BANNING FOREIGN PHARMACIES FROM SPONSORED SEARCH:
THE ONLINE CONSUMER RESPONSE

Matthew Chesnes
Weijia (Daisy) Dai
Ginger Zhe Jin

Banning Foreign Pharmacies from Sponsored Search: The Online Consumer Response
Matthew Chesnes, Weijia (Daisy) Dai, and Ginger Zhe Jin
NBER Working Paper No. 20088
May 2014, Revised February 2018
JEL No. D83,I18,K32,L81

ABSTRACT

Increased competition from the internet has raised concerns about the quality of prescription drugs sold online. Given the pressure from the Department of Justice, Google agreed to ban pharmacies not certified by the National Association of Boards of Pharmacy (NABP) from sponsored search listings. Using comScore click-through data originated from health-related queries, we study how the ban affects consumer search and click behavior in a difference-in-differences framework using the synthetic control method. 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 (other-certified websites), experience an increase in organic clicks that partially offsets the loss in paid clicks after the ban. In contrast, pharmacies not certified by any certification agencies experience much lower rates of substitution in organic clicks. These results suggest that the ban has increased the search cost for other-certified websites, but at least some consumers overcome the search cost by switching from sponsored to organic links. The lower substitution for uncertified websites may be explained by the rising consumer concerns about the quality of drugs sold on uncertified websites after the ban.

Matthew Chesnes
Federal Trade Commission
mchesnes@ftc.gov

Weijia (Daisy) Dai
University of Maryland
Department of Economics
3114 Tydings Hall
College Park, MD 20742
daisy.w.dai@gmail.com

Ginger Zhe Jin
University of Maryland
Department of Economics
3115F Tydings Hall
College Park, MD 20742-7211
and NBER
jin@econ.umd.edu

1 Introduction

The growth of the internet has led to a dramatic rise in the number of retailers available to consumers in many industries. While the proliferation of competition may benefit consumers in several ways including lower prices, there is also the concern that the quality of 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.

Partially due to the high price of brand-name prescription drugs, U.S. consumers have been motivated to search online for cheaper supplies from foreign pharmacies, despite the fact that personal importation is illegal.[1] Search engines, such as Google, are an important gateway to accessing online pharmacies. An investigation by the DOJ revealed that, as early as 2003, Google was allowing unapproved pharmacies to purchase sponsored links and target U.S. consumers.[2] Due to the heightened concern to protect consumers, Google agreed to ban pharmacies not certified by the National Association of Boards of Pharmacy (NABP) from their sponsored search listings in February 2010.[3] In August 2011, Google settled with the DOJ, forfeiting $500 million in revenue earned from the ads.[4] Because the search engine ban only applied to sponsored links for non-NABP-certified pharmacies and these websites potentially continued to appear in the organic search results, the ban provides an excellent opportunity to study how consumers substitute between organic and sponsored links and shed light on the effectiveness of sponsored search advertising. More generally, this paper examines

---

[1]The Federal Food, Drug, and Cosmetic Act (FD&C Act) prohibits the importation of unapproved drugs into the U.S. See http://www.fda.gov/RegulatoryInformation/Legislation/FederalFoodDrugandCosmeticActFDCAct.

[2]When a user submits a query into a search engine, in addition to providing a list of relevant (i.e., organic) links based on the engine's search algorithm, additional links are returned that are based on their relevance to the query and a payment made by the link's owner. These latter results are called sponsored or paid links.

[3]While the announcement of the settlement singles out Canadian pharmacies, which face regulations within Canada, importation of drugs from Canada into the U.S. is still illegal because the FDA cannot ensure their safety and effectiveness. In addition, some pharmacies that claim to be based in Canada are actually selling drugs from other foreign countries that may lack sufficient regulation.

[4]See http://www.justice.gov/opa/pr/2011/August/11-dag-1078.html, retrieved December 28, 2013. Other search engines adopted a similar ban on sponsored search listings soon after Google's ban, but the settlement was only between Google and the DOJ.

how consumers' search and click activities have changed after the ban on non-NABP-certified pharmacies from sponsored search advertising.

While the FDA does not vigilantly enforce the ban on personal drug imports,[5] the NABP emphasizes the illegality of buying foreign drugs and highlights the dangers of rogue pharmacies. Independent research has confirmed some of the NABP's concerns. In particular, a review article by Orizio et al. (2011), suggests that geographic characteristics were concealed in many uncertified websites, some websites sold drugs without a prescription, and an online questionnaire was a frequent tool used to replace a prescription. Regarding 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.

Following the ban, the extent of organic substitution depends on a number of factors. The ban may raise concerns among consumers about the safety of drugs purchased from non-NABP-certified pharmacies. Consumers may be discouraged from clicking on these pharmacies, even if their organic link is readily available. We expect that this effect may be more salient after the Google-DOJ settlement, as the settlement was reported in the news more widely than the search engine ban itself. In comparison, the ban increases the technical difficulty of accessing a banned pharmacy website via the search engine: a website that appeared in the sponsored search before the ban but did not rank high in the organic results may lose its visibility and consumer attention. Other factors depend on how search engine users change their behavior to combat the increased search cost. Some consumers may be willing to exert effort to look for the organic links of the banned websites if the expected benefits from these banned websites (e.g., lower cost drugs and more privacy protection) exceeds the perceived risk (e.g., drug safety concerns, drug efficacy and the stigma of doing something illegal). In addition, banned pharmacy websites may have strong incentives to "manage" their organic ranks, especially those that disproportionately relied on sponsored ads before the ban.

---

[5]The FDA defines personal drug imports as those that represent a reasonable risk and are intended for personal use of no more than a 3-month supply.

Keeping these factors in mind, we classify pharmacy websites into three tiers: NABP-certified (tier-A), other-certified (tier-B), and uncertified (tier-C). NABP-certified websites refer to U.S. pharmacies that receive approval from NABP or the NABP-endorsed certifier, LegitScript.[6] NABP-certified websites are allowed to advertise in sponsored search listings before and after the ban. Other-certified websites refer to foreign or domestic pharmacies that are certified by PharmacyChecker.com or the Canadian International Pharmacy Association (CIPA), but not by NABP or LegitScript. All other online pharmacies are classified as uncertified websites.[7] Although both other-certified and uncertified websites are banned from Google's sponsored search listings after February 2010, we distinguish them for two reasons: first, uncertified websites 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 websites may be subject to a higher safety standard in the eyes of consumers that purchase drugs online (as evidenced in Bate et al. 2013) and therefore the ban could trigger different organic substitution patterns for tier-B and tier-C websites. The substitution pattern can also differ between tier-B and tier-C websites because, in general, tier-B websites receive