ABSTRACT

Title of dissertation: ESSAYS ON PHARMACEUTICAL ADVERTISING

Weijia Dai
Doctor of Philosophy, 2015

Dissertation directed by: Professor Ginger Z. Jin
Professor Andrew Sweeting
Department of Economics

The dissertation focuses on two distinctive issues in pharmaceutical advertising. One on the matching choices between advertisers and advertising agencies, and the other on the effect of paid-link advertising on consumer search for online pharmacies. The goal of this dissertation is to empirically uncover the underlying economic mechanisms. Moreover, the analysis of matching problem provides new insights on the formation of vertical relationships between clients and professional service agencies and has implications for professional service market consolidations. And the examination of consumer searches for pharmaceuticals online sheds lights on consumers' concerns over quality and affordability of prescription drugs and draws attention on advertising regulation.

In the first two chapters, I focus on two essential features of the market for professional services. One is the necessary mutual agreement in forming relationships, and the other is that a client perceives conflict when hiring the same service agency as his product market competitor. To incorporate these two features, I construct and estimate a two-sided matching model and allow agents' choices to depend on conflict. The results show that conflict does indeed reduce match surplus, and the reduction is greater for a

pair of agents who have matched with each other in the previous period. Also, preserving previously formed matches yields much higher surplus than forming new matches. Based on these estimates, I conduct a counterfactual exercise to illustrate the effect of conflict on allocation of matches and another counterfactual exercise to illustrate the effect of a merger between advertising agencies on market equilibrium.

In the third chapter, coauthored with Matthew Chesnes and Ginger Jin, we examine how government's sudden ban of foreign online pharmacies from paid search on Google and other search engines changes consumer searches for the banned websites. Using click-through data from comScore, we find that non-NABP-certified pharmacies receive fewer clicks after the ban, and this effect is heterogenous. In particular, pharmacies not certified by the NABP but certified by other sources, referred to as tier-B sites, experience a reduction in total clicks, and some of their lost paid clicks are replaced by organic clicks. These results have implications for the change in consumer search cost and health concern.

ESSAYS ON PHARMACEUTICAL ADVERTISING

by

Weijia Dai

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Ginger G. Jin, Chair
Professor Andrew Sweeting, Co-Chair
Professor Roger R. Betancourt
Professor Peter Cramton
Professor David Godes

Dedication

To my parents, L. Dai and S. Xiang.

Acknowledgments

First and foremost, I wish to express my deep gratitude to my advisor Professor Ginger Jin. My dissertation would not have been possible without her most generous support and advice. I greatly appreciate all her contributions of time and ideas to make my Ph.D. experience productive and stimulating. I would also like to thank my co-advisor Professor Andrew Sweeting for his immeasurable help on my dissertation. I am very grateful of his willingness to helping me solve my dissertation related questions at all times, and I very much appreciated that he guided me to overcome the major difficulties in the first two chapters of my dissertation with his remarkable intellectual insights.

For this dissertation, thanks are due to my dissertation committee members, Professor David Godes, Professor Roger Betancourt, and Professor Peter Cramton, for their invaluable time, interest, and guidance.

For the third chapter of the dissertation, I am especially grateful for my coauthor Dr. Matthew Chesnes for the wonderful experience of working together with him.

My colleagues at the department of economics have made my graduate study and work a memorable experience and have provided me with helpful comments on my dissertation. I have benefited from the discussion with Larry Ausubel, Roger Betancourt, Ron Chan, Peter Cramton, Jingting Fan, Haishan Yuan, Soohyung (Soo) Lee, Ethan Kaplan, Lixin Tang, Lesley Turner, Shulin Wee, Bobby Zhou, Ben Zou, and seminar participants at the and department of economics and Robert Smith School of Business.

This dissertation has also been benefited greatly from the constructive comments and suggestions made by Panle Jia Barwick, Jason Chan, Jeremy Fox, Qiang Fu, Matthew Gentzkow, Avi Goldfarb, Andrew Hagiu, Bart Hamilton, Ben Handel, John Hatfield, Daniel Hosken, Susumu Imai, Karam Kang, Bying-Cheol Kim, Michael Luca, Shiko Maruyama, Fiona Scott Morton, Minjae Song, William Vogt, Weng Xi and attendants

at the 2013 WHITE Conference, the 2013 and 2014 Southern Economics Association Annual Conference and the 2014 American Economic Association Conference.

For the data and institutional knowledge, I would like to thank George Chressanthis for suggesting the data sources, and Robert Girondi from the Fox School of Business, Drew Guiteras from McKinney, Prasun Subedi from Pfizer for providing institutional detail on advertising agencies and pharmaceutical marketing.

I gratefully acknowledge the funding sources that made my dissertation work possible. I would like to acknowledge financial support from the department of Economics for purchasing the research data, and the Betancourt Scholarship that supports my dissertation writing.

Contents

List of Tables

List of Figures

Introduction

The dissertation studies two factors in the advertising production process, choice of advertising agencies and consumer response to paid link advertisements on search engines. The former concerns whether an advertiser can match with a desirable advertising agency, and the latter determines if consumers are responsive to the form of advertisement they are exposed to.

Advertising is often outsourced to one or several advertising agencies that provide both creative and media buying services. While outsourcing advertising is a common practice, product market competitors often want to avoid the conflict of using the same advertising agency. At the same time, since it is costly for advertising agencies to ensure confidentiality when contracting with product market competitors, they may also avoid serving direct competitors simultaneously. Given the limited number of advertising agencies that specialize in a particular type of product category, conflict affects the allocation of matches in the advertising market.

The goal of the first two chapters is to model the choice problem, to use a revealed preference approach to back out the effect of conflict on agents' preferences, and to examine how the existence of conflict impacts the overall allocation of contracts. Moreover, since conflict is determined by the relationship structure in the market, in Chapter 2, I also examine how a change in market structure, such as a merger between advertising agencies, changes the market equilibrium.

The nature of the choice problem described above suggests that we need to account for the interdependence of decisions made by both sides of the market, and it is important to take into account all market participants. This differs from the previous literature that

ignores the strategy interdependence and the two-sidedness of the decision process (Rogan (2013); Asker & Ljungqvist (2010); Aobdia (2012)).

To solve this problem, Chapter 1 lays out a two-sided matching the model that analyzes the relationship choices and allow for negative externality generated by conflict. Chapter 2 estimates the latent matching surplus function using observed matches between pharmaceutical advertisers and advertising agencies. After backing out the preference parameter, in Chapter 2, I am able to simulate the market equilibrium effect of conflict and the effect of a merger between the advertising agencies.

Drug advertising is particularly suited to a study of agency-client relationships since there is a clear definition of product market competition by therapeutic class. Also, prescription drug advertisers are important clients for the advertising agencies, contributing roughly 10% of total agency revenue.

I find that conflict is costly to matching relationships, and crowds out low return matches in the market. The merger of advertising agencies further reduces choice alternatives of agencies and in general reduces total surplus in the market. Although this study does not explicitly examine commissions paid to advertising agencies, the fact that merger significantly reduces the substitution of agencies due to conflict has important implications for antitrust policy. This study is a first attempt to directly model the effect of conflict on agent preferences and the mechanism of conflict in affecting matching allocation in the market. While I focus on a particular set of agents, namely advertising agencies and their pharmaceutical clients, there are many more markets where vertical relationships can be characterized by two-sided decisions and are suitable to be analyzed in a matching framework.

Chapter 3 studies consumer response to paid link advertisements on search engines.

Search engine is one major gateway for consumers to find information, hence advertisers bid on paid links of relevant queries (keywords) in the hope that their links will be seen and clicked by consumers. This form of advertising often attracts online retailers, including foreign pharmacies that place paid links and directly sell pharmaceuticals to US consumers online. The high drug price in the US makes buying online from foreign pharmacies an attractive option for US consumers and makes it appealing for foreign pharmacies to reach US consumers through online paid link advertisements.

The concern has since grown because poor online prescription drugs quality may lead to adverse health outcomes. The Food and Drug Administration (FDA) prohibits the importation of unapproved drugs into the US, and the National Association of Boards of Pharmacy (NABP) emphasizes their illegality and cites examples of unsafe drugs from rogue pharmacies. Because of heightened concern to protect consumers, Google agreed to ban non-NABP-certified pharmacies from their sponsored search listings in February 2010 and settled with the Department of Justice (DOJ) in August 2011. This creates an exogenous change to the paid-link advertisements consumers are able to see through online searches, and we study how the ban on non-NABP-certified pharmacies from sponsored search listings affects consumer search on the Internet.

Using click-through data from comScore, we find that non-NABP-certified pharmacies receive fewer clicks after the ban, and this effect is heterogeneous. In particular, pharmacies not certified by the NABP but certified by other sources – referred to as tier-B sites – experience a reduction in total clicks, and some of their lost paid clicks are replaced by organic clicks. These effects do not change significantly after the DOJ settlement. In contrast, pharmacies not certified by any of the four major certification agencies – referred to as tier-C sites – suffer a greater reduction in both paid and organic clicks, and

the reduction was exacerbated after the DOJ settlement. These results suggest that the ban has increased search cost for tier-B sites, but at least some consumers overcome the increased search cost by switching from paid to organic links. In addition to search cost, the ban may have increased health concerns for tier-C sites and discouraged consumers from reaching them via both paid and organic links.

The remainder of this dissertation is divided as follows. Chapter 1 presents the matching model and its challenges to define stable equilibrium when externality exists. In doing so, Chapter 1 articulates the assumptions made on the model and their implications. Chapter 2 first estimates the model using data on matches formed by pharmaceutical advertisers and advertising agencies, and then uses the estimated parameters to simulate matching allocations when market structure changes. Lastly, Chapter 3 studies how consumer find information on the internet by examining consumer search and click responses after a ban of paid link advertisements from foreign pharmacies on US search engines.

Chapter 1 – Matching with Conflicts: An Application to the Advertising Industry

1.1 Introduction

In many professional service markets, such as auditing, advertising, banking, and consulting, clients may prefer that the service agency they work with does not work for their product market competitors. This is mainly due to concerns about conflicts of interest. For example, strategically important information may be leaked to competitors, or the act of maintaining relationships with conflicting clients may lower the service agency's performance. In the advertising industry, an oft-cited example is that of Colgate-Palmolive, who terminated its long-standing relationship with Saatchi and Saatchi when the advertising agency started to work for Procter & Gamble following a merger.[1] To address potential conflict, the industry association American Association of Advertising Agencies (4A's) has advised advertising agencies to be vigilant in protecting clients' confidential information either by avoiding simultaneously serving clients who compete head-to-head or by setting up information firewalls (so called "Chinese walls") to actively prevent information leakage within the agency.[2]

Conflict lowers expected surplus of relationships between service agencies and clients and affects both sides' choices of who to contract with. The goal of the first two chapters

[1] Advertising agencies Saatchi & Saatchi and Ted Bates merged in 1987. Before their merger, Colgate-Palmolive hired Saatchi & Saatchi, and Procter & Gamble hired Ted Bates. The merger between the two agencies brought Colgate- Palmolive and Procter & Gamble under the same roof and led to the immediate departure of Colgate-Palmolive (Siman 1989; Goldman 1997). In service industries, a merger of top service agencies with broad client bases almost inevitably leads to competition overlap among their clients. For example, the merger between Arthur Young and Ernst & Whinney in 1989 brought their clients PepsiCo and Coca-Cola together, resulting in Arthur Young's resignation from PepsiCo's contract. It is an aphorism within the service industry that secrets are harder to keep in advertising agencies (The Economist 1997).

[2] See American Association of Advertising Agencies (2009), "Confidentiality Of Marketing Services Information". Retrieved from http://goo.gl/hHGQ9i

is to model the contract choice problem, to estimate the effect of conflict on agents' preferences, and to quantify how the existence of conflict impacts the overall allocation of contracts. Moreover, since conflict is determined by relationship structure in the market, I examine how a change to the market structure, such as a merger between service agencies, changes the market equilibrium.

Compared to the classical matching literature, my setting is non-standard in two ways. First, conflict introduces an externality of one player's matching choice on other players' preferences.[3] Second, a many-to-many matching model is necessary as both drug advertisers and advertising agencies commonly match with multiple partners. In order to obtain a stable equilibrium solution when advertising agencies' preferences may not be substitutable, I made an assumption that drug advertisers do not face a binding capacity constraint in terms of how many matches they can form in a specific year. This assumption implies that a drug advertiser's decision to match with each agency is independent and leads to stable equilibrium. However, I am able to maintain the feature that the number of agencies a drug advertiser eventually hires is still limited since the match value of a drug-agency pair must overcome a fixed cost of signing up a contract and the cost of conflict if conflict exists. There is no explicit capacity constraint for advertising agencies in my model either, but since conflict increases the cost to match with another drug in the same therapeutic class, this negative externality is similar to a capacity constraint specific to the class.[4] The model yields the result that the equilibrium can be characterized by

[3]The negative externality in this model directly affects agents' preferences. This is different from the externality in the typical matching models in which capacity constraint places a negative externality on agents' choice set.

[4]More strictly speaking, since I assume that conflict cost does not depend on the number of drugs in a conflicting relationship, negative externality in my model arises when the number of drugs an agency serves in a therapeutic class rises from one to two. But when an agency already has two conflicting clients, adding the third client in the therapeutic class only creates the conflict cost to the extra match itself, and does not impose a negative externality to existing matches.

pairwise deviations that are independent of transfers made between contracting parties. This allows me to identify the surplus function using only information on matches.

A two-sided model with conflict helps us better understand the role of conflict in contractual relationships. Rogan (2013), Asker & Ljungqvist (2010), Aobdia (2012) model and identify conflict from clients' switching behavior when their service agencies are involved in exogenous mergers. Such an identification strategy ignores the fact that all clients' contracting decisions, whether or not they are currently in a contract, are affected by the market structure change. I am able to incorporate the market-wide interdependency through conflicts. In particular, I simulate market equilibrium when conflict cost is reduced to zero and when mergers reduce the number of service agencies in the market. My findings are consistent with Asker & Ljungqvist (2010) in that conflict is costly for matches, and mergers reduce the number of matches formed in the market. My work also complements Asker & Ljungqvist (2010) by constructing a counterfactual that shows the magnitude of equilibrium effect of a change to the market structure on all agents in the market, no matter whether they have a contract with the merging agencies or not at the time of merger.

My work has important implications for competitiveness of professional service markets, especially when the number of viable service agencies is limited for each type of client due to client-required experience and expertise. A typical antitrust analysis concerns whether a merger may generate price inefficiency because consumers have fewer suppliers to choose from. Later in Chapter 2, I show that a merger between professional service providers can exacerbate conflict of interest among clients and further limit a client's choice set. Whether this warrants additional antitrust concern depends on the pre-merger market structure and additional knowledge of the mechanism by which matches and payments

are negotiated.

The rest of the chapter is organized as follows: In the next section, I describe background of the advertising agency industry and explain the dataset I construct for this study. In section 3, I explain the matching model and characterize the model equilibrium. In sections 4, I discuss my model elements and compares them to the classic two-sided matching models.

1.2 The Advertising Agency Business and the Concern of Conflict

Advertising agencies provide a number of professional services to clients, including planning and creating advertising content, delivering finished creative content to consumers through media outlets, advising clients on overall marketing and branding strategies. In fact, with a total US revenue of $35.6 billion in 2012,[5] advertising agencies are in charge of almost all of the advertising activities of the majority of large advertisers, an arrangement that has existed since the 1920s (Mohammad et al., 2012). Among the 1,746 firms who spend more than $1 million annually on advertising and responded to the 2003 *Advertising Age* survey, 86% outsource all of their advertising services to advertising agencies (Horsky, 2006). This figure that is lower for advertisers with smaller advertising budgets. And in the advertising industry, larger advertising agencies dominate: in 2003, advertisers spent $245 billion on advertising in the US, and 61% of the spending are billed through the largest 1000 advertising agencies (*Advertising Age* 2004).[6]

Client concerns about conflict have long been recognized in the advertising industry.

[5] See Advertising Age (2013). Retrieved from http://adage.com/article/agency-news/digital-media-drive-u-s-agency-revenue/241114/

[6] Billing is the total amount of marketing dollars given to the advertising agencies to spend for the advertisers.

One widely documented case was the merger between Saatchi & Saatchi and Ted Bates in the late 1980s. In this merger, the two agencies lost accounts worth more than $300 million, and at least one of the clients in a conflict relationship with his product market competitors dismissed the agency (Siman, 1989). The 4A's has also issued conflict policy guidelines that cautions "head-to-head conflict at the brand or category level".[7]

To date, conflict continues to attract attention in agency merger announcements, since mergers may directly bring conflicting clients into the same agency. The importance of client conflicts came to light again when Publicis and Omnicom, the second and third largest advertising holding companies ranked by 2012 revenue, announced plans to merge in July 2013. Given the size of these advertising agency holding companies, the potential competition overlap among clients was big. The proposed Publicis-Omnicom merger brought together high-profile competing advertisers, such as PepsiCo and Coca-Cola, Volkswagen and Mercedes, General Motors and Toyota, and AT&T and Verizon.[8] The merger was ultimately cancelled due to a variety of reasons, but the CEO of Publicis Maurice Levy mentioned part of the reason be the merger diverts agencies' attention to "best serve our clients". Inspired by this merger event, I will analyze the impact of a hypothetical Publicis-Omnicom merger on allocation of matches for pharmaceutical advertisers in the counterfactual exercise.

The costs of conflict arise from three main sources: clients concern about information leakage, the cost to the agency of maintaining client relationships when conflict occurs (Asker & Ljungqvist (2010), Aobdia (2012)), and the disinclination of conflicting clients or forgo access to an agency's limited product-specific specialty resources (e.g., experts,

[7]See American Association of Advertising Agencies (2009), "Conflict Policy Guidelines". Retrieved from http://www.aaaa.org/agency/compensation/positionpapers/pages/070700_conflict.aspx

[8]See Adweek (2013), "Publicis-Omnicom Merger Set". Retrieved from http://www.adweek.com/news/advertising-branding/how-likely-actual-omnicom-publicis-merger-151468

marketing opportunities).

Advertising agencies access clients' confidential business information and marketing plans and act on their behalf to execute their advertising strategy. Advertising agencies' failure to protect clients' confidential marketing plans may hinder advertising effectiveness, so clients have the expectation that agencies will carefully guard their confidential information. The promise to prevent client information from leaking becomes particularly fragile when a client shares the same advertising agency as his product market competitors. One reason is that the information barrier within an organization is much lower than across organizations. The other reason is strategic: when there is a large prospective gain for an advertising agency that uses one client's marketing strategy to benefit his competitor, the agency may have an incentive problem in committing to protect confidential client information.

The second conflict concern is the negative impact on an advertising agency's performance when it is trying to maintain relationships with conflicting clients. One direct cost is the "Chinese wall" advertising agencies build to serve conflicting clients. This includes creating separate office spaces, separate staffs, and separate communication channels to reduce communication between teams of employees who work on competing accounts. Another indirect cost is that the agency may sacrifice efficiency for fairness where the agency balance resources devoted to competing clients.[9] Sometimes, maintaining such conflicting relationships is so costly that the agency is willing to let the less valuable client go.[10]

[9]Anecdotal evidence can be seen in discussions of industry magazines. Marc Brownstein, president of the Brownstein Group, whose clients includes IKEA, eBay, and Comcast mentioned in the article he writes for the *Advertising Age* magazine that serving competing clients at the same time is "utopian". He argues that maintaining relationships with competing clients and keeping every party happy is very hard. See Advertising Age (2012), "Can We, Should We, Serve Competing Clients at the Same Time". Retrieved from http://adage.com/article/small-agency-diary/serve-competing-clients-time/235777/

[10]In 2012, Omnicom resigned its contract with Farmers Insurance just weeks after winning the account when its other client State Farm, "a much bigger client by marketing spending", learned about Omnicom's new relationship with Farmers. See Advertising Age (2012), "Farmers Insurance, Omnicom's PHD Part Ways Due to Client Conflict". Retrieved from http://adage.com/article/agency-news/farmers-insurance-

Conflict can also occur when clients compete for advertising agency's market-specific resources. For example, if an agency has only one expert in the client's field and can assign that expert to only one client due to the "Chinese wall," clients in conflict matched with the agency should expect a lower chance that they will benefit from access to that expert.

1.3 A Two-Sided Matching Model

To allow both drug advertisers and advertising agencies to choose whether to sign a contract, I analyze their choice problem using a two-sided matching model. Because advertising agency commissions are negotiated in the contracting process, I allow side payments to be endogenously determined in the matching game. In this section, I outline the model and necessary assumptions, and I then establish that the socially optimal allocation belongs to the set of pairwise stable equilibrium. Using pairwise stable equilibrium, I will derive conditions that characterize the equilibrium and use them to identify and estimate the model.

1.3.1 Model Setup and Assumptions

I define a matching market as a treatment class in a given year in which drugs and advertising agencies seek matches. I assume that each side of the market fully observes the characteristics and preference shocks of all drugs and agencies. Each drug and agency maximizes its own surplus for the current period. I also assume that agency decisions are independent across treatment classes. Because these two assumptions imply independence

omnicom-s-phd-part-ways-due-conflict/237158/

of agent decisions across markets, I will focus on a single market in the following analysis.

There are m drugs in the market, indexed by $i = 1, .., m \in I$ and n advertising agencies, indexed by $j = 1, .., n \in J$. Each drug can hire multiple agencies and each agency can serve multiple drugs. A *matching* μ is a mapping from the set $I \cup J$ into the set of all subsets of $I \cup J$, where $\mu(i) \in 2^J$ and $\mu(j) \in 2^I$, and $i \in \mu(j)$ if and only if $j \in \mu(i)$. Denote $\mathcal{M}$ as the set of all possible matchings.

Denote the transfer that a drug i pays to an agency j as t_{ij}, I can write the drug and agency surpluses as

$$S^i(i, j|\mu) = y(X_i, Z_j) - \kappa_f - t_{ij} + \epsilon^i_{ij} \quad (1)$$

$$S^j(i, j|\mu) = -\mathbf{1}(|\mu_j| > 1)\kappa_c + t_{ij} + \epsilon^j_{ij} \quad (2)$$

where $y(X_i, Z_j)$ is the expected return that drug i yields from advertising agency j's service based on observed drug characteristics X_i and agency characteristics Z_j. κ_f is the fixed cost associated with establishing a match, which I assume to be common across all matches. For simplicity, I will refer to $y(X_i, Z_j)$ as the "return of a match", and to $y(X_i, Z_j) - \kappa_f$ as the "net return of a match". κ_c is the cost incurred when the pair (i, j) is involved in a client conflict. I assume κ_c is constant for any matched pair involved in a conflict. Conflict arises when an agency serves two or more drugs in the same market, that is, if $|\mu_j| > 1$. In the surplus function, I express that the conflict cost accrues to the agencies only for illustration; in the equilibrium I define, matching will depend only on the total net surplus, not on which side pays the conflict cost. ϵ^i_{ij} and ϵ^j_{ij} are match specific preference shocks observable to all agents in the market, and shocks to drugs and to agencies are each distributed $i.i.d$ across all possible match pairs.

I make four modeling assumptions as follows:

Assumption 1 Each drug and agency's total surplus is a simple summation of surplus from all matches the drug and agency obtain, i.e.

$$S_i = \sum_{j \in \mu(i)} S^i(i, j|\mu), \; S_j = \sum_{i \in \mu(j)} S^j(i, j|\mu). \tag{3}$$

More specifically, the model assumes that, for each drug, there are no complementarities among the set of agencies it hires. In particular, drug preference is responsive and satisfies substitutability given matching allocation of its competitors. However, preference of advertising agencies is not substitutable,[11] which may challenge the existence of a stable equilibrium in a usual matching setting.

Assumption 2 When agents make matching decisions, externalities arise only through client conflict.

It is common to expect that drug advertisers' surpluses may be influenced by externalities arising from product market competition. For example, one drug's advertising effort may affect his competitors' product revenue. This model does not preclude this type of externality, but it assumes that this type of externality does not interact with the characteristics and identity of the advertising agency, and hence does not affect matching choices. More specifically, think of the overall drug-agency matching and advertising production process as a two-stage game. In the first stage, drugs and agencies take preference draws, form expectations about the returns of possible matching opportunities, and decide

[11] Consider the following example in which agent preference fails substitutability. For illustration, let advertising agencies have full bargaining power and let drugs obtain reservation value. Assume that an advertising agency obtains net return of 10, 4, 4 with drug D_1 , D_2 and D_3, and that the conflict cost is 2.5 for each match. If the agency is given the set $\{D_1, D_2\}$, it will choose to match with D_1 alone, which gives a surplus of 10, rather than matching with both drugs, which gives a total surplus of 9. When given the set $\{D_1, D_2, D_3\}$, the agency will choose the full set of drugs, which yields a surplus of 10.5. Hence the $\{D_1, D_2\}$ is chosen only when D_3 is chosen.

what contracts to propose and to accept to maximize their expected surplus. In the second stage, the matches are formed and each drug's advertising strategy is determined based on his competitor's strategies. However, I assume that strategic interactions between drugs in the second stage are not affected by the first-stage hiring decisions. This assumption is made based on the literature that argues that relational characteristics, such as identity and impression, are important in determining relationships (Morgan & Hunt (1994), Alvesson (1994)). This is not surprising, given that the effectiveness of an advertising agency is hard to predict. So choices made at the matching stage can be perceived as drugs and agencies seeking matches that minimize the cost of doing business.

Assumption 3 The capacity constraints for both drugs and advertising agencies do not bind.

Given that the number of matching opportunities that yield positive returns is limited in a market, this assumption suggests that drugs and advertising agencies are willing to add a match as long as the additional match increases the total surplus of the agent. In fact, I can relax this assumption to allow the market-specific capacity constraint for advertising agencies to bind since conflict cost is a form of constraint on an advertising agency's ability to take on more drugs in given market. On the drug side, a capacity constraint can appear as an increasing cost for the drug to match with additional advertising agencies. I estimate a version of the model that allows a matching cost to increase as a drug takes on additional agencies. The result shows that, for a given drug, each additional match slightly reduces, rather than increases, the matching cost, but the magnitudes are very small, suggesting that capacity constraint for drugs is unlikely to bind.

Assumption 4 The opportunity cost of forming a match is R for drugs, and zero

for advertising agencies.

Note that the opportunity cost of forming a match in this situation compares a drug and an agency's decision to take or leave each matching opportunity. This is different from the "acceptance" notion that compares the agent's decision to participate in the matching market with the decision to stay out.

1.3.2 Equilibrium Characterization

In this section, I define the socially optimal allocation, and show that it is pairwise stable given my model setting.

For notation compactness, I follow contract notations in Hatfield and Milgrom (2005). Denote a bilateral contract between an drug-agency pair (i, j) as $a_{ij} = ((i, j), t_{ij})$. a_{ij} contains two elements. The first element describes the match between a pair of agents from each disjoint set that signs the contract, $a_I = i$ and $a_J = j$, where it corresponds to the prior notation of $i \in \mu(j)$ and $j \in \mu(i)$. The second element describes the transfer t_{ij} paid from drug i to agency j. Only one contract can be signed between i and j. An market allocation $\mathcal{A}$ is a collection of contracts that specify matching allocation over all agents and a set of transfers, $\mathcal{A} \equiv I \times J \times T$. Given a set of contracts $A \subset \mathcal{A}$, I denote $Ch_i(A)$ as the most preferred subset of contracts chosen by drug i among contracts that names i as a partner ($a_I = i$), and I denote $Ch_j(A)$ as the most preferred subset of contracts chosen by agency j among contracts with $a_J = j$.

Definition 1 A *socially optimal allocation* is a matching $\mu \in \mathcal{M}$ that maximizes the aggregate social surplus of all matches:

$$S(\mu) = \sum_{i} \sum_{j' \in \mu(i)} S^i(i, j') + \sum_{j} \sum_{i' \in \mu(j)} S^j(i', j).$$

This concept only considers the total surplus generated from each match, ignoring any transfers designated in the contact. Note also that, in the market I define, there is no capacity constraint on how many matches an agent can obtain. However, the socially optimal allocation does not imply that every possible pair of matches will be formed. Adding an additional match is costly for two reasons. The first cost is the cost of forming a match (κ_f), common to all matches. The second cost is the cost of conflict if an additional match increases the amount of conflict in the market. So at the socially optimal allocation, any additional match will only reduce the total surplus.

Definition 2 A set of contracts $A \subset \mathcal{A}$ constitutes *a pairwise stable equilibrium* if

(i) $Ch_I(A) = Ch_J(A) = A$, and

(ii) there does not exist a drug i and an agency j' and a new contract $a' = ((i, j'), t_{ij'})$, $a' \notin Ch_i(A)$, such that$a' \in Ch_i(A \cup a') \subset Ch_J(A \cup a')$, and

(iii) and there does not exist an advertising agency j and a set of contracts $a'' = ((i', j), t_{i'j})$, $a'' \neq Ch_j(A)$ such that

$$a'' \in Ch_j(A \cup a'') \subset Ch_I(A \cup a'').$$

Condition (i) for the pairwise stable equilibrium says that the equilibrium set of contract A is the best subset of contracts among those assigned to them in A, or that

each agent is unwilling to drop the contract(s) it has at hand according to allocation A. Conditions (ii) and (iii) stipulate that, given allocation A, an agent cannot construct a new contract with an opponent it is not contracting with that makes the agent strictly better off and all other agents weakly better off.

Claim Given the above-defined surplus function S and all possible matchings $\mathcal{M}$, if a matching μ^* is socially optimal, there exists a set of contracts $A \subset \mathcal{A}$ that is pairwise stable and implements μ^*.

Proof Assume that advertising agencies have full bargaining power. The proof will proceed as follows: Given a matching μ, I will show that the set of transfers is endogenously determined by the bargaining rule, so each allocation A is reduced to a matching μ and the agent surplus can be written as a function of matching μ. I will then prove by contradiction that the social stable allocation μ^* is pairwise stable.

Since drugs are not capacity constrained and they evaluate contracts with different advertising agencies independently, when agencies have full bargaining power, each drug will accept a contract as far as its surplus is higher than that drug's reservation value for a contract, R. So for a given matching μ, $S^i = R$ and $S^j = \sum_{i \in \mu(j)} S(i,j) - |\mu_j| R$.

