exploit the natural hierarchy in the Bayesian learning process to incorporate it within the hierarchical Bayesian model. We use data augmentation, coupled with the Metropolis Hastings algorithm to make inferences about individual-level parameters of learning. We conduct this analysis on a unique panel dataset of physicians, where we observe prescription decisions and detailing (salesforce efforts) at the individual physician-level for a new prescription drug category.

Our results show that there is significant heterogeneity across physicians in their rates of learning about the quality of new drugs. We also find that there are asymmetries in the temporal evolution of responsiveness of physicians to detailing – physicians who are more responsive to detailing in early periods are less responsive later on and vice versa. These finding have interesting implications for targeting of detailing across physicians and over time. We find that firms could increase their revenues if they took these temporal and cross-sectional differences in responsiveness into account while deciding their allocations of detailing.

Keywords: Resource Allocation, Pharmaceutical Markets, Learning Models, Markov Chain Monte Carlo Methods
1. Introduction

New products are the lifeblood of firm performance – they account for about a quarter of all sales and revenue growth. Firms also spend about half their marketing budgets promoting new products (see Urban and Hauser 1993). A major concern of firms is the allocation of marketing resources during the launch and rollout of a new product. The mechanism that governs this resource allocation over time is the response of the market to this new product. This response is driven by the considerable uncertainty about the product quality of the new product. A major role of marketing activity is to disseminate information about the new products in a manner that consumers reduce their uncertainty about its quality via a “learning” process. Besides the evolution of this learning process over time, consumers may also differ in their learning behavior i.e., there may be heterogeneous learning. For example, some consumers may learn faster than others about the quality of these new products. Firms could then use their knowledge of this heterogeneous learning to allocate their marketing resources both over time and across consumers.

In this research, we use data on the launch of new (ethical) pharmaceutical drugs to estimate individual physician level rates of learning. Our key research contribution is to provide a method by which these rates can be computed. This is in contrast to the previous literature where these rates have been typically been computed at the pooled level. In addition, our approach is able to estimate these rates after controlling for other individual level behaviors such as risk aversion and responsiveness to patient requests. We then use these estimated rates to examine firms’ marketing resource allocation during the time of product rollout.

The pharmaceutical industry in general (and prescription drug categories in particular) is particularly well suited to study this problem for three reasons. First, uncertainty about drug quality is particularly relevant in the case of prescription drugs. In spite of an extensive process of clinical trials before the launch of new drugs, there is still considerable uncertainty about their quality, their side effects and any risks associated with administering the drug. For instance, the Food and Drug Administration (FDA), without whose approval prescription drugs cannot be marketed, has the following to say about new drugs (CDER 2000)

“The practical size of pre-marketing clinical trials means that we cannot learn everything about the safety of a drug before we approve it. Therefore, a degree of uncertainty always exists about the risks of drugs.”
Furthermore, while the pre-marketing clinical trials provide information at the mean level (efficacy, incidence of side effects etc.), physicians with heterogeneous patient bases are often uncertain about the quality of a new drug for their specific patient bases.

Second, pharmaceutical firms spend a large amount of money on marketing communication directed towards physicians. For example, the industry spent $8.5 billion on marketing communication directed at physicians in 2000 (Wittink 2002, Neslin 2001). Most of this is spent on detailing, which involves personal sales calls made by salespersons of the pharmaceutical firm on physicians. Third, since detailing is a personal interaction between a physician and the firm’s representatives, it is allocated at the individual physician level. Firms need to decide how many calls to make to each individual physician and when to make these calls. These three factors - true uncertainty about a new product, large expenditure and individual level allocation make the pharmaceutical industry an excellent setting for our problem.

In this study, we use a unique panel dataset of physicians, which contains prescription and detailing information for all drugs in a category to estimate individual physician-level effects in the presence of learning. We develop a methodology to estimate individual physician level effects by specifying a Bayesian learning model for each physician and estimating the model in a hierarchical Bayesian framework using Markov Chain Monte Carlo methods. Our results show that there is considerable heterogeneity in learning rates across physicians. For example, the number of details required for the uncertainty to drop to ten percent of its initial value ranges from one to forty-six across physicians. We then show that this heterogeneity is economically significant for the firm. In fact, heterogeneity in learning is about twice as important as other individual physician level effects in enhancing firm revenue.

The rest of the paper is organized as follows. We first discuss the extant literature related to this study briefly. Then we discuss the data. Before discussing the model itself, we present some model free evidence for the presence of learning and heterogeneity in learning. We then present the model in detail and discuss our estimation strategy. Subsequently, we discuss the identification of our model parameters. We then discuss the results of our analysis and present the counterfactual experiments we conduct. Finally, we conclude, discussing the model, the results and the limitations of the study.
2. Related Literature

This study is broadly related to three streams of research in the literature – pharmaceutical promotions, learning and targeted marketing. While there have been many studies that have looked at pharmaceutical promotions (some with physician learning), none of these have looked at targeting of these pharmaceutical promotions in the presence of learning. Our contribution will be to study the heterogeneity in physician learning about new drugs and use this to analyze the targeting of detailing expenditure for new drug introductions.

There has been a large amount of research in the area of pharmaceutical demand in marketing, economics as well as medical sciences (see Manchanda and Honka 2005 for an extensive review of detailing studies). Early studies in the marketing literature (Parsons and Vanden Abeele, 1981; Lilien, Rao and Kalish 1981) studied the effect of sales force effort on sales using aggregate data. Montgomery and Silk (1972) is an early study that investigates the effect of marketing communication expenditures on sales of prescription drugs. Recent research (Kamakura, Kossar and Wedel 2004; Manchanda and Chintagunta 2004; Gönül et al. 2001; Wosinska 2002 and Manchanda, Rossi and Chintagunta 2004) has used panel data to investigate the effect of detailing and/or DTC advertising on pharmaceutical demand. There has also been research that has specifically investigated the role of pharmaceutical promotion and classified it as informative and persuasive (Leffler 1981; Hurwitz and Caves 1988; Rizzo 1999; Currie and Park 2002; Narayanan, Manchanda and Chintagunta 2005). The broad consensus in this literature is that detailing positively affects prescriptions by physicians. In addition, some of these studies find evidence for significant amount of heterogeneity in physician response to detailing.

The second stream of research that is relevant to this study is the literature on Bayesian learning. Early studies that used a Bayesian learning process include Stoneman (1981), Jensen (1982), Meyer and Sathi (1985) and Roberts and Urban (1988). Erdem and Keane (1996) was a pioneering paper in marketing that used a model of Bayesian learning to incorporate informative effects of advertising. Since then, there has been a growing interest in problems involving learning. Crawford and Shum (2005), Coscelli and Shum (2004), Ching (2002), Anand and Shachar (2001), Currie and Park (2002), Ackerberg (2003), Narayanan, Manchanda and Chintagunta (2005), Byzalov and Shachar (2004) and Mukherji (2002) all apply Bayesian learning models to a variety of contexts. These studies find evidence that there is significant amount of learning and uncertainty reduction through advertising and product experience. Ackerberg (2003), Currie and Park (2002), Narayanan, Manchanda and Chintagunta (2005) and
Byzalov and Shachar (2004) specifically address the issue of informative and non-informative roles of advertising or promotional activity and attempt to estimate both of them from the data. Narayanan, Manchanda and Chintagunta (2005) find evidence for the presence of both these roles in the case of a new pharmaceutical category, while the other studies find evidence for only the informative effect. Heilman, Bowman and Wright (2000) and Ackerberg (2001) also suggest that the role of marketing communication is different for new and familiar products. Akçura, Gönül and Petrova (2004) adopt an alternative approach to learning, using a Kalman-filter based model instead of Bayesian learning model to study patients’ learning about over-the-counter drugs through their own usage experience. They find variation in learning rates across patients and drugs.

Targeted promotions have interested researchers in marketing in recent years. Targeting promotions to segments of consumers has long been an industry practice as well as a topic for research. Numerous studies have looked at price discrimination (cf. Villas-Boas 1999; Fudenberg and Tirole 2000), targeted coupons (Shaffer and Zhang 1995) and the targeting of advertising (cf. Iyer, Soberman and Villas-Boas 2005) to different segments of consumers. Rossi, McCulloch and Allenby (1996) and Manchanda, Rossi and Chintagunta (2004) are instances of research that study the targeting of promotional activities towards individual consumers. These involve the estimation of individual-level response parameters using Bayesian methods. This stream of research demonstrates that firms can obtain significant benefits by targeting their promotions.

There has been a significant amount of study about the allocation of salesforce efforts in the salesforce literature. This study is also related to that literature, since we focus on the targeting of detailing (sales calls). Early studies like Lodish (1971); Montgomery, Silk and Zaragoza (1971) and Lodish (1980) used decision calculus based methods to arrive at efficient allocation of salesforce efforts across customer accounts. Lodish (1975); Zoltners (1976); Zoltners and Sinha (1983); Rangaswamy, Zoltners and Sinha (1990) have studied how to design optimal salesforce territories. Horsky and Nelson (1996) study the problem of optimal salesforce size and productivity. Mantrala, Sinha and Zoltners (1992) study the effect of resource allocation rules on profitability.

Thus, there is a relatively large stream of literature that studies allocation of resources, in particular of salesforce efforts, another stream that studies learning and a third stream that studies the responsiveness of physician prescription behavior to marketing communication. However, the study of new products and specifically new drugs, on which firms spend a significant part of
their promotional budget is still an open area for research in the literature. This study attempts to fill this gap by studying the targeting of detailing to individual physicians in the case of new drugs.