Let μ^* be a matching that is not pairwise stable. If μ^* violates assumption (i), or that

$$\sum_{i \in \mu(j)} S(i,j) < \sum_{i \in \mu(j) \setminus \{i'\}} S(i,j),$$

dropping i' from $\mu(j)$ improves surplus for j and does not affect other agents' payoffs, so μ^* is not socially optimal.

If μ^* violates assumption (ii), then there exists a pair (i', j), for which $i' \notin \mu(j)$ and

$$S(i', j) = R,$$

$$\sum_{i \in \mu(j) \cup i'} S(i, j) - (|\mu_j| + 1)R > \sum_{i \in \mu(j)} S(i, j) - |\mu_j|R.$$

By adding (i', j) to the current allocation μ^*, the social total surplus strictly increases, and therefore, μ^* is not socially optimal. Hence by contradiction, there exists a transfer schedule such that the socially optimal allocation is pairwise stable. □

One of the key assumptions that leads me to this simple proof is that drug advertisers have no binding capacity constraint and make independent decisions on forming matches with each agency. This assumption is tested and justified in the empirical section in the second chapter.

1.4 Model Discussions

To determine key characteristics of drug and advertising agency matching market, I allow features of the matching model to differ from canonical models of matching with transfers (Shapley & Shubik, 1971) and matching with contracts (Kelso Jr & Crawford (1982); Hatfield & Milgrom (2005)). In order to maintain model stability and obtain identification from the data I observe, I make the simplification assumptions discussed in the last section. Next, I will discuss further some of the model assumptions and their implications.

Firstly, in classical matching models, substitutability is the key concept that allows

characterization of the one-to-one and many-to-one stable matchings by the fix-point set, and hence by a lattice (Adachi (2000); Hatfield & Milgrom (2005)), and it is also extended into the many-to-many settings (Echenique & Oviedo, 2004). In my model, substitutability fails because of the simple assumption that conflict cost does not change with the number of matches in the conflicting relationship. One could argue for defining the conflict cost such that it increases at some minimum rate when the number of conflicting drugs matching with an advertising agency increases. However, in an unreported version of the estimation, I allow the conflict cost for each drug in the conflicting relationship to rise when the number of conflicting drugs increases. The coefficient on the change of conflict cost is not significant, which rejects the above argument. Examples in the literature show that the situations in which stability fails in the absence of substitutability occur when agents on both sides of the market are capacity-constrained.[12] Lifting capacity constraint solves the problem since it makes an agent's decisions to match with each opponent independent. This does not get rid of all interesting externalities in the market, because advertising agency's decisions of which drugs to take are not independent because of the conflict.

Since transfers are not observed in the model, the concept of a socially optimal allocation allows me to identify preferences by focusing only on the matching allocation. Assuming away capacity constraints also plays a role in identifying the cost of conflict without needing transfer information. This will become clear when I discuss the model identification.

[12]The singular preference is the key where stability fails in the example given by Hatfield & Milgrom (2005). Assume that there are two hospitals h_1, h_2 and two drugs d_1, d_2. Preference of h_1 fails substitutability that $\{d_1, d_2\} \succ_{h_1} \{d_1\} \succ_{h_1} \emptyset \succ_{h_1} d_2$, assume also that h_2 has singular preference and $\{d_2\} \succ_{h_2} \{d_1\} \succ_{h_2} \emptyset$. Both doctors have singular preference: $\{h_2\} \succ_{d_1} \{h_1\} \succ_{d_1} \emptyset$ and $\{h_1\} \succ_{d_2} \{h_2\} \succ_{d_2} \emptyset$. Assume a matching A_1 in which both d_1 and d_2 are matched with h_1, and h_2 is unmatched. The pair (d_1, h_2) can block A_1, and forms the new matching A_2 in which d_1 is matched with h_2 and both d_2 and h_1 are unmatched. The pair (d_2, h_2) blocks A_2 and form a matching A_3 in which d_2 and h_2 are matched and d_1 and h_1 are matched. The pair (d_2, h_1) blocks A_3 and the matching is back to A_1. Hence stable matching does not exist.

Secondly, note that the above proof shows that socially optimal matching can be characterized by pairwise stable equilibrium. However, socially optimal matching is not the only pairwise stable equilibrium. For example, in the non-substitutability example I gave in according to footnote 11, advertising agency j's preference is $D_1 = Ch_j(D_1, D_2)$, $D_1 = Ch_j(D_1, D_3)$ and $(D_1, D_2, D_{3)} = Ch_j(D_1, D_2, D_3)$. Agency j's matching with D_1 is pairwise stable since neither pairwise deviation between (A_j, D_2) or (A_j, D_3) can block (A_j, D_1), but it is not socially optimal. However, socially optimal matching is setwise stable since D_2 and D_3 together can block (A_j, D_1). Although non-unique pairwise stable equilibrium can arise in this model, I assume that the socially optimal allocation is reached since it is robust to setwise deviations.

Thirdly, another feature of the model that arises from conflict externality is that a drug advertiser has preferences not only about the contracts that are assigned to it, but also about contracts that are assigned to its competitors. Sasaki & Toda (1996) and Hafalir (2007) point out that in a matching game with negative externalities, the outcome of a stable equilibrium depends on agents' expectation of blocking matching allocations. In the stability definition above, I have made the behavioral assumption that each agent expects that other matches will remain the same if it deviates.

Fourthly, the static model has two implications. One is that drug advertisers only try to maximize the current period's surplus, and the other is that drugs and agencies will not make negative transfer payments or demand payments that exceed the surplus of the match.

Lastly, although I assume that agencies have full bargaining power in the proof of results, there exist other mechanisms that could produce socially optimal matching. For example, in the conclusion section, I will discuss a mechanism by which drugs propose

matches and transfers that also reach the socially optimal equilibrium. Such mechanisms resemble the market practice more closely.

Chapter 2 – An Empirical Examination of Matching Between Pharmaceutical Advertisers and Advertising Agencies

2.1 Introduction

In this chapter, I investigate the importance of conflict in the context of advertising agencies working for prescription drug advertisers. Drug advertising is particularly suited to a study of agency-client relationships for several reasons. First, prescription drug advertisers are important clients for the advertising agencies, contributing roughly 10% of total agency revenue. Second, there is a clear definition of product market competition by therapeutic class. Third, although many pharmaceutical companies produce more than one drug, advertising agencies chosen for different drugs within the same pharmaceutical company are often scattered. Conversations with the marketing directors of Pfizer and AstraZeneca suggest that advertising decisions are made independently by the marketing team of each drug, which allows me to assume that advertising agency choices are made at the drug level.

Based on a new dataset I construct of agency-drug relationships from 2002 to 2010 (combined from multiple industry sources), I build an empirical many-to-many matching model in which conflict acts as a cross-match externality when one advertising agency works for multiples drug advertisers in the same therapeutic category. More specifically, at the beginning of each year, advertising agencies and drug advertisers are assumed to observe drug/agency characteristics, match history of the previous year, and the draw of idiosyncratic shocks for each possible drug-agency pair. They then make matching choices to maximize their own surplus for the current year. The matching framework accounts

for the fact that an advertising agency and a drug advertiser will only contract when both parties find it in their interests to do so, a feature often missed in a more standard one-sided choice model.

To identify the relative contribution of conflict and characteristics of drugs and agencies on matching surplus, I rely on the revealed preference approach that compares the observed matching with its counterfactual deviations. More specifically, I construct preference inequalities using the model prediction that the stable matching yields higher total surplus than the alternative matching from a pair-wise deviation. Following J. T. Fox (2008), I estimate the model using a maximum score estimator. Intuitively, parameters of the surplus function are identified by the relative frequency of characteristics observed in equilibrium and in its deviations. In particular, the effect of conflict on surplus is identified when a deviation from equilibrium changes the amount of conflict in the market relative to other characteristics of the match.

There are two main advantages to using the semi-parametric maximum score in this setting. First, maximum score does not require me to solve the matching model explicitly and write down the probability of equilibrium allocation. This reduces the computation burden, especially when the number of participants is large and equilibrium is slow to compute. Second, the maximum score estimator does not require specifying the error distribution, so it avoids error misspecification in choice models that may lead to inconsistent estimators (Hausman et al., 1998).[13]

[13]Also, normal error assumptions may be problematic in the vertical contract setting. Vertical relationships are observed to be persistent but can dissolve without sharp changes on observables. One possibility is that the relationship dissolves as a result of experiencing a rare but large negative shock. This suggests that the error distribution may be asymmetric with a long left tail. In such cases, maximum score estimation is preferred to parametric error assumptions. For the counterfactual exercise, I assume that the error follows a mixture of two normal distributions, and calibrate it by matching simulated moments given the error distribution and the data moments. I find that the distribution is bimodal with one normal component having a much more negative mean than the positive component, confirming the above big negative story.

Using the maximum score estimator, I find that conflict does indeed reduce the match surplus and is especially costly for a pair of agents who have matched with each other in the previous period. Also, preserving a previously formed match yields a much higher joint surplus than forming matches with new partners. Moreover, conflict crowds out matches whose returns are positive but too low to offset the conflict cost. In a counterfactual, I find that 14% of all matches that would have formed in a non-conflict environment are unable to form when conflict is present. Among them, 95% are matches with new partners. In another counterfactual, I simulate the effect of a hypothetical merger between two large advertising holding companies, Omnicom and Publicis.[14] With a combined 32% market share in the US,[15] Omnicom and Publicis proposed to merge in August 2013 but canceled the deal ten months later for various reasons. With the hypothetical merger, I find that the number of matches with new partners is reduced by 7%; most previously formed matches survive conflict but lose 6.8% of the surplus due to the increased amount of client conflict in equilibrium.

This chapter is also closely related to the empirical matching literature that has recently advanced tools for estimating matching models (J. T. Fox (2008); J. T. Fox (2010); Agarwal (2012)). This line of work deals with the endogeneity that arises from capacity constraint that each agent's choice set is endogenous and unobserved. These models and empirical methods mainly use the sorting pattern of agents who are matched to identify matching preferences. However, the canonical matching models do not consider

[14]Mergers of service providers raise particular interest in previous studies of conflict since mergers directly change the relationship structure. Mergers increase the probability of conflict by reducing the number of service provider choices, and in some cases, by directly bringing conflicting clients to the same agency. The advertising agency industry is also of antitrust policy interest since the industry has been gradually consolidating since the merger waves in the mid 1980s.

[15]The estimates of total U.S. agency revenue come from *Advertising Age*, which collects data from the 1,000 biggest agencies (http://adage.com/article/news/publicis-omnicom-group-facts/243346/). Until 2012, the four biggest agency holding companies in the US ranked by revenue captured 60% of the advertising service market share.

other types of externalities arising from client choices, such as conflict. One exception is Uetake & Watanabe (2012) that studies bank merger as a matching decision and allows post-match market competition as a source of negative externality. I provide an application of the matching model that incorporates conflicts.

The rest of the chapter is organized as follows: In the next section, I describe market of the advertising agencies for pharmaceutical advertisers and explain the dataset I construct for this study. I then summarize the data and use descriptive regressions to show matching patterns in this market. In sections 3 and 4, I discuss identification and estimation of the model and explain the estimated structural parameters. In section 5, I conduct two sets of counterfactual exercises to analyze the effect of conflict and the effect of advertising agency mergers on the allocation of matches. Section 6 concludes.

2.2 Market Description and Data

2.2.1 Agency Market for Pharmaceutical Advertisers

Pharmaceutical companies spend heavily on marketing. Marketing expenditures on pharmaceuticals amount to $26-36 billion annually, representing 9-12% of the total drug sales between 2002 and 2010 (Kornfield et al., 2013). Among other promotional activities, advertising agencies participate in direct-to-consumer (DTC) advertising, which accounts for 12-15% of their marketing spending. Advertising agencies also collaborate with pharmaceutical companies on promotional activities aimed at providers (e.g., journals, meetings and conference), which account for another 27-32% of the total marketing spending, excluding face-to-face sales representative detailing. As a result, around $10-16 billion of pharmaceutical advertising spending is billed through advertising agencies

annually.

Two features of the matching between drug advertisers and advertising agencies facilitate the study of conflict. First, product market competitors can be defined precisely. I define that clients are competitors if they seek advertising agencies to advertise for the same treatment indication (e.g. high cholesterol). The second feature is that drugs manufactured within the same company usually have independent marketing teams. This practice is also confirmed in the data. The advertising agencies hired by different drugs manufactured by the same pharmaceutical firm are not clustered. For example, in 2009, AstraZeneca has 22 drugs hiring advertising agencies, and the drugs hired 13 unique agencies; Sanofi-Aventis has 39 drugs hiring 17 unique agencies; and Pfizer has 70 drugs hiring 32 unique agencies. This practice allows me to model each drug (advertiser) as the decision maker without worrying about the complementarity in the matching decision for different drugs within the same manufacturer.

In terms of commissions paid by drug advertisers to advertising agencies, I assume in the model that the commissions are determined endogenously by the bargaining rules between drugs and advertising agencies. In earlier years, the advertising industry's fee model compensated advertising agencies with 15% of the total billing, but since the late 1990s, the advertising agency fee model has gradually increased the proportion of incentive pay in the total amount of fee. Hence, it is reasonable to assume that the commission is not a fixed percentage of the billing.[16]

2.2.2 Data Sources

I collect data on (1) matches between drug advertisers and advertising agencies,

[16]In 2003, advertising agencies' total revenue represented approximately 13% of the capitalized billings from clients (Horsky, 2006).

(2) drug advertisers and advertising agencies characteristics, and (3) the organizational structure of advertising agencies between 2002 and 2010.

I compile data on matches from advertising agency surveys conducted independently by two industry magazines, *Pharmalive* and *Medical Marketing and Media*, and one data vendor, Advertising Redbooks. To my knowledge, *Pharmalive* and *Medical Marketing and Media* are novel sources of data used for researching relationships between drug advertisers and advertising agencies. Each conducts annual surveys on US advertising agencies with a major client base in healthcare.[17] Advertising Redbooks (also known as the Standard Directory of Advertising Agencies) is a comprehensive data source on advertising agency characteristics and clients and has been used previously in research on advertising agencies (Horsky (2006); Rogan (2013)). The Redbooks data are used to complement the list of non-healthcare advertising agencies that form matches with drug advertisers.[18] Data on matches contain information on names of drugs and agencies that have a contract at a given year, the year the match was first established, and the specific treatment indication (e.g. High Cholesterol, Crohn's Disease) of the drug advertised.[19]

Conflict is defined as the situation in which advertisers for drugs of the same treatment class hire the same advertising agency in the same year. To define treatment class, I obtain class information from the Medical Expenditure Panel Survey (MEPS)

[17]To ensure that the list of agencies working for pharmaceutical companies is comprehensive, I compare the agency data separately collected by the two industry magazines and find that their coverage overlaps and includes a similar list of 100 healthcare agencies.

[18]Although Advertising Redbooks is a standard data set for studying client-agency relationships, the documentation for matches between drug and advertising agencies is much less complete compared with data sourced from the industry magazine. Redbooks usually lacks information on product name (e.g., Lipitor) and only records the advertiser's name (e.g., Pfizer). It misses many healthcare specialized agencies, and it does not include information on when the account was first established.

[19]One caveat of the match data published by the above sources is that it lacks detailed documentation of the types of advertising services engaged for each drug-agency pair. Most of the services are marketing project assignments aimed at reaching consumers and healthcare professionals. In these instances, conflict is a concern. But there are also accounts for which conflict is not a major consideration when forming matches, such as media-buying agencies who pools clients' accounts together in order to negotiate better media prices. For this reason, I exclude media-buying agencies.

and Drugs.com, both of which use the coding system provided by Cerner Multum.[20] For any drug advertised for multiple indications, I only keep the major indication.[21]

In addition to treatment class, I collect information on drug approval date from the FDA Orange Book and DTC advertising expenditures from the Kantar Media. I also construct a US blockbuster drug indictor with a dummy set equal to one if the drug was ranked in the top 200 in annual US sales, based on yearly information from IMS Health by way of Drugs.com.

An important component in the data construction is deciding at what agency organization level the clients' concerns about conflict are relevant (i.e., whether product competitors only avoid contracting with the same advertising agency, or if they also avoid contracting with agencies that are associated via a parent holding company). This consideration is especially relevant in the advertising industry since many advertising agencies exist in a tree-like organizational structure in which advertising agencies are affiliated with one giant parent holding company (e.g., Publicis and Omnicom), and the agencies themselves have branch offices in different locations. The effect of conflict may be underestimated if conflict is defined at the holding company level when clients, in fact, only care about conflict within an agency. Conversely, the conflict effect may be overestimated if conflict is defined at the agency branch level. In principle, the conflict concern depends on how closely different units within the holding company are related in terms of communicating and decision-making.

I choose to define conflict at the level of the brand name agency, which is one level below the advertising holding company. The rationale is that brands within the

[20] I do not use drug indication provided by *Pharmalive*. Although they are available, they are usually narrow and the codings of disease names are not comparable across drugs.

[21] Drug characteristics are collected at the drug level rather than at the drug-treatment level: I collect the FDA drug approval date for the main treatment indications, but not for the minor treatment indications, so I drop matches indicating advertising for minor indications.

same holding company run their own websites, on which they make account-winning and award-winning announcements under their own brand names. Industry news sources often quote agencies using their brand names, and agencies within the same holding company sometimes compete for the same client.[22]

As an example of the construction, I treat Area23, DraftFCB, McCann, and ICC Lowe as unique advertising agencies even though they are all directly affiliated with the holding company Interpublic. But I do not treat branch offices under the same brand name as independent agencies. So I treat ICC Lowe Pace, ICC Lowe Thermal, and ICC Lowe Trio as the same agency, ICC Lowe. I determine organizational structures and affiliations using the Corporate Affiliation data of Thomson Reuters, bundled with Redbooks.

Lastly, to identify organizational change, I collect merger information on all advertising agencies using the advertising agency SCI code 7311 and the data sources Bloomberg, CapitalIQ, and SDC Platinum from 2001 to 2011.

2.2.3 Sample Construction

I treat each treatment class in a given year as an independent matching market since advertising agency contracts are reviewed and renewed annually. I construct data on prescription drugs and agencies, their characteristics, and their matching statuses each year between 2002 and 2010.

More specifically, I first determine the set of drug advertisers that potentially seek matches in a given year and treatment class. In theory, any prescription drug advertiser may conduct marketing activities and consider hiring an advertising agency for assistance. But since a complete directory of all drugs currently marketed for each treatment class

[22]The data sources on drug-agency relationships are also most comprehensive at the brand level as defined above.

is not available, I compile the list of likely-to-match drugs using two criteria: a drug is included if it has ever appeared in the Medical Expenditure Panel Survey (MEPS)'s Prescribed Medicines File between 2002 and 2010, or if it has appeared in the DTC advertising data between 1996 and 2011.[23] This twofold inclusion criteria allows me to capture drugs that are either commonly prescribed by doctors or explicitly advertised to consumers. Drugs missing from my sample tend to be those that treat rare diseases. In that case, the entire treatment class is not included in my sample. There are also drugs with small and niche consumer bases that are captured by neither MEPS nor DTCA data. If such drugs also hire advertising agencies, I assume that they are not good substitutes for mass market drugs in their classes and hence do not impose conflict concerns.

To be conservative, I assume drugs may start looking for advertising agencies 5 years before surplus market approval. In the data, I observe that about 10% of the matches were formed by drugs before they get the FDA approval.[24] Majority of them have started hiring agencies up to 5 years before market approval.[25]

To determine the set of agencies that potentially serve a treatment class, I include all agencies that have ever done advertising work for any drug in the treatment class. Agencies that are newly formed during the sample period are assumed to have started serving the treatment class from the first year of their establishment. I assume that an agency has exited the market when it stops being observed in any data set that I use to compile matching and agency information in a given year.

[23]MEPS is a nationally representative survey of the U.S. civilian non-institutionalized population conducted by the Agency for Healthcare Research and Quality (AHRQ) and covers drugs prescribed to the general US population.

[24]Many pharmaceutical companies start to inform medical professionals about the existence of a promising new drug when it is in the last phase of clinical trials, so we observes in the data that drugs starts to hire agencies before it gets the FDA approval.

[25]During the period between 2002 and 2010, there were only 20 matches formed more than 5 years before the drug was approved.

2.2.4 Market Descriptive Statistics

There are an average of 627 unique prescription drugs in 152 treatment classes to be matched in each year between 2002 and 2010. The mean number of drugs in a treatment class market is 4, with a maximum of 28. Among the 1,214 class-years in the data, 46% of the treatment classes have only one drug in a given year and do not have any potential conflict concerns.[26] As I mention in the data construction, I treat each agency brand as a unique agency. There are 149 unique advertising agency brands in the sample, and each agency can work in multiple markets, generating a total of 1,525 unique agency-class combinations in the sample. The mean number of agencies in a treatment class market is 8, with a maximum of 38.

Table 1 summarizes the size of each class-year, the number of matches formed, and the number of conflicts. On average, a class-year contains 4-5 drugs and 8 agencies. In a typical class-year, 68% of the drugs and 46% of the agencies find at least one match in the market, which results in around 4-5 matches per class per year. The median size of the market is small with 2 drugs and 5 agencies. But for the 300 classes that form the top quartile in terms of market size, each class has more than 6 drugs and 11 agencies. Drugs can match with more than one agency in a market. Restricting the sample to only those drugs that obtain at least one match, the mean number of advertising agencies a drug hires is 1.6. Of all the drugs seeking matches, 59% hire a single agency in a year, 25% hire two agencies, and 15% hire three or more agencies.

Agencies may also match with more than one drug in a market, which results in a client conflict. In a given market, 7% of agencies are involved in a conflict (roughly

[26] Although these drugs do not help identify conflict, they help to identify preferences of drugs and agencies.

one agency in each class), and each conflict contains 2 drug advertisers, on average. An average of 2.5 matches, or 25% of all matches, are caught in conflict in a market. In the following descriptive regressions, I will show which drugs and agencies are more likely to be matched and to be caught in conflict.

In addition to the static market statistics, Table 2 shows a high degree of state dependence among matches, which is commonly observed in vertical relationships. Among nearly 700 matches formed each year, around 73% of the drug-agency pairs will match with each other again in the next year. It should also be noted that the probability of match dissolution is similar regardless of how many years the relationship has been formed before. The probability of match dissolution appears independent of the relationship duration, remaining at a constant rate of 23 to 25 percent. This suggests that the value of preserving previous matches versus starting new matches is similar regardless of how many years the relationship has formed. Meanwhile, new drug-agency pairs form around 28% of all matches formed in a given year.

I use descriptive regressions to examine which drugs and agency characteristics are associated with a higher probability of forming matches. We expect that drugs have more incentives to hire agencies when they have a large marketing budget and when they conduct the types of promotional activities that they lack the specialty and resources to market on their own. For example, drugs often delegate DTC advertising, consumer and professional educational programs, and conference organization tasks to advertising agencies. And given that an advertising agency's capacity in dealing with each client is limited, we will also expect that a drug advertiser with a bigger marketing budget is also more likely to find it profitable to hire more than one agency. The lower panel of Table 3 presents simple comparisons by drug characteristics that do and do not hire any

advertising agencies in a class-year and Table 4 shows the results of linear probability regressions of a drug's agency hiring decision. Both tables confirm the expectations. A drug that spends more on DTC advertising, in particular DTC advertising on TV, and a drug approved by the FDA and enjoying market exclusivity[27] is more likely to hire one or more advertising agencies. The result is similar if I compare across drugs in column (1), or within drugs in column (2) of Table 4.

I also examine whether drug advertisers with a larger marketing budget hire more agencies. (See columns (3) and (4) of Table 4.) Conditional on having any match in a given market, I regress the dummy variable of a drug having multiple matches on drug characteristics. Both regressions, with and without drug fixed effects, show that drugs with higher levels of DTC advertising and drugs that are between approval and patent expiration are more likely to hire multiple agencies.

The summary statistics in the upper panel of Table 3 and linear probability regressions of Table 5 compare the characteristics of advertising agencies that are hired or not hired in a market. The differences in characteristics are very small, and regression results are weak with a small R^2 of 0.01 if I do not control for agency fixed effects. The signs of the differences are still as expected: an advertising agency with a bigger size in terms of total billing and total number of employees is more likely to get matches, and more likely to have conflicting clients. In addition, an agency claiming consumer specialty is less likely to get a match, and an agency claiming specialty in handling health accounts is more likely to get a match. This is reasonable because advertising agencies with a consumer focus usually only deal with a few blockbuster drugs for which they create TV ads, but do not targets pharmaceutical clients in general.

[27]Patent expiration data is approximated by 12 years as documented in the literature. The publicly accessible FDA Orange Book is incomplete in dating the exclusivity period for each individual drug in my sample.

In tracing the identity of the advertising agencies, I also identified nine merger cases that affect 20 advertising agencies in my sample period. Although the number of mergers is small, the agencies involved in the mergers worked in a wide range of treatment classes, and a single merger affects many classes at the same time. During the sample period, treatment classes that have ever had an advertising agency affected by the mergers is 127, or 80% of all classes. Counting the number of class-years that are affected by an agency at the year of its merger, 389 are affected. And if counting all class-years with a merged agency post its merger event, there are 826 such class-years, or 54% of all class-years in sample. This gives us enough variations to identify the average merger effect of matching surplus.

2.3 Model Estimation and Identification

I use the maximum score estimator proposed by J. T. Fox (2010) to estimate the matching surplus function by assuming that the observed matches in the data represent the socially optimal equilibrium. Using pairwise deviation conditions that characterize the socially optimal allocation and a rank order property, I identify the model by comparing the total surplus of observed matches to counterfactual matches following pairwise deviation.

2.3.1 Maximum Score Estimator

Applying pairwise deviation conditions to the socially optimal allocation, I derive a system of inequalities that can be used to estimate the surplus function without requiring

transfer information. The objective function of the maximum score estimator is

$$Q(\theta) = \frac{1}{K} \sum_{k \in K} \sum_{\tilde{\mu}_k \in \tilde{\mathcal{D}}_k} \sum_{(i,j) \in \mu_k^e} \mathbf{1}\{S(i,j|\mu_k^e, \theta) - S(i',j'|\tilde{\mu}_k, \theta) \geq 0\} \tag{4}$$

where k denotes each independent matching market, K is the total number of markets, $\tilde{\mathcal{D}}_k$ is the set of pairwise deviation matchings from allocation μ_k^e , and μ_k^e is the equilibrium matching in market k. The objective function of the estimator is to find a parameter vector $\hat{\theta}$ to maximize the number of times that the observed equilibrium yields greater surplus than the pairwise deviated allocations based on deterministic characteristics of the model.

Given that agents in the markets behave according to the set of random preference shocks they draw and that the shocks are unobserved by the econometrician, surplus inequalities constructed based on deterministic characteristics are not guaranteed to hold. In order to apply the maximum score estimator, J. T. Fox (2010) proves a rank-order property that is key to model identification. The rank-order property states that the match allocation that yields higher surplus based on observables is more likely to be observed in the data, under the condition that the errors are *i.i.d.* distributed for each match. Denote A as the realization of matching allocation, W as the observable, S as the surplus function, and G as the stochastic structure of the model. The probability of the allocation A given observables W is

$$Pr(A|W; S, G) = \int_{\epsilon} 1(A|W, \epsilon; S) dG(\epsilon).$$

The rank order property states that

$$Pr(A_1|W;S,G) > Pr(A_2|W;S,G)$$

if and only if

$$\sum_{(i,j)\in A_1} S(i,j) > \sum_{(i,j)\in A_2} S(i,j). \tag{5}$$

With the rank order property, the estimator is constructed to punish the wrong prediction of inequalities based on the deterministic part of the surplus. The crucial assumption for identification using maximum score is that $\epsilon_{ij} = \epsilon_{ij}^i + \epsilon_{ij}^j$ is $i.i.d.$ across all matches. Specifically, heteroskedastic errors based on match characteristics are not allowed.

Table 7 listed the full set of pairwise deviation conditions used. In general, the deviations are of three types. The first involves replacing one partner in the current match with another opponent. This type of deviation does not change the total number of matches in a market. The second type involves adding a new match, and the third type involves dissolving a current match. This list exhausts all possible pairwise deviations. In estimation, I use all inequalities derived from these deviations.

My model setup, combined with the ability to observe all agents in the market, allows me to use a larger set of inequalities than that used in J. T. Fox (2010). These inequalities also help me to identify the impact of agent-specific characteristics. One major difference in my model setup from classic matching models is that I assume away capacity constraints for drugs, hence removing competition among advertising agencies contracted by drug advertisers. This assumption allows a drug marketing team's decision to add or drop a match to be independent of its decisions about other matches. Hence, inequalities based on unilateral deviation of a drug does not concern transfers. Such inequalities

cannot be constructed by the model in J. T. Fox (2010) or in matching models with capacity constraints since a one-sided deviation involves an agent exceeding its quota of matches. To construct such a deviation, one needs to know the transfer paid to one agent in order to induce it to replace a partner.

There are two main advantages to using maximum score estimation in my setting. First, the maximum score estimator is free from error distribution assumptions, so it avoids the concern that misspecified error in choice models could lead to inconsistent estimators (Hausman et al., 1998). It is particularly useful in analyzing vertical relationships since the errors may be asymmetric. Given that vertical relationships are observed to be persistent, the adverse shock must be large to dissolve the relationship without sharp changes on observables. This suggests that an error may be associated with large downside risk disproportional to upside shocks in size. Without strong ex-ante information about the error distribution, maximum score estimation is preferred to making parametric error assumptions. I calibrate the error distribution assuming a mixture of two normal distributions, and I find that the distribution is bimodal, with one normal distribution having a much more negative mean than the positive mean of the other normal distribution. Second, maximum score estimation does not require solving the matching model fully. When the number of participants in the markets is large, the maximum score estimator reduces the computation burden relative to parametric models.

The estimation approach of using preference inequalities is known to lack point identification if there is not a match-specific densely distributed continuous variable. Although I include agency size as a continuous variable in the model, I do not have a comparable densely distributed variable for drugs, so I can only set-identify the parameters. In order to conduct inference, I estimate confidence regions for set-identified parameters following

Shaikh (2006) and the estimation manual of Santiago & Fox (2008).

2.3.2 Model Specification

Descriptive regressions in section 2 show that some drug and agency characteristics are important for predicting matches, that past matching relationships matter, and that conflict may affect the match surplus. Assuming linearity in parameters, I specify the matching surplus function of an advertiser-agency pair as

$$S_{ij} = X_i\alpha + Z_j\beta + X_iZ_j\gamma + \lambda L_{ij} - 1(|\mu_j| \geq 2) \times \kappa_c - \kappa_f + \epsilon_{ij}$$

where X_i is a vector of drug characteristics, including drug life-cycle, an indicator for DTC advertising and DTC TV advertising, and an indicator if the drug was a top-200 selling prescription drug in the US in the previous year. I use an indicator of whether a drug does direct-to-consumer advertising for the current period because I assume that the drug advertiser has decided if it will do any DTC advertising before forming matches.[28] Z_j contains characteristics of the agency, including number of employees, total billing, age, and claimed specialty in the past year. I use agency information for the previous year since this is the information drug advertisers have when making matching decisions. Also, I include the interaction between drug and agency characteristics to determine whether synergy is created when drugs and agencies of specific characteristics match together.

Z_j also contains information on agency mergers. Since I do not explicitly model the endogenous merger decision, I add a dummy to control the unobserved differences

[28] One can argue that drug's observed current year advertising spending depends on if the drug is matched to an agency, and is endogenous. I also tried in the reduced form to include the previous year's advertising spending, assuming that past spendings predicts the amount the drug plans to spend in the current year. The results are similar.

of merging agencies in contributing to the match surplus. I add one dummy variable for an advertising agency that has ever had a merger during the sample period, and another dummy variable for an advertising agency during periods after its merger. Adding merger dummies to capture the unobserved synergy also prevents me from over-stating the adverse effects of mergers in the counterfactual analysis. Although mergers reduce matching opportunities, synergy makes each matching opportunity more attractive and hence could lead to more matches and matches with higher surplus. Endogeneity of merger decisions could lead to an over-estimation of merger synergy since only advertising agencies that stand to gain decide to merge. Since I argue that mergers reduce total surplus by reducing matching opportunities, if endogeneity of mergers is a concern, mergers' adverse effects on total surplus will only be understated in the counterfactual analysis.

From the persistence of contract, it is expected that preserving a match formed by partners matched in $t-1$ is much more valuable than a match formed by new partners in period t, so a dummy L_{ij} is added that is equal to 1 if $Match_{ij,t-1} = 1$. In some specifications, I inspect whether the value of past matches differs for matches with different advertising agency and drug characteristics by interacting L_{ij} with X_i and Z_j.

Matching cost is denoted as κ_f and is common across matches. Conflict cost is denoted as κ_c. Given that previously formed matches play an important role in predicting matches, I also interact L_{ij} with conflict to examine whether it is more or less costly for previously formed matches to be caught in conflict compared to newly formed matches.

2.3.3 Identification

To restate the key identifying assumption, I assume that errors are drawn independently from the same distribution. This assumes away heteroskedastic error, and the

concern that there may be unobservables correlated with the controls.[29] In addition, assumptions made on the matching model are maintained in the estimation for identification.

The maximum score estimator applies the idea of revealed preference: the match allocation we observe in the market is more likely to yield higher aggregate surplus than alternative allocations. This suggests that the characteristics more often observed in the matches are likely have a positive impact on the match surplus. By checking each type of deviation from the observed equilibrium, the estimator compares the relative effects of the sets of characteristics affected by the deviation. In the following section, I explain the pairwise deviations derived from the surplus optimization conditions and show how different sets of deviations help identify different parameters of the model.

Comparisons of surpluses between observed allocation and deviation should be written in expectation. For convenience of illustrating which coefficients are identified from different sets of deviations, I will only focus on the deterministic part of the surplus function.

The cost of conflict is identified when a deviation causes the number of matches formed by an advertising agency to change. For example, consider the deviation in which a drug i replaces one of its hired advertising agencies $j \in \mu(i)$ with another agency $j' \notin \mu(i)$, causing a change in the amount of conflict. The change in total surplus is illustrated in the following formula,

$$(Z_j - Z_{j'})\beta + X_i(Z_j - Z_{j'})\gamma + \lambda \Delta L_i + \Delta ConflictCost \geq 0$$

[29] One advantage of using inequalities that only involve the comparison between different matching opportunities is that this allows for the presence of market level unobservables that might be correlated with the controls. Depending on the set of inequalities used, we can even allow agent-specific unobservables when they are differenced away with the construction of inequalities. However, in order to recover all parameters for the counterfactual exercise, I need to use inequalities that compare opportunities of being matched with unmatched, and hence need to make a strong exogenous assumption.

where $\Delta L_i = Match_{ij,t-1} - Match_{ij',t-1}$, and

$$\Delta ConflictCost = \begin{cases} 0 & if\ N_j = 1, N_{j'} = 0\ or\ N_j \geq 2, N_{j'} \geq 2 \\ \kappa_c & if\ N_j = 1, N_{j'} \geq 2 \\ 2\kappa_c & if\ N_j = 1, N_{j'} = 1 \\ -\kappa_c & if\ N_j > 2, N_{j'} = 0\ or\ N_j = 2, N_{j'} \geq 1 \\ -2\kappa_c & if\ N_j = 2, N_{j'} = 0. \end{cases} \tag{6}$$

In this type of deviation, the relative value (cost) of agency characteristics, agency and drug synergy, matches formed in the past, and conflict is identified. The above comparison between equilibrium and its deviation illustrates the properties associated with the identification from preference inequalities.

First, because of the sign of the conflict cost and the sign of the inequality, I must identify the upper and lower bounds of the conflict cost using different inequalities. For example, assume that the conflict is indeed costly and has a negative impact on surplus. I identify the lower bound of the magnitude of the conflict cost when the amount of conflict increases in the deviation, and I identify the upper bound of the absolute value of conflict cost when the amount of conflict decreases in the deviation. So it is necessary to include full sets of inequalities to identify both bounds.

Second, note that the one-sided deviation is necessary to identify conflict cost. If deviation involves two pairs of matches switching partners with each other (as in J. T. Fox (2008)), it causes no change in the number of drug advertisers an advertising agency works for and does not help identify conflict.

Third, it becomes clear in the inequality that the model can only be identified up to

scale. Mathematically, scaling all coefficients up or down does not change the probability of the allocation being observed. Intuitively, I can only measure the relative effect of the impact of different characteristics on the match surplus. In the estimation, I choose to normalize coefficients by $||\theta|| = 1$.

The above deviation also hints at which coefficients are not identified if I focus only on a subset of the inequalities. Switching the advertising agencies a drug hires only involves changes in agency characteristics, and the drug's own characteristics are not identified since they are differenced away in the inequalities. Intuitively, the impact of drug characteristics on surplus can only be identified when deviation involves replacing a drug. I am also unable to identify the cost of matching since there is no change in the total number of matches formed. I can only identify the cost of matching by adding or dropping matches from the equilibrium.

The absolute "location" of the surplus function is not identified, but it is relative to the outside option of not forming a match. To be precise, when I discuss the benefits and costs of forming matches, these are actually the benefits and costs of forming matches relative to the outside option. For example, since the "cost" per match κ_f is the constant term for the surplus function, it represents the net cost (benefit) of forming a match relative to the outside option for the baseline match.[30] As for the coefficients on other characteristics involved in a match, I interpret them as the effect of each characteristic on the net surplus relative to the outside option of a match. In the case of a positive coefficient, for example, the model cannot separately identify whether the effect comes from increasing the relative value of a match or decreasing the relative cost of a match.

[30] In the estimation, continuous variables are centered at their sample mean.

2.4 The Impact of Conflict on Surplus

In this section, I first present suggestive evidence of the existence of conflict using reduced form regressions and then present the conflict parameters estimated in the structural model.

2.4.1 Descriptive Evidence on Conflict

I run descriptive regressions to look for evidence that matches are less likely to form when two or more drug advertisers in the same class contract with the same agency.

Ideally, I can condition on each drug's probability of forming a match and examine whether one drug's decision to match with an agency is affected by its competitor's probability of matching with the same agency. However, intention to match is not observed; only matches formed in equilibrium are observed. Replacing the intention to match with observed matches obviously causes endogeneity. In the same spirit but using a slightly revised regression, I define that a match with an agency causes "potential conflict" when the agency was observed to match with the drug's competitor in the previous period. When errors are not temporally correlated, matches in the past period are uncorrelated with error in the current period. Also, if an agency matched with a drug in the previous period, about 73% chance this match will form again, so the potential conflict defined in the past period is likely to be a good predictor of potential conflict in the present.

In the regression, I pool all possible matching opportunities between drugs and agencies in each class-year and use the linear probability model to describe which characteristics are associated with a higher probability of matches. Besides "potential conflict," I include the same set of controls as in the structural specification, including characteristics

of drugs and agencies and past match relationships. The evidence in Table 8 is robust to adding heterogeneous effects of previously formed matches.

Table 8 also shows heterogeneous impacts of previously formed matches on current matching choices. Specifically, results show that matches with certain drug and agency characteristics are associated with a higher probability of matching again the next period, the implications of which fit both a story of sunk cost and one of value created in the previously formed relationship. On the one hand, matches with drug advertisers utilizing DTC and DTC TV spending are more likely to continue relationships with agencies, which can be explained by the high sunk cost association with TV ad production.[31] On the other hand, matches with agencies with more employees are more likely to continue into the next year, which is consistent with a value creation story in which advertising agencies with larger teams may be more likely to provide a diverse and more effective service in the long term than the more specialized smaller agencies.

2.4.2 Maximum Score Estimates of Conflict

The baseline specification of surplus function includes drug and agency characteristics, merger indicators for agencies, and an indicator for previously formed matches. It also allows conflict cost to vary between previously formed matches and new matches. Given that the estimators are identified up to scale, I choose to normalize coefficients by restricting the sum of coefficients squared to be one, or $||\theta|| = 1$, and exclude the coefficient for the indicator for the previously formed matches given the normalization restriction.

[31] The regression also suggests that drugs that haven't yet received FDA approval are more likely to continue their contracts. This may be driven by the censoring of the number of years drugs start looking for agencies before approval. Drugs on average will contract with an agency for 4 years (with match dissolution rate at about 27%), but the majority (90%) of observed matches are formed at less than or equal to 3 years before approval.

Maximum score estimates for the baseline specification are shown in specification (1) of Table 9.

The cost of conflict is relative to the value of the matches. The value of preserving a previously formed match relative to the cost of forming a match with a new partner is the most important contributor to the match surplus. Comparing the midpoint of the estimates, the cost of conflict to a previously formed match relationship is equal to roughly 25% the value of such a relationship. Conflict is also costly in matches with new partners, but the magnitude is much smaller, at only 2% of the cost associated with the new relationship. Although the cost of conflict is large only for previously formed matches, these matches represent 72% of all matches. Conflict affecting these matches also plays an important role in determining the overall matching allocation as I will show in the counterfactual exercise.

It is not surprising that conflict is much more costly for previously formed relationships. One reason is that an advertising agency's knowledge about a drug advertiser grows quickly in the first year of the relationship, so the agency puts a greater amount of confidential information at risk when it signs with the established client's competitor. Drugs will be more worried about information leakage after the initial year of a contract than when the match was first formed. The other reason is explained in Rogan (2013). She finds that conflicts result in higher probabilities of client switching when the relationships have existed for a longer time because the advertiser's trust is betrayed when the advertising agency he previously worked with takes on his product market competitor.

Given the important role of previously formed matches in affecting current matches, I discuss two main explanations - sunk cost and match-specific unobservables - in the matching market between drug advertisers and advertising agencies. The persistence of

contracts may result from the fixed cost invested in the initiation of a relationship, for example, the cost of learning about the product, becoming familiar with advertising regulations specific to the drug, and conducting consumer research. In other words, these investments also become relationship-specific assets accumulated in the previous contractual period (Riordan & Williamson, 1985). Another interpretation is that this captures the unobserved time-invariant match-specific heterogeneity specific to the pair. If this is the case, we should expect that the pair that matched in any previous period but not in the past period would have a higher probability of forming a match again than a pair that has never contracted before. However, I do not find such evidence in the data. The probability of a pair that matched in the previous year matching again is 0.77. This probability is 0.028 for a pair who did not match in the previous year, but matched years before, and the probability is 0.20 for a pair who has never matched before. This evidence supports a sunk cost mechanism as the main driver of contract persistence. The sunk cost explanation may imply dynamic incentives that this study does not directly address, but I will be careful in interpreting the effect of conflict on allocation of matches between previously formed matches and new matches.

Moreover, a merger produces a big synergy effect that is one-third the size of the cost of conflict for previously formed matches. In addition, agencies that have ever been involved in a merger produce a higher surplus in their matches. Merger synergy will make matching opportunities more desirable, counteracting the matching opportunities it takes away for drug advertisers, so counterfactual exercises are necessary to evaluate the merger effect. Other drug and agency characteristics appear to be much less important since the indicator for past relationship is included, and year-to-year variations of drug and agency characteristics are small.

I also run alternative specifications for robustness checks and find that the relative magnitudes of conflict costs and benefits of the previous matches and common matching costs are similar across specifications. One specification deserves special note: I investigate whether there is a capacity constraint for drug advertisers by allowing the matching cost to change when drugs form additional matches. The result is shown in column (2) of Table 9. There is no evidence that it is more costly for the drug to take on additional matches. In an unreported table, I also allow conflict cost to vary when the number of drugs in a conflicting relationship grows. The differential conflict cost is not significant, suggesting that it is reasonable to make the constant conflict cost assumption. Because the presence of previously formed matches has an important effect on surplus, I try allowing this effect to differ by drug and advertising agency characteristics and allowing the effect to differ by the duration of the established relationship. In the first model with heterogeneous effects of past matches, the coefficients are all insignificant, and the magnitudes of the main coefficients are unchanged. In the second model of heterogeneity with time, I add dummy variables to indicate the different durations of the established relationship.[32] There is no evidence that a longer relationship is more or less valuable, which is consistent with the data pattern: match dissolution rates are roughly the same conditional on number of years observing the relationship.

2.5 Counterfactual

I conduct counterfactual exercises to examine the effect of conflict on equilibrium matches given the current market structure, and I then examine the effect of a change in market structure when conflict is costly. I will explain how conflict affects the allocation

[32] I add dummies for matches that have lasted one, two, three, and four or more years.

of matches conceptually and then discuss details and results of the counterfactuals.

2.5.1 Conflict Crowds Out Matches

When there is additional cost for an advertising agency to simultaneously serve directly competing advertisers, any pair of drug and advertising agencies with a positive net return should form a match, denoted as $r_{ij} = y_{ij} - \kappa_f + \epsilon_{ij} > 0$. However, when concern about conflict exists, it lowers the return of conflicting matches and makes some matches no longer profitable to form. This crowding out effect depends on the match return and number of matches caught in conflict.

A match with a net return lower than the conflict cost it incurs, $r_{ij} < \kappa_{c,ij}$, will not sustain conflict since it is never socially optimal to add a match whose net surplus is negative. When two or more matches in this range try to match with the same advertising agency, only the single match with highest return will form with the agency.

When the net return of the match is greater than its conflict cost $r_{ij} > \kappa_{c,ij}$, two situations may arise. In the first situation, the match is crowded out when there exists another match with the agency whose return is higher. Compared with having one match, adding an additional match not only incurs an extra conflict cost for the added match but also creates a conflict cost for the existing match. So in order to form, the additional match must be valuable enough to overcome both its own conflict cost and the externality it creates for the other match in order to form the match, $r_{ij} > \kappa_{c,ij} + \kappa_{c,i'j}$. In the second situation, the match will always form. This happens when the match's return is high enough to overcome the conflict cost externality, or that the match is of higher value for the advertising agency, so it crowds out other matches.

The above mechanisms through which matches are crowded out is interesting since

they imply that even though conflict cost for new matches may be low, it is still difficulty for new matches to form because, when in conflict with a previously formed match, the new match must overcome the previously formed match's conflict cost in addition to its own conflict cost.

Since estimation does not support conflict cost as an increasing function of the number of matches in conflict, this negative externality only exists for the first marginal match added. Conditional on having two matches for an agency, the third match only adds a conflict cost for itself. This suggests that if the agency already has two matches in equilibrium with sufficiently high net returns to sustain conflict, a new match's return only needs to be higher than its own conflict cost for it to form.

2.5.2 Error Calibration

Simulating the market equilibrium requires knowledge of the error distribution. Since the maximum score estimation strategy does not impose assumptions on the error distribution, I calibrate the error distribution by matching simulated moments to the data moments using different distribution parameters. I explain the details of the calibration exercise in this section.

I choose a mixture of normal distributions because the error distribution is likely to be asymmetric based on the pattern of match formation and dissolution observed in the data. The previous data description indicates that most matches with partners matched in the previous year will form again in the current year. The parameter estimates reflect that preserving previously formed matches yields a much higher surplus than forming new matches. Without any preference shocks, the magnitude of the surplus gain for previously

formed matches is so large that no observed characteristics, including conflict, can have enough impact to dissolve such matches. The data shows, however, that previously formed matches dissolve one fourth of the time in a given year. This suggests that there must exist some large unobserved downside shock that dissolves previously formed matches. Compared to previously formed matches, most matches not formed in the previous year will not form in a new year, as the cost of forming a match is larges. So there must be some large positive shocks to match the data pattern that one fourth of all matches formed in a given year are new matches. Since the mean value for an existing match is much higher in absolute value than the matching cost, I expect the shocks to have a longer left tail than right tail.

Besides the ratio of match formation and dissolution that helps to identify the error distribution, the amount of conflict in which previously formed matches and new matches are involved in also provides information on error distribution. Imagine the case in which the error distribution is a mixture of two normal distributions, each with mean fixed. The variances of the two normal components determine the distribution of surplus conditional on deterministic characteristics of the matches. The distribution of the match surplus, in turn, determines how many new matches with relatively low value will be crowded out because of conflict, and how many can sustain a relationship with conflict. At the same time, it determines how many previously formed matches crowd out new matches and remain matched without conflict and how many are matched while in conflict with new or previously formed matches. Hence I add two more moments in calibration, the percentage of previously formed matches, and the percentage of new matches involved in a conflicting relationship, to identify the variance of the two distribution components.

To summarize, I use four moments to identify four parameters governing the mixture

of two normal distributions. The moments are, in each period, the percentage of previously formed matches that form again, the percentage of new matches among the currently formed matches, the percentage of previously formed matches caught in conflict, and the percentage of new matches caught in conflict. The lower panel of Table 11 shows the distribution of errors. The two normal components have one positive mean and one negative mean, and the negative mean much greater in absolute value than the positive mean. Since the error is centered at zero, the component with the negative mean has a much lower weight. It confirms the intuition that relationships can have very large downside shocks, but with some relatively small probability of occurring.

2.5.3 The Impact of Conflict on Market Equilibrium

In the first counterfactual, I compare the the match allocation with conflict cost as estimated with the match allocation that would arise if I assume the conflict cost is zero. Both are based on market observables and 500 paths of match-specific error draws. The results are shown in Table 12.

Taking the scenario without conflict cost as the baseline, the first result of note in Table 12 is that conflict crowds out 14% of matches that would have formed without conflict. Among 876 matches crowded by conflict, 849 are new matches. This is consistent with my earlier analysis that even though the conflict cost for new matches is small, their chances of finding matches are the most affected. As the conceptual analysis illustrates, two types of matches are crowded out. The first type is a match that has a return lower than its conflict cost and cannot survive in a conflicting relationship. The second type is a match that has a high enough return to overcome its own conflict, but cannot compensate

the negative externality it causes in conflict. The majority (75.5%) of crowded out matches belong to the second type. This illustrates the importance of taking into account the negative externality that conflict places on choices.

The crowding out effect also suggests a distributional effect on the market equilibrium. When conflict reduces the overall surplus for an agency, a socially optimal equilibrium dictates that only the match that yields the largest surplus can be kept. In the static framework, this suggests that new drugs or drugs not previously matched will find it harder to match with agencies. In a dynamic framework, however, matches may be formed based on the expected future return. That being said, we may still expect a similar result in a dynamic setting: the drug that is expected to remain a fringe product with lower market share and small marketing budget will find it harder to match with desirable advertising agencies.

The high-return matches are less affected by the conflict. Even in the event of conflict, they can either crowd out other lower return matches, or stay in the conflicting relationship if they are willing to accept a lower surplus due to affected agency performance or to spare enough resources to ensure that the agency firmly maintains information confidentiality. For the second type of matches, even though the drug advertiser knows that forming a conflicting match will lower the productivity of the relationship, it will still find it more profitable to form the match with conflict than to give up the matching opportunity. Comparing to the scenario without conflict, the mean surplus of matches in conflict is reduced by 11.5% due to conflict cost.

2.5.4 The Impact of Merger on Market Equilibrium

My second counterfactual simulates a market structure change resulting from a successful the merger between Publicis and Omnicom. The two advertising holding companies proposed to merge in July 2013, but the deal was called off in May 2014. The proposed merger would have affected pharmaceutical clients if Publicis and Omnicom had consolidated some of their health agencies. I consider a hypothetical integration of Publicis' Publicis Healthcare Communications Group(PHCG), and Omnicom's Cline, Davis and Mann (CDM). Because my data ends with 2010, I will use observed characteristics in 2010, and previous match relationship status at the end of 2009 to predict matches affected by the merger. At the end of 2009, both PHCG and CDM are sizable agencies that are available to match in a wide range of markets. PHCG and CDM overlap in 31 treatment classes overall. In 2009, they actively serve for drug advertisers in 15 of these treatment classes.

Conceptually, a merger has four kinds of effects, each of which affects the direction of surplus changes differently.

The first effect is that the synergy created by mergers helps to increase total surplus. Synergy raises the mean return of all matching opportunities with the merged agencies This creates a positive return for a greater share of matching opportunity and increase the surplus of matches that form.

The second effect runs counter to the first one. Mergers reduce the number of advertising agencies from which a drug marketer can choose and reduce the expected number of matches in the market. This is similar to the concept of welfare loss when product variety is reduces in a market with differentiated goods but differs in that drugs' demand

for advertising agencies is not restricted to one unit. Because of the limited number of agencies working within a treatment class and the even fewer number of matching opportunities with a positive return, the model assumes that a drug marketing team always enjoys variety and would like to match with any agency with which it yields a positive return. With this assumption, when two agencies in a market consolidate into one, the expected number of matches is halved.

The third effect comes from the increasing probability of conflict. I will focus on the analysis in a single treatment market and explain two cases that raise potential conflicts. In the first case, a merger involves one agency that did not match with any drug in the previous period and another agency that matched with one drug in the market. When the agencies merge, not only does the matching opportunity decrease, but the match between the merged agency and the single previously matched drug also crowds out new matches with lower returns. In this case, merger exacerbates the crowding out effect.

In the second case, the merger involves two agencies that were both previously matched with one or more drugs in the market. Because of the value of the past relationship, both drugs brought into the merged agency are very likely to bear the conflict costs and continue the contracts. Because the negative externality only applies to the first marginal match that moves an agency from no conflict to some conflict. any other drug that wishes to match with the merged agency only needs to overcome its own conflict cost. Hence, in this case, the merger actually weakens the crowding out effect.

The last effect is on the previously formed high-return matches that would be brought to the same agency through merger. Whether the matches dissolve depends on the synergy and the variance of the downside shock. However, even when the conflicting matches remain in the same agency, the merged agency's performance is affected by

having to deal with the rising number of conflicting clients.

The simulated hypothetical merger combines the above mechanisms, and Table 13 shows the results. Although the value of merger synergy is equal to 25% of the common matching cost (Table 11), suggesting that mergers significantly improve the matching opportunity for all matches, the number of matches and total surplus are reduced overall. The merger affects both the number and the surplus of equilibrium matches. First, the total number of matches decreases, and all of the decrease comes from a reduced number of new matches. Although only four matches are not formed because of the merger, this represents 10% of the new matches that would form in the absence of the merger. Second, previously formed matches are 68% more likely to be brought into actual conflict after the merger. Although none of these matches will dissolve due to the merger, their surplus is reduced by 6.8%.

PHCG and CDM both have contracts with at least one drug in each of the overlapping markets they serve. When the mergers bring these competing drugs into the same agency, the conflict cost is not large enough to dissolve these high-return matches and therefore increases the number of actual conflicts in the market. After these existing matches pay the big initial cost of conflict, additional conflicts only pay their own conflict cost and hence find it easier to match even when there is conflict. The second panel of Table 13 shows that the number of crowded out matches is actually reduced after the merger. Therefore, this merger reduces the number of matches mainly through disrupting matching opportunities of differentiated services rather than increasing the crowding out effect.

2.6 Conclusion

Using a matching model with conflict, I analyze contract decisions between drug advertisers and advertising agencies and examine the market equilibrium effect of client conflict. The model shows that it is important to take into account the negative externality generated from conflict for both the agents that were matched in the previous period and those that are seeking new partners. Focusing on the outcome of contract allocation, I show that conflict crowds out a significant amount of contracts with lower returns. The consolidation of advertising agencies reduces the number of equilibrium contracts by eliminating differentiated products (agencies). The crowding out effect through merger may increase or decrease depending on the contract allocation at the time of the merger.

Conflict raises new angles in merger analysis. Since I analyze a market equilibrium without pricing inefficiencies, I cannot analyze advertising agencies' market power beyond the above discussions. But the implication for merger analysis of the existence of conflict is clear and is easily extended to a wide range of professional service markets. Conflict reduces the alternatives for clients, and when service agencies merge, it further reduces substitution among service agencies. This limitation to the substitution pattern can be especially acute in professional service industries since the number of service agencies working for one type of client is usually limited due to the client's demand for expertise. And because experience is highly valued in service agencies, the entry barrier in these markets is also considerable. This suggests that a more thorough analysis is necessary to take into account conflict, market segmentation due to the expertise, and entry barriers when conducting merger analysis.

This study bears several limitations due to data constraints. First, the assumption

that the capacity constraint does not bind for drug advertisers is not directly testable but is necessary because I need a characterization of the equilibrium using only observed match allocations. Even with stability, conflict is not identified in canonical matching models that require knowledge of both transfer and matching allocation in characterizing the equilibrium. Second, due to the importance of sunk cost and the crowding out effect created by conflict, it would be interesting to investigate dynamic incentives and pricing powers in dynamic games. Without transfer data, however, this work cannot be undertaken empirically, but theoretical analysis would be the first step toward drawing a complete picture in a dynamic framework. Third, the study takes the market structure as given and examines the effect of conflict on equilibrium matches due to the limited time span of the observed data. With historical data, the exact entry date of advertising agencies into each therapeutic market may be observed. Endogenizing entry and merger decision in the model and analyzing the impact of conflict on market structure will be interesting directions for future research.

2.7 Tables

Table 1: Statistics on Matching Markets

	Mean	StdDev	25th	Median	75th	N
Summary of # of Drug and Agency per Class-year[1]						
# of Drugs	4.4	11.1	1	2	6	1,214
# Drugs Matched	2.8	3.3	1	1	4	1,214
% of Drugs Matched	68%	36%	50%	76%	100%	1,214
# of Agencies	7.8	7.4	2	5	11	1,214
# of Agencies Matched	3.7	4.1	1	2	5	1,214
% of Agencies Matched	46%	29%	50%	28%	60%	1,214
Summary of # of Matches per Class-year						
# of Matches	4.5	5.6	1	2	6	1,214
# of Matches per Drug	1.1	1.1	0	1	2	5,012
# of Matches per Matched Drug	1.6	0.9	1	1	2	3,703
# of Matches per Agency	0.6	0.7	0	0	1	9,470
# of Matches per Matched Agency	1	0.6	1	1	1	4,995
Overview of Amount of Conflicts per Class-year[2]						
# of Class-year with Conflicts	322					
# of Matches in Conflicts	2.5	3.9	0	2	4	657
% Matches in Conflicts	23%	27%	0	13%	43%	657
# of Drugs in Conflicts	2.1	3.0	0	0	3	657
% of Drugs in Conflicts	23%	28%	0%	0%	40%	657
# of Agencies in Conflicts	1.1	1.5	0	0	2	657
% of Agencies in Conflicts	7%	11%	0%	0%	12.5%	657
# Drugs per Conflict Relationship[3]	2.3	0.5	2	2	2.5	657

[1] There are 158 unique treatment classes, and there are a total of 1,214 class-year (markets) in our analysis between 2002 and 2010. The number of classes is unbalanced across years due to new drugs and classes with all agencies exiting the market. [2] Among 1,214 class-year, there are 557 classes have only one drug in a given year, and there is no conflict in such classes. In producing statistics for conflict, I only focus on 657 class-year with two or more drugs in a given class-year. There are 322 class-year with actual conflicts. [3] A "conflict relationship" is one among multiple drugs in a class-year that hire the same agency, or one that an agency serves multiple drugs in a same class-year.

Table 2: Turnover of Matches

Year	# of Matches	# of New Matches	New%	# End Matches[1]	End%
2002	586	225	38.4%	150	25.6%
2003	651	187	28.7%	195	30.0%
2004	635	179	28.2%	154	24.3%
2005	715	227	31.7%	176	24.6%
2006	779	225	28.9%	248	31.8%
2007	731	202	27.6%	208	28.5%
2008	734	184	25.1%	204	27.8%
2009	673	148	22.0%	171	25.4%
2010	631	128	20.3%	.	.

[1] # of end matches is the number of matches that are in their the last year of the relationship.

Note: This table pools all matches together regardless of markets to examine the turnover rate of matches overall.