3. Data

The dataset used for the empirical analysis in this study is from the category of prescription drugs known as ‘Erectile Dysfunction Drugs’. The drugs in this category are prescribed to treat ‘Erectile Dysfunction’ (ED) amongst adult men. About 15 to 30 million men in the United States are believed to be affected by the condition. There is only one category of oral drugs that can treat this condition and currently, there are three drugs that have been approved by the Food and Drug Administration (FDA). The first drug to be approved in the category was Viagra, marketing by Pfizer, and was approved by FDA in March 1998. Levitra, marketed jointly by GSK Pharmaceuticals and Bayer was approved in August 2003 and Cialis, marketed by Eli Lilly, was approved in November 2003. No further drugs have been approved after Cialis.

The data are at the physician level and consists of a panel of 900 physicians in the United States. These data were obtained from a data-vending firm – ImpactRx, based in New Jersey, which has set up this panel and sells the data to pharmaceutical firms. The panel is a representative sample of the universe of physicians, balanced across geographic regions, physician specialties and prescription volume. For these physicians, we have observations of prescriptions written by them for their patients and also of detailing calls made by representatives of pharmaceutical firms (detailers). These data are collected directly from the physician, using a Personal Digital Assistant based diary method. All detailing calls made to the physician are recorded. Totally, we have 15320 prescription observations and 16700 observations of detailing calls in the dataset.

For each physician, we have information about their specialty (General Practitioner or specialist and the specific specialty). In each prescription observation, we observe the drug that was prescribed to the patient. Unlike most existing data sources, which collect prescription data from pharmacies or insurance companies, our dataset actually captures the physician’s decision. This is because pharmacy data, for instance, may not be able to capture the fact that the drug actually filled out in the pharmacy may sometimes be different from that prescribed by the physician. Thus, this captures the physician’s intended prescription and is not susceptible to
“slippage” or missing data. A unique feature of the prescription data in our dataset is that we observe if a patient requested a drug. This is recorded by the physician at the time of the consultation by the patient. In terms of detailing, the physician records the drug that is detailed in that call on every detailed visit.

The share of prescriptions across all physicians and across all time periods for Cialis, Levitra and Viagra are respectively 21.0%, 28.8% and 50.2% respectively. However, this analysis over all time periods does not give the full picture, since the shares of these drugs are changing over time. In the last month for which we have data, the largest share drug is Cialis, with a share of 38.1%. Viagra and Levitra have market shares of 33.4% and 28.5% respectively. All three drugs are heavily detailed, with each physician receiving an average of 18.6 details during the 9 month period in our data, i.e. more than 2 detailing calls per month on average. The number of patient requests were relatively small – on average patients requested any one of the three drugs only 15.4% of the time. The largest number of requests in aggregate are for Viagra (51.7% of the total requests), followed by Levitra (25.5%) and Cialis (22.8%). However, in the last month in which data are available, about 46.3% of the requests were for Cialis, with Viagra (32.6%) and Levitra (21.1%) making up the rest of the requests.

4. Model Development

In this section, we discuss the model used in this study. We first describe the model structure and lay out the underlying assumptions. We then describe the model specification in detail.

4.1 Prescription Decision

We assume that the physician is the sole decision maker for which drug is to be prescribed to a patient. However, we shall allow the patient to influence the prescription decision of the physician. The data allows us to incorporate this influence of the patient and we shall describe this in more detail in the discussion on the model specification. In general, the prescription decision is also influenced by intermediaries like insurance firms, HMOs, Medicare, Medicaid etc through the inclusion or exclusion of specific drugs in their formularies. However, in the time period for our data, Erectile Dysfunction drugs were not on the formularies of most HMOs. Also, there were no generic substitutes to the brand name drugs. Hence, it is reasonable to assume in the case of this category that these intermediaries did not play a role in the decision of which drug to prescribe and that it was the physician’s decision alone.
The physician is assumed to value the health of the patient and her preferences are assumed to map into a utility function over the space of treatment options. The physician may value the health of the patient both out of a sense of professional integrity and also to avoid malpractice litigation in the future. A desire to build and maintain a reputation may also motivate a physician to desire the best treatment for her patients. Thus, the physician is assumed to choose the option that provides the greatest utility. We also allow the physician to take the patient’s preferences into account (conditional on the patient expressing these preferences). The drugs in this Erectile Dysfunction category are seen as substitutes and there are no instances when multiple drugs within the category are prescribed to the same patient. Thus, the physician makes a discrete choice amongst the drugs within the category.

The physician is assumed to be a utility maximizer, i.e. her preferences over the space of drugs are assumed to map into a utility function defined on these drugs. Further, she is assumed to be uncertain about the quality of a new drug. The term ‘quality’ refers to a scalar that summarizes aspects of the drug like its efficacy (how well it treats the condition for which it is prescribed), side effects, etc. This quality enters the utility function of the physician (details are in the section on model specification). Since the physician is uncertain about this quality, she is in turn uncertain about the utility function also. Thus, we also make the assumption that she is an expected utility maximizer, i.e. she chooses the alternative that provides her the greatest expected utility. Finally, we assume that physicians are not forward looking.2

4.2 Learning
As described earlier, physicians are assumed to be uncertain about the quality of drugs. At this stage, it is important to clarify that there could be two levels of uncertainty about drug quality. The first level of uncertainty is regarding the mean quality of the drug across the patient base of the physician. A second level of uncertainty could be regarding the match between a specific patient and a specific drug. In this study, we focus on the uncertainty about the mean quality of the drug and learning about this mean quality. There are two reasons for this focus. First, uncertainty about the mean quality of the drug is likely to be important in the case of new drugs. Second, we do not have data that would allow us to model the learning about the patient-drug match for a specific patient, since we do not have repeated observations of prescription decisions for the same patient. We only observe the first prescription occasion for a particular patient.

2 A model without this assumption is analytically intractable. This is because the state-space implied by such a model will be very large – its dimension will be equal to the product of the number of physicians and the state variables for each physician
Therefore, when we refer to quality, we would be referring to the mean quality of the drug across the patient base of the physician.

Physicians are assumed to learn about the quality of a new drug through two main sources – the feedback they receive from their own patients who were prescribed the drug in the past (which we shall henceforth refer to as patient feedback), and marketing communication directed towards them by pharmaceutical firms. Marketing communication directed towards physicians is primarily in the form of detailing.3

Physicians are assumed to be Bayesian updaters. Thus, at any given period of time, they have some prior belief about the quality of the drug, which they update with information from patient feedback and marketing communication received at that time using Bayes Rule to form their posterior belief. This posterior belief in turn becomes the prior belief in the next time period and the process is repeated then. Whenever a patient walks in and a prescription decision is to be made, the physician uses the most updated belief about the quality of the drug in the decision.

We make the following assumptions regarding the learning process. In our dataset, we do not observe when a particular patient gives feedback to the physician. We assume therefore, that every prescription results in feedback and that the physician receives this feedback one month after the prescription is written. The fact that every prescription results in a feedback is not a critical assumption. The results of the analysis will not be affected as long as we are willing to believe that a fixed proportion of past prescriptions would result in feedback. If that were the case, the parameters related to this feedback would be scaled by this proportion of past prescriptions. However, marginal effects and elasticities would remain unaffected, as we shall see in subsequent sections.

Patient feedback and marketing communication are assumed to provide unbiased but noisy signals about the true quality of a drug. In our operationalization, we assume that these signals are normally distributed, with the mean being the true quality. Physicians are assumed to see the realizations of these signals and know the variances of these signals but do not know the true mean. The assumption that detailing provides an unbiased signal is based on the following two reasons. First, the signal is what the physician perceives from the interaction with the detailer. Given that a physician has repeated contact with detailers, the notion that physicians are able to discount any bias in detailers’ messages would make the assumption of unbiasedness a

3 Physicians could also learn via using the free samples provided to them by pharmaceutical firms. Our data unfortunately do not contain information on sampling. In general, sampling and detailing is highly correlated and therefore this is less of a concern.
valid one. Second, the information provided during a call is regulated by the Food and Drug Administration (FDA). Any visual aids and brochures used during the call have to be pre-approved by the FDA and have to include negative as well as positive aspects of the drug. Thus, the information has to be factual and to that extent, the detailing call has to be unbiased.

Finally, we make the assumption that at the initial period (i.e. before they prescribe any drug or receive any detailing calls), all physicians start off believing that the quality of the drug is the population distribution of the true quality of the drug across all physicians. We further make an assumption that this is a normal, with its mean equal to the population mean of the true quality distribution across physicians and its variance is equal to the population variance of this distribution. In other words, physicians are assumed to know the distribution of qualities across physicians, but are uncertain about the true quality of the drug for their specific patient base. This is a reasonable assumption since drugs go through an extensive approval process overseen by the Food and Drug Administration (FDA) before they can be introduced to the market. Our assumption is that physicians know about the mean and variance of the effect for a representative sample of patients at drug launch. The one-on-one interaction provided by the detailing call as well as feedback from patients who try the new drug then helps the physician learn about the true quality of the drug for her unique patient base.

4.3 Risk Aversion
We assume that physicians are risk averse with respect to their uncertainty about the true quality of a new drug for their specific patient base. We further make the specific assumption that they have a constant absolute risk aversion (CARA). Thus, quality is assumed to enter the utility function in the CARA form. In other words, our assumption is that irrespective of the absolute quality level of the drug, their risk aversion remains the same. This is a reasonable assumption since the drugs in the category we study do not differ dramatically in their quality levels. Hence, it is likely that the utility function would have similar levels of concavity in the local region of the quality line where all the drugs are located. If however, there were dramatic differences between the drugs in terms of their absolute quality levels, one may be potentially concerned about assuming a constant absolute risk aversion utility function.

4.4 Informative and Persuasive effects of marketing communication
We allow for marketing communication to affect physician utility in two ways. We shall refer to the effect on utility through the learning process as the informative effect. Additionally, marketing communication can directly affect the utility of the physician. This effect, which we
shall refer to as the *persuasive effect* may represent any prestige or image effects (cf. Becker and Murphy 1993, Ackerberg 2003) or reminder effects. We shall capture these effects in a reduced form manner by allowing a linear stock of detailing calls to enter the utility function. This separation of the two effects of marketing communication is in line with prior research (Anand and Shachar 2001; Narayanan, Manchanda and Chintagunta 2005). In our empirical specification, we assume that the count of detailing calls in the preceding month constitutes this stock for the direct effect. It is important to note that these terms – “informative” effect and “persuasive” effect - are just convenient labels to differentiate between the two effects and do not refer to the underlying constructs of information and persuasion.

4.5 Heterogeneity

Physician-level heterogeneity is a critical aspect of our study, especially with regards to learning. This is because different physicians have different levels of training and ability. Also, they operate in different environments (e.g., some are in research hospitals, while others have rural practices). Prior research on learning has typically assumed homogenous learning effects across consumers (or physicians). This has been led both by data limitations, particularly in the case of studies that have used aggregate data as well as limitations in the available estimation methodologies. For example, in pharmaceutical categories, past research on learning has been limited by the absence of panel data of physicians with detailing efforts recorded for all the drugs in a category. We allow physician response heterogeneity to manifest itself over four different aspects of marketing communication and individual differences. We describe these in detail below.

As mentioned earlier, in our setup, detailing (or more generally marketing communication) has two effects – the informative effect and persuasive effect. As noted before, learning is related to the informative effect. We incorporate heterogeneity in this effect, in our specification by allowing the detailing signals to have different variances for different physicians. This implies that different physicians see these signals with different levels of noise. This is a plausible assumption, considering that different physicians have different levels of ability and training and pay different levels of attention to information they receive from detailing calls. It is important to note that a signal is the physician’s perception of the information given in a detailing call. Thus, the same signal may be perceived by different physicians with different levels of noise. Some physicians may be able to perceive the information received in the detailing call precisely while others may not be able to perceive it as precisely. The implication of this is that
the former would learn through fewer detailing calls than the latter. This is in contrast to the models used in the extant literature on learning, where all agents are assumed to learn at the same rate.

Heterogeneity in the persuasive effect implies that different physicians have different levels of response to prestige or reminder effects for instance, and in our specification, this would manifest itself in heterogeneity in the coefficient for the detailing stock variable.

Similar to the heterogeneity in the informative effect of detailing, we also allow heterogeneity in the learning through patient feedback by allowing the variance on the feedback signals to differ by physician. Again, the underlying assumption is that physicians differ in their abilities and training levels and therefore, the degree of noise in patient feedback differs by physicians. The physicians who receive more precise signals of patient feedback learn faster than physicians who receive less precise signals.

Finally, the true quality is also allowed to be heterogeneous. The reason why different physicians may have different true quality levels is that the patient base may be different for different physicians. For instance, some physicians may have typically older patients than other physicians. Since drugs differ on their efficacy and side effects for different types of patients, the true quality levels may also differ by patient base of physician.

5. Model Specification

5.1 Utility Function

When physician $i$ has to make a decision on which drug to prescribe at occasion $t$, she chooses the alternative $j$ that provides the greatest utility, with the utility function defined as

$$
\hat{U}_{ijt} = f(\hat{Q}_{ijt}) + X_{ijt}\beta_i + \epsilon_{ijt}
$$

where

$\hat{Q}_{ijt}$ is physician $i$’s belief about the true quality of drug $j$ at time $t$ and is stochastic from the point of view of the physician

$f(\hat{Q}_{ijt})$ is a function through which the quality belief enters the utility function

$X_{ijt}$ is a row vector ($I \times K$) of physician, drug and time (patient) specific variables,

$\beta_i$ is a column vector ($K \times I$) of physician specific sensitivity to these variables.
\( \varepsilon_{ijt} \) is an i.i.d. physician, drug and time specific shock and could include a patient and drug specific match value.

Note that since we have an observation every time a patient walks into the office and since we do not have repeated observations for the same patient, the prescription occasion \( t \) is identical to patient \( t \) for the physician.

This utility is stochastic from the physician’s perspective because the belief about the quality \( \hat{Q}_{ijt} \) is stochastic. But it must be remembered that \( \varepsilon_{ijt} \) is not stochastic from the point of view of the physician (i.e., the physician observes the \( \varepsilon_{ijt} \) on each prescription occasion). The physician is assumed to be an expected utility maximizer. This expected utility is

\[
U_{ijt} = E \left[ \bar{U}_{ijt} \right] = E \left[ f(\hat{Q}_{ijt}) \right] + X_{ijt} \beta_I + \varepsilon_{ijt}
\]  

(2)

Whether the physician is risk averse or risk neutral with respect to the uncertainty in quality beliefs depends on the exact specification of the function \( f(\hat{Q}_{ijt}) \). In particular, if the physician were assumed to be risk-neutral, the function would be linear in \( \hat{Q}_{ijt} \). If \( f(\hat{Q}_{ijt}) \) is specified to be non-linear, the physician would not be risk-neutral. Here, we assume that the physician is risk averse with respect to the uncertainty in the quality belief and specifically assume that the function \( f(\hat{Q}_{ijt}) \) takes the CARA form. Thus,

\[
f(\hat{Q}_{ijt}) = -\exp(-r\hat{Q}_{ijt})
\]  

(3)

5.2 Quality Evolution

Physicians are assumed to update their quality belief in each period based on signals they receive through detailing and through patient feedback (via past prescriptions). These signals are assumed to be normally distributed around the physician-specific true mean quality of the drug.

Assume that there are \( nd_{ijt} \) detailing signals at time \( t \) and the \( m^{th} \) signal is assumed to be given as

\[
\hat{D}_{ijtm} \sim N(Q_{ij}, \sigma_{D_{ij}}^2)
\]  

(4)

and that there are \( nf_{ijt} \) patient feedback signals feedback signal, and the \( m^{th} \) signal is given by

\[
\hat{F}_{ijtm} \sim N(Q_{ij}, \sigma_{F_{ij}}^2)
\]  

(5)
A series of unobserved signals that are normally distributed can be summarized by their sample mean, which is also normally distributed. We define these sample means as follows

$$\bar{D}_{ijr} = \frac{\sum D_{ijw}}{nd_{ijr}} \sim N\left(Q_{ijr}, \frac{\sigma_{D_{ij}}^2}{nd_{ijr}}\right) \quad (6)$$

$$\bar{F}_{ijr} = \frac{\sum F_{ijw}}{nf_{ijr}} \sim N\left(Q_{ijr}, \frac{\sigma_{F_{ij}}^2}{nf_{ijr}}\right) \quad (7)$$

The quality belief at time $t=0$ is assumed to be a normal distribution whose mean is the mean of the population distribution of the true quality and whose variance is the variance of this population distribution. Thus, the assumption is that physicians know the distribution of the true quality across all physicians but are uncertain about the true quality for their own specific patient base.

The physician is assumed to update his beliefs in a Bayesian manner, i.e. at any given period of time, he combines his prior belief about the quality of the drug with the information obtained through both detailing and feedback signals and applies Bayes Rule to form his posterior belief. Since the prior belief at time $t=0$ and all signals are assumed to be normally distributed, it turns out that the posterior belief at every time period is also a normal distribution. This posterior belief is given by

$$\tilde{Q}_{ijr} \sim N\left(Q_{ijr}, \sigma_{\tilde{Q}_{ijr}}^2\right) \quad (8)$$

where

$$Q_{ijr} = \frac{\sigma_{Q_{ijr}}^2}{\sigma_{Q_{ijr}(t-1)}^2} Q_{ijr(t-1)} + nf_{ijr} \frac{\sigma_{Q_{ijr}}^2}{\sigma_{F_{ijr}}^2} \bar{F}_{ijr} + nd_{ijr} \frac{\sigma_{Q_{ijr}}^2}{\sigma_{D_{ijr}}^2} \bar{D}_{ijr} \quad (9)$$

and

$$\sigma_{\tilde{Q}_{ijr}}^2 = \frac{1}{\sigma_{Q_{ijr(t-1)}}^2 + \frac{nd_{ijr}}{\sigma_{D_{ijr}}^2} + \frac{nf_{ijr}}{\sigma_{F_{ijr}}^2}} \quad (10)$$

It is important at this stage to point out how heterogeneity in learning manifests itself in the model. Note that the variances of the detailing and feedback signals ($\sigma_{D_{ij}}^2$ and $\sigma_{F_{ij}}^2$, respectively), are physician specific. It can be seen from equation (9) that for a physician for
whom $\sigma^2_{D_i}$ is a large value, relatively lower weight is placed on the detailing signal $\tilde{D}_i$ than for a physician with a low value of $\sigma^2_{D_i}$, everything else remaining the same. The former physician is a slow learner from detailing than the latter. Similarly, a physician with a higher value of $\sigma^2_F$ would be a slower learner from feedback than a physician with a lower value of this parameter. Thus, the variances of the detailing and feedback signals summarize the heterogeneity in learning across physicians. By contrast, in traditional learning models, these variances are assumed to be homogenous across agents. Allowing for these variances to be individual specific is a key contribution of our approach.

Given that the quality belief in any period is a normal distribution with mean $Q_{ijt}$ and variance $\sigma^2_{Q_{ijt}}$, we can now evaluate the expected utility of the physician as

$$U_{ijt} = E[U_{ijt}] = -\exp \left( -rQ_{ijt} + \frac{r^2\sigma^2}{2} \right) + X_{ijt} \beta_i + \epsilon_{ijt}$$

(11)

5.3 Informative Effect of Detailing

The physician’s belief at the initial period (t=0) is given by the population distribution of true qualities. As he receives detailing and feedback signals, the belief changes. In particular, if we see equations (9) and (10), it will be clear that the mean of the quality belief, $Q_{ijt}$, evolves to the physician-specific true mean quality, $Q_i$, and the variance of the belief, $\sigma^2_{Q_{ijt}}$, evolves to 0. Focusing on the detailing signal, every detailing signal thus has an effect on the quality belief. This in turn affects the expected utility in any given period, which in turn influences the probability of prescribing the drug. Thus, detailing has an effect on prescription behavior through this process of learning. We shall refer to this effect as the informative effect of detailing.