Table 3: Summary Statistics of Advertising Agency and Drug Characteristics

	Matched[1]		Unmatched[1]	
Variables	Mean	StdDev	Mean	StdDev
Advertising Agency Characteristics				
Billings(millions)	475	999.0	418	944.0
#Employees	342.2	667.4	338.7	739.1
Agency Age in # of Years	29.2	20.1	28.2	20.2
TV Specialty	0.2	0.3	0.2	0.3
DTC Specialty	0.2	0.3	0.2	0.3
Health Specialty	0.7	0.4	0.7	0.4
Drug Characteristics				
DTCA Expenditure (million)	9.4	29.4	0.23	3.3
DTCA TV Expenditure (million)	5.6	20	0.09	2.1
I(Top 200 Sales)	0.2	0.4	0.1	0.3
$I(DrugAge < 0)$	0.1	0.3	0.4	0.5
$I(0 \leq DrugAge \leq 5)$	0.4	0.5	0.2	0.4
$I(6 \leq DrugAge \leq 12)$	0.3	0.5	0.2	0.4
$I(DrugAge \geq 13)$	0.2	0.4	0.2	0.4

[1] The "matched" column summarizes characteristics of drugs and agencies that obtain any match in a class-year. And the "unmatched" column summarizes characteristics of drugs and agencies that does not obtain any match in a class-year. Among all agency-class-year, 46.6% are matched. Among all drug-(class)-year 66.3% are matched.

Note: 1 The table summarize characteristics of all drugs and agencies in the market between 2002 and 2010. The unit of observation is a drug-year, or an agency-class-year.

Table 4: Probability of Matching for A Drug

	$I(Match_{it} > 0)$		$I(Match_{it} \geq 2 \mid I(Match_{it} > 0)$	
	(1)	(2)	(3)	(4)
$Ln(DTCA_t)$	0.0167***	0.0116***	0.00763***	0.00414***
	(0.00107)	(0.00119)	(0.00152)	(0.00158)
$I(DTCATV_t)$	0.0368**	0.0210	0.173***	0.0550
	(0.0162)	(0.0233)	(0.0294)	(0.0359)
$I(0 \leq DrugAge_t \leq 5)$	0.326***	0.336***	0.0951***	0.155***
	(0.0184)	(0.0205)	(0.0258)	(0.0281)
$I(6 \leq DrugAge_t \leq 12)$	0.250***	0.182***	0.113***	0.116***
	(0.0194)	(0.0265)	(0.0268)	(0.0377)
$I(DrugAge_t \geq 13)$	0.114***	-0.0419	0.0208	-0.0258
	(0.0218)	(0.0348)	(0.0295)	(0.0508)
$I(Top200Sales_{t-1})$	0.00731	0.0516**	0.0708***	-0.00926
	(0.0150)	(0.0255)	(0.0208)	(0.0362)
$Constant$	0.378***	0.439***	0.243***	0.280***
	(0.0146)	(0.0190)	(0.0211)	(0.0276)
FE	–	Drug	–	Drug
Observations	5,584	5,584	3,703	3,703
R^2	0.1745	0.4616	0.0687	0.3573

Note: 1 Specification in Column (1) and (2) regress the indicator of if a drug is matched with any agency at a given year on drug characteristics. Column (3) and (4) condition on the sample of drugs in years with any match and regress the indicator of if a drug has more than one match on drug characteristics. 2 Categorical dummy variable $I(DrugAge < 0)$ is excluded in the regression. 3 Standard errors are clustered at the drug level.

Table 5: Probability of Matching for An Agency

	$I(Match_{jt} > 0)$		$I(Match_{jt} \geq 2 \mid I(Match_{jt} > 0)$	
	(1)	(2)	(3)	(4)
$AgencyAge_t$	0.000457	0.0189***	0.00119***	0.00833***
	(0.000416)	(0.00262)	(0.000318)	(0.00319)
$Ln(Billing_{t-1})$	0.0174***	0.0483***	0.0138***	-0.00430
	(0.00521)	(0.0112)	(0.00382)	(0.0121)
$Ln(Emp_{t-1})$	0.0328***	0.0117	0.00915*	0.0331**
	(0.00697)	(0.0164)	(0.00539)	(0.0164)
$DTC\ Specialty_{t-1}$	-0.0621***	-0.209***	0.0108	-0.0520
	(0.0204)	(0.0534)	(0.0149)	(0.0433)
$TV\ Specialty_{t-1}$	-0.0408*	-0.0723	-0.0195	-0.0770
	(0.0209)	(0.0513)	(0.0155)	(0.0525)
$Health\ Specialty_{t-1}$	0.0976***	0.175***	0.0327**	0.0116
	(0.0182)	(0.0460)	(0.0134)	(0.0438)
$Constant$	0.504***	-0.0390	0.0949***	-0.0722
	(0.0196)	(0.0803)	(0.0147)	(0.0968)
FE	–	Agency	–	Agency
Observations	10,716	10,716	4,993	4,993
R^2	0.0098	0.4722	0.0162	0.6152

Note: 1 Column (1) and (2) specifications regress the indicator of if an advertising agency is matched with any drug at a given year on agency characteristics. Column (3) and (4) condition on the sample of agencies in years with any match and regress the indicator of if an agency has more than one match on agency characteristics. 2 Standard errors are clustered at the agency level.

Table 6: Types of Pairwise Deviation

Parameter of Surplus function:

$$S_{ij} = X_i\alpha + Z_j\beta + X_iZ_j\gamma + \lambda L_{ij} - 1(|\mu_j| \geq 2) \times \kappa_c - \kappa_f + \epsilon_{ij}$$

Type	**Identified Parameters**	κ_c,κ_f
Replacing partners		
1 Swapping partners between two pair	γ, λ	/
2 Drug replacing an agency partner	β, γ, λ	κ_c
3 Agency replacing a drug partner	α, γ, λ	–
Adding new matches		
4 A new match added between a drug and an agency that already have matches with others	α, β, γ, λ	κ_c,κ_f
5 A new match added between a unmatched drug and an agency that already have matches with others	α, β, γ, λ	κ_c, κ_f
6 A new match added between an unmatched agency and a drug that already have matches with others	α, β, γ, λ	κ_f
7 A new match added between an unmatched agency and an unmatched drug	α, β, γ, λ	κ_f
Drop a current match		
8 A current match is dropped	α, β, γ, λ	κ_c, κ_f

Table 7: Descriptive Regression: Probability of a Forming Match

Dependent Variable:	$I(Match_{ijt}=1)$	
	(1)	(2)
$PotentialConflict_{ij,t-1}$	-0.0295***	-0.0144***
	(0.00191)	(0.00144)
$I(Match_{ij,t-1}=1)$		0.745***
		(0.00253)
$Ln(Emp_{t-1})$	0.00495***	0.00198***
	(0.000977)	(0.000700)
$Ln(Billing_{t-1})$	0.000486	-0.000503
	(0.000755)	(0.000561)
$AgencyAge_t$	-0.0000654	-0.000128***
	(0.0000590)	(0.0000412)
$DTC\ Specialty_{t-1}$	-0.00881***	-0.00412**
	(0.00279)	(0.00205)
$TV\ Specialty_{t-1}$	-0.0122***	-0.00314
	(0.00294)	(0.00220)
$Health\ Specialty_{t-1}$	0.0157***	0.00202
	(0.00250)	(0.00182)
$AnyMerge_j$	0.0331***	0.00122
	(0.00260)	(0.00170)
$PostMerger_{jt}$	0.0272***	0.00677***
	(0.00254)	(0.00228)
$I(0 \leq DrugAge_t \leq 5)$	0.00458*	-0.000508
	(0.00263)	(0.00201)
$I(6 \leq DrugAge_t \leq 12)$	-0.00244	-0.00467**
	(0.00260)	(0.00202)
$I(DrugAge_t \geq 13)$	0.0305***	-0.0132***
	(0.00242)	(0.00219)
$I(Top200Sales_{t-1})$	0.0292***	-0.00418**
	(0.00397)	(0.00168)
$I(DTCA_t > 0)$	0.0159***	0.0107***
	(0.00253)	(0.00156)
$I(DTCATV_t > 0)$	0.0240***	0.00639***
	(0.00369)	(0.00222)
$Constant$	0.0301	0.0269***
	(0.00323)	(0.00249)
Observations	76,887	76,887
R^2	0.0185	0.5389
$Mean(I(Match_{ijt}=1))$	7.13%	

Note: 1 This regression pools observations of all possible combination of pairwise matches in each class-year and examine what characteristics predict current period match. 2 Potential conflict is defined by if a match between a drug and an agency involves an agency that also contracts with the drug's competitor. Since current period matches are interdependent, we use potential conflict in the past year instead. Potential conflict in the past year is defined as for a given drug-agency pair, if the agency contracted with any competitors of the drug in the past year. 3 The probability of any pair to form a match 7%, so the potential conflict and potential conflict reduces the chance of match 20-40% of the time.

Table 8: Descriptive Regression: Heterogeneous Effect of Previously Matching Relationships

Dependent Variable:	$I(Match_{ijt} > 0)$		
$PotentialConflict_{t-1}$	-0.0133***	$Match_{t-1}$	0.928***
	(0.00123)		(0.0279)
$Ln(Emp_{t-1})$	-0.000385	$Match_{t-1} \times Ln(Emp_{t-1})$	0.0344***
	(0.000547)		(0.00689)
$Ln(Billing_{t-1})$	-0.00000205	$Match_{t-1} \times Ln(Billing_{t-1})$	-0.0114**
	(0.000433)		(0.00551)
$AgencyAge_t$	-0.0000318	$Match_{t-1} \times AgencyAge_t$	-0.00148***
	(0.0000323)		(0.000386)
$DTC\ Specialty_{t-1}$	-0.00158	$Match_{t-1} \times DTC\ Specialty_{t-1}$	-0.0329
	(0.00155)		(0.0206)
$TV\ Specialty_{t-1}$	-0.00211	$Match_{t-1} \times TV\ Specialty_{t-1}$	-0.0351
	(0.00167)		(0.0232)
$Health\ Specialty_{t-1}$	0.00335**	$Match_{t-1} \times Health\ Specialty_{t-1}$	-0.0160
	(0.00145)		(0.0178)
$AnyMerge_j$	0.00522***	$Match_{t-1} \times AnyMerge_j$	-0.0482***
	(0.00148)		(0.0161)
$PostMerger_{jt}$	0.00600***	$Match_{t-1} \times PostMerger_{jt}$	0.0137
	(0.00213)		(0.0194)
$I(0 \leq DrugAge_t \leq 5)$	0.00350**	$Match_{t-1} \times I(0 \leq DrugAge_t \leq 5)$	-0.128***
	(0.00169)		(0.0235)
$I(6 \leq DrugAge_t \leq 12)$	-0.000873	$Match_{t-1} \times I(6 \leq DrugAge_t \leq 12)$	-0.128***
	(0.00160)		(0.0238)
$I(DrugAge_t \geq 13)$	-0.00708***	$Match_{t-1} \times I(DrugAge_t \geq 13)$	-0.166***
	(0.00162)		(0.0265)
$I(Top200Sales_{t-1})$	-0.00299**	$Match_{t-1} \times I(Top200_{t-1})$	-0.0129
	(0.00138)		(0.0141)
$I(DTCA_t > 0)$	0.00704***	$Match_{t-1} \times I(DTCA_t > 0)$	0.0519***
	(0.00137)		(0.0139)
$I(DTCATV_t > 0)$	0.00194	$Match_{t-1} \times I(DTCATV_t > 0)$	0.0340**
	(0.00212)		(0.0157)
$Constant$	0.0187***		
	(0.00197)		
Observations	76,887	$Mean(I(Match_{ijt} = 1))$	7.13%
R	0.5417		

Note: 1 This regression adds the interaction terms between indicator of matches previously formed and characteristics of advertising agencies and drugs. The regression is trying to produce suggestive evidences for what pervious formed matches are more likely to be continued. 2 This regression pools observations of all possible combination of pairwise matches in each class-year and examine what characteristics predict current period match. 3 Potential conflict is defined by if a match between a drug and an agency involves an agency that also contracts with drug's competitor. Since current period matches are interdependent, we use potential conflict in the past year instead. Potential conflict in the past year is defined as for a given drug-agency pair, if the agency contracted with any of the drug's competitors in the past year.

Table 9: Baseline Maximum Score Estimates

	(1)		(2)	
Parameters	L.B.	U.B.	L.B.	U.B.
$\kappa_c(Conflict)$	0.00235	0.00679	0.00026	0.03238
$\kappa_c(Conflict) \times Match_{t-1}$	0.17793	0.32090	0.07360	0.22472
$\kappa_f(FC)$	0.18334	0.24986	0.11156	0.54223
$\kappa_f \times \Delta DrugMatch$[1]			-0.00034	0.00000
Drug Characteristics				
$I(0 \leq DrugAge_t \leq 5)$	0.00000	0.00001	0.00000	0.00001
$I(6 \leq DrugAge_t \leq 12)$	-0.00137	0.00197	-0.00101	-0.00000
$I(DrugAge_t \geq 13)$	-0.01381	-0.00656	-0.26384	-0.00022
$I(Top200Sales_{t-1})$	-0.00073	-0.00046	-0.00117	0.00070
$I(DTCA_t > 0)$	0.00386	0.00609	0.00057	0.30943
$I(DTCATV_t > 0)$	-0.00046	-0.00020	-0.00101	0.30892
Agency Characteristics				
$AgencyAge_t$	-0.00019	-0.00009	-0.00909	0.00031
$Ln(Billing_{t-1})$	-0.01206	-0.00389	-0.02084	-0.00002
$Ln(Emp_{t-1})$	-0.00423	-0.00120	-0.02943	0.01094
$DTC\ Specialty_{t-1}$	-0.00348	0.00010	-0.04255	0.00084
$TV\ Specialty_{t-1}$	0.00001	0.00005	-0.13799	0.00906
$Health\ Specialty_{t-1}$	-0.00340	0.00420	-0.00166	0.01297
$AnyMerge_j$	0.00987	0.01209	0.00456	0.13832
$PostMerger_{jt}$	0.00790	0.16711	0.00000	0.14076
$Match_{t-1}$ *bounds implied*	0.69894	0.94476	0.75005	0.86697
OptObj		201,312		201,695
Total Ineq		221,044		221,044
%Correct		91.07%		91.25%

[1] $\kappa_f \times \Delta DrugMatch$ measures value(cost) of an drug to add an additional match.

Note: 1 This table presents the main specifications of maximum score estimator. Specification (2) differs from (1) in allowing the matching cost to change with the number of agencies a drug hires. 2 L.B. and U.B. represents lower and upper bounds of the 95th percentile confidence region. It is derived using subsampling method. 3 Parameters are normalized by $||\theta|| = 1$, or that square of coefficients sum to one. $Match_{t-1}$ is excluded for normalizing the sum of squares. Inference for freely varying coefficients is directly derived from the subsampling method. For illustration, bounds for the excluded coefficient (impact of previously formed matches) shown above are $[(1 - \tilde{\theta}_{lb})^{0.5}, (1 - \tilde{\theta}_{ub})^{0.5}]$ where $\tilde{\theta}$ represents all coefficients other than the excluded one.

Table 10: Maximum Score Estimates: Controlling for Heterogeneous Effect of $Match_{t-1}$

Parameters	**Lower Bound**	**Upper Bound**
$Match_{t-1}$	0.8988	0.9910
$Conflict\ Cost$	0.0006	0.0099
$Conflict\ Cost \times Match_{t-1}$	0.0898	0.3903
$Match\ Cost$	0.0989	0.1418
$AnyMerger$	0.0086	0.0462
$PostMerger$	0.0019	0.0932
Other Controls		
Drug Characteristics	Yes	
Agency Characteristics	Yes	
Het. Past Match Effect	Yes	
OptObj	201,395	
Total Ineq	221,044	
% Correct Inequalities	91.20%	

Note: This table reports key parameters in the specification that includes interactions of indicator of matches previously formed and drug and agency characteristics. The coefficients for the interaction terms are all insignificant.

Table 11: Parameters in Counterfactual Simulations

Parameters	**Value**	**Parameters, cont.**	**Value**
$Match_{t-1}$	0.8793	$AgencyAge_t$	-0.0011
$\kappa_c(Conflict)$	0.0064	$Ln(Billing_{t-1})$	-0.0038
$\kappa_c(Conflict) \times Match_{t-1}$	0.2465	$Ln(Emp_{t-1})$	-0.0050
$\kappa_f(FC)$	0.3319	$DTC\ Specialty_{t-1}$	-0.0057
$I(0 \leq DrugAge_t \leq 5)$	0.0332	$TV\ Specialty_{t-1}$	-0.0119
$I(6 \leq DrugAge_t \leq 12)$	-0.0016	$Health\ Specialty_{t-1}$	0.0004
$I(DrugAge_t \geq 13)$	-0.2105	$AnyMerge_j$	0.0103
$I(Top200Sales_{t-1})$	-0.0013	$PostMerger_{jt}$	0.0877
$I(DTCA_t > 0)$	0.0364		
$I(DTCATV_t > 0)$	0.0328		

Calibrated Error Distribution Parameters

$F(\epsilon) = \omega_1\Phi_1(x) + \omega_2\Phi_2(x)$

$\Phi_1 \sim N(-0.7383, 0.0563)$

$\Phi_2 \sim N(0.1819, 0.0024)$

$(\omega_1, \omega_2) = (0.1977.0, 8023)$

Note: 1 The parameter in the upper panel is the midpoint of each parameter range in its 95% confidence region according to results of the first column of Table 9. 2 The lower panel is the error distribution that I calibrate 4 parameters of the distribution using four data moments: in each period, the percentage of matches dissolving, the percentage of matches forming, the percentage of previously formed matches in conflict, and the percentage of new matches in conflict.

Table 12: Counterfactual: The Impact of Conflict on Equilibrium

	Costly-Conflict[1]	**Costless-Conflict**[1]	**Δ(Costly-Costless)%**
Total # of matches	5193.05	6068.99	-14.4%
Total surplus	2877.05	3205.59	-10.2%
Matches with conflict[2]			
Number	1765.91	3106.76	-43.1%
Mean surplus	0.33	0.37	-11.5%
Matches without conflict			
Number	3427.14	2962.23	15.6%
Mean surplus	0.67	0.69	-3.1%
Matches continued from $t-1$			
Number	4220.52	4247.90	-0.6%
Mean surplus	0.68	0.75	-9.2%
Matches new at t			
Number	972.53	1821.09	-46.6%
Mean surplus	0.01	0.02	-13.8%
Matches continued from $t-1$ & with conflict			
Number	1118.84	1481.06	-24.5%
Matches new at t & with conflict			
Number	647.07	1625.70	-60.2%
Crowded Out Matches[3] $(r_{ij} = y_{ij} - \kappa_f + \epsilon_{ij})$			
$0 < r_{ij} < \kappa_{c,ij}$			
Number	214.27	0	-
Mean surplus	0.01	-	-
$r_{ij} > \kappa_{c,ij}$			
Number	661.67	0	-
Mean surplus	0.02	-	-

[1] I simulate two scenarios using characteristics of the agents and their matching history as observed in the data. The column "Costly-Conflict" assumes that there is an additional cost for an advertising agency to serve two competing drug clients simultaneously. And the column "Costless-Conflict" assumes that such additional cost is zero. [2] "with conflict" describes the situation when two or more drugs in the same class match with the same advertising agency in the same year. [3] r_{ij} denotes the net return of a match between drug i and advertiser j without taking into account the cost of conflict. In the r_{ij} expression, y_{ij} is the return of a pair of match, κ_f is the fixed cost of matching, common to all matches. $\kappa_{c,ij}$ is the cost of conflict incurred to the match itself, and ϵ_{ij} is the random shock of the match surplus.

Table 13: Counterfactual: The Impact of Mergers on Equilibrium

	With-Merger[1]	**Without-Merger**[1]	**Δ(With-Without)%**
Total # of matches	84.352	88.164	-4.3%
Total surplus	31.3877	33.7157	-6.9%
Matches with conflict[2]			
Number	71.39	54.428	31.2%
Mean surplus	0.3291	0.2415	36.3%
Matches without conflict[2]			
Number	12.962	33.736	-61.6%
Mean surplus	0.6105	0.6107	0.0%
Matches continued from $t-1$			
Number	50.92	50.918	0.0%
Mean surplus	0.6089	0.6533	-6.8%
Matches new at t			
Number	33.432	37.246	-10.2%
Mean surplus	0.0119	0.0125	-4.8%
Matches continued from $t-1$ & with conflict			
Number	40.982	24.33	68.4%
Matches new at t & with conflict			
Number	30.408	30.098	1.0%
Crowded Out Matches[3] ($r_{ij} = y_{ij} - \kappa_f + \epsilon_{ij}$)			
$0 < r_{ij} < \kappa_{c,ij}$			
Number	2.376	1.478	60.8%
Mean surplus	0.0275	0.0144	91.0%
$r_{ij} > \kappa_{c,ij}$			
Number	15.69	24.526	-36.0%
Mean surplus	0.0167	0.0156	7.1%

[1] I simulate two scenarios using characteristics of the agents and their matching history as observed in the data. The column "With-Merger" simulates matches that will form in 2010 if PHCG and CDM merge. And the column "Without-Merger" simulates matches that will form in 2010 if PHCG and CDM do not merge. [2] "with conflict" describes the situation when two or more drugs in the same class match with the same advertising agency in the same year. [3] r_{ij} denotes the net return of a match between drug i and advertiser j without taking into account the cost of conflict. In the r_{ij} expression, y_{ij} is the return of a pair of match, κ_f is the fixed cost of matching, common to all matches. $\kappa_{c,ij}$ is the cost of conflict incurred to the match itself, and ϵ_{ij} is the random shock of the match surplus.

Chapter 3 – Consumer Protection or Consumer Frustration? The Impact of Banning Foreign Pharmacies from Sponsored Search

3.1 Introduction

The Internet has led to a dramatic increase in the number of retailers available to consumers in many industries. The proliferation of competition may benefit consumers in several ways including lower prices. However, there is also the concern that the quality of the new product offerings may be lower, though difficult to discern by consumers. The concern is particularly acute for online prescription drugs, a market where poor product quality may lead to adverse health outcomes.

The high price of brand name prescription drugs has motivated US consumers to search for cheaper supplies from foreign pharmacies, despite the fact that personal importation is illegal. The Federal Food, Drug, and Cosmetic Act (FD&C Act) prohibits the importation of unapproved drugs into the US.[33] In particular, section 355(a) states: "No person shall introduce or deliver for introduction into interstate commerce any new drug, unless an approval of an application ... is effective with respect to such drug."[34] The FDA further states that interstate shipment includes importation and the FD&C Act applies to "any drugs, including foreign-made versions of U.S. approved drugs, that have not received FDA approval to demonstrate they meet the federal requirements for safety and effectiveness."[35]

[33] See http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct.

[34] See http://www.gpo.gov/fdsys/pkg/USCODE-2010-title21/pdf/USCODE-2010-title21-chap9-subchapV-partA-sec355.pdf.

[35] See http://www.fda.gov/ForIndustry/ImportProgram/ucm173743.htm.

Based on data from IMS Health, Skinner (2006) estimated that sales to US consumers from 278 confirmed or suspected Canadian-based Internet pharmacies reached CDN$507 million in the 12 month periods ending June 2005.[36] More than half of the sales were on top-selling brand-name prescription drugs consumed primarily by seniors. According to Skinner (2005), Canadian prices for the 100 top-selling brand-name drugs were on average 43% below US prices for the same drugs. Consistently, Quon et al. (2005) compared 12 Canadian Internet pharmacies with 3 major online US drug chain pharmacies and found that Americans can save an average of approximately 24% per unit of drug if they purchase the 44 most-commonly purchased brand-name medications from Canada. The large price difference between US and Canada has motivated not only individual Americans to order brand name prescription drugs from foreign pharmacies but also a large number of bills introduced by state or federal legislators in favor of legalizing or facilitating the cross-border drug trade with Canada.[37]

While drug sales from foreign pharmacies have been growing, the National Association of Boards of Pharmacy (NABP) emphasizes the illegality of buying foreign drugs and highlights the danger of rogue pharmacies. In particular, NABP (2011) reviewed 7,430 Internet pharmacies as of December 2010 and found 96.02% of them operating out of compliance with US state and federal laws and/or NABP patient safety and pharmacy practice standards. Among these non-NABP-recommended pharmacies, 2,429 (34%) had server locations in a foreign country, 1,944 (27%) had a physical address out of US, 4,005 (56%) did not provide any physical address, 5,982 (84%) did not require a valid prescription, 4,397 (62%) issued prescriptions via online consultation, 3,210 (50%) offered foreign

[36] This number was measured in standardized manufacturer-level prices and did not include "foot traffic" sales to US consumers through regular "brick-and-mortar" border pharmacies in Canada. Sales measured by final retail prices to US customers was not available but is certainly higher than CDN$507.

[37] According to Skinner (2006), the number of state and federal bills on this topic increased from 3 in 2002 to 84 in 2005.

or non-FDA-approved drugs, 5,928 (83%) did not offer medical consultation, and 1,129 (16%) did not have secure sites. Independent research, mostly from medical researchers rather than economists, confirmed some of the NABP concerns, although the data gathered for these studies were often of a much smaller sample size. In particular, Orizio et al. (2011) reviewed 193 articles about Internet pharmacies, of which 76 were based on original data. The articles with original data suggested that geographic characteristics were concealed in many websites, at least some websites sold drugs without a prescription and an online questionnaire was a frequent tool used to replace a prescription. On drug quality, researchers often found inappropriate packaging and labeling, however, the chemical composition was found to differ from what is ordered in only a minority of studied samples.

Internet search engines, such as Google, are one avenue consumers use to reach Internet pharmacies. Upon submitting a query, a user is presented with two types of results. The first are organic results whose ranks are solely a function of search engine's relevance algorithm. The second type are called paid or sponsored links, which appear based on both the relevance of the link to the query and a monetary bid placed by the owner of the link. If the user clicks on a sponsored link, the link owner pays the search engine their bid. An example of a Google search results page is shown in Figure 1. Because of heightened concern to protect consumers, Google agreed to ban non-NABP-certified pharmacies from their sponsored search listings in February 2010. Eighteen months later (August 24, 2011), Google settled with the US Department of Justice (DOJ) by "forfeiting $500 million generated by online ads & prescription sales by Canadian online pharmacies."[38]

[38]http://www.justice.gov/opa/pr/2011/August/11-dag-1078.html, retrieved December 28, 2013.

At first glance, the ban is a form of a minimum quality standard. Both Leland (1979) and Shapiro (1986) showed that a minimum quality standard (and its variant forms such as occupational licensing) can eliminate poor quality products, encourage high quality sellers to enter the market, and expand consumer demand because consumers anticipate higher quality under the regulation. These effects tend to benefit consumers who appreciate high quality. However, a minimum quality standard can also increase barriers to entry and reduce competition (Stigler (1971), Peltzman (1976)). Even if the standard improves average quality on the market, it raises the market price and potentially hurts price-sensitive consumers by denying them access to low quality products. If the minimum quality standard is set by the industry, the harm can be even greater as the industry has incentives to set too high a standard in order to reduce competition (Leland, 1979).

A number of empirical studies have attempted to test the theory of minimum quality standards by examining price, quantity, quality, and market structure, but all of them assumed that the standard is well enforced in reality.[39] This assumption does not hold for online pharmacies: after the ban, consumers can still access non-NABP-certified pharmacies through organic search.[40] Moreover, the ban contains only limited safety information on specific online pharmacies and consumers can gather safety information through other

[39]Law and Kim (2005) explored the effects of occupational licensing in the Progressive Era and showed that the licensing regulation had improved markets when consumers faced increasing difficulty in judging the quality of professional services. Law and Marks (2009) examined the introduction of state-level licensing regulation during the late nineteenth and mid-twentieth centuries and found that licensing laws often helped female and black workers, particularly in occupations where worker quality was hard to ascertain. On the negative side, Pashigian (1979) reported that state-specific occupational licensing had a quantitatively large effect in reducing the interstate mobility of professionals; Shepard (1978) estimated that the price of dental services and mean dentist income were between 12 and 15 percent higher in non-reciprocity jurisdictions when other factors are accounted for; Adams et al. (2003) compared state-by-state regulation on midwifery licensing and found that more stringent licensing regulation led to fewer births by midwifery, which led them to conclude that licensing regulation had a detrimental effect by restricting entry and competition.

[40]Organic search refers to links returned by a search engine due to their relevance to the search terms and not due to an advertising campaign by the link owner. In contrast, paid or sponsored search refers to links returned by a search engine as a result of both relevance to the search terms and advertising.

channels. Other channels of information includes consumer experience, word of mouth, and alternative certification agencies. Specifically, Google used a private certification agency – PharmacyChecker.com – to filter rogue pharmacies before the ban. This abandoned practice is more lenient than the ban because PharmacyChecker certifies both US and foreign pharmacies while NABP automatically disqualifies any foreign pharmacies.[41] Even after the ban, Google uses the Canadian Internet Pharmacy Association (CIPA) to screen sponsored ads that target Canadian consumers, but the CIPA-certified pharmacies are not NABP-certified for US customers because they are foreign. According to Leland (1979) and Shapiro (1986), one welfare loss from a minimum quality standard is the denial of low quality products to price-sensitive consumers. With organic links and alternative information channels, this denial is likely incomplete for online pharmacies, which offers us an excellent opportunity to study how pharmacies comply with the minimum quality standard coexist or even compete with non-NABP-certified pharmacies. A few papers have examined the effect of tighter law enforcement restricting illicit drugs such as heroin and cocaine, but all of them focus on price, production, or crime rather than search activities on the consumer side.[42] Our results on consumer search will shed new light on the interaction between two competing marketplaces: one legal (NABP-certified pharmacies) and one illegal (non-NABP-certified pharmacies).

[41] In this sense, Google adoption of the NABP standard is similar to a switch from certification to a minimum quality standard, on which Shapiro (1986) argued that certification can be more welfare-improving because it allows the whole spectrum of quality to be known and available to consumers.

[42] Via a theoretical model, Becker et al. (2006) showed that optimal enforcement on illegal drug suppliers depend on demand and supply elasticities. When demand and supply are not too elastic, it does not pay to enforce any prohibition unless the social value of drug consumption is negative. Dobkin & Nicosia (2009) examined a large and abrupt government intervention in the supply of methamphetamine. They found that the intervention had a large effect in increasing the price of methamphetamine sold illegally, in reducing related hospital and treatment admissions, and in reducing arrests related to methamphetamine use; but all these effects were temporary. Miron (2003) also found that cocaine and heroin were substantially more expensive than they would be in a legalized market. Looking at the problem in an opposite direction, Chaudhuri et al. (2003) examined how the WTO enforcement of pharmaceutical patents would affect the Indian market of Quinolones. They estimated that the withdrawal of all domestic products in this segment was associated with substantial welfare losses to the Indian economy, even in the presence of price regulation and the overwhelming portion of this welfare loss derived from the loss of consumer welfare.