It is important to note that the informative effect of detailing could positively or negatively affect the probability of prescribing a drug. If the true quality of the drug for the patients of a particular physician is higher than the population average, detailing would increase the probability of prescribing the drug. On the other hand, if the true quality for the physician is lower than the population average, the probability of prescribing the drug would reduce with detailing due to the informative effect.
It can also be seen from equations (9) and (10) that the effect of detailing on the physician’s quality belief is highest initially and reduces with every subsequent updation. This can be seen from the fact that the variance of the detailing signal $\sigma^2_{o}$ converges towards zero with every updation and thus, the coefficient of the detailing signal in equation (9) also converges to zero. Thus, the informative effect of detailing is highest at the introductory phase of a drug and after the physician has learnt about the drug and reduced her uncertainty, this effect is negligible.

### 5.4 Persuasive Effect of Detailing

In the previous section, we discussed the informative effect of detailing that influences the physician’s prescription decision through the learning process. In particular, we found that this is highest initially and is negligible in later phases of a new drug’s life cycle. However, it is well documented that detailing has an effect on physicians’ prescription behavior even in the case of mature products (cf. Gönül et al 2001; Manchanda and Chintagunta 2004). It has been suggested that this effect in the case of mature products may be because of an image or prestige role of detailing, or perhaps due to reminder effects. We shall refer to all effects of detailing except the informative effect defined in the previous section as the persuasive effect of detailing. As noted earlier, these terms – *informative effect* and *persuasive effect* – are labels to differentiate between the two effects and are not meant to convey that one effect exclusively involves information and the other persuasion.

The persuasive effect of detailing is captured in a reduced form manner by including a stock of detailing counts in the linear $X_{ijt}$ variable in the utility function (equation 1). The coefficient of this variable measures the persuasive effect of detailing. This effect would capture any role of detailing that remains unchanged over the product life cycle of the drug.

### 5.5 Patient Influence

The data allows us to account for patient influence on the prescription decision of the physician. We observe in the data if the patient requested a specific drug or not. We allow a dummy variable indicating whether the patient requested the drug or not in the linear $X_{ijt}$ variable in the utility function in equation (1). The coefficient of this variable captures the influence of the patient’s request on the prescription decision of the physician. In an indirect way, this also captures the effect of direct to consumer advertising (DTC), since DTC often asks a patient to talk to the doctor about the drug.
5.6 Hierarchical Bayesian Model

The objective of our empirical analysis is to suggest revenue-enhancing allocation plans for detailing. Since detailing has to be allocated at the individual physician level, the estimates of the effect of detailing also have to be obtained at the individual physician level. Past studies using Bayesian learning models have used frequentist methods for estimation, making the estimation of individual physician level parameters infeasible. A Bayesian approach is a natural way to estimate individual physician-level parameters. We provide a new modeling approach by estimating a Bayesian learning model under a Hierarchical Bayesian framework. We use Markov Chain Monte Carlo (MCMC) methods to estimate the individual-level parameters.

However, the challenge in specifying the model as a Hierarchical Bayesian Model is that the quality beliefs are unobserved. In a standard frequentist estimation of learning models (Erdem and Keane 1996), one could integrate out these unobserved quality beliefs by simulation methods. However, for using MCMC methods, we make use of a simple observation. From equation (9), it is clear that not just is the quality belief $\tilde{Q}_{ijt}$, a stochastic variable, but even its mean $Q_{ij}$ is stochastic. This is because, from equation (9), $Q_{ij}$ is a function of two stochastic variables – the realizations of the detailing and feedback signals, $\tilde{D}_{ij}$ and $\tilde{F}_{ij}$ respectively. Further, since these two variables are assumed to have normal distributions, $Q_{ij}$ is a linear combination of normal variables, and therefore is also a normal variable.

In particular, we can derive the distribution of $Q_{ij}$, conditional on $Q_{ij(t-1)}$ as

$$Q_{ij} \mid Q_{ij(t-1)} \sim N\left(\tilde{Q}_{ij}, v^2_{ij}\right)$$  \hspace{1cm} (12)

where

$$\tilde{Q}_{ij} = \frac{\sigma^2_{Q_{ij}}}{\sigma^2_{Q_{ij(t-1)}}} Q_{ij(t-1)} + \left(\frac{nf_{ij}}{\sigma^2_{F_{ij}}} + nd_{ij} \frac{\sigma^2_{Q_{ij}}}{\sigma^2_{D_{ij}}}\right) Q_{ij}$$  \hspace{1cm} (13)

and

$$v^2_{ij} = nf_{ij} \frac{\sigma^4_{Q_{ij}}}{\sigma^4_{F_{ij}}} + nd_{ij} \frac{\sigma^4_{Q_{ij}}}{\sigma^4_{D_{ij}}}$$  \hspace{1cm} (14)
Note that from equation (10), the variance of the quality belief $\sigma^2_{Q_{ijt}}$ is not a stochastic variable, conditional on the parameters of the model. Given the parameters of the model, it is known deterministically. The main difference between the mean $Q_{ijt}$ and variance of the quality belief $\sigma^2_{Q_{ijt}}$ is that while the former depends on the (unobserved) realizations of the detailing signal $\bar{D}_{ijt}$ and the feedback signal $\bar{F}_{ijt}$, the latter does not. Hence, in a frequentist estimation of a learning model, we integrate out the mean but not the variance of the quality belief. Analogously, in a hierarchical model, we specify the mean of the quality belief as a level of the hierarchy, but not the variance.

Given that we can write the unobserved mean of the quality belief in any period as a random variable, conditional on the mean of the quality in the previous period, we thus have a natural hierarchy of quality beliefs

$$Q_{ijt} \mid Q_{ij(t-1)} \sim N(\bar{Q}_{ijt}, \nu_{ijt}^2)$$
$$Q_{ij(t-1)} \mid Q_{ij(t-2)} \sim N(\bar{Q}_{ij(t-1)}, \nu_{ij(t-1)}^2)$$
$$\vdots$$
$$Q_{ij1} \mid Q_{ij0} \sim N(\bar{Q}_{ij1}, \nu_{ij1}^2)$$

(15)

We can then specify the rest of the hierarchical model as follows. We assume that the prescription choice follows a probit process. Thus, the random errors $\varepsilon_{ijt}$ of the utility function in equation (1) follow a multivariate normal distribution.

$$\begin{bmatrix} \varepsilon_{ij1} \\ \vdots \\ \varepsilon_{ijb} \end{bmatrix} \sim MVN(0, \Sigma)$$

(16)

The alternative that provides the greatest utility is chosen. Hence, $U_{ijt}$ follows a truncated multivariate distribution, conditional on choice. If choice is given by the indicator variable $I_{ijt}$ (which is 1 if brand $j$ is chosen and 0 otherwise), the truncation is such that

$$U_{ijt} > U_{ikt}, I_{ijt} = 1, I_{ikk} = 0 \quad \forall k \neq j$$

(17)
Let the vector $\gamma_i$ (dimension K x 1) denote the individual level parameters of the model. These parameters are specified as a function of physician-level characteristics, as follows

$$\gamma_i = \left[ \beta_i' \ln\left(\sigma^2_i\right) \ln\left(\sigma^2_{D_i}\right) \ln\left(\sigma^2_{F_i}\right) \ln(r_i) \right]^T \sim MVN\left(\Lambda Z_i, V_{\gamma}\right) \quad (18)$$

where

$\gamma_j$ is a M x 1 column vector of physician characteristics, including a first element which has the value 1; and

$\Lambda$ (K x M matrix) and $V_{\gamma}$ (K x K matrix) are parameters

Thus, the hierarchical model can be specified as follows

$$U_{ijt} \mid I_{ijt}, X_{ijt}, Q_{ijt}, \sigma^2_{ijt}, \sigma^2_{Q_{ijt}}, r_i, \Sigma$$

$$Q_{ijt} \mid Q_{ij(t-1)}, nd_{ijt}, I_{ij(t-1)}, \sigma^2_{ijt}, \sigma^2_{Q_{ij}}, r_i$$

$$\gamma_i \mid \Lambda, Z_i, V_{\gamma}$$

In order to complete the model, the priors for the parameters are specified as follows:

$$\Sigma = \begin{pmatrix} \sigma^2_1 & \ldots & 0 \\ \vdots & \sigma^2_j & \vdots \\ 0 & \ldots & 1 \end{pmatrix}, \quad \sigma^2_j \sim IG\left(s_j, s_{2j}\right)$$

$$\lambda = \text{vec}\left(\Lambda'\right) \sim N\left(\bar{\Lambda}, V_{\lambda}\right)$$

$$V_{\gamma} \sim \text{Wishart}(g, G)$$

$$Q_{j0} \sim N\left(\bar{Q}_0, \theta_0\right)$$

In the current version of the model, the individual-level parameters are not related to the physician demographics, i.e. $Z_i$ is a scalar with the value 1. Thus, M=1. However, we shall conduct some post-estimation analysis to find relationships between the individual-level parameters and physician demographics.
6. Estimation

As described earlier, the purpose of this study is to be able to estimate individual physician-level parameters of the model. One approach that may be considered is to specify a model of learning and then estimate such a model for each individual physician, given the data for that physician. However, this approach is not feasible since the data required for such estimation is not available for each of the physicians in the dataset. Hence, a Hierarchical Bayes approach is a natural alternative, which utilizes shrinkage to get usable estimates for each of the physicians in the data, irrespective of the amount of data available for these physicians.