How easy is it to switch to organic links when sponsored links of the same website are no longer available? A rising literature has shown that sponsored links accounted for 15% of all clicks (Jansen et al., 2007), consumers had a preference against sponsored links (Jansen and Resnick 2006), consumers appreciated sponsored links as advertisements if they were relevant (Jansen et al., 2007), and organic and sponsored links from the same website of a national retailer were complements in consumer clicks (S. Yang & Ghose, 2010). Two studies released by Google painted a somewhat different picture. Chan, et al. (2012) found that 81% of sponsored impressions and 66% of sponsored clicks occurred in the absence of an associated organic link on the first page of search results. This suggests that most sponsored links are from websites that are not easy to find in organic search. Chan et al. (2012) examined 446 incidences where sponsored ads were paused between October, 2010 to March, 2011. From these incidences, they found that 89% of the traffic generated by sponsored ads was not replaced by organic clicks (leading to the same destination website) when the ads were paused. This suggests that organic and sponsored traffic are not necessarily substitutes. If many non-NABP-certified pharmacies do not appear in high ranked organic results, the ban of their appearance in sponsored listings could be an effective tool to minimize consumer clicks on them in organic search.

It is worth noting that the organic-sponsored substitution is not necessarily the only margin for the ban to take effect. The ban could have other market-wide effects depending on how consumers digest the information conveyed by the ban. Apparently, the ban tells consumers that NABP-certified pharmacies are believed to be safer than non-NABP-certified pharmacies, and this message should be more salient after the Google-DOJ settlement. However, the ban may also send an indirect message about the overall danger of the online prescription drug market, or inform consumers that some alternative

and potentially cheaper pharmacies exist although they are not allowed to advertise in sponsored search. Moreover, the ban groups all other certified pharmacies with uncertified pharmacies, making it more difficult for consumers to differentiate quality among the non-NBAP-certified websites. These economic forces, as well as the technical difficulty of substituting sponsored clicks for organic clicks, may affect consumer search in different directions. This leaves the net effect and the source of the net effect an empirical question.

Overall, the goal of this study is to examine how consumer search on the Internet changes after the ban of non-NABP-certified pharmacies from sponsored advertising. In particular, we classify pharmacy sites into three tiers: NABP-certified (tier-A), other-certified (tier-B), and uncertified (tier-C). NABP-certified sites refer to US pharmacies that receive approval from NABP or the NABP-endorsed certifier, LegitScript.[43] By search engines' policy, they are free to advertise in sponsored search listings before and after the ban. Other-certified sites refer to foreign or domestic pharmacies that are certified by PharmacyChecker.com or CIPA, but not by NABP or LegitScript. All the rest are classified as uncertified sites. Although both other-certified and uncertified sites are banned from Google's sponsored search after February 2010, we distinguish them for two reasons: first, uncertified sites were prohibited from sponsored listings even before the ban, but the screening was imperfect. In comparison, other-certified websites were allowed to bid for sponsored ads until the ban. Second, other-certified sites are subject to different safety information in the eyes of consumers and therefore the ban could have different effects on them as compared to the other two types of pharmacy sites.

Using 2008-2012 comScore data, we find that the banned pharmacies experience a reduction in the number of total clicks after the ban but the effect is heterogeneous. In

[43] As detailed in Section 2, NABP endorses LegitScript to act on its behalf in screening websites for search engines, so we treat approval from LegitScript the same as certification from NABP.

particular, tier-B sites experience a smaller reduction in total clicks with some of the lost paid click-throughs replaced by organic clicks. These effects do not change significantly after the Google-DOJ settlement. In contrast, tier-C sites receive fewer traffic in both paid and organic clicks, and the reduction is even greater after the DOJ settlement. We also explore whether the effect of the ban depends on what drug names consumers search for on the Internet. Drug queries that led to more clicks on non-NABP-certified pharmacies before the ban are most affected by the ban, but chronic drug queries are less affected by the ban than non-chronic drugs. Overall, we conclude that the ban has increased search cost for tier-B sites but at least some consumers overcome the search cost by switching from paid to organic links. In addition to search frustration, the ban may have increased health concerns for tier-C sites, which explains why consumers are discouraged from reaching them via both paid and organic links.

This chapter proceeds as follows. In section 2, we provide background on the online market for prescription drugs as well as changes to Google's policy regarding sponsored search ads from online pharmacies. We lay out our econometric framework in section 3 including a model we use to separate the effects of the ban on consumer beliefs and search costs. Section 4 describes the data provided by comScore and results are presented in section 5. Section 6 concludes.

3.2 Background

3.2.1 The Online Market of Prescription Drugs

According to IMS, prescription drug sales in the US has grown from $135 billion in 2001 to $307 billion in 2010 (IMS 2011). A literature review by Orizio et al. (2011)

found that the percent of general population using online pharmacies was often reported to be between 4% and 6%. Although the percentage is small, the total volume of sales can be huge. According to Skinner (2006), sales to US consumers from 278 Canadian or seemingly-Canadian pharmacies reached CDN$507 million in the 12 month periods ending June 2005. The US$500 million fine that Google agreed to pay in 2011 also indicates the size of the online prescription drug market, as the fine is calculated by the revenue received by Google for selling sponsored ads to Canadian pharmacies and the estimated revenue that Canadian pharmacies got from their sales to US consumers.[44]

One major concern of online purchase is drug safety. As described in NABP (2011) and Orizio et al. (2011), drug safety can be potentially compromised by a relaxed prescription requirement, insufficient medical consultation, incorrect packaging and labeling, wrong ingredients, or no delivery at all. Some rogue websites also aim to steal consumer credit card information for identity theft. Although the FD&C Act prohibits the importation of unapproved drugs, when determining the legality of personal shipments, "FDA personnel may use their discretion to allow entry of shipments of violative FDA regulated products when the quantity and purpose are clearly for personal use, and the product does not present an unreasonable risk to the user."[45] Therefore, a consumer who purchases a drug from a foreign pharmacy for personal use faces some uncertainty regarding the likely reaction by the FDA.

To address safety concerns, the FDA also publicizes anecdotes of unsafe pharmaceuticals on the Internet and warns consumers against rogue websites (which could be foreign or domestic). They also advise consumers to avoid any foreign websites and only make

[44] CNN report August 24, 2011, accessed at http://money.cnn.com/2011/08/24/technology/google_settlement/index.htm.

[45] See . The FDA defines personal shipments as containing no more than 90-days supply for personal use and does not involve a controlled substance. A controlled substance is a drug that has a high potential for abuse, does not have an accepted medical use, and/or does not meet accepted safety requirements.

online purchases from the US websites certified by the NABP. The NABP certification ensures that a US website comply with laws in both the state of their business operation and the states to that they ship medications. As of February 29, 2012, NABP has certified 30 online pharmacies, 12 of which are run by large PBM companies (open to members only) and the rest include national chain pharmacies (such as cvs.com and walgreens.com) and large online-only pharmacies (such as drugstore.com).

Another private certification agency, LegitScript.com[46], is similar to the NABP in terms of only approving US-based websites and endorsed by the NABP to screen pharmacy websites after the Google ban. As of March 5, 2012, the home page of LegitScript announced that they monitored 228,419 Internet pharmacies among which 40,233 were active. Within active websites, LegitScript found 221 legitimate (0.5%), 1,082 potentially legitimate (2.7%) and 38,929 not legitimate (96.8%). Their certification criterion includes a valid license with local US jurisdictions, valid registration with the US Drug Enforcement Administration (DEA) if dispensing controlled substances, valid contract information, valid domain name registration, requiring a valid prescription, only dispensing FDA approved drugs, and protecting user privacy according to the HIPAA Privacy Rule (45 CRF 164). There are more LegitScript-certified websites than NABP-certified websites, probably because the NABP requires interested websites to apply and pay verification fees while LegitScript's approval is free and does not require website application. Because the NABP praises the work of LegitScript and endorses the use of LegitScript by domain name registrars to assist in identifying illegally operating websites, throughout this chapter we treat LegitScript the same as NABP and label websites certified by either agency as NABP-certified.

[46]LegitScript was founded by a former White House aide named John Horton.

The other two private certifiers – PharmacyChecker.com and the Canadian International Pharmacy Association (CIPA) – are fundamentally different from NABP/LegitScript. CIPA is a trade association of Canadian pharmacies and only certifies Canadian websites that comply with Canadian laws, while PharmacyChecker.com covers US, Canada, and many other countries. Upon voluntary application (with a fee), PharmacyChecker certifies that any approved website has a valid pharmacy license from its local pharmacy board, requires a prescription for US purchase if the FDA requires a prescription for the medication, protects consumer information, encrypts financial and personal information, and presents a valid mailing address and phone number for contact information. As of March 9, 2012, PharmacyChecker has approved 73 foreign websites and 51 US websites. PharmacyChecker also charges fees for an approved website to be listed on PharmacyChecker.com beyond a short period of initial approval. Consequently, those listed on PharmacyChecker's Pharmacy Ratings page are only a selected list of PharmacyChecker-approved websites. Because PharmacyChecker is unwilling to share their complete list of approvals, we are not able to conduct a full comparison between approvals by PharmacyChecker and those by the NABP, LegitScript or the CIPA. Of the 37 websites listed on the Pharmacy Ratings page of PharmacyChecker.com, only three are labeled US while all the others are either listed under one foreign country or a number of foreign countries plus US. This list is incompletely overlapped with the list of approval from the NABP, LegitScript and the CIPA. Among the four certification agencies, PharmacyChecker is the only one that provides head-to-head drug price comparison across online pharmacies.

As detailed in the next subsection, Google used to contract with PharmacyChecker to filter websites listed in its sponsored search page but switched to NABP/LegitScript after it agreed to ban non-NABP-certified pharmacies in February 2010.

Before we focus on the Google policy regarding online pharmacies, it is important to understand why US consumers buy prescription drugs online. According to Mullner & Gurau (2005), the most frequent reasons quoted by interviewees for buying or intending to buy online were convenience and saving money, followed by information anonymity and choice. Skinner (2005) estimated that Canadian prices for the 100 top-selling brand-name drugs were on average 43% below US prices for the same drugs.[47] Quon et al. (2005) compared 12 Canadian Internet pharmacies with 3 major online US drug chain pharmacies and found that Americans can save an average of approximately 24% per unit of drug on the 44 most-commonly purchased brand-name medications from Canada. In an audit study, Bate et al. (2013) purchased samples of five popular brand-name prescription drugs from NABP/LegitScript-certified websites (tier-A), PharmacyChecker/CIPA-certified websites (tier-B), and websites that were not certified by any of the four certifiers (tier-C). After comparing the purchased samples with authentic versions, they found similar drug quality between tier-A and tier-B samples, but the cash price of tier-B samples were 49.2% cheaper than tier-A samples after controlling for other factors.[48] These findings suggest that a lower price for brand-name prescription drugs is an important incentive for US consumers to shop online.

As for what type of drugs are purchased online, S. Fox (2004) reported that the most frequently bought drugs were for chronic conditions (75%), followed by weight loss and sexual performance substances (25%). Consistently, Skinner (2006) found resemblance between the top five therapeutic categories used by US seniors and the top five therapeutic categories in the cross-border online sales from Canada to US. This suggests that seniors

[47]This number has adjusted for currency equivalency. Skinner (2005) also reported that the 100 top-selling generic drugs are on average priced 78% higher in Canada than in the US. This explains why most cross-border sales from Canada to US concentrated on brand-name drugs.

[48]The price difference was mostly driven by non-Viagra drugs. There was no significant price difference across tiers for Viagra.

are an important source of demand for Canadian pharmacies. Bate et al. (2013) reported an online survey of RxRights members. Because RxRights is a non-profit organization that pays special attention to the cost of prescription drugs, their members are likely more price sensitive than the general population. Among 2,907 respondents who purchase prescription medication for either themselves or family members, 54.8% admitted to purchasing at least one category of the drugs online at some time in the past year, 72.4% of online shoppers purchased from foreign websites only, and an overwhelming majority (91.1%) cited cost savings to be one of the reasons for buying from foreign websites. Surprisingly, most respondents had medical insurance and/or some prescription drug coverage, and the percentage of being insured was not lower among online shoppers. Comments left by respondents suggested that incomplete coverage on prescription drugs, in the form of high deductible, high coinsurance rate, or the donut hole of the Medicare Part D coverage, was one of the factors that motivated the insured to shop online. The survey reported in Bate et al. (2013) also highlighted how respondents searched for pharmacies. Conditional on shopping online, 53.1% used Internet search, 40.4% checked with a credentialing agency such as PharmacyChecker, 22.4% used personal referrals, and only 12.7% looked for the cheapest deal. Consistently, most online shoppers restrict themselves to one primary website, sometimes with supplements from other websites.

3.2.2 Google Policy on Online Pharmacies

As summarized in Table 14, Google used to contract with PharmacyChecker to ensure that every pharmacy website listed in Google's sponsored search page is legitimate according to PharmacyChecker's certification standard. Despite this policy, the FDA found in July 2009 that some online pharmacies advertising on Google had not been

approved by PharmacyChecker.[49] Shortly after (November 2009), the FDA issued 22 warning letters to website operators.[50] At about the same time (August 2009), a study published by LegitScript.com and KnuhOn.com criticized Microsoft Bing for allowing rogue online pharmacy to advertise on its search engine. The study found that "89.7% (of the advertising websites) led to 'rogue' Internet pharmacies that do not require a prescription for prescription drugs, or are otherwise acting unlawfully or fraudulently."[51] While 89.7% is an impressive number, one should note that LegitScript emphasizes the illegality of personal importation and classifies *all* foreign websites as unlawful. In contrast, PharmacyChecker certifies foreign pharmacies and therefore some foreign websites that are unlawful in the eye of LegitScript can be legitimate by the PharmacyChecker standard.

Figure 1 presents a screen shot of Google search page following the query "Lipton" in 2008. On the left hand side are organic links featured by brand-name website (lipitor.com) and information oriented websites such as wikipedia.org. On the right hand side are sponsor links, the top two of them are clearly foreign pharmacies (canadapharmacy.com and canadadrugpharmacy.com). The manufacturer (Pfizer) also placed a sponsored link of lipitor.com at the top of the whole page.

In response to the highlighted concern of drug safety, on February 9, 2010, Google announced two changes regarding its pharmacy advertising policy. The first change is to only accept ads from US online pharmacy websites that are certified by the NABP and from Canadian websites that are certified by CIPA. The second change is that the NABP-certified websites can only target their ads to Google users in the US and the

[49] http://www.nytimes.com/2011/05/14/technology/14google.html?_r=0, retrieved December 25, 2012.

[50] http://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm191330.htm, retrieved December 25, 2012. The current FDA website hosting safety information of online purchase of drugs: http://www.fda.gov/Drugs/ResourcesForYou/Consumers/BuyingUsingMedicineSafely/BuyingMedicinesOvertheInternet/de

[51] The report http://www.cnn.com/2009/TECH/08/20/internet.drugs/index.html posts the link http://www.legitscript.com/BingRxReport.pdf, but it is unavailable to access on December 25, 2012. The report is also available here: http://www.legitscript.com/download/BingRxReport.pdf.

CIPA-certified websites can only target Google users in Canada. The new policy is only applicable to US and Canada.[52] Two months later (April 21, 2010), LegitScript announced assistance to Google in implementing Google's Internet pharmacy advertising policy in place of PharmacyChecker.[53] On June 10, 2010, both Microsoft and Yahoo! started to require NABP certification for online pharmacy advertisers.[54]

In May 2011, Google announced in its quarterly report that "in connection with ... an investigation by the United States Department of Justice into the use of Google advertising by certain advertisers, we accrued $500 million for the three month period ended March 31, 2011."[55] On August 24, 2011, the DOJ made it official that "Google Forfeits $500 Million Generated by Online Ads & Prescription Drug Sales by Canadian Online Pharmacies".[56]

Figure 2 presents a screen shot of Google search page following the query "lipitor" in 2013. In contrast to Figure 1, there are no sponsored links on the page except for lipitor.com at the top. The void of sponsored search on the right hand side is filled by a drug fact label of lipitor with links to official information about the drug's side effects, warnings and user guidance from the National Library of Medicine. The drug fact label started on June 22, 2010 under a partnership between Google and the National Institute of Health (NIH)[57], and probably has diverted some click traffic following drug name queries after the ban.

[52]http://adwords.blogspot.com/2010/02/update-to-pharmacy-policy-in-us-and.html, retrieved December 24, 2012.

[53]http://blog.legitscript.com/2010/04/legitscript-to-help-google-implement-internet-pharmacy-ad-policy/. retrieved December 24, 2012.

[54]https://www.nabp.net/news/microsoft-and-yahoo-now-require-vipps-accreditation-for-online-pharmacy-advertisers

[55]http://sec.gov/Archives/edgar/data/1288776/000119312511134428/d10q.htm, retrieved December 24, 2012.

[56]http://www.justice.gov/opa/pr/2011/August/11-dag-1078.html, retrieved December 24, 2012.

[57]http://venturebeat.com/2010/06/22/google-health-search-adds-drug-info-upping-pharma-ad-spend/, retrieved December 23, 2013.

In light of these events, we define three regimes for our empirical analysis as shown in Table 15. Regime 0 refers to a 17-month period up to January 2010, right before Google adopted the ban. Regime 1 ranges from March 2010 to July 2011, covering a period after the Google ban but before the Google-DOJ settlement. The 13-month period after the Google-DOJ settlement is referred to as Regime 2. Because our data are monthly but both the Google ban and the Google-DOJ settlement occurred in the middle of a month, our sample excludes the two event months (February 2010 and August 2011). As mentioned in Section 1, we classify pharmacy websites into three tiers: tier-A refers to NABP/LegitScript-certified US websites that are always allowed to advertise in Google sponsored search. Tier-B refers to the pharmacy websites that are not certified by NABP/LegitScript, but certified by PharmacyChecker or CIPA. All the pharmacy websites that are not certified by any of the four certification agencies are referred to as tier-C. By definition, only tier-C websites were blocked (imperfectly) from sponsored listings in regime 0, whereas both tier-B and tier-C websites are blocked in regime 1 and regime 2. Throughout the chapter, we use "NABP-certified" exchangeably with "tier-A", "other-certified" exchangeably with "tier-B", and "uncertified" exchangeably with "tier-C".

3.3 Conceptual and Econometric Framework

While consumers have many ways to reach drug-related websites, here we focus on searches through search engines due to data limitations. For simplicity, this section assumes that there is only one search engine available and therefore abstracts from substitution between search engines.[58] Conditional on a consumer using a search engine, her

[58] Our data contain search and click volumes for each of the five largest search engines. According to comScore's new release, Google has a 64-67% market share in organic search during our sample period. Because some comScore data on searchers are not engine specific, our empirical results pool all engines.

search consists of entering a query in the search box and clicking into website link(s) offered in the search results page.[59] As detailed below, most clicks into pharmacy sites come from queries related to pharmacies (e.g., canadapharmacy, pharmacychecker, or "cheap drug Canada"), queries containing a drug name (e.g., lipitor), or queries related to health conditions, drug manufacturers, drug regulators, etc. Organic and paid clicks are recorded separately in the comScore data. To examine how paid, organic or total clicks change after the ban, we assess the effects on both the extensive and intensive margins using a two-part model. [60] The extensive margin is whether a website receives any positive clicks in a given month,[61] while the the intensive margin is the number of clicks a website receives, conditional on receiving some (non-censored) clicks.

Defining $Y_{it}^{AllQueries}$ as paid/organic/total clicks that website i received in month t, we investigate the extensive margin using a probit regression:

$$\begin{aligned} Prob(Y_{it}^{AllQueries} > 0) &= \Phi(\alpha + \sum_{k\in\{B,C\}} \beta_k * Tier_k + \sum_{r=1}^{2} \gamma_r * Regime_r \\ &\quad + \sum_{k\in\{B,C\}} \sum_{r=1}^{2} \theta_{kr} * Tier_k * Regime_r). \end{aligned} \tag{7}$$

Tier and *Regime* are indicator variables for the type of pharmacy (tier A, B, or C) accessed at website i and the time period to which month t belongs (regime 0, 1, or 2).

[59] We use the term "query" to denote the actual text the user enters into the search box on the search engine and the term "click" to denote the subsequent clicks by the user on organic or paid links that result from the search. The data include the number of times a certain query was entered into a search engine and the number of clicks on each link, conditional on the query. A query with no subsequent clicks is recorded by comScore as one query and zero clicks.

[60] The distribution of clicks per website is characterized by a spike at zero and a bell-shape positive distribution skewed to the right, and the two-part model with log-normal positive distribution best captures the data pattern.

[61] The number of clicks is coded as censored if the website receives too few clicks. We do not have specific information on the censoring rule, so we code the censored clicks as zero. In one specification, we analyze the extensive margin as whether a website receives any positive or censored clicks, and the results are similar.

The intensive margin is assessed using a simple OLS model conditional on a website receiving positive clicks:

$$(ln(Y_{it}^{AllQueries})|Y_{it}^{AllQueries} > 0) = \alpha_i + \sum_{r=1}^{2} \gamma_r * Regime_r + \sum_{k \in \{B,C\}} \sum_{r=1}^{2} \theta_{kr} * Tier_k * Regime_r + \epsilon_{it}, \quad (8)$$

where α_i denotes website fixed effects. Because website fixed effects absorb the tier dummies, $Tier_k$ only appears in the interaction with $Regime_r$. We do not include website fixed effects in equation (1) because a probit regression with fixed effects may introduce an incidental parameter problem. In both specifications (1) and (2), θ_{kr} measures the conditional differential effect of regime 1 and regime 2 for tier-B and tier-C websites compared with the control group tier-A pharmacies in regime 0.

A priori, when $Y_{it}^{AllQueries}$ represents total clicks, one may expect θ_{kr} to be negative for tier-B and tier-C websites after the ban, either because the ban has sent a negative message about the safety of these websites or because the ban has made it more difficult to find tier-B and tier-C sites even if consumers' beliefs remain unchanged. The challenge is how to distinguish these two explanations. One strategy is to explore the timing difference: arguably, the massive media coverage on the Google-DOJ settlement (regime 2) may have increased the salience of the negative message about the safety of tier-B and tier-C websites, while the difficulty to find these websites should have increased in regime 1, right after Google started to ban these websites from sponsored search. Moving from regime 1 to regime, consumers' perceptions about the safety of the sites may have been affected by the settlement. This suggests that we can differentiate the above two explanations by

comparing the effects of the ban in regime 1 and regime 2.

The second strategy is to compare the changes in total and organic clicks on tier-B and tier-C websites. Because tier-C websites were prohibited from sponsored listings even before the ban[62], the ban should be a greater shock to clicks on tier-B websites than on tier-C websites, if the main effect of the ban is informing consumers of the danger of other-certified websites. This implies that the organic clicks on tier-B websites should drop more after the ban than those on tier-C websites. In contrast, if the main effect of the ban is adding consumer search cost in reaching non-NABP-certified websites, the drop in the organic clicks on tier-B websites may be smaller than those on tier-C websites, either because tier-B websites were on average easier to find in organic search (proxied by their organic clicks before the ban) or because tier-B websites were perceived safer than tier-C websites thanks to their non-NABP certification.

The above regressions summarize all search behaviors including what query to search for and what link to click into. Assuming the ban has different effects on tier-B and tier-C pharmacy sites (which turns out to be true in our data), we can further examine which consumer behavior leads to the difference: is it because the ban motivates differential search intensity on pharmacy queries that spell out the names of tier-B or tier-C sites, or because searchers are more or less likely to click into tier-B or tier-C sites conditional on the same pharmacy queries? Taking tier-A pharmacy name queries as the baseline (excluded from j), the effect on query intensity can be studied in the following specification:

$$ln(Y_{jt}^{Pharmacy}) = \alpha_j^P + \alpha_t^P + \beta_1^P \cdot X_j^P \cdot Regime_1 + \beta_2^P \cdot X_j^P \cdot Regime_2 + \epsilon_{jt}^P, \qquad (9)$$

[62]Paid clicks are observed on tier-C websites due to imperfect screening by the search engines.

where $Y_{jt}^{Pharmacy}$ denotes the number of searches for pharmacy query j in month t.[63] X_j is a set of dummies indicating the type of query j. The coefficients $\{\beta_1^P, \beta_2^P\}$ denote the difference-in-differences estimates of how the two regimes affect various pharmacy queries as compared to the queries on tier-A pharmacy names.

As detailed in Section 4.2, we can distinguish pharmacy name queries (e.g. "cvs"), discount pharmacy queries (e.g. "cheap drug") and general pharmacy queries (e.g. "pharmacy at"). Different pharmacy query types may indicate different intention of search and therefore respond differently to the ban. To capture the effect of the ban on clicks into website i conditional on pharmacy query type j, let X_j to be the dummy variable for each pharmacy query type, we extend equations (1) and (2) to allow key parameters $\{\gamma_r, \theta_{kr}\}$ to vary by the type of query:

$$\begin{aligned}
Prob(Y_{ijt}^{Pharmacy} > 0)) &= \Phi(\alpha + \sum_j \alpha_j X_j + \sum_{k\in\{B,C\}} Tier_k + \sum_{r=1}^{2} Regime_r \\
&\quad + \sum_j \sum_{k\in\{B,C\}} \beta_{kj} Tier_k * X_j + + \sum_j \sum_{r=1}^{2} \gamma_{rj} Regime_r * X_j \\
&\quad + \sum_j \sum_{k\in\{B,C\}} \sum_{r=1}^{2} \theta_{krj} * Tier_k * Regime_r * X_j \\
ln(Y_{ijt}^{Pharmacy} | Y_{ijt}^{Pharmacy} > 0) &= \alpha + \sum_j \alpha_j * X_j + \sum_{r=1}^{2} Regime_r \\
&\quad + \sum_j \sum_{k\in\{B,C\}} \beta_{kj} Tier_k * X_j + + \sum_j \sum_{r=1}^{2} \gamma_{rj} Regime_r * X_j \\
&\quad + \sum_j \sum_{k\in\{B,C\}} \sum_{r=1}^{2} \theta_{krj} * Tier_k * Regime_r + \epsilon_{ijt}
\end{aligned}$$

The relationship between a user's query and resulting click destinations sheds light on the economic effects of the ban. If a query for "discount pharmacy" directs more traffic away from both tier-B and tier-C websites after the ban, it suggests that consumers

[63] We also estimate equation 9 using the number of searchers that submit query j in month t.

have heightened safety concerns for all non-NABP-certified websites. In comparison, if the query directs traffic away from tier-C sites but not from tier-B sites, it is probably because consumers are willing to tolerate the risk of tier-B sites and/or find a way to get around the ban of tier-B sites in sponsored search. Pharmacy name queries may provide more direct evidence. If we find a tier-C pharmacy name query leads to fewer organic clicks on tier-C sites but a tier-B pharmacy name query does not lead to fewer organic clicks on tier-B sites, one explanation is that the ban has different effects in conveying the safety risk for these two types of pharmacy sites.

We also explore how the effect of the ban differs by the types of drugs consumers search for on the Internet. Existing literature suggests that consumers that target chronic or privacy-oriented drugs will be affected the most by the ban because cost saving and privacy are dominant reasons for using online/foreign pharmacies before the ban. Non-NABP-certified websites may be more attractive for lifestyle drugs, either because users of these drugs appreciate privacy or because they do not have a formal prescription and prefer websites with a less rigid prescription requirement.[64]

However, as the ban cannot prohibit consumers from reaching non-NABP-certified pharmacies via organic links, it is unclear whether the ban leads to more or less of a click reduction for these drug queries. To examine this question, we classify drug queries according to (1) whether drug query j attracted a high fraction of clicks into non-NABP-certified pharmacies before the ban, (2) whether drug query j targets lifestyle drugs or controlled substances, and (3) whether drug query j targets chronic drugs.[65] Defining

[64]The term "lifestyle drug" does not have a precise definition, but one author describes a drug in this category as "one used for 'non-health' problems or for problems that lie at the margins of health and well being." Of course, some lifestyle drugs are at times used to treat serious medical conditions. See Gilbert, Walley and New (*British Medical Journal*, November 2000).

[65]For robustness, we also considered drugs for whom the searchers were more likely to be elderly or low-income before the ban.

each classification variable as X_{g_j}, we estimate the differential effects of the ban on the extensive margin of clicks into pharmacy site i from drug query type g_j in month t (Y_{ijt}), by:

$$\begin{aligned} Prob(Y_{igt}^{Drug} > 0)) &= \Phi(\alpha + \sum_g \alpha_g X_g + \sum_{k\in\{B,C\}} Tier_k + \sum_{r=1}^{2} Regime_r \\ &\quad + \sum_g \sum_{k\in\{B,C\}} \beta_{kg} Tier_k * X_g + + \sum_g \sum_{r=1}^{2} \gamma_{rg} Regime_r * X_g \\ &\quad + \sum_g \sum_{k\in\{B,C\}} \sum_{r=1}^{2} \theta_{krg} * Tier_k * Regime_r * X_g \\ ln(Y_{igt}^{Pharmacy} | Y_{igt}^{Pharmacy} > 0) &= \alpha + \sum_g \alpha_g * X_g + \sum_{r=1}^{2} Regime_r \\ &\quad + \sum_g \sum_{k\in\{B,C\}} \beta_{kg} Tier_k * X_g + + \sum_g \sum_{r=1}^{2} \gamma_{rg} Regime_r * X_g \\ &\quad + \sum_g \sum_{k\in\{B,C\}} \sum_{r=1}^{2} \theta_{krg} * Tier_k * Regime_r + \epsilon_{igt} \end{aligned}$$

The coefficients of the interaction terms with X_{g_j}, denoted as $\{\gamma_{rg}, \theta_{krg}\}$, indicate whether the ban has differential effects on clicks by the type of drug query.

3.4 Data Summary

Our primary datasource is comScore.[66] ComScore tracks the online activity of over two million persons worldwide, one million of whom reside in the US. ComScore extrapolates the observed activity in the households it tracks and by using various demographic weights, it determines the aggregate activity of all US Internet users. We obtained access to click-through data from US households. ComScore data have been used to study internet search behavior by a number of economists, for example, Waldfogel & Chen (2006),

[66] http://www.comscore.com/.

Chiou & Tucker (2011), and George & Hogendorn (2013).

3.4.1 Click and Search Data

We use data from comScore's Search Planner suite of tools, which provides click-through data on queries submitted to five large search engines - Google, Yahoo!, Bing, Ask, and AOL. The click data (available on comScore's "term destinations" report) are organized by query-month-engine and include the number of queries (searches), searchers, and clicks in a given month. In addition, clicks are also broken down into organic versus paid and by destination URL.[67] At times, due to small sampling of some queries, click activity is censored because comScore is unable to reliably extrapolate the observed activity to the whole population.[68] We observe 49 months of data from September 2008 to September 2012.

In addition to click activity following each query, we also download from comScore a demographic profile (comScore's "term profile" report) of searchers who perform each query in each month. The profile includes a distribution of age, income, household size, the presence of children, and the geographic location of the searchers. We also observe the share of clicks following a query that are received by each of the five search engines.

As an example, Figure 3 shows an example of these reports for Lipitor in January 2012. The term destination report lists the total clicks, divided between organic and paid, following queries for Lipitor in January 2012. Because we selected "match all forms", the

[67] A query is the actual text that a searcher enters on a search engine. Our data include click activity on websites following the exact query, but also clicks following queries where the text appears somewhere in the search box, potentially along with other words. Plural forms of the query are also included. comScore refers to this as "match-all-forms" queries as opposed to "exact" queries that return the clicks on the query text exactly as entered on the search engine.

[68] Our data has a limitation in regard to censoring. When a click count is censored by comScore, the name of the website entity appears in the database with a click count of -1. This means there were positive clicks on the website during that month, but extrapolation to the population would not produce a reliable estimate. We treat these websites as having zero clicks in our analysis.

click counts include queries for Lipitor alone as well as Lipitor plus other keywords. This report shows clicks on all five search engines combined, but separate reports were also run on individual search engines. The click counts under the key metrics section is comScore's estimate of the total number of clicks by users in the US on all websites following the query. In addition, the clicks are broken down by specific entity.[69] Each entity name is also assigned to one or more categories, such as, health, government, or pharmacy. It is important to note that the clicks we observe on an entity all originate from a search engine. We do not know how many clicks a website receives via direct navigation, bookmarks, etc.

In addition, the term profile report provides information about searchers for Lipitor in January 2012. While the report is not engine-specific, it provides the total number of searches and searchers, irrespective of clicks following those searches. The report also provides demographic information on the households that searched for Lipitor in January 2012. A few examples are shown in the table, but demographics are provided for age, income, geographic region, location (home/work/school), household size, and the presence of children. Finally, the report tells us the share of searches on each of the five search engines.[70]

3.4.2 Query List and Website Classification

A list of queries must be submitted to comScore in order to extract query-level data. To create a list of drug and pharmacy related terms, we use several resources. The first one is a list of brand names from the FDA's Orange Book of all approved drugs.[71] The

[69] Usually an entity name is a URL, but comScore also aggregates clicks on websites with common ownership and lists them under a different entity level (e.g., property, media title, channel, etc). We collect click data at the finest level available to avoid double counting.

[70] From the share, we can determine the number of searches that were performed on each engine, however the demographics are only available for searchers across all engines.

[71] http://www.accessdata.fda.gov/scripts/cder/ob/default.cfm.

second resource is a list of drug manufacturers from Kantar Media[72] We also include three government website names that provide drug information (FDA, NIH, and CDC), and four website names that certify online pharmacies (NABP, LegitScript, PharmacyChecker, and CIPA). The resulting list of queries is supplemented by the names of online pharmacies, which is based on comScore's own categorization of the websites in their data. Running our list of drug names on comScore, we can identify the top pharmacy website names in the comScore "Pharmacy" category.[73] This list, plus any pharmacy names that we can find on any of the four certifying websites, comprise our preliminary list of pharmacy websites.

To address the possibility that searchers may reach drug and pharmacy related websites by searching for a medical condition, symptom, or another non-drug and non-pharmacy term, we supplement the query list with data from Keywordspy.com. This website collects information on keywords that companies bid on for sponsored ads on a search engine. It also reports a list of keywords that more likely lead to organic clicks on a certain website.[74] This allows us to identify a list of organic keywords that are popular searches when the destination is ultimately an online pharmacy. We also add all keywords that the FDA bid on to appear in an engine's sponsored ads.

The combination of all these sources led to over 8,000 queries, far too many to download from comScore given time constraints. Therefore, we restricted the list of drugs to only those that were advertised (in the Kantar media data) and/or prescribed by a physician from 2006-2009.[75] We also ran the complete list of queries through comScore twice on two time windows in 2009 and 2012 and restricted our sample to queries that

[72]http://kantarmediana.com/intelligence.

[73]The "Pharmacy" category ID on comScore is 778268. A website may have multiple classifications, but any site with this ID we classify as a pharmacy.

[74]This is similar to the Keyword Tool in Google's Adwords.

[75]The latter comes from the National Ambulatory Medical Care Survey (NAMCS).

accounted for the top 90% of clicks in either window. This left us with 690 queries. Because comScore reports the clicks both for the query exactly as it appears and variations of the query (e.g., clicks following a search for "canada online pharmacy" are included in a search for "canada pharmacy"), we only use queries that are not variations of another to avoid double counting. This further restricts our sample to 528 queries. Each query was then submitted to comScore and monthly reports from each search engine were downloaded for the analysis.

Each of the 528 queries are then classified into different query types (see Table 16). Along with drug queries, pharmacy queries are further classified according to their certify-status (tier A, B, or C) as well as general and discount pharmacy keywords. Queries that are not drug or pharmacy related are classified as other.

Table 16 shows the total query count in each category of query. Within each broad group of queries (drug, pharmacy, and other), we further classify the queries by their intention to search for online pharmacies. We expect that the effect of the ban will be most significant on the searches and clicks of queries that are used to reach non-tier-A online pharmacies before the ban. In particular, for the pharmacy query group, we first separate out the queries that are the exact name of the online pharmacy websites and classify them according to the pharmacy tiers. Queries that target pharmacies that sell cheap or discount drugs, and those operate in foreign countries, which more likely lead to clicks on non-tier-A pharmacies, are classified into discount pharmacy search terms.[76] The remaining pharmacy queries are all general search terms for pharmacies.[77] As discussed in

[76]Among 46 discount pharmacy queries, 11 contain words "canada, international and europe", 5 contain words "online", 17 contains word "cheap, discount, low cost, free, deal, coupon".

[77]In the general pharmacy terms, there are three queries "pharmacy in", "pharmacy on" and "the pharmacy" carrying exactly the same observations, so we dropped the first two. To check if "the pharmacy" counts all clicks from the query that contains only the word "pharmacy", we calculate the total number of clicks by all queries with "pharmacy" in it except for "the pharmacy". We find that "the pharmacy" always records a larger number of clicks and conclude that "the pharmacy" includes all clicks for queries

the previous section, the sample of queries in our study are chosen if they lead to a sufficient volume of traffic that can be captured by comScore. Among 528 queries, we choose to focus on drug and pharmacy queries because they are more likely to lead to online pharmacy websites and thus better reflect the changes in consumer search behavior.[78] Figure 4 shows that the number of searchers and searches evolve similarly by broad query groups. Pharmacy search queries experience a spike in the last few months of each year because some pharmacy queries include large retail stores (e.g., walmart and target) with seasonal demand. We control for seasonality in robustness checks of our results.

The last step in processing the data is to classify the destination websites in the database into various categories. We analyze the click data only for pharmacy websites so we classify online pharmacy websites according to their certify-status (tier A, B, or C).[79] The destination website classification is used in the results shown in the regression tables.

Because some of the comScore data are not engine specific, all empirical results present below pool data from all five search engines.

3.5 Descriptive Statistics

Table 17 summarizes the number of searches and clicks by query type. The ratio of online pharmacy clicks to searches (column 3) is associated with the search cost of finding a certain website. If the desired pharmacies do not appear in the paid links or high in the organic results, this may lead consumers to not click on any website and subsequently.

with "pharmacy" in it. We kept the query "the pharmacy", but subtract the from it the total number of clicks by queries containing the complete word "pharmacy".

[78] In regime 0, only 2.3% of the clicks on pharmacy websites are led from queries other than drug and pharmacy queries, so we choose to not to focus on these queries.

[79] Since the search engine ban only applies to online pharmacies that sell prescription drugs, our analysis restricts to this set of pharmacies. We cannot directly infer whether a pharmacy sells prescription drugs from its site name or comScore classification, so we check by clicking into each pharmacy website to verify that prescription drug is sold in the website at the time of our study.

This would result in a low pharmacy clicks-to-searches ratio.

The ratio of pharmacy clicks to total clicks (column 4) show how paid and organic clicks vary on each type of pharmacies led from different query types. Pharmacy queries lead to many more clicks on pharmacy websites than drug queries. Tier-B names are very likely to lead to pharmacy websites (93-98%) followed by tier-A names (78-81%) and discount pharmacy keywords (59-67%).[80] Tier-C pharmacy names are associated with the lowest percentage of pharmacy clicks among all pharmacy name queries and this percentage drops sharply from 39.8% in regime 0 to 31.4% in regime 1 and 7.1% in regime 2. In contrast, the percentage of pharmacy clicks is stable or even increasing for Tier-B pharmacy names after the ban. Compared with pharmacy queries, drug queries have a much lower percentage of pharmacy clicks (22.1%) and that percentage plummets after the ban (to 2-4%). This is probably because many drug queries target information websites rather than pharmacies and the searchers targeting a pharmacy website using a drug query cannot find the pharmacy sites via sponsored links following the ban. The remaining columns of Table 4 report paid and organic clicks separately. The organic clicks to Tier-B and Tier-C sites have increased after the ban for almost all pharmacy and drug queries, suggesting substitution to organic results when sponsored links are no longer available.

Focusing on pharmacy websites, Table 18 also summarizes the organic and paid click volume on pharmacy websites by tier and by regime. For tier-A pharmacies, the number of organic and paid clicks grows from regime 0 to regime 2. Tier-B pharmacies in regime 0 are accessed mostly via paid clicks, with an average of 6,338 monthly paid clicks and 1,795 monthly organic clicks. The ban results in almost 100% loss in paid clicks, but

[80]The average clicks per search and the percent pharmacy clicks are first calculated at the query level and then averaged.

part of the loss is offset by a large increase in organic clicks, suggesting that searchers are substituting organic for paid links. For tier-C websites, the average number of paid clicks falls as expected and the average organic clicks rises in regime 1, but then falls in regime 2, consistent with substitution to organic links in regime 1 and more awareness of the risks associated with these sites in regime 2. The differential change in organic clicks on tier-B and tier-C websites is evident in Figure 5, where we plot the monthly trends of paid and organic clicks by tier. Part of the reduction in organic clicks on tier-C pharmacies may be attributable to fewer tier-C pharmacy queries after the ban, as shown in Figure 6.

The last three columns of Table 18 show the distribution of number of websites active in each regime. With the same set of queries in each regime, the number of online pharmacy websites that are recorded as having any clicks in comScore is relatively stable for tier-A and tier-B pharmacies, but declines 33% for tier-C from 138 to 92. This decline could be due to both health concerns and search costs. The decline in the number of tier-C websites may have several implications. For pharmacy competition, this may benefit the remaining tier-C pharmacies if consumers preferring tier-C pharmacies continue to buy from them. However, if consumers are shifting from tier-C to tier-B or tier-A pharmacies, we will observe clicks on tier-C websites decline as a whole.

The top panel of table A1 in the appendix lists examples of drug queries that led to a high proportion of clicks into tier-B and tier-C websites in the first 9 months of our sample (September 2008 to May 2009) before the ban. Five of the top 10 drug queries on list are controlled substances. The bottom panel lists drugs with a low proportion of clicks into tier-B and tier-C websites. Only one query in the tier-B list is controlled substance and it also includes more drugs that target chronic diseases such as high blood pressure pain. These patterns are not surprising as tier-C sites are less likely to require prescriptions and

controlled substances are subject to closer screening by the FDA at customs enforcement. In an unreported table, we also rank drug queries by the absolute count of total clicks into tier-B or tier-C sites. These alternative ranks are similar to the ranks presented in Table A1, except that some high-volume drug queries are ranked higher in the tier-B list if they target chronic conditions (e.g., lipitor and insulin) or ranked higher in the tier-C list if they target lifestyle drugs or controlled substances.

Overall, these statistics suggest a similar trend in searches across broad query groups, but different click patterns into tier-A, tier-B and tier-C websites. In general, we observe more paid and organic clicks on tier-A pharmacies, a greater substitution from paid clicks to organic clicks for tier-B pharmacies after the ban, a reduction in organic clicks for tier-C pharmacies as well as a reduction in search intensity for tier-C pharmacy names. The drug queries that led to tier-B and tier-C clicks before the ban are also different: tier-B sites were more likely to receive clicks from searches for chronic drugs, while tier-C sites were more likely to receive clicks from queries for lifestyle drugs or controlled substances.

3.6 Regression Results

3.6.1 Total Clicks from All Queries

Our first set of regressions focus on clicks received by pharmacy website i in month t from all queries. As detailed in Section 3, this is our broadest specification and it summarizes all search behavior leading to pharmacy websites.

Table 19 reports pharmacy websites clicks results for total and organic clicks. Within total clicks, column (1) examines whether website i received any clicks in month t; Column (2) examines whether website i received any positive clicks in month t, where positive

clicks refers to non-censored click counts in the comScore data. Both columns (1) and (2) refer to the extensive margin, following the probit specification in equation (1). On the intensive margin, column (3) uses equation (2) to examine the log of the number of clicks, conditional on a website receiving positive clicks in the month. Because click traffic of many websites is too low to have non-censored positive clicks, the number of observations drops 72% from columns (1) and (2) to column (3). The results for "any click" and "any positive click" are similar, so for organic clicks we only report regressions for "any positive organic click" (column 4) and log positive organic clicks conditional on having positive organic clicks (column 5). All columns use tier-A sites as the excluded baseline group.

The first three columns suggest that, after the ban, tier-C sites suffer on the extensive margin while tier-B sites suffer on the intensive margin. In particular, the probability of a tier-C site receiving any positive clicks falls 6.69 percentage points in regime 1 and the net effect grows to 10.92 percentage points by regime 2. In comparison, there is no significant change in the probability of a tier-B site receiving any positive click. Conditional on receiving any positive clicks, the amount of total clicks received by a tier-B site falls 61.7% in regime 1 and by a similar magnitude (58.3%) in regime 2. Recall that the ban on sponsored search was effective in both regimes 1 and 2, but the Google-DOJ settlement at the beginning of regime 2 had broader media coverage and likely heightened the health concerns of uncertified pharmacies. The larger drop in tier-C clicks in regime 2, together with the lack of a further drop of tier-B clicks in regime 2, suggests that consumers may have had more health concerns with tier-C sites than with tier-B sites after the Google-DOJ settlement. Another possible explanation is that tier-C websites were ranked low in organic results and their organic ranks became even lower in regime 2 as consumers had difficulty finding them in regime 1.

Focusing on organic clicks only, the last two columns of Table 19 show that tier-B sites enjoy an 88.2% increase of organic clicks in regime 1 from regime 0 and 113.6% increase in regime 2 relative to tier-A. Combined with the fall in total clicks on these sites, this suggests that the loss of paid clicks on tier-B sites was offset with an increase in organic clicks, although total clicks still fall. In contrast, tier-C sites suffer a reduction in traffic via both organic and total clicks, and the reduction is greater in regime 2 than in regime 1. These differential effects suggest that the ban generates search frustration and some, but not all, consumers switch from paid to organic links for tier-B sites. This does not rule out health concerns for tier-B sites, but the Google-DOJ settlement may have raised more health concerns for tier-C sites than for tier-B sites.

We also estimate auxiliary models to assess the robustness of these results. To control for the possibility of a pre-treatment trend in clicks, we include a trend term that was allowed to vary separately in each regime. We also checked for the impact of seasonality by including a dummy variable for the holiday months of November and December for tier-A sites. Neither of these specifications impacted the qualitative results.[81] Because the ban on tier-B and tier-C pharmacies from sponsored links was imperfect (as shown in figure 5), we also conducted robustness checks on the cut-off date of regime 1 (the date of the ban) in two ways. First, we used a new regime 1 cut-off corresponding to the actual month when paid clicks on non-NABP certified pharmacies fell to nearly zero (September 2010). Second, we performed a placebo check by placing the regime cut-off in June 2009, well before the ban. The first strategy does not affect the qualitative results and the second shows no change in organic and paid clicks in the hypothetical regime 1 treatment period before the actual ban. In the first strategy, we also tried cutting the

[81] Estimates for all robustness checks are available from the authors upon request.

regime1 into two halves corresponding to before and after September 2010.We find the coefficients similar for these two halves, except that the drop of tierC websites total clicks at the extensive margin in deepened relative to tier-A in the second half of regime1.

3.6.2 A Closer Look at Pharmacy Queries

We next investigate whether the click reduction on tier-B/tier-C sites is driven by consumers searching less intensively for tier-B/tier-C pharmacy names or a lower likelihood to click on tier-B/tier-C sites, conditional on a particular type of pharmacy query. To answer this question, Table 20 reports regressions of log (searchers) and log (searches) of pharmacy queries. Taking tier-A pharmacy queries as the baseline, we look into general pharmacy queries, discount queries, tier-B queries and tier-C queries separately. The only significant effects in this table are the drop of searches and searchers in tier-C pharmacy queries. The similar magnitudes of the effect on searches and searchers suggest that fewer consumers search for tier-C pharmacy names after the ban and even fewer after the Google-DOJ settlement.

Table 21 examines how the ban changed total clicks into website i from a pharmacy query of type j. We report the extensive margin (total clicks > 0) and the intensive margin (log(total clicks), if positive) separately. Within each margin, we organize columns by destination: 1$\times$ denotes the baseline destination (tier-A), tier-B$\times$ denotes additional effects into tier-B destinations, and tier-C$\times$ denotes additional effects into tier-C destinations. The rows are organized by pharmacy query types: general, discount, tier-B and tier-C relative to tier-A queries. The most noticeable result is that tier-B and discount queries more likely lead to tier-B destinations after the ban but a tier-C query is less likely

to lead to a tier-C destination. One possible explanation is that tier-B websites appear high in organic ranks when consumers search for the tier-B names but tier-C websites are ranked lower when consumers search for the tier-C names. Although we do not know the exact organic ranks of each result in our sample period, we have searched tier-B and tier-C pharmacy names in Google in 2013 and found the pharmacy websites appear highly ranked in all cases. If the organic results in our sample period are similar to what we got in 2013, this does not explain the differential effect on tier-B and tier-C queries from our regression. These results, combined with a lower search intensity for tier-C queries, suggest that consumers may shy away from tier-C websites due to health concerns but are persistent in searching for and clicking into tier-B websites despite potentially higher search costs.

3.6.3 Heterogeneous Effects of Drug Queries

Pharmacy queries are often associated with clicks on pharmacy websites, however we do not observe which drug or condition the searchers are interested in once they click on the website. In contrast, each drug query focuses on a particular drug, which is suggestive of the drug cost, drug type, and searcher demographics, and allows us to explore heterogeneous effects across different drugs or across different types of searchers.[82]

The existing literature suggests that consumers tend to use online pharmacies for chronic or privacy-sensitive conditions. Foreign online pharmacies can offer large cost savings if a brand name drug is expensive in the US and consumers need it frequently. Some foreign pharmacies, especially those in tier-C, offer online consultation and have less

[82]We are not able to explore heterogeneous effects across different types of searchers for pharmacy queries because the search volume on each pharmacy query is not large for comScore to provide searcher demographics both before and after the ban.

restrictive prescription requirements than pharmacies in other tiers. These features can be attractive to consumers who are reluctant to obtain a prescription because of privacy concerns or because of perceived stigmas associated with some lifestyle drugs. In light of this literature, we explore heterogeneous effects of the ban in four directions.

First, we characterize drugs according to what percentage of clicks before the ban were on tier-B or tier-C sites. For a particular drug that had non-censored total clicks in the first nine months of our data before the ban (September 2008 to May 2009, a total of 233 drugs), we compute the fraction of total clicks into tier-B and tier-C sites. The distribution of this fraction is very skewed, ranging from 100% (for two queries that only led to tier-C clicks) to 0% (for 110 queries that only led to tier-A clicks). 79 drugs are defined as H-drug if this fraction is greater than 3%, and 112 drugs as L-drug queries if this fraction is below 0.1%.[83] In the regressions for both extensive and intensive margins, we take L-drug queries as the baseline and examine whether H-drug queries have a differential effect on the interactions between the destination tier and regime dummies. The regression sample excludes the first nine months of our data because they are used to define the H and L drugs.

Estimates of equations (6) and (7) are shown in table 22. The results show that H-drug queries are associated with a greater loss in clicks on tier-B or tier-C sites after the ban. Specifically, H-drug queries experience more of a reduction in tier-B and tier-C total clicks on the intensive margin. However, organic clicks for tier-B sites following H-drug queries are unaffected while they fall for tier-C sites.

In contrast to the recovery of organic clicks following pharmacy query searches after the ban, the lack of substitution to organic clicks following H-drug queries is possibly

[83]The other 42 drugs had a fraction of total clicks into tier-B and tier-C sites ranging between 0.1% and 2.72%. We omit these queries in the regressions. Appendix Table A1 provides a list of the top 10 H-drug queries and top 10 L-drug queries, ranked by the total clicks on pharmacy websites.

because tier-B sites rarely show up as high-ranked organic links when one searches for a specific drug. In contrast, tier-B sites often appear on the first page of organic results if one enters pharmacy queries. These losses in total and organic clicks on tier-C sites are larger and more significant after the Google-DOJ settlement, which is consistent with the previous finding that consumers shy away from tier-C sites due to not only increased search cost after the ban but also heightened health concerns after the settlement.

Our second analysis of heterogeneous effects focuses on lifestyle drugs. We define lifestyle drugs as those that target ED (5 queries), birth control (11 queries), weight loss (3 queries), facial skin problems (11 queries), or smoking cessation (3 queries). We also include drugs that are designated as controlled substances by the US government (23 queries).[84] In total, 50 drug queries are classified as lifestyle drugs.[85] As we expect, lifestyle drug queries are more likely to result in clicks into tier-C sites before the ban.[86] Taking non-lifestyle drug queries as the baseline, Table 23 reports regression results for the differential effects of lifestyle drug queries. In general, the differential effect is insignificant, except for a greater reduction in total clicks from lifestyle queries into tier-B sites on the intensive margin and a greater reduction in total clicks into tier-C sites on the extensive margin, both after the Google-DOJ settlement.

A third type of heterogeneous effect could exist between chronic and non-chronic drug queries. A drug query is defined as chronic if the drug was on average prescribed five or more times a year per patient in the nationally representative 2010 Medical Expenditure Panel Survey (MEPS). A query is defined non-chronic if the average prescription frequency is below 3.5 per patient per year. In total, we have 73 chronic drug queries and 83

[84]Some, but not all, sleep aid, ADHD and muscle relaxant drugs are controlled substances.

[85]Appendix Table A2 provides a list of top 10 lifestyle queries and top 10 non-lifestyle queries, ranked by the number of pharmacy-related clicks following each query.

[86]The fraction of total clicks into tier-C sites in the first nine months of our data is 6.9% for lifestyle drug queries, and 2.81% for non-lifestyle drugs.

non-chronic drug queries.[87] Those with no representation in the MEPS data or with prescription frequency between 3.5 and 5 are dropped from regressions.

Taking non-chronic queries as the baseline, Table 24 shows that chronic queries suffer less of a reduction in total and organic clicks into tier-B and tier-C sites on the intensive margin. These effects are larger and more significant after the Google-DOJ settlement. In comparison, there is no significant differential effect between chronic and non-chronic queries on the extensive margin. Because the intensive margin captures larger websites by definition, this suggests that the ban has less (and in fact close to zero) effect on clicks from chronic queries to large tier-B and tier-C websites. These differential effects are impressive if we consider the facts that the banned pharmacies have a low chance to appear high in organic results following a drug query and the percent of clicks on pharmacy websites following drug queries has plummeted from 22% to 2-3% after the ban.[88]

Our results show that organic and paid clicks on tier-A pharmacies increase after the ban on non-NABP certified pharmacies. Total clicks on tier-B pharmacies fall after the ban, though consumers substitute to organic links to partially offset of the fall in paid clicks. Clicks on tier-C sites fall as well, and we find very little substitution to organic links after the ban. This is consistent with health concerns driving consumers away from non-tier-A pharmacies, though are still willing to click (potentially with higher search costs) on other-certified tier-B sites after their ban. it is also consistent with the possibility that tier-B sites are ranked higher than tier-C sites in organic results and therefore are easier to find when sponsored links disappear from the search page. Our analysis of heterogeneous

[87] Appendix Table A3 provides a list of the top 10 chronic queries and top 10 non-chronic queries ranked by the number of pharmacy-related clicks following each query.

[88] Although we do not present the results here, we also investigated if the average demographics of each drug searcher had a heterogeneous impact on how the ban affected clicks on pharmacy websites. We find that the ban has no differential effect on queries that had on average older searchers or lower-income searchers. These tables are available upon request.

impacts shows that the effects on tier-B and tier-C websites are larger for H-drugs, lifestyle drugs, and drugs that treat non-chronic conditions.

3.7 Conclusion

We have shown that following the ban on non-NABP-certified pharmacies from sponsored search, there is a reduction in total clicks into the banned pharmacies. However, this effect is differential in several dimensions.

First, the websites certified by non-NABP agencies – referred to as tier-B sites – experience a reduction in total clicks, and some of their lost paid clicks are replaced by organic clicks. These effects do not change significantly before or after the Google-DOJ settlement. In contrast, pharmacies not certified by any of the four major certification agencies – referred to as tier-C sites – suffer the greatest reduction in both paid and organic clicks, and the reduction is exacerbated after the Google-DOJ settlement.

Second, we explore whether the effect of the ban depends on what drug names consumers search for on the Internet. Drug queries that led to more clicks on non-NABP-certified pharmacies before the ban are most affected by the ban, but chronic drug queries are less affected by the ban than non-chronic drugs.

Overall, we conclude that the ban has increased search cost for tier-B sites, but at least some consumers overcome the search cost by switching from paid to organic links. In addition to search frustration, our results suggest that the ban may have increased health concerns for tier-C sites and discouraged consumers from reaching them via both paid and organic links. It is also possible that tier-C sites are buried deeper in organic results than tier-A and tier-B sites, and the extra obscurity adds difficulty for consumers to switch to

organic links for tier-C sites.

Unfortunately, comScore data do not contain the rank information of search results following a specific query. Hence we cannot distinguish the effects of heightened health concerns from organic rank changes after the Google-DOJ settlement.

More generally, our study is limited to consumer search via search engines, as recorded in the comScore data. Due to the lack of individual click-through data, we do not know whether a consumer switches between drug, pharmacy and other queries after the ban of non-NABP-certified pharmacies from sponsored search. Nor do we know whether the banned pharmacies have engineered their organic results or the NABP-certified pharmacies have increased price or changed their advertising strategy after the ban. These supply side questions warrant further study.