The Bayesian method of inference for the parameters of the model, including the individual-physician level parameters, involves obtaining draws from the joint posterior distribution of these parameters. However, the joint posterior distribution does not correspond to any known distribution family. Hence, we use the Gibbs Sampler to obtain draws from this joint posterior distribution of the parameters by sequentially drawing from the full conditional distributions of sub-vectors of the full parameter vector. These sub-vectors are chosen such that it is easy to draw from their respective full conditional distributions. We iterate this process of making draws from the respective full-conditional distributions until we attain convergence to the true joint posterior distribution of the parameters (cf. Gelfand and Smith 1990). Some of these full conditional distributions are known distributions and hence drawing from them is trivial. However, the full conditional distributions for some of the parameters are in turn not from any of the known distribution families. Hence, we use the Metropolis Hastings Algorithm (Chib and Greenberg 1995) to make draws from these full conditional distributions. Additionally, using the Data Augmentation approach (Tanner and Wong 1987), we treat the unknown utilities \( U_{ijt} \) and mean efficacies \( Q_{ijt} \) as parameters and make draws for them from their own full conditional distributions. The full details of the likelihood, the full conditional distributions and the details on the algorithm are given in the Appendix.

The empirical estimation of the model was done using a program written in ‘C’. The chain for the Gibbs Sampler was run for a total of 100,000 observations. The first 50,000 draws were discarded as ‘burn-in’ before convergence was attained. The subsequent draws were used for inference.
7. Identification

The identification of the parameters of the model is discussed in this section. In particular, it is important to point out how we are able to identify individual-specific parameters of learning and also how we are able to separate the informative and persuasive effects of detailing. Additionally, we also conducted simulations to assess the identification of the parameters of our model and found that we were able to recover the parameters of the model with a reasonable degree of accuracy.

The individual level parameters in the model are – the variances of the detailing and feedback signals ($\sigma_D^2$ and $\sigma_F^2$ respectively), the risk aversion parameter $r_i$, the true mean quality of each drug $Q_{ij}$ and the linear coefficients $\beta_i$.

It must also be noted that we need multiple observations of new drugs in order to infer that the learning rates are systematic to the physician. It would otherwise be hard to separate from random events like specific realizations of the random error $\varepsilon_{ijt}$, which would give us similar prescription patterns. However, if we observe the patterns of evolution of prescription behavior for the same physician for multiple drugs, we would be able to make inferences about the learning rate and thus about the parameters that summarize this learning rate ($\sigma_D^2$ and $\sigma_F^2$).

The true quality $Q_{ij}$ of the drug is identified out of the steady state prescription behavior of the physician. We have already seen that as the physician learns, the quality belief $Q_{ijt}$ evolves to the true quality of the drug $Q_{ij}$. At the extreme, at steady state, the quality belief is indeed the true quality. Thus, we need to observe a long enough time series of prescriptions for the physicians to be able to correctly identify the true mean quality. In the data we use for our empirical analysis, the data is observed for about 9 months after launch. Hence, the prescription behavior of the physician towards the end of the dataset tells us about the true mean quality of the drug.

Coming to the identification of the risk aversion parameter, we first note from equation (11) that the expected utility of the physician depends not just on the mean of the quality belief $Q_{ijt}$ in any time period, but also on its variance $\sigma_{Qijt}^2$. We have seen earlier that this variance declines monotonically as the physician learns about the drug. It is easy to see from equation (11)
that the expected utility of the physician is inversely related to this variance. Additionally, this variance is interacted with the risk aversion parameter – the higher the risk aversion, the lower will be the expected utility for a given value of this variance. The lower the expected utility, the lower would be the probability of prescribing a drug. Since all physicians are assumed to start with the same quality belief and this variance is the same for all physicians, systematic differences in shares of prescribing the new drug in early periods after introduction would be explained only by differences in risk aversion. Since we observe two new drugs being introduced in our data, the identification is even stronger. Physicians who are more risk averse would have lower initial probability of prescription for every new drug compared to the average physician. Thus, correlation in such behavior across initial periods after introduction of at least two new drugs is helpful in identifying an individual specific risk aversion coefficient, while it can, in principle be identified even without such observations of multiple drug introduction. In our dataset, we observe two new introductions, thus strengthening the identification of this parameter.

The variances of the detailing and feedback signals are identified from the evolution patterns of physician prescription behavior and how they are related with detailing and feedback signals. We have seen earlier that with every signal, the physician’s quality belief is updated. As a result, the mean of this quality belief evolves from the mean of the initial quality belief towards the mean of the final quality belief. Equation (9) summarizes the process of updating of the mean of the quality belief. As the quality belief for a drug evolves, the probability of prescribing that drug also evolves over time. We have already discussed, when discussing the model specification, how a greater variance of the detailing signal implies a slower learning rate. Thus, the rate of learning helps identify the variance parameter for the detailing signal. A similar argument holds in the case of the variance of the feedback signal.

The linear coefficients $\beta_i$ are identified from the covariance of the prescription behavior of the physician with the linear variables (detailing stock variable and dummy variable for patient requests). An important concern could be about the separate identification of the informative and persuasive effects of detailing. The identification of the persuasive effect, which is the coefficient of the detailing stock variable in the utility equation (equation 1), is aided by the fact that in our dataset, the incumbent drug – Viagra – has existed for about 5 years before the first observation in the data. Hence, we make the a priori assumption that physicians have learnt fully about Viagra and hence the informative effect for this drug is zero. Hence, any effect of detailing of Viagra on the prescription behavior of physicians is entirely due to the persuasive effect. For the new
entrants – Levitra and Cialis - both the informative and persuasive effects are present. Hence, we are able to identify both these effects.

A potential concern could be about endogeneity. Specifically, if firms optimally decide on their detailing levels to individual physicians based on a complete knowledge of their behavior, then the detailing would be endogenous and hence parameter estimates would be biased. There is an institutional feature of the pharmaceutical industry because of which this may not be a significant concern for the purpose of our study. The typical rule that firms use to decide on their detailing levels to physicians is a volumetric decile-based rule. Physicians are typically classified into deciles based on their total category volumes and detailing allocations are made at the decile levels. We find that this is consistent with the patterns of detailing in our data. Figure 6 illustrates the allocation of detailing for Levitra across deciles of physicians over the first seven months after launch of the drug. It is clear that the allocation of detailing for each decile is relatively unchanged over this period. Manchanda, Rossi and Chintagunta (2004) have a discussion on the rules used by firms and also show empirically that firms are not optimal in their detailing decisions for a mature product. Manchanda, Xie and Youn (2004) also find similar decile-based detailing patterns by firms. If concerns on endogeneity were to persist, it might be possible to specify a model in which the detailing decision is also modeled and parameters of the detailing equation and the physician prescription equation are estimated simultaneously, like in Manchanda, Rossi and Chintagunta (2004).

8. Results
We first present the parameter estimates of the model. In Table 1, we report the parameter estimates for the individual-level parameters for the model. The individual-level parameters are the detailing signal variance (\( \sigma^2_D \)), feedback signal variance (\( \sigma^2_F \)), the absolute risk aversion (\( r_i \)), the coefficients for the detailing stock (\( \beta_1 \)) and patient requests (\( \beta_2 \)); and the true mean qualities for Cialis (\( Q_1 \)) and Levitra (\( Q_2 \)). In our Bayesian inference, we obtain a distribution for each individual-level parameter for each physician. We compute the mean parameter value for each physician and then report the mean and standard deviation across physicians of this individual-level mean parameter value in Table 1.
The estimate for detailing signal variance is 1.3505. This parameter value implies that it takes about 12.1545 detailing calls for the uncertainty of a physician to reduce to one-tenth of its initial value. This is an estimate of the *informative* effect of detailing. Similarly, the parameter estimate of the feedback signal variance is 1.5629, and this corresponds to 14.0661 feedback signals to reduce the uncertainty to one-tenth of its initial value. This also suggests that an average detailing call is more informative than an average feedback signal since it requires a smaller number of detailing calls than feedback signals to reduce the physician’s uncertainty about drug quality. This is consistent with the findings of prior research (cf. Narayanan, Manchanda and Chintagunta 2005). The standard deviations for these estimates suggest that there is considerable heterogeneity across physicians in these parameters. We shall return to a discussion on heterogeneity in these parameters later in this section.

The parameter estimates for the coefficient of the linear detailing stock ($\beta_{1i}$) suggest that there is a positive *persuasive* effect of detailing. Thus, even after a physician has reduced his uncertainty about the drug, there is still a positive effect of detailing. There is also a strong positive effect of patient feedback, as indicated by the parameter estimates for the coefficient for patient requests ($\beta_{2i}$). To the best of our knowledge, this is the first time that the effect of patient requests has been quantified. Given the data description earlier, it is clear that the incidence of patient requests is very low at about 15%, but if a patient makes a request, it has a strong effect on the physician’s prescription behavior.

The parameter estimates for the mean true qualities for Cialis and Levitra are both close to one (the true quality of Viagra, which we fix for identification purposes), suggesting that the qualities of the three drugs are similar, on average, albeit at the individual physician level, they may differ significantly. Cialis is on average marginally of higher quality than Viagra and Levitra is of marginally lower quality than Viagra. This is consistent with industry reports that Cialis is a higher quality drug. Note that in steady state, the share of prescriptions of the drug with the highest true quality would be largest.

In Table 2, we report the parameter estimates for the pooled parameters. The pooled parameters in this model are the variances of the normal errors in the utility function ($\sigma_i^2$ and $\sigma_j^2$). The additional pooled parameters include the parameters that relate the individual-level parameters to physician demographics, i.e. $\lambda$ and $V_\gamma$ - which are currently not included in the
model. Thus, the vector of means of the individual-level parameters is the same as \( \lambda \) and the variance-covariance matrix is the same as \( V_{\gamma} \).