3.8 Figures and Tables

Figure 1: Google Search Screenshot, Before the Ban

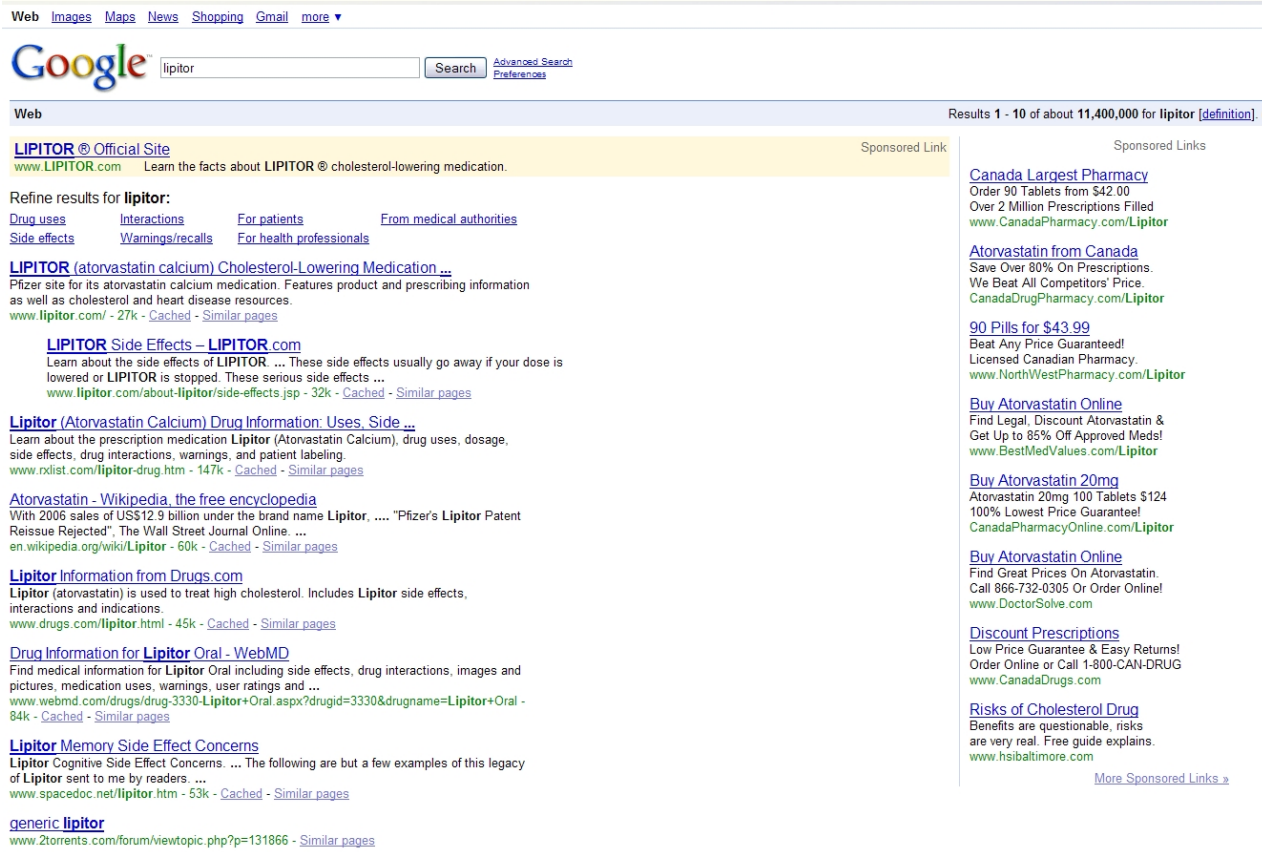


Figure 2: Google Search Screenshot, After the Ban

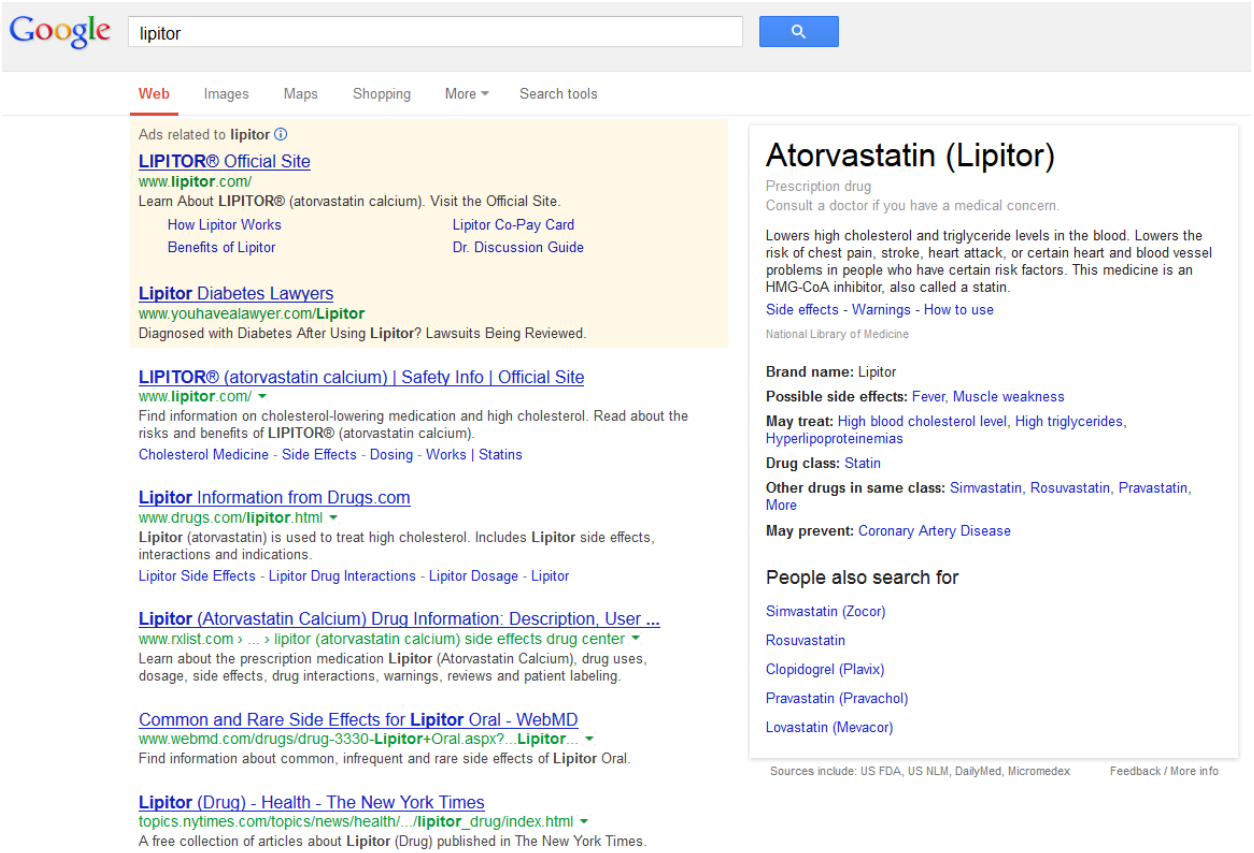


Figure 3: Example ComScore Data

Report:	Term Destinations
Query:	Lipitor
Date:	January 2012
Engine:	All
Match Option:	Match All Forms

Key Metrics

Total Clicks	169,156
Paid Clicks	38,670
Organic Clicks	130,486

Site Clicks

Entity Name	lipitor.com	Wal-Mart	walmart.com	…
Entity Level	Property	Property	Media Title	…
SubCategory	778218	778230	778230,778281	…
Organic Clicks	27,228	10,713	10,713	…
Paid Clicks	34,420	2,861	2,861	…

Report:	Term Profile
Query:	Lipitor
Date:	January 2012
Engine:	n/a
Match Option:	Match All Forms

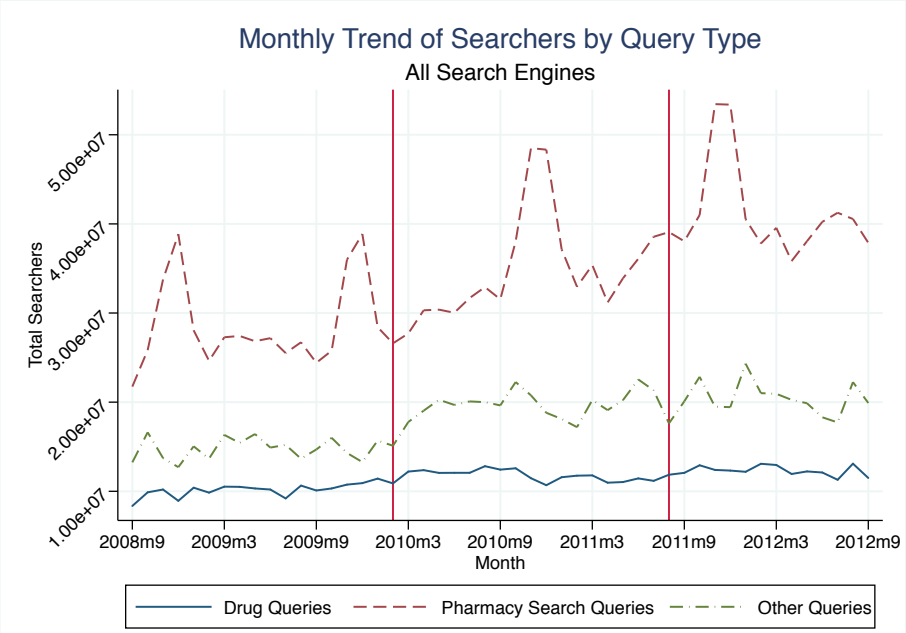
Key Metrics

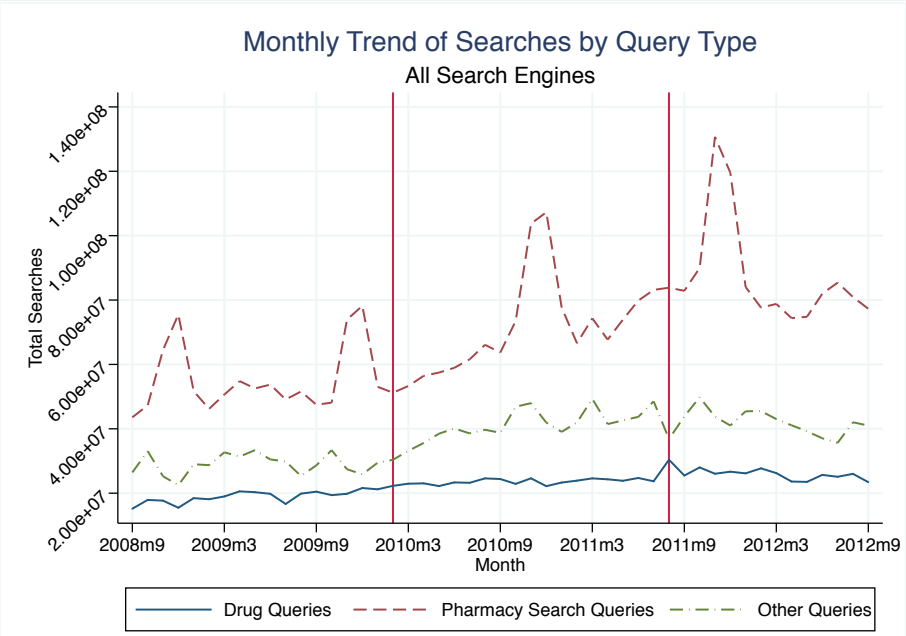
Searches	293,240
Searchers	219,414
Searches per Searcher	1.34

Demographics

Title	HoH Age	Income	Region	…
Level	45-54	$75k-99k	New England	…
Reach	40.15	15.65	2.21	…

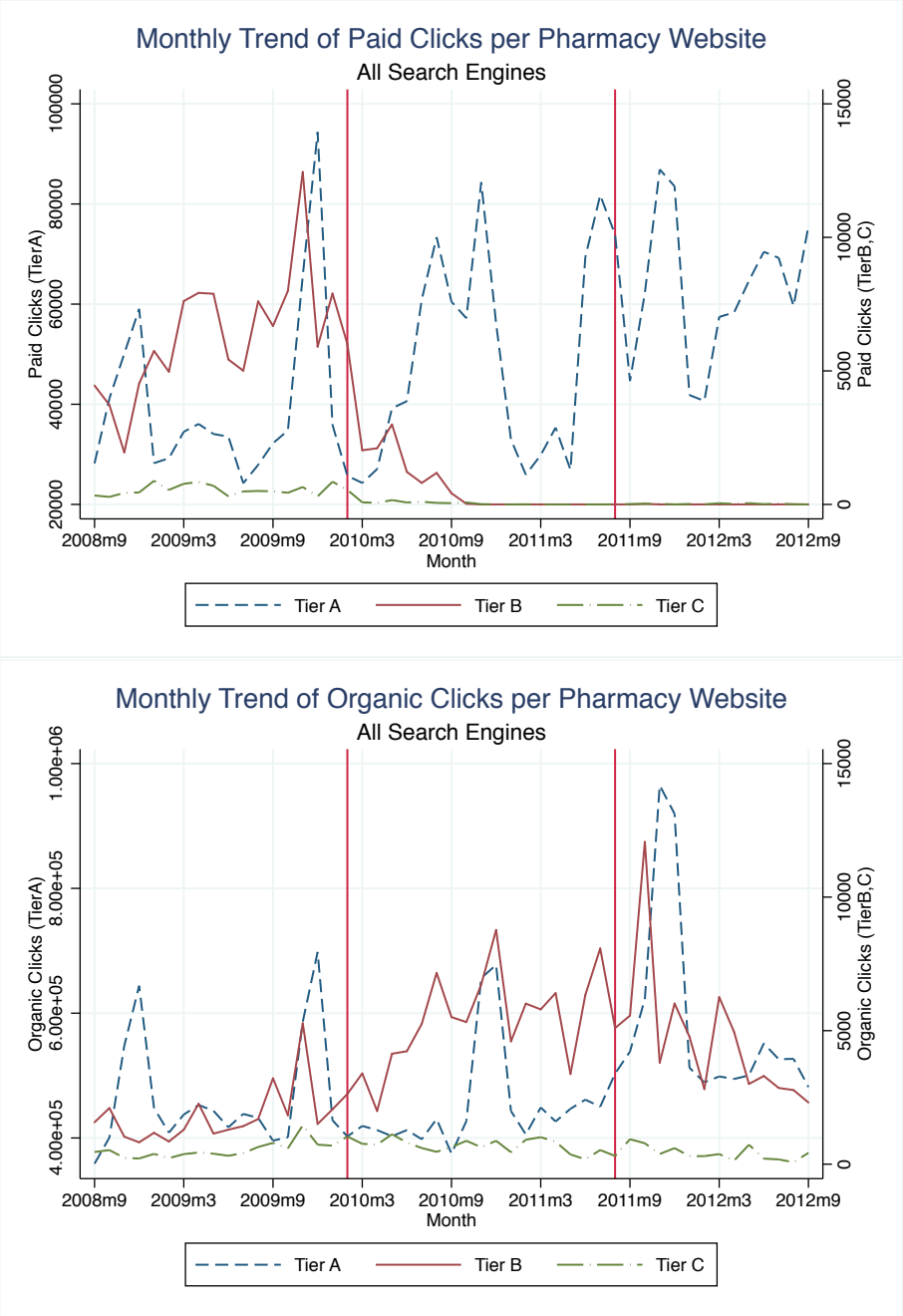
Figure 4: Searchers and Searches by Broad Query Type





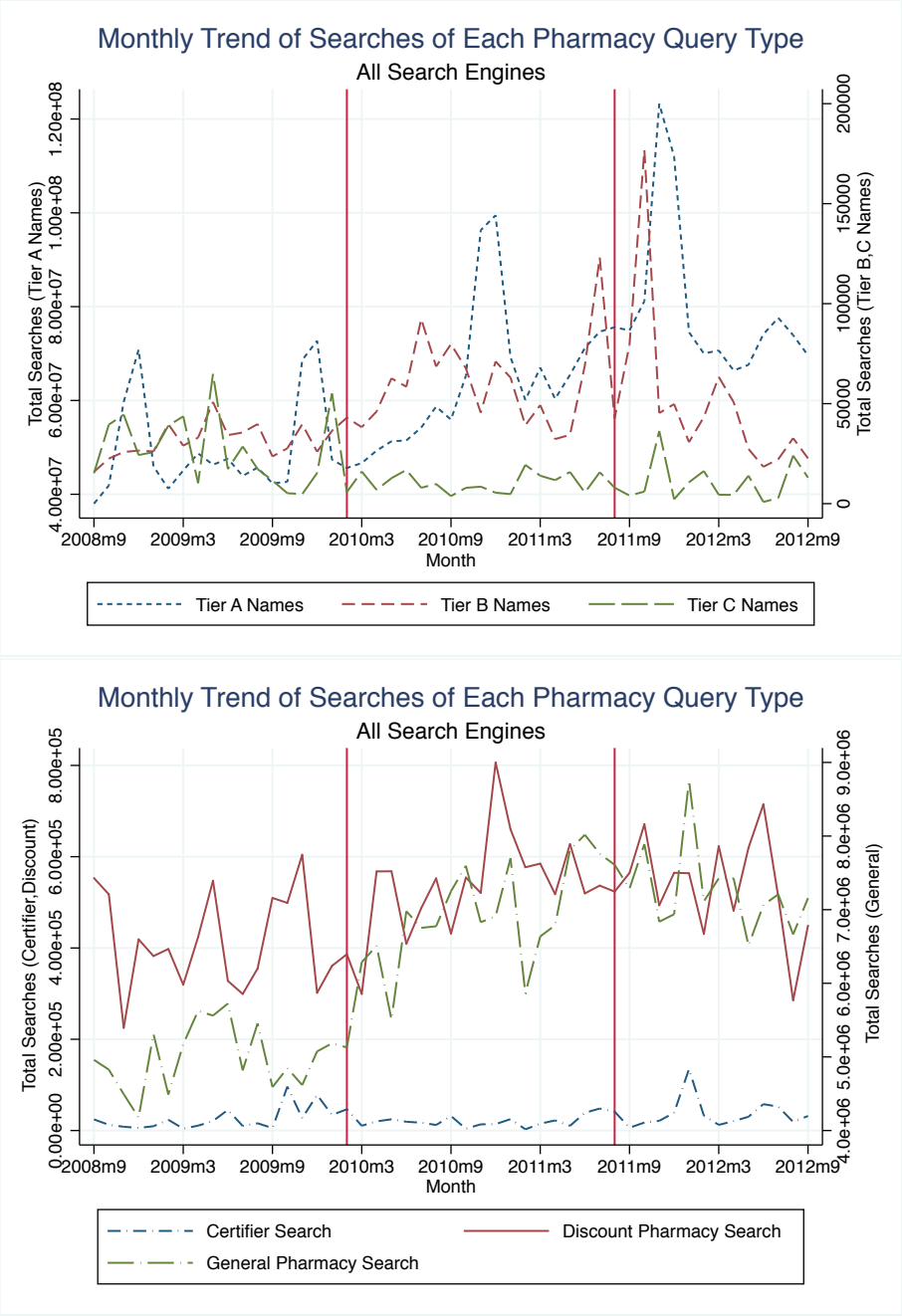
Notes: The top figure plots the total number of searchers of each query type in each month. The bottom figure plots the total number of searches of each query type in each month.

Figure 5: Clicks On Pharmacy Websites



Notes: 1. The figures plot the total monthly paid and organic clicks of each tier of online pharmacy website. The total clicks sum over all types of queries that lead to clicks on online pharmacies. 2. If the ban on sponsored links has been perfectly implemented, we should observe zero paid clicks from Tier-B and Tier-C websites in regime 2. The positive paid clicks on Tier-B websites are on "canadapharmacy.com" in November 2011, and on "northwestpharmacy.com" in August 2012. The positive paid clicks on Tier-C websites are from "freemedicine.com" and "albertsonssavonpharmacies.com".

Figure 6: Pharmacy Searches by Query Type



Notes: The top figure plots the total number of searches for each pharmacy tier in each month. The bottom figure plots the total number of searches for other pharmacy-related queries in each month.

Table 14: List of Events

Time	Event
before 2009	Google contracted with PharmacyChecker to filter out uncertified websites
July 2009	Some pharmacies advertising on Google were found to be uncertified by PharmacyChecker
August 2009	LegitScript.com and KnuhOn.com criticized Microsoft for allowing rogue pharmacies to advertise on Bing
November 2009	FDA issued 22 warning letters to website operators
February 9, 2010	Google began to ban non-NABP-certified pharmacies from sponsored ads for US consumers
April 21, 2010	Google contracted with LegitScript to implement the ban
June 10, 2010	Microsoft and Yahoo! started to ban non-NABP-certified pharmacies from sponsored ads for US consumers.
June 22, 2010	Google partnered with the National Institute of Health (NIH) and expanded its search tool to include drug facts with NIH links. This is only available to US consumers.
August 24, 2011	DOJ announced its settlement with Google

Table 15: Regimes

Regime	Time	Policy
Regime 0	September 2008 - January 2010	Google used PharmacyChecker to filter online pharmacy ads
Regime 1	March 2010 - July 2011	Google required NABP-certification and switched to LegitScript in place of PharmacyChecker
Regime 2	September 2011 - September 2012	Google reached an official settlement with DOJ

Notes: February 2010 and August 2011 are excluded because the imposition of the ban and the announcement of the settlement occurred in the middle of these two months.

Table 16: Query List

Query Group	Query Type	Count	Examples	Source
Pharmacy	General Pharmacy Keywords	6	pharmacy at	Keywordspy.com
	Discount Pharmacy Keywords	46	cheap drugs	Keywordspy.com
	TierA Pharmacy Names	9	cvs	comScore, cert. websites
	TierB Pharmacy Names	13	jandrugs	comScore, cert. websites
	TierC Pharmacy Names	19	canadamedicineshop	comScore, cert. websites
	Certifier Search	8	vipps	cert. websites
Drug	Prescription Drug Names	263	lipitor	FDA Orange Book, Keywordspy.com
Other	Drug Manufacturer	59	pfizer	Kantar Media
	Information/Gov.	5	fda	comScore
	Information/Info Sites	17	webmd	comScore
	Information/Health Terms	8	panic-anxiety	comScore
	Other Drugs/Non-Online Rx	17	renvela	FDA Orange Book
	Other Drugs/OTC Related	58	prevacid	FDA Orange Book
	Total Count	528		

No

February 2010 and August 2011 are excluded because the imposition of the ban and the announcement of the settlement occurred in these two months.

Table 17: Query Statistics: Overall Number of Searches and Clicks

		Total	PharmClicks/	%Pharmacy	Paid Clicks			Organic Clicks		
Query Type	Reg	Searches*	Search	Clicks	Tier-A	Tier-B	Tier-C	Tier-A	Tier-B	Tier-C
Pharmacy Queries										
General Pharmacy Search	*0*	832.6	9.6%	27.9%	94,325	20,843	6,692	306,419	6,312	13,792
	1	1,156.6	8.3%	39.7%	72,707	2,483	1,390	259,706	16,445	18,972
	2	1,208.7	6.5%	21.0%	88,117	0	222	268,329	10,373	17,160
Discount Pharmacy Search	*0*	9.0	38.9%	66.5%	932	5,889	776	3,673	2,900	3,815
	1	11.8	33.4%	58.5%	1,825	815	19	3,097	10,353	5,184
	2	11.7	26.3%	62.4%	1,512	1	0	3,571	10,370	3,166
Tier-A Pharmacy Names	*0*	5,546.1	49.8%	80.6%	230,232	71	20	2,883,102	55	183
	1	7,167.0	51.1%	78.2%	283,555	0	0	2,794,803	105	217
	2	8,853.2	45.1%	78.8%	380,141	0	0	3,793,243	794	568
Tier-B Pharmacy Names	*0*	2.4	50.2%	92.9%	632	366	98	2,088	652	96
	1	4.7	52.9%	93.0%	721	64	0	1,695	3,319	0
	2	3.9	50.2%	97.9%	958	0	0	740	3,543	0
Tier-C Pharmacy Names	*0*	1.4	47.2%	39.8%	0	0	160	0	0	250
	1	0.6	47.8%	31.4%	0	0	104	113	0	684
	2	0.6	0.0%	7.1%	0	0	0	0	0	15
Certifier Search	*0*	2.8	117.0%	6.5%	59	0	0	77	0	0
	1	2.2	0.9%	1.3%	0	0	0	44	0	0
	2	4.1	3.9%	1.5%	109	0	0	0	0	0
Drug Queries	*0*	71.9	14.1%	22.1%	273	1,039	1,092	6,348	63	578
	1	89.9	2.2%	2.6%	329	238	121	1,750	535	1,439
	2	97.6	2.6%	3.5%	559	2	111	2,171	713	1,344

* in thousands

Notes: 1. All statistics in this table are averaging across queries within each query type×month, and the statistics related to clicks are conditional on queries that led to any clicks on any pharmacy website. "Total Searches" is the average monthly searches per query. "PharmClicks/Search" is the average monthly (Pharmacy Website Clicks/Searches) ratio per query. "%Pharmacy Clicks" is the average monthly ratio of clicks on pharmacy websites to all clicks led from each query. Columns for paid clicks and organic clicks show the number of monthly clicks that land on each tier of pharmacy led from each query. 2. The large number of searches on Tier-A pharmacy names is due to the discount chains that also sell general products besides drugs. 3. The pharmacy clicks to search ratio for Tier-C queries in regime 2 is not precisely zero, but we cannot calculate the ratio due to censoring.

Table 18: Pharmacy Website Statistics

		Mean		Median		StdDev		25 percentile		75 percentile		N	N	N
	Regime	*paid*	*organic*	*paid*	*organic*	*paid*	*organic*	*paid*	*organic*	*paid*	*organic*	*active*	*(Paid>0)*	*(Organic>0)*
TierA	0	40,538	466,980	0	627	138,298	2,078,990	0	0	412	7,566	47	23	36
	1	48,571	452,544	0	680	206,487	2,075,955	0	0	132	8,071	50	19	39
	2	62,696	586,653	0	567	228,356	2,820,957	0	0	175	5,119	48	19	34
TierB	0	6,338	1,795	735	217	10,168	3,640	0	0	7,929	2,058	26	17	17
	1	633	5,476	0	824	1,105	10,870	0	108	1,137	3,712	27	13	24
	2	2	4,652	0	1,078	8	7,376	0	0	0	5,201	25	2	17
TierC	0	544	522	0	0	2,593	1,495	0	0	0	189	138	28	74
	1	39	694	0	0	244	2,932	0	0	0	56	132	14	59
	2	18	417	0	0	223	1,787	0	0	0	0	92	2	40

Notes: 1. The clicks number in the table are at each month×website level and the statistics are calculated among the sample within each website type×regime. We keep the balanced sample of websites, (57 tier-A websites, 28 tier-B websites, and 181 tier-C websites) in calculating the statistics. 2. We define *active websites* as websites having received either censored or positive clicks from the set of queries in our data. The last three columns report the number of websites in each regime that are active, have positive (non-censored) paid clicks, and have positive (non-censored) organic clicks.

Table 19: Regression Results: Clicks on Online Pharmacy Websites (from All Queries)

	(1)	(2)	(3)	(4)	(5)
	I(AnyClicks>0)	*I(TtlClicks)*	*Ln(TtlClicks)*	*I(OrgClicks>0)*	*Ln(OrgClicks)*
TierB	0.128	0.0990		-0.0780	
	(0.231)	(0.253)		(0.250)	
TierC	-0.534***	-0.788***		-0.895***	
	(0.159)	(0.170)		(0.168)	
Regime1	0.0520	0.0158	0.176	0.0158	0.199*
	(0.0484)	(0.0450)	(0.104)	(0.0449)	(0.108)
TierB×Regime1	0.0960	-0.144	-0.617**	0.0114	0.882***
	(0.160)	(0.134)	(0.253)	(0.122)	(0.245)
TierC×Regime1	-0.230***	-0.260***	-0.140	-0.172**	0.130
	(0.0769)	(0.0897)	(0.198)	(0.0843)	(0.186)
Regime2	-0.0231	-0.0871	0.151	-0.0924	0.146
	(0.0747)	(0.0692)	(0.130)	(0.0685)	(0.121)
TierB×Regime2	0.0668	-0.0384	-0.583**	0.149	1.136***
	(0.171)	(0.146)	(0.255)	(0.134)	(0.255)
TierC×Regime2	-0.480***	-0.424***	-0.0197	-0.323***	0.247
	(0.111)	(0.127)	(0.230)	(0.119)	(0.222)
Constant	0.0790	-0.189	9.043***	-0.194	8.508***
	(0.141)	(0.146)	(0.0489)	(0.146)	(0.0484)
Marginal Effect					
TierB×Regime1	0.0328	-0.037		0.0028	
	(0.0546)	(0.0345)		(0.0302)	
TierC×Regime1	-0.0785***	-0.0669***		-0.0426**	
	(0.0251)	(0.0228)		(0.0206)	
TierB×Regime2	0.0228	-0.0099		0.037	
	(0.0583)	(0.0376)		(0.0332)	
TierC×Regime2	-0.164***	-0.1092***		-0.08***	
	(0.0378)	(0.0329)		(0.0297)	
Observations	12,502	12,502	2,698	12,502	2,552
FE	-	-	Website	-	Website

Notes: Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
1. Dummy variables for Tier-A pharmacies, regime 0, and their interactions are excluded from the regression. 2. This table examines the differential changes in total and organic clicks outcome in each regime. Dependent variable in column (1) is if a website had any clicks, paid or organic, including censored clicks in a given month. Dependent variables in columns (2) and (4) are if a website has any non-censored positive total or organic clicks in a given month, respectively. Dependent variables in columns (3) and (5) are the number of non-censored positive total and organic clicks (respectively) on a website when the number of clicks is non-censored and positive. 3. Standard errors are clustered at the website level for all regressions. 4. In counting the total number of clicks into each website, we included clicks from all types of queries - pharmacy queries, drug queries and other queries.

Table 20: Regression Results: Searchers and Searches of Pharmacy Queries

	Ln(Searchers)	*Ln(Searches)*
Regime1×TierBQuery	-0.258	-0.260
	(0.585)	(0.598)
Regime1×TierCQuery	-1.487*	-1.550*
	(0.616)	(0.628)
Regime1×Certifier	-0.415	-0.426
	(0.482)	(0.485)
Regime1×General	-0.329	-0.252
	(0.555)	(0.573)
Regime1×Discount	-0.188	-0.151
	(0.498)	(0.504)
Regime1	0.612	0.624
	(0.468)	(0.472)
Regime2×TierBQuery	-0.687	-0.749
	(0.722)	(0.729)
Regime2×TierCQuery	-1.916**	-2.085**
	(0.659)	(0.663)
Regime2×Certifier	0.367	0.333
	(0.731)	(0.755)
Regime2×General	0.129	0.0982
	(0.687)	(0.699)
Regime2×Discount	-0.242	-0.281
	(0.619)	(0.623)
Regime2	0.418	0.475
	(0.583)	(0.585)
Constant	4.273***	4.456***
	(0.0758)	(0.0781)
Observations	4,794	4,794
Fixed Effects	Query	Query

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Notes: 1. Tier-A pharmacy names and regime 0 are excluded. 2. An observation is at the query×month level, and outcome variable is the log level of the total searchers and searches for a query in a month. 3. Standard errors are clustered at the query level.

Table 21: Regression Results: Total Clicks on Online Pharmacy Websites (from Pharmacy Queries)

	$I(TotalClicks > 0)$			$Ln(TotalClicks)$		
Covariates	1×	TierB ×	TierC ×	1×	TierB ×	TierC ×
Marginal Effect						
Regime1	0.0078	-0.0498***	-0.0215**	0.305	-0.108	-0.230
	(0.0063)	(0.0254)	(0.0146)	(0.170)	(0.311)	(0.395)
Regime2	-0.0017	-0.0451**	-0.0238	0.466**	1.925*	0.799*
	(0.0069)	(0.029)	(0.0181)	(0.147)	(0.761)	(0.323)
TierB Query	-0.112***	0.2005***	0.0709**	-6.382***	7.578***	6.809***
	(0.0085)	(0.0177)	(0.0168)	(0.779)	(0.842)	(0.923)
TierC Query	-0.5412***		0.5608***	-6.981***		7.741***
	(0.0135)		(0.0063)	(0.776)		(0.679)
Discount	-0.0644***	0.2385***	0.1635***	-4.294***	6.898***	5.832***
	(0.0072)	(0.0165)	(0.0123)	(0.998)	(1.078)	(1.039)
General	0.0375***	0.14***	0.0864***	-1.228	2.585**	1.639*
	(0.0062)	(0.0161)	(0.0116)	(0.725)	(0.775)	(0.783)
TierBQuery×Regime1	-0.0289***	0.0675***		-0.312	0.942	
	(0.0124)	(0.0296)		(0.238)	(0.507)	
TierCQuery×Regime1	0.2878***		-0.2946***	0.475		
	(0.0329)		(0.0338)	(0.626)		
Discount×Regime1	-0.0136**	0.0315	0.0143	-0.000350	0.155	0.0803
	(0.0103)	(0.028)	(0.0178)	(0.243)	(0.442)	(0.471)
General×Regime1	-0.0081	0.0187	0.0029	-0.181	-0.0185	0.484
	(0.0087)	(0.0275)	(0.0167)	(0.184)	(0.380)	(0.422)
TierBQuery×Regime2	-0.0539***	0.0814***		0.123	-1.254	
	(0.0165)	(0.0349)		(0.332)	(0.721)	
TierCQuery×Regime2	0.002***		-0.0689**			-2.351***
	(0)		(0.0339)			(0.341)
Discount×Regime2	-0.0229**	0.057**	0.0108	0.303	-2.456**	-1.434**
	(0.0116)	(0.0318)	(0.0216)	(0.387)	(0.766)	(0.496)
General×Regime2	-0.0071	0.003	-0.0291	-0.504**	-1.944**	0.104
	(0.0095)	(0.0312)	(0.0204)	(0.170)	(0.656)	(0.435)
Constant		-0.1471***	-0.1947***	8.424***		
		(0.013)	(0.0102)	(0.275)		
Observations	51,465			6,700		
FE	-			Website		

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We used a subsample of clicks on pharmacy websites following pharmacy-related queries. Dummy variables for query type "TierA Names", TierA pharmacies, regime 0, and their interactions are excluded in the regression. 2. The regressions examine the differential changes in the total clicks in each regime from different types of pharmacy queries. In the extensive margin specification, the dependent variable is whether a website recorded any non-censored clicks from one type of pharmacy query in a given month. In the intensive margin specification, the dependent variable is the number of clicks on a website from one type of pharmacy query at a given month, conditional on positive clicks. 3. Coefficients for the extensive margin regression are in the first three columns and the intensive margin regression are in the last three columns. The coefficients for the cross product with a TierB destination website are in columns (2) and (5) and the cross product with a TierC destination website are in columns (4) and (6). 4. Some coefficient estimates were not identified due to too few observations (e.g., comScore recorded no clicks on TierB pharmacies following a query for a TierC pharmacy name). 5. Standard errors are clustered at the website level for all regressions.

Table 22: Regression Results: Online Pharmacy Clicks from H-Drug Vs. L-Drug Queries

	(1)	(2)	(3)	(4)
	I(Ttlclicks>0)	*Ln(TtlClicks)*	*I(OrgClicks>0)*	*Ln(OrgClicks)*
Regime1	0.0095	-0.990	0.0046	-1.336**
	(0.0077)	(0.617)	(0.0083)	(0.591)
Regime2	-0.0088	-0.990***	-0.0071	-0.908
	(0.0108)	(0.566)	(0.009)	(0.748)
H-Drug	0.0593***	0.0287	0.0437***	-0.00259
	(0.0194)	(0.397)	(0.0166)	(0.318)
H-Drug×Regime1	-0.0223***	1.204**	-0.009	1.091
	(0.0095)	(0.524)	(0.0081)	(0.690)
H-Drug×Regime2	0.0025	1.623*	0.0121	1.017*
	(0.0167)	(0.301)	(0.0152)	(0.312)
TierB	-0.0104		-0.0957*	
	(0.0355)		(0.049)	
TierB×Regime1	-0.0526**	1.324	0.044	0.173
	(0.0249)	(0.895)	(0.0361)	(0.691)
TierB×Regime2	-0.0634***	1.716	0.0392	-0.0910
	(0.0263)	(1.095)	(0.0306)	(1.073)
H-Drug×TierB	0.0918***	1.464***	0.1206***	-1.622*
	(0.0304)	(0.819)	(0.0425)	(0.389)
H-Drug×TierB×Regime1	-0.0207	-2.425**	-0.0624	0.734
	(0.0247)	(1.029)	(0.0388)	(0.817)
H-Drug×TierB×Regime2	-0.0377	-3.554*	-0.0745**	0.620
	(0.0272)	(1.088)	(0.0358)	(0.842)
TierC	-0.0806**		-0.0797**	
	(0.039)		(0.039)	
TierC×Regime1	-0.0348*	2.330*	-0.009	2.845*
	(0.0182)	(0.859)	(0.0173)	(0.791)
TierC×Regime2	-0.0563*	2.598*	-0.0412	3.137*
	(0.0308)	(0.878)	(0.0311)	(0.936)
H-Drug×TierC	0.0776***	0.708	0.0816***	0.630
	(0.0293)	(0.566)	(0.0296)	(0.531)
H-Drug×TierC×Regime1	0.0006	-2.727*	-0.0189	-2.517*
	(0.0203)	(0.819)	(0.0196)	(0.901)
H-Drug×TierC×Regime2	-0.0145	-3.452*	-0.0213	-3.320*
	(0.0323)	(0.799)	(0.0341)	(0.722)
Constant		7.668*		7.747*
		(0.269)		(0.245)
Observations	14,060	921	14,060	754
FE	-	Website	-	Website

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Notes: 1. Dummy variables for Tier-A pharmacies, regime 0, and their interactions are excluded from the regression. 2. This table examines the heterogeneous changes in total and organic clicks in each regime resulting from H-Drug and L-Drug queries. The dependent variables in columns (1) and (3) are indicators if a website had any non-censored total or organic clicks in a given month, and the columns report the marginal effects of a probit regression. The dependent variables in columns (2) and (4) are the number of non-censored total and organic clicks on a website when the number of clicks is non-censored and positive. 3. H-Drug and L-Drug are defined by their ratio of clicks into Tier-B and Tier-C websites in the first nine months of the sample (2008/09 - 2009/05). 4. We exclude the first 9 months of observations from the sample as clicks during that time were used to define H and L drugs queries. 5. Some coefficient estimates were not identified due to too few observations. 6. Standard errors are clustered at the website level for all regressions.

Table 23: Regression Results: Online Pharmacies Clicks from Lifestyle Vs. Non-lifestyle Drug Queries

	(1)	(2)	(3)	(4)
	I(Ttlclicks>0)	*Ln(TtlClicks)*	*I(OrgClicks>0)*	*Ln(OrgClicks)*
Regime1	-0.0032	-0.207	0.0065	-0.713
	(0.0176)	(0.526)	(0.0128)	(0.555)
Regime2	-0.0173	0.00661	0.001	-0.515
	(0.0208)	(0.574)	(0.0163)	(0.596)
Lifestyle (LS)	-0.0359*	-0.308***	-0.0109	-0.320
	(0.019)	(0.171)	(0.0082)	(0.256)
LS×Regime1	0.0257*	0.116	0.0066	0.174
	(0.0151)	(0.241)	(0.0065)	(0.319)
LS×Regime2	0.0537***	0.290	0.0231	0.376
	(0.0211)	(0.270)	(0.0158)	(0.253)
TierB	0.0955***		0.0149	
	(0.038)		(0.03)	
TierB×Regime1	-0.114***	-0.0200	-0.0278	1.863*
	(0.0317)	(0.621)	(0.0218)	(0.693)
TierB×Regime2	-0.116***	-0.403	-0.0234	1.765**
	(0.0394)	(0.651)	(0.0289)	(0.791)
LS×TierB	0.0041	0.557	0.0138	0.583
	(0.0324)	(0.366)	(0.0285)	(0.369)
LS×TierB×Regime1	0.0172	-0.681	0.0026	-0.646
	(0.0305)	(0.541)	(0.0193)	(0.708)
LS×TierB×Regime2	-0.019	-0.860***	-0.0236	-0.704
	(0.0442)	(0.484)	(0.031)	(0.526)
TierC	-0.0436		-0.0332	
	(0.0346)		(0.0293)	
TierC×Regime1	-0.0657***	0.713	-0.0439**	1.291**
	(0.0264)	(0.568)	(0.0197)	(0.584)
TierC×Regime2	-0.0588	0.474	-0.0512*	0.900
	(0.0362)	(0.644)	(0.0278)	(0.667)
LS×TierC	0.0733***	0.760*	0.0392*	0.613***
	(0.0274)	(0.283)	(0.02)	(0.349)
LS×TierC×Regime1	0.0035	-0.626	0.0171	-0.366
	(0.0248)	(0.470)	(0.0189)	(0.490)
LS×TierC×Regime2	-0.0656*	-0.708	-0.0257	-0.437
	(0.0354)	(0.592)	(0.0288)	(0.633)
Constant		7.901*		7.390*
		(0.141)		(0.179)
Observations	18330	1439	18330	1064
FE	-	Website	-	Website

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Notes: 1. Dummy variables for Tier-A pharmacies, regime 0, and their interactions are excluded from the regression. 2. This table examines the heterogeneous changes in total and organic clicks in each regime led by lifestyle and non-lifestyle drug queries. The dependent variables in columns (1) and (3) are if a website has any non-censored positive total or paid clicks in a given month, and the columns report the marginal effects of the probit regression. The dependent variables in columns (2) and (4) are the number of non-censored positive total and paid clicks on a website when the number of clicks is non-censored and positive. 3. Some coefficient estimates were not identified due to too few observations. 4. Standard errors are clustered at the website level for all regressions.

Table 24: Regression Results: Online Pharmacy Clicks from Chronic Vs. Non-chronic Drugs Queries

	(1)	(2)	(3)	(4)
	I(Ttlclicks>0)	*Ln(TtlClicks)*	*I(OrgClicks>0)*	*Ln(OrgClicks)*
Regime1	0.0142	-0.137	0.0178	-0.730
	(0.0205)	(0.746)	(0.0174)	(0.802)
Regime2	0.0259	0.178	0.0303	-0.544
	(0.0254)	(0.914)	(0.0195)	(0.901)
Chronic	-0.0183	0.264	-0.0156	0.0257
	(0.0156)	(0.191)	(0.0101)	(0.370)
Chronic×Regime1	-0.0025	-0.857**	-0.0102	-0.553
	(0.0094)	(0.393)	(0.008)	(0.629)
Chronic×Regime2	-0.0187	-0.742*	-0.0169	-0.274
	(0.0197)	(0.278)	(0.0128)	(0.202)
TierB	0.0936***		0.0292	
	(0.0376)		(0.0333)	
TierB×Regime1	-0.1021***	-0.536	-0.0372	1.337
	(0.0306)	(0.801)	(0.0235)	(0.896)
TierB×Regime2	-0.1339***	-1.079	-0.0563**	1.380
	(0.0377)	(0.953)	(0.0258)	(0.948)
Chronic×TierB	-0.0118	-0.640	-0.0221	-0.409
	(0.0276)	(0.428)	(0.027)	(0.479)
Chronic×TierB×Regime1	-0.038	1.558**	0.0006	1.228
	(0.0233)	(0.758)	(0.0199)	(0.900)
Chronic×TierB×Regime2	0.0134	1.373*	0.026	1.009***
	(0.0364)	(0.516)	(0.0278)	(0.520)
TierC	0.0143		0.0092	
	(0.032)		(0.0276)	
TierC×Regime1	-0.0628***	0.452	-0.0415**	1.209
	(0.0265)	(0.801)	(0.0209)	(0.850)
TierC×Regime2	-0.1053***	0.181	-0.0789***	1.052
	(0.0327)	(0.948)	(0.0245)	(0.939)
Chronic×TierC	-0.0567***	-0.695*	-0.057***	-0.323
	(0.0239)	(0.239)	(0.0212)	(0.419)
Chronic×TierC×Regime1	-0.0012	1.325**	0.0196	0.791
	(0.021)	(0.521)	(0.0176)	(0.730)
Chronic×TierC×Regime2	0.0295	1.877*	0.0283	1.158***
	(0.0367)	(0.438)	(0.0265)	(0.596)
Constant		8.035*		7.639*
		(0.141)		(0.154)
Observations	16920	1171	16920	853
FE	-	Website	-	Website

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Notes: 1. Dummy variables for Tier-A pharmacies, regime 0, and their interactions are excluded from the regression. 2. This table examines the heterogeneous changes in total and organic clicks in each regime led by chronic and non-chronic drug queries. The dependent variables in columns (1) and (3) are if a website has any non-censored positive total or paid clicks in a given month, and the columns report the marginal effects of the probit regression. The dependent variables in columns (2) and (4) are the number of non-censored positive total and paid clicks on a website when the number of clicks is non-censored and positive. 3. Some coefficient estimates were not identified due to too few observations. 4. Standard errors are clustered at the website level for all regressions.

3.9 Appendix Tables

Table A1: Examples of H-Drugs and L-Drugs

Top 10 H-Drugs by Total Clicks				
Rank	Query	Total Clicks[a]	Tier-BC Ratio[b]	May Treat
1	viagra	2,890,258	88%	ED*
2	phentermine	2,140,199	52%	over weight, controlled substance
3	xanax	1,866,525	21%	depression, insomnia, controlled substance
4	cialis	1,056,012	87%	ED*
5	oxycodone	829,212	5%	pain, controlled substance
6	insulin	744,736	15%	diabetes
7	ambien	697,907	6%	sleep aid, controlled substance
8	effexor	656,777	6%	depression
9	cymbalta	648,823	10%	depression
10	oxycontin	553,726	16%	pain, controlled substance
Top 10 L-Drugs by Total Clicks				
Rank	Query	Total Clicks[a]	Tier-BC Ratio[b]	May Treat
1	coumadin	729,570	0%	blood clots
2	metoprolol	516,298	0%	high blood pressure
3	flexeril	409,765	0%	pain
4	keflex	307,195	0%	bacterial infections
5	skelaxin	243,452	0%	pain
6	bystolic	224,755	0%	high blood pressure
7	omnicef	184,677	0%	infections
8	strattera	138,808	0%	attention-deficit/hyperactivity disorder
9	zyprexa	133,542	0%	psychotic mental disorders
10	lupron	132,092	0%	advanced prostate cancer

* ED stands for erectile dysfunction.

Notes: [a] Total Clicks is the total number of clicks on online pharmacy websites following each search query from September 2008 to September 2011. The drugs in each category are ranked by this total number of clicks. [b] Tier-B,C ratio is the percentage of total clicks from each query that led to Tier-B and Tier-C sites in the first nine months of the sample (2008/09 - 2009/05). A drug query is defined as an H-Drug is the Tier-B,C ratio is greater than 3%, and is defined as L-Drug when the Tier-B,C ratio is smaller than 0.1%. In total, we have 79 H-Drug queries and 112 L-Drug queries.

Table A2: Examples of Lifestyle and Non-Lifestyle Drugs

Top 10 Lifestyle Drugs				
Rank	Query	Total Clicks[a]	Tier-BC Ratio[b]	May Treat
1	viagra	2,890,258	36.6%	ED*
2	phentermine	2,140,199	51.7%	over weight, controlled substance
3	xanax	1,866,525	20.3%	depression, insomnia, controlled substance
4	cialis	1,056,012	23.3%	ED*
5	oxycodone	829,212	5.1%	pain, controlled substance
6	ambien	697,907	6.4%	sleep aid, controlled substance
7	oxycontin	553,726	15.9%	pain, controlled substance
8	botox	420,769	0.7%	wrinkle, face lift
9	levitra	367,965	13.9%	ED*
10	soma	327,303	6.9%	pain and stiffness of muscle spasms
Top 10 Non-Lifestyle Drugs				
Rank	Query	Total Clicks[a]	Tier-BC Ratio[b]	May Treat
1	lexapro	1,053,639	0.0%	depression
2	zoloft	817,323	0.1%	depression
3	suboxone	811,330	1.6%	chronic pain
4	insulin	744,736	1.0%	diabetes
5	coumadin	729,570	0.0%	blood clots
6	effexor	656,777	0.5%	depression
7	cymbalta	648,823	0.3%	depression
8	prozac	639,980	1.5%	depression
9	synthroid	529,037	0.4%	hypothyroidism
10	metoprolol	516,298	0.0%	high blood pressure

* ED stands for erectile dysfunction.

Notes: [a] Total Clicks is the total number of clicks on online pharmacy websites following each search query from September 2008 to September 2011. The drugs in each category are ranked by the total number of clicks. [b] Tier-BC Ratio is the percentage of total clicks from the query that landed on TierB and TierC sites in the first nine months of the sample.

Table A3: Examples of Chronic and Non-Chronic Drugs

Top 10 Chronic Drugs					
Rank	Query	Total Clicks[a]	Tier-BC Ratio[b]	Prescription Freq.[c]	May Treat
1	lexapro	1,053,639	0.0%	5.5	depression
2	zoloft	817,323	0.1%	5.1	depression
3	effexor	656,777	0.5%	5.3	depression
4	cymbalta	648,823	0.3%	6.3	depression
5	oxycontin	553,726	15.9%	5.1	pain, controlled substance
6	synthroid	529,037	0.4%	5.7	hypothyroidism
7	metoprolol	516,298	0.0%	5.7	high blood pressure
8	gabapentin	507,686	1.0%	5.6	seizures
9	pristiq	440,084	2.3%	5.0	depression
10	seroquel	438846	0.8%	6.2	schizophrenia
Top 10 Non-Chronic Drugs					
Rank	Query	Total Clicks[a]	Tier-BC Ratio[b]	Prescription Freq.[c]	May Treat
1	viagra	2,890,258	36.6%	3.2	ED*
2	xanax	1,866,525	20.3%	2.5	depression, insomnia, controlled substance
3	cialis	1,056,012	23.3%	2.6	ED*
4	oxycodone	829,212	5.1%	3.4	pain, controlled substance
5	celexa	459,163	0.2%	1.0	depression
6	flexeril	409,765	0%	2.2	pain and stiffness of muscle spasms
7	levitra	367,965	13.9%	3.2	ED*
8	metronidazole	340,345	14.5%	1.9	bacterial infections
9	keflex	307,195	0%	1.5	bacterial infections
10	zithromax	295,800	45.6%	1.2	bacterial infections

* ED stands for erectile dysfunction.

Notes: [a] Total Clicks is the total number of clicks on online pharmacy websites following the search query from September 2008 to September 2011. The drugs in each category are ranked by the total number of clicks. [b] Tier-B,C ratio is the percentage of total clicks from each query that led to Tier-B and Tier-C sites in the first nine months of the sample (2008/09 - 2009/05). [c] Prescriptions Freq.(frequency) is the average number of prescriptions for each patient in a given year. It is calculated from 2010 Medical Expenditure Panel Survey and is weighted to reflect the national representative statistics. When the average number of prescriptions is higher than 5, we define the drug as chronic, while if it is below 3.5, we define the drug as non-chronic.

References

Ackerberg, D. A., & Botticini, M. (2002). Endogenous Matching and the Empirical Determinants of Contract Form. *Journal of Political Economy*, *110*(3), 564–591.

Adachi, H. (2000). On a Characterization of Stable Matchings. *Economics Letters*, *68*(1), 43–49.

Adams III, A. F., Ekelund Jr, R. B., & Jackson, J. D. (2003). Occupational Licensing of a Credence Good: the Regulation of Midwifery. *Southern Economic Journal*, 659–675.

Agarwal, N. (2012). An Empirical Model of the Medical Match. *Unpublished working paper, Harvard*.

Alvesson, M. (1994). Talking in Organizations: Managing Identity and Impressions in An Advertising Agency. *Organization Studies*, *15*(4), 535–563.

Anand, B. N., & Galetovic, A. (2006). Relationships, Competition and The Structure of Investment Banking Markets*. *The Journal of Industrial Economics*, *54*(2), 151–199.

Aobdia, D. (2012). Proprietary Information Spillovers and Auditor Choice. *Unpublished working paper, UCLA*.

Asker, J., & Ljungqvist, A. (2010). Competition and the Structure of Vertical Relationships in Capital Markets. *Journal of Political Economy*, *118*(3), 599–647.

Azoulay, P. (2004). Capturing Knowledge Within and Across Firm Boundaries: Evidence from Clinical Development. *American Economic Review*, 1591–1612.

Baccara, M. (2007). Outsourcing, Information Leakage, and Consulting Firms. *The Rand Journal of Economics*, *38*(1), 269–289.

Baccara, M., & Imrohorolu, A. (2012). A Field Study on Matching with Network Externalities. *The American Economic Review*, *102*(5), 1773–1804.

Bate, R., Jin, G. Z., & Mathur, A. (2013). In Whom We Trust: The Role of Certification Agencies in Online Drug Markets. *The BE Journal of Economic Analysis & Policy*, *14*(1), 111–150.

Becker, G. S., Murphy, K. M., & Grossman, M. (2006). The Market for Illegal Goods: the case of Drugs. *Revista de Economía Institucional*, *8*(15), 17–42.

Blake, T., Nosko, C., & Tadelis, S. (2014). *Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment* (Tech. Rep.). National Bureau of Economic Research.

Chan, D. X., Kumar, D., Ma, S., & Koehler, J. (2012). *Impact Of Ranking Of Organic Search Results On The Incrementality Of Search Ads*. Citeseer.

Chaudhuri, S., Goldberg, P. K., & Jia, P. (2003). *Estimating the Effects of Global Patent Protection in Pharmaceuticals: a Case Study of Quinolones in India* (Tech. Rep.). National Bureau of Economic Research.

Chiou, L., & Tucker, C. (2011). How Does Pharmaceutical Advertising Affect Consumer Search? *Available at SSRN 1542934*.

Dobkin, C., & Nicosia, N. (2009). The War on Drugs: Methamphetamine, Public Health, and Crime. *The American economic review*, *99*(1), 324.

Echenique, F., & Oviedo, J. (2004). A Theory of Stability in Many-to-Many Matching Markets.

Fox, J. T. (2008). *Estimating Matching Games with Transfers* (Tech. Rep.). National Bureau of Economic Research.

Fox, J. T. (2010). Identification in matching games. *Quantitative Economics*, *1*(2), 203–254.

Fox, S. (2004). Prescription Drugs online. Pew Internet and American Life Project. *October.: www. pewinternet. org*.

George, L. M., & Hogendorn, C. (2013). Local News Online: Aggregators, Geo-Targeting and the Market for Local News. *Geo-Targeting and the Market for Local News (November 1, 2013)*.

Ghose, A., & Yang, S. (2009). An Empirical Analysis of Search Engine Advertising: Sponsored Search in Electronic Markets. *Management Science*, *55*(10), 1605–1622.

Gilbert, D., Walley, T., & New, B. (2000). Lifestyle Medicines. *BMJ: British Medical Journal*, *321*(7272), 1341.

Hafalir, I. E. (2007). Efficiency in Coalition Games with Externalities. *Games and Economic Behavior*, *61*(2), 242–258.

Hafalir, I. E. (2008). Stability of Marriage with Externalities. *International Journal of Game Theory*, *37*(3), 353–369.

Hatfield, J. W., & Kojima, F. (2010). Substitutes and Stability for Matching with Contracts. *Journal of Economic Theory*, *145*(5), 1704–1723.

Hatfield, J. W., & Milgrom, P. R. (2005). Matching with Contracts. *American Economic Review*, 913–935.

Hausman, J. A., Abrevaya, J., & Scott-Morton, F. M. (1998). Misclassification of the Dependent Variable in a Discrete-response Setting. *Journal of Econometrics*, *87*(2), 239–269.

Horsky, S. (2006). The Changing Architecture of Advertising agencies. *Marketing Science*, *25*(4), 367–383.

Jansen, B. J., Brown, A., & Resnick, M. (2007). Factors Relating to the Decision to Click on a sponsored Link. *Decision Support Systems*, *44*(1), 46–59.

Jansen, B. J., & Spink, A. (2009). Investigating Customer Click through Behaviour with Integrated Sponsored and Nonsponsored Results. *International Journal of Internet Marketing and Advertising*, *5*(1), 74–94.

Kelso Jr, A. S., & Crawford, V. P. (1982). Job Matching, Coalition Formation, and Gross Substitutes. *Econometrica: Journal of the Econometric Society*, 1483–1504.

Kleinrock, M. (2011). The Use of Medicines in the United States: Review of 2010 Parsippany. *NJ: IMS Institute for Healthcare Informatics*.

Kóczy, L. Á. (2009). Sequential Coalition Formation and the Core in the Presence of Externalities. *Games and Economic Behavior*, *66*(1), 559–565.

Kornfield, R., Donohue, J., Berndt, E. R., & Alexander, G. C. (2013). Promotion of Prescription Drugs to Consumers and Providers, 2001–2010. *PloS one*, *8*(3), e55504.

Law, M. T., & Kim, S. (2005). Specialization and Regulation: The Rise of Professionals and the Emergence of Occupational Licensing Regulation. *The Journal of Economic History*, *65*(03), 723–756.

Law, M. T., & Marks, M. (2008). The effects of occupational licensing laws on minorities: Evidence from the progressive era. *Available at SSRN 943765*.

Leland, H. E. (1979). Quacks, Lemons, and Licensing: A Theory of Minimum Quality Standards. *The journal of political economy*, 1328–1346.

Miron, J. A. (2003). The Effect of Drug Prohibition on Drug Prices: Evidence from the Markets for Cocaine and Heroin. *Review of Economics and Statistics*, *85*(3), 522–530.

Mohammad, A., Berndt Ernst, R., Davis James, C., et al. (2012). The Unbundling of Advertising Agency Services: An Economic Analysis. *Review of Marketing Science*, *10*(1), 1–55.

Morgan, R. M., & Hunt, S. D. (1994). The commitment-trust theory of relationship marketing. *the journal of marketing*, 20–38.

Mullner, R., & Gurau, C. (2005). Pharmaceutical Marketing on the Internet: Marketing Techniques and Customer Profile. *Journal of Consumer Marketing*, *22*(7), 421–428.

Orizio, G., Merla, A., Schulz, P. J., & Gelatti, U. (2011). Quality of Online Pharmacies and Websites Selling Prescription Drugs: a Systematic Review. *Journal of medical Internet research*, *13*(3).

Pashigian, B. P. (1977). *Occupational Licensing and the Inter-State Mobility of Professionals.* National Bureau of Economic Research Cambridge, Mass., USA.

Peltzman, S. (1976). *Toward a More General Theory of Regulation.* National Bureau of Economic Research Cambridge, Mass., USA.

Pycia, M. (2012). Stability and Preference Alignment in Matching and Coalition Formation. *Econometrica*, *80*(1), 323–362.

Quon, B. S., Firszt, R., & Eisenberg, M. J. (2005). A Comparison of Brand-name Drug Prices between Canadian-based Internet Pharmacies and Major US Drug Chain Pharmacies. *Annals of Internal Medicine*, *143*(6), 397–403.

Riordan, M. H., & Williamson, O. E. (1985). Asset Specificity and Economic Organization. *International Journal of Industrial Organization*, *3*(4), 365–378.

Rogan, M. (2013). Too Close for Comfort? The Effect of Embeddedness and Competitive Overlap on Client Relationship Retention Following an Acquisition. *Organization Science*, *25*(1), 185–203.

Romano, J. P., & Shaikh, A. M. (2010). Inference for the Identified Set in Partially Identified Econometric Models. *Econometrica*, *78*(1), 169–211.

Santiago, D., & Fox, J. T. (2008). A Toolkit for Matching Maximum Xcore Estimation and Point and Set Identified Subsampling Inference. *Unpublished paper (October). National Bureau of Economic Research and University of Chicago*.

Sasaki, H., & Toda, M. (1996). Two-sided Matching Problems with Externalities. *Journal of Economic Theory*, *70*(1), 93–108.

Shaikh, A. M. (2006). *Inference for Partially Identified Econometric Models* (Unpublished doctoral dissertation). Stanford University.

Shapiro, C. (1986). Investment, Moral Hazard, and Occupational Licensing. *The Review of Economic Studies*, *53*(5), 843–862.

Shapley, L. S., & Shubik, M. (1971). The Assignment Game I: The core. *International Journal of Game Theory*, *1*(1), 111–130.

Shepard, L. (1978). Licensing Restrictions and the Cost of Dental Care. *Journal of Law and Economics*, 187–201.

Silk, A. J., & Berndt, E. R. (1994). Costs, Institutional Mobility Barriers, and Market Structure: Advertising Agencies as Multiproduct Firms. *Journal of Economics & Management Strategy*, *3*(3), 437–480.

Silk, A. J., & King III, C. (2013). How Concentrated is the US Advertising and Marketing Services Industry? Myth Versus Reality. *Journal of Current Issues & Research in Advertising*, *34*(1), 166–193.

Siman, S. J. (1989). *Agency-client Relationships Before and After Mergers in the Advertising Industry* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Skinner, B. J. (2005). *Canada's Drug Price Paradox: The Unexpected Losses Caused by Government Interference in Pharmaceutical Markets*. Fraser Institute.

Skinner, B. J. (2006). *Price Controls, Patents, and Cross-border Internet Pharmacies: Risks to Canada's Drug Supply and International Trading Relations*. Fraser Institute.

Sørensen, M. (2007). How Smart is Smart Money? A Two-sided Matching Model of Venture Capital. *The Journal of Finance*, *62*(6), 2725–2762.

Stigler, G. J. (1971). The Theory of Economic Regulation. *The Bell journal of economics and management science*, 3–21.

Uetake, K., & Watanabe, Y. (2012). Entry by Merger: Estimates from a Two-Sided Matching Model with Externalities. *Available at SSRN 2188581*.

Villas-Boas, J. M. (1994). Sleeping with the Enemy: Should Competitors Share the Same Advertising Agency? *Marketing Science*, *13*(2), 190–202.

Waldfogel, J., & Chen, L. (2006). Does Information Undermine Brand? Information Intermediary Use and Preference for Branded Web Retailers. *The Journal of Industrial Economics*, *54*(4), 425–449.

Yang, S., & Ghose, A. (2010). Analyzing the Relationship between Organic and Sponsored Search Advertising: Positive, Negative, or Zero Interdependence? *Marketing Science*, *29*(4), 602–623.

Yang, Y., Shi, M., & Goldfarb, A. (2009). Estimating the Value of Brand Alliances in Professional Team Sports. *Marketing Science*, *28*(6), 1095–1111.