Next, we move to the heterogeneity in the individual-level parameter estimates. Figure 1 depicts the histograms of the individual-level parameters across physicians. Figures 1(a) and 1(b) respectively show the histograms of the detailing and feedback signal variances. These figures suggest that there is considerable heterogeneity in these parameters across physicians. There is much greater heterogeneity in the detailing signal variance than the feedback signal variance. Figure 1(c) shows the distribution of the risk aversion parameter. Physicians are fairly heterogeneous in their risk aversion levels as well. We shall, in the counterfactual experiments presented in the next section, assess the economic significance of heterogeneity in learning as well as risk aversion.

We find from observing Figure 1(d) that for all physicians, the mean of the detailing stock coefficient is positive. Furthermore, for an overwhelming majority of these physicians, the 95% credible intervals of the individual level parameter does not include zero. Thus, it can be concluded that the detailing stock has a positive effect on their choice probabilities of the detailed brand. Figure 1(e) shows the distribution across physicians of the patient request coefficient. It is positive for all physicians and the 95% credible interval does not include the value zero for any physician.

In Figures 1(f) and 1(g), we show the distributions across physicians of the true mean qualities of Cialis and Levitra respectively. Once again, on average, Cialis has a higher quality than Viagra and Levitra a lower quality than Viagra, but there is considerable heterogeneity in the true quality levels. Therefore, for a particular physician, it is not necessary that this rank ordering of drugs be maintained.

When talking about heterogeneity, a valid concern could be whether the individual-level parameters are significantly different from each other. One could compare the individual-level parameters for each pair of physicians to see if they differ from each other in statistical terms. We present a more informal analysis of heterogeneity of these individual parameters across physicians. In Table 3, we compare the across-physician and within-physician standard deviations of these parameters. If the across-physician standard deviation for a parameter is smaller than or similar to the within-physician standard deviation, it would suggest that the 95% credible intervals for the physicians overlap and hence they are not significantly different from each other in terms of that parameter. On the other hand, if the across-physician standard deviation is larger
than the within physician standard deviation, it would provide greater support to the claim that the individual-level parameters are significantly different for different physicians.

We find that the across-physician variation is much higher than the within-physician variance in the case of the two signal variances. Since the heterogeneity in the detailing signal variance summarizes the heterogeneity in learning as discussed in the model specification section, we can conclude that there is significant amount of heterogeneity in learning across physicians.

The coefficient of the detailing stock also has a higher across-physician standard deviation than the within-physician standard deviation. Thus, there is a significant amount of heterogeneity in this parameter as well. In the case of the other individual-level parameters – the risk aversion parameter, the coefficient for patient requests and the true mean qualities for Cialis and Levitra – the within-physician standard deviation is larger than the across-physician standard deviation, suggesting that physicians do not significantly differ in these parameters.

In order to explore the heterogeneity in learning across physicians even further, we plot a histogram of the number of detailing calls required to reduce the physician’s uncertainty about a new drug to one-tenth its initial value. This is computed using the parameter estimates of the detailing signal variance for each physician. Figure 2, which shows this plot, suggests that heterogeneity in the detailing signal variance parameter indeed manifests itself in significant heterogeneity in the number of calls required to reduce the physicians’ uncertainty.

An interesting fact about the parameter estimates is the negative correlation between the informative and persuasive effects of detailing for physicians. What this means is that physicians who have a high informative effect of detailing are likely to have a low persuasive effect and vice versa. The informative effect of detailing is summarized by the detailing signal variance. We have already seen that a high informative effect is equivalent to saying that the learning rate for the physician is high. And this manifests itself in the parameter estimates in terms of a low value of the detailing signal variance. Similarly, a low informative effect manifests itself in a high detailing variance. The persuasive effect is summarized by the coefficient of the detailing stock variable. Thus, a physician for whom the value of this coefficient is low would have a low persuasive effect and vice versa. Therefore, a negative correlation between the informative and persuasive effects of detailing implies a positive correlation between the detailing signal variance and the detailing stock coefficient. In Figure 3, we plot the detailing signal variance for each physician against the detailing stock coefficient to show this positive correlation between the parameters and consequently the negative correlation between the informative and persuasive
effects. In Figure 4, we plot the detailing signal variance against the risk aversion parameter, to assess whether these parameters are correlated.

In order to understand what explains the heterogeneity in the key parameters of interest, we regress the individual level parameters observed physician characteristics. We have two physician characteristics, one of which is the specialty of the physician (whether a General Physician or a Urologist or some other specialty) and the volume-based decile of the physician. The volume-based decile is obtained by computing the total category volume of the physician in the 3-month period immediately before the first month in our dataset and then categorizing all physicians into deciles based on this volume. Thus, we have sets of dummy variables for specialty and decile. We conduct three such regressions for three sets of individual-level parameters – the detailing signal variance ($\sigma_{D_i}^2$), the risk aversion parameter ($r_i$), and the coefficient of the linear detailing stock variable ($\beta_{D_i}$). The regression estimates are reported in Table 4.

The key finding of this analysis is that the persuasive effect of detailing, summarized by the detailing stock coefficient is significantly correlated with the decile of the physician. All but one of the decile dummy variables are significant at the 95% level and one is significant at the 90% level. Furthermore, there is a pattern to the persuasive effect. The detailing stock coefficient is lower for the physicians in the higher deciles. Thus, physicians who prescribe more on average are less influenced by detailing in the steady state. However, the risk aversion and detailing signal variance do not seem to be explained by decile. Three of the decile dummies are significant at the 90% level or above in the case of the detailing signal variance, and four dummies are significant in the case of the risk aversion parameter. The coefficients for physician specialty are not significant even at the 90% level in any of the three regressions.

9. Managerial Implications
The negative correlation between the informative and persuasive effects has implications for the allocation of detailing efforts across physicians and over time. To illustrate this, let us consider a situation with two physicians, one with a high informative effect and a low persuasive effect, and the other with the opposite. Let us, for the sake of convenient refer to the former as the fast learner and the latter as the slow learner. For both these physicians, the informative and persuasive effects are present in the introductory phases of the new drug’s life cycle. In later
stages, only the persuasive effect plays a role. For the fast learner, the total effect of detailing starts off at a very high level, but rapidly reduces till it converges to the persuasive effect and then remains constant at that level. For the slow learner, everything remaining the same, the total effect again starts at a high level (perhaps not as high as for the fast learner) but falls more slowly. It converges to a persuasive effect that is higher than for the fast learner. Since this total effect denotes the responsiveness of the physician to detailing and the optimal detailing level depends on responsiveness, it would be optimal for firms to allocate higher amounts of detailing initially to the fast learner but then rapidly reduce this allocation. On the other hand, for the slow learner, we might expect to see a corresponding increase in allocation over time.

We have also seen that the informative and persuasive effects are related to the decile of the physician. Specifically, physicians in higher deciles have a lower persuasive effect and those in lower deciles have a higher persuasive effect. Thus, if firms were to optimally allocate their resources to physicians, one would expect to see them allocating a high proportion of resources to physicians in high deciles in the early stages after the introduction of the drug and then reducing this proportion over time. However, discussions with practitioners in the industry reveal that firms do not allocate their detailing efforts in this manner. The most commonly used rule for allocating resources to physicians is a decile-based rule, which remains constant over time. We find the same patterns in our data as well, with detailing for every decile remaining approximately constant over the entire period in our dataset.

We conduct three counterfactual experiments to see how firms could increase their revenues if they took into account learning by physicians and heterogeneity in this learning. In each of these experiments, we conduct separate experiments for each of the two new drugs – Levitra and Cialis – one by one. We keep the total amount of detailing by firms in the first three months after the launch of the respective drug as fixed. We only alter the allocation of detailing across physicians or over time and then compute the predicted expected revenues. In all these experiments, detailing calls are rounded off, i.e. a physician can only get an integer number of calls. All these calls are assumed to be made in the beginning of the month, so as to abstract away from the problem of timing of detailing calls within a month. Competitors’ detailing is kept unchanged in all these experiments. The ‘optimal’ allocations are obtained using a numerical optimization routine. We conduct all these experiments for a subset of 100 physicians randomly chosen from the sample. This is done in order to keep the optimization feasible. We compare the predicted revenues in the counterfactual case with those using the actual allocation plans.
In Experiment 1, we change only the temporal allocation of detailing, but keep the cross-sectional allocation unchanged, i.e. we vary the allocation of detailing to each of the three months in the launch quarter, but do not vary the proportions of detailing within that month to individual physicians. Thus, this is a two-dimensional optimization exercise, with the unknown variables being the allocation for the first two months (and the allocation of the third month automatically known from these since the proportions add up to 1). We find that by varying the month-to-month allocations of detailing, Levitra can get revenue gains of 7.6% compared to the current allocation plan, while Cialis could obtain an even higher gain of 10.3%. This reflects the fact that firms could gain by frontloading their detailing to the period immediately after launch. This is because in this early period, all physicians have both an informative effect and a persuasive effect and are hence most responsive to detailing. As they learn about the drug, the informative effect asymptotes down to zero and hence only the persuasive effect is present. Thus, for all physicians, the responsiveness to detailing reduces over time. This result is similar to that reported in Narayanan, Manchanda and Chintagunta (2005).

In Experiment 2, we change the cross-sectional allocation of detailing, but keep the temporal allocation unchanged. Thus, we keep the total amount of detailing within each month fixed, but vary how much of that is allocated to each of the physicians. In each month, we have an independent optimization, with the dimension of the unknown vector being one less than the number of physicians (i.e. 99 since the experiment was conducted for 100 physicians). We find that there are relatively modest revenue increases that can result through this exercise, 4.6% for Levitra and 6.3% for Cialis. The reason for revenue increases being lower in this case is perhaps that the firms are using some decile-based rule for allocation and while this is not optimal, it does inherently take into account heterogeneity across physicians (since the informative and persuasive effects of detailing are explained to some extent by decile).

In Experiment 3, we allow both the temporal and cross-sectional allocation of detailing to change, i.e. we find the detailing level for each physician for each month that maximizes revenues. We find that this gives us a revenue increase of 12.1% for Levitra and 16.3% for Cialis.

The results of these three counterfactual experiments are summarized in Table 5. To sum up, it appears that firms could significantly increase their revenues during the launch period by simply reallocating their existing expenditure on detailing. An interesting extension could be study what would be the optimal levels of detailing and not just the optimal allocations of
detailing. Note that in all the three experiments we have conducted, the total amount of detailing in the first three months after launch was kept fixed. However, firms may be interested in knowing if they could reduce their total detailing expenditure during launch. This could also be addressed by future research.

While these counterfactual experiments give us a sense of the revenue gains that firms could make by taking heterogeneity into account, the extent to which these gains arise because of heterogeneity in learning, in risk aversion and in the persuasive effect is not yet obvious. In order to assess the contribution of heterogeneity in these three parameters on the revenue gains, we sequentially repeat experiment 3 taking into account heterogeneity in only one of these parameters at a time. Thus, we first assume that the risk aversion parameter is homogenous across physicians and set it to the population mean of this parameter. We also fix the detailing stock coefficient at the population mean for all physicians. We then compute the revenue gains in the three experiments described earlier. This gives us a measure of the revenue gains by optimally allocating detailing calls, taking into account only the heterogeneity in learning. We then fix the detailing signal variance and detailing stock coefficient at their respective population means to assess the revenue gains from heterogeneity in risk aversion. The same exercise is repeated for the persuasive effect as well (i.e. detailing stock coefficient). The results of this exercise are summarized in Table 6.

The results in Table 6 suggest that between 40% and 50% of the revenue gains in experiment three (where detailing is reallocated both temporally and cross-sectionally) come through accounting for the heterogeneity in learning. Between 20% and 30% of the gains come from accounting for heterogeneity in the risk aversion parameter. Accounting for heterogeneity in the persuasive effect gives a gain of about 30%.

10. Conclusion
We started off with a problem of optimal allocation of marketing communication for new products, with the specific focus on allocation of these resources over time and across consumers. Heterogeneous learning was recognized as a factor that would affect the temporal as well as cross-sectional allocation of these resources. We specified a structural model of heterogeneous learning and developed a methodology to estimate such a model at the individual physician level. We estimated this model using a unique panel dataset consisting of physician prescriptions and
detailing calls. We allowed for detailing to have both an informative and a persuasive effect and estimated both these effects at the individual physician level. We then conducted a set of counterfactual experiments to find the implications of heterogeneous learning on optimal allocations of detailing over time and across agents (physicians in our case).

Our parameter estimates indicate that there is considerable heterogeneity across physicians in terms of their learning rates. Some physicians require only a few detailing call to substantially reduce their uncertainty about a new drug, others require a large number of repeated detailing calls in order to reduce their uncertainty to the same extent. Physicians also differ significantly in their persuasive effect of detailing, which is the only effect of detailing after they have learnt about the drug. Because of both these effects, there is a significant amount of variation across physicians in terms of how their responsiveness to detailing varies over time.

We also find that volume based deciles explain the variation in the persuasive effects of detailing to some extent. Specifically, physicians who are heavy prescribers in the category are more likely to have a low persuasive effect. Thus, their responsiveness to detailing reduces rapidly to a low level. The responsiveness of light prescribers reduces more slowly and settles down at a relatively high responsiveness after they have learnt about the drug.

We conducted three counterfactual experiments to see if firms could increase their revenues in this category by changing their detailing allocation patterns. We find that if they change just their temporal allocation without changing the cross-sectional allocation of detailing, they get a 7.6% to 10.3% increase in revenues in the first three months after launch. This reflects gains in revenues by front-loading their detailing to early periods after the launch of the drug. If they change their cross-sectional allocation without altering the temporal allocation, firms obtain a more modest 4.6% to 6.0% increase in revenues. If they change both their temporal and cross-sectional allocations, they get a substantial 12.1% to 16.3% increase in revenues. Furthermore, accounting for heterogeneity in learning accounts for 40% to 50% of these gains, while accounting for heterogeneity in risk aversion accounts for about 20-25% of the gains and accounting for heterogeneity in the persuasive effect accounts for about 30% of these gains.

Note that while our empirical analysis is conducted for a category of prescription drugs, our model structure and estimation method is generalizable. Specifically, it would be valid in any situation where consumers learn about new products through some means of marketing communication. For instance, in the case of industrial goods, firms often send out salespersons to their customers not just to sell their products but also to inform them about new products or
services. This is virtually identical to the situation in this paper. Even in the case of consumer goods, firms may need to allocate their resources for informative advertising across markets or segments of consumers even if they do not necessarily target individual consumers. Accounting for differences in learning rates across different segments of consumers or different markets may be important in arriving at optimal allocation plans for advertising.

We finally list some of the limitations of this study. In specifying this model of learning about new drugs, we have assumed away other potentially important sources of learning, for instance, learning from other physicians. This assumption of no learning through other sources is due to the absence of appropriate data and might overstate the degree of learning through detailing that we infer. Another potentially important concern could be about endogeneity. Firms could be at least partially optimizing their detailing allocations already and the presence of the resulting endogeneity may bias our estimates. While there are reasons to believe that this may not be a very big concern for the specific category we study, this could be addressed by including a detailing supply equation in the model and jointly estimating parameters of demand and supply. A complication that would arise in this case is that the detailing supply equation cannot be static due to the presence of learning. Learning causes detailing to have persistent effects over time and hence firms are likely to take into account the effect of their detailing on future prescription behavior of physicians. This would complicate the problem substantially. We also assume away any forward-looking behavior of physicians. Physicians, if they are aware that they learn through detailing, may have incentives to be more willing to see detailers early on in order to learn about the drug quickly. These are challenging problems and the methodology to account for these phenomena are not fully developed yet. Future research could potentially address these questions.
References


### Table 1: Individual Level Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailing signal variance</td>
<td>$\sigma^2_{D_i}$</td>
<td>1.3505</td>
</tr>
<tr>
<td>Feedback signal variance</td>
<td>$\sigma^2_{F_i}$</td>
<td>1.5629</td>
</tr>
<tr>
<td>Absolute risk aversion</td>
<td>$r_i$</td>
<td>0.1833</td>
</tr>
<tr>
<td>Coefficient – detailing stock</td>
<td>$\beta_{1i}$</td>
<td>0.6083</td>
</tr>
<tr>
<td>Coefficient – patient request</td>
<td>$\beta_{2i}$</td>
<td>1.5772</td>
</tr>
<tr>
<td>True Mean Quality – Cialis</td>
<td>$Q_{1i}$</td>
<td>1.0080</td>
</tr>
<tr>
<td>True Mean Quality – Levitra</td>
<td>$Q_{2i}$</td>
<td>0.9991</td>
</tr>
</tbody>
</table>

Notes:
1. Since these parameters are at the individual level, for each individual physician, the parameter has a mean and a standard deviation. The reported parameters are the mean and standard deviation of the physician-specific parameter means.

### Table 2: Pooled Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Mean</th>
<th>Std. Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utility Error Variance</td>
<td>$\sigma^2_1$</td>
<td>5.7249</td>
</tr>
<tr>
<td>Utility Error Variance</td>
<td>$\sigma^2_2$</td>
<td>5.3632</td>
</tr>
</tbody>
</table>

Notes:
1. Other pooled parameters include the elements of the matrix $V_\gamma$ and the vector $\lambda$. These are not reported here for the sake of brevity.
Table 3: Across-physician vs. within-physician standard deviations of the individual-level parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Across-physician standard deviation</th>
<th>Within-physician standard deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Detailing signal variance</td>
<td>$\sigma^2_{D_i}$</td>
<td>0.7795</td>
</tr>
<tr>
<td>Feedback signal variance</td>
<td>$\sigma^2_{F_i}$</td>
<td>0.8793</td>
</tr>
<tr>
<td>Absolute risk aversion</td>
<td>$r_i$</td>
<td>0.1521</td>
</tr>
<tr>
<td>Coefficient – detailing stock</td>
<td>$\beta_{1i}$</td>
<td>0.1032</td>
</tr>
<tr>
<td>Coefficient – patient request</td>
<td>$\beta_{2i}$</td>
<td>0.0688</td>
</tr>
<tr>
<td>True Mean Quality – Cialis</td>
<td>$Q_{i1}$</td>
<td>0.1142</td>
</tr>
<tr>
<td>True Mean Quality – Levitra</td>
<td>$Q_{i2}$</td>
<td>0.1186</td>
</tr>
</tbody>
</table>

Notes:
1. Across-physician standard deviation: the mean parameter value for each physician is first computed and then the standard deviation of these mean values is reported in this table.
2. Within-physician standard deviation: the within-physician standard deviation of the parameter is computed for each physician and then the mean of these standard deviations is reported in this table.
## Table 4: Regression of individual-level parameters on physician characteristics

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Detailing Signal Variance $\sigma^2_{d_i}$</th>
<th>Absolute Risk Aversion $r_i$</th>
<th>Detailing Stock Coefficient $\beta_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1.4589**</td>
<td>0.2412**</td>
<td>0.5895**</td>
</tr>
<tr>
<td>Specialty – GP</td>
<td>-0.0156</td>
<td>-0.0266</td>
<td>-0.0205</td>
</tr>
<tr>
<td>Specialty – Urologist</td>
<td>-0.1835</td>
<td>-0.0285</td>
<td>-0.0196</td>
</tr>
<tr>
<td>Decile 1</td>
<td>-0.1211</td>
<td>-0.0392*</td>
<td>0.0527**</td>
</tr>
<tr>
<td>Decile 2</td>
<td>-0.2995**</td>
<td>-0.0252</td>
<td>0.0541**</td>
</tr>
<tr>
<td>Decile 3</td>
<td>-0.1885*</td>
<td>-0.0354</td>
<td>0.0438**</td>
</tr>
<tr>
<td>Decile 4</td>
<td>-0.0885</td>
<td>-0.0612**</td>
<td>0.0543**</td>
</tr>
<tr>
<td>Decile 5</td>
<td>-0.1665</td>
<td>-0.0667**</td>
<td>0.0521**</td>
</tr>
<tr>
<td>Decile 6</td>
<td>-0.0933</td>
<td>-0.0257</td>
<td>0.0404**</td>
</tr>
<tr>
<td>Decile 7</td>
<td>-0.1202</td>
<td>-0.0236</td>
<td>0.0409**</td>
</tr>
<tr>
<td>Decile 8</td>
<td>-0.0524</td>
<td>-0.0707**</td>
<td>0.0300*</td>
</tr>
<tr>
<td>Decile 9</td>
<td>0.1452**</td>
<td>-0.0072</td>
<td>0.0286**</td>
</tr>
</tbody>
</table>

* Significant at the 90% level
** Significant at the 95% level
Table 5: Counterfactual Experiments: Revenue gains through reallocation of detailing

Temporal Allocation

<table>
<thead>
<tr>
<th>Cross – Sectional Allocation</th>
<th>No Change</th>
<th>Change</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Current Situation</td>
<td>Experiment 1</td>
</tr>
<tr>
<td></td>
<td>Cialis: 10.3%</td>
<td>Cialis: 10.3%</td>
</tr>
<tr>
<td></td>
<td>Levitra: 7.6%</td>
<td>Levitra: 7.6%</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>Experiment 2</td>
</tr>
<tr>
<td></td>
<td>Cialis: 6.0%</td>
<td>Cialis: 6.0%</td>
</tr>
<tr>
<td></td>
<td>Levitra: 4.6%</td>
<td>Levitra: 4.6%</td>
</tr>
<tr>
<td></td>
<td>Change</td>
<td>Experiment 3</td>
</tr>
<tr>
<td></td>
<td>Cialis: 16.3%</td>
<td>Cialis: 16.3%</td>
</tr>
<tr>
<td></td>
<td>Levitra: 12.1%</td>
<td>Levitra: 12.1%</td>
</tr>
</tbody>
</table>

Table 6: Revenue gains from reallocation of detailing: relative contributions of heterogeneity in learning, risk aversion and persuasive effect

<table>
<thead>
<tr>
<th>Drug</th>
<th>% Contribution of revenue gain by heterogeneity in</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Learning</td>
</tr>
<tr>
<td>Cialis</td>
<td>42.6%</td>
</tr>
<tr>
<td>Levitra</td>
<td>49.3%</td>
</tr>
</tbody>
</table>
Figure 1: Histograms of means of individual level parameters

1(a): detailing signal variance $\sigma^2_{D_i}$

1(b): feedback signal variance $\sigma^2_{F_i}$

1(c): coefficient of absolute risk aversion $r_i$
Figure 1: Histograms of Means of Individual Level Parameters

1(d): detailing stock coefficient $\beta_{1i}$

1(e): patient request coefficient $\beta_{2i}$

1(f): true mean quality for Cialis $Q_{i1}$

1(g): true mean quality for Levitra $Q_{i2}$
Figure 2: Histogram of the number of calls required to reduce uncertainty (variance of quality belief) about a new drug to one-tenth of the initial value

Figure 3: Plot of Informative Effect vs. Persuasive Effect
Figure 4: Plot of Informative Effect vs. Risk Aversion
Appendix: Full Conditional Distributions

Given the model described in the model section, the joint posterior distribution of all the parameters conditional on the data is given by is given by the following expression

\[ L \propto \prod_{i=1}^{r_i} \left[ f(U_i \mid \{I_{i \mu}\}, \{X_{i \mu}\}, \{Q_{i \mu}\}, r_i, \beta, \Sigma) \prod_{j=1}^{n_{ij}} \left[ f(Q_{ij} \mid Q_{ij-1}, n_{ij-1}, I_{ij-1}, \sigma^2, \sigma^2_i, Q_{i \mu}, Q_{i \mu}) \right] \right] \cdot f(\gamma \mid \Lambda, Z, V) \cdot f(\Lambda, V) f(V \mid g, G) \prod_{j=2}^{\frac{1}{2}} \left[ f(\sigma^2_j \mid s_{ij}, s_{ij}) \right] \]

We shall now derive the full conditional distributions of all the parameters (and augmented parameters like \( U_{ijt} \) and \( Q_{ijt} \)), so that we can use a Gibbs sampling method. For those parameters for which the full conditional distributions are not from known distribution families, we shall use the Metropolis Hastings algorithm to draw from the respective full conditional distributions.

Suppose we define:

\[
U_{\mu} = \begin{pmatrix} U_{1\mu} \\ \vdots \\ U_{r_{ij}\mu} \end{pmatrix}, \quad \Omega_{\mu} = \begin{pmatrix} -\exp\left(-r_{ij}Q_{ij\mu} + \frac{r^2\sigma^2_{ij\mu}}{2} \right) \\ \vdots \\ -\exp\left(-r_{ij}Q_{ij\mu} + \frac{r^2\sigma^2_{ij\mu}}{2} \right) \end{pmatrix}, \quad X_{\mu} = \begin{pmatrix} X_{1\mu} \\ \vdots \\ X_{r_{ij}\mu} \end{pmatrix}
\]

The full expression for the joint posterior distribution can then be written as
\[
L \propto \prod_{i=1}^{N} \left( \prod_{j=1}^{T} \frac{1}{\sqrt{2\pi\Sigma}} \exp \left( -\frac{1}{2} \left( \left( U_{ij} - \Omega_{j} - X_{ij} \beta_{j} \right)^{T} \Sigma^{-1} \left( U_{ij} - \Omega_{j} - X_{ij} \beta_{j} \right) \right) \prod_{j \neq k}^{T} \left( 1 \left( U_{ij} > \max \left( U_{ik} \right), k \neq j \right) \right) \right) \right)^{\frac{-1}{2}}
\]

\[
\left| V_{i} \right|^{\frac{-1}{2}} \exp \left( -\frac{1}{2} \left( \gamma_{i} - \Lambda Z \right)^{T} V_{i}^{-1} \left( \gamma_{i} - \Lambda Z \right) \right)
\]

\[
\left| V_{\gamma} \right|^{\frac{-1}{2}} \exp \left( -\frac{1}{2} \left( \lambda - \bar{\Lambda} \right)^{T} V_{\gamma}^{-1} \left( \lambda - \bar{\Lambda} \right) \right)
\]

\[
\prod_{j=1}^{T} \left( \exp \left( -\frac{1}{2} \left( G^{(j)} V_{\gamma} \right) \right) \right)
\]

From this joint posterior, we can derive the full conditional distributions as follows:

1. \( \sigma_{j} \mid \left\{ U \right\}_{i}, \left\{ X \right\}_{i}, \left\{ \Omega \right\}, \left\{ \beta \right\}, \left\{ r \right\} \sim IG \left( \frac{\Sigma T}{2} + s_{j}, \frac{2 s_{j}}{2 + s_{j} \sum_{i=1}^{T} \left( U_{ij} + \exp \left( -r \frac{\sigma_{j}}{2} \right) - X_{ij} \beta_{j} \right)} \right) \)

2. \( V_{\gamma} \mid \gamma_{i}, \Lambda, Z, g, G \sim \text{Inverse Wishart} \left( g + N, \left[ \sum_{j=1}^{N} \left( \gamma_{j} - \Lambda Z \right) \left( \gamma_{j} - \Lambda Z \right)^{T} \right] + G \right) \)

3. \( \lambda \mid \left\{ \gamma \right\}_{i}, V_{\gamma}, Z, V_{\Lambda} \sim N \left( \left[ V_{\gamma}^{-1} \otimes Z'Z + V_{\Lambda}^{-1} \right] \left[ V_{\gamma}^{-1} \otimes Z'Z + V_{\Lambda}^{-1} \right] \left[ V_{\gamma}^{-1} \otimes Z'Z + V_{\Lambda}^{-1} \right] \right) \)

where

\[
\hat{\lambda} = \text{vec} \left[ (Z'Z)^{-1} Z' \Gamma \right], \quad Z = \begin{pmatrix} Z'_{1} \\ \vdots \\ Z'_{N} \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \gamma_{1}' \\ \vdots \\ \gamma_{N}' \end{pmatrix}
\]

4. Each physician is independent conditional on the \( X \) and \( Z \) matrices. And each observation for the physician is independent conditional on the vector of \( Q_{ijt} \). Thus, we can draw the latent utilities for a particular physician and a particular observation separately from the other
observations. This involves sequentially drawing from a truncated multivariate normal distribution for each time period.

\[
\begin{bmatrix}
U_{ijt} \\
\vdots \\
U_{i_kt}
\end{bmatrix}
\sim \text{Truncated MVN}
\begin{pmatrix}
-\exp\left(-r_i^2 \sigma_{\gamma_i}^2 + \frac{r_i^2 \sigma_{\gamma_i}^2}{2} + X_i \gamma_i \right) \\
\vdots \\
-\exp\left(-r_i^2 \sigma_{\gamma_i}^2 + \frac{r_i^2 \sigma_{\gamma_i}^2}{2} + X_i \gamma_i \right)
\end{pmatrix}, \Sigma
\]

with the truncation such that \( U_{ijt} > U_{ikt}, \forall k \neq j, I_{ijt} = 1 \)

5. The full conditional distributions for the individual level parameters, \( \gamma_i \), and the quality means \( Q_{it} \) are not from known families of distributions. Hence, draws from the distribution of these parameters for each individual physician are obtained using the Metropolis Hastings algorithm. We use a Random Walk Metropolis Hastings algorithm (Chib and Greenberg, 1995) with a normal candidate density to make these draws. The variances of these densities were obtained from the hessian of the pooled maximum likelihood estimates for these parameters, which was scaled up to obtain the best numerical efficiency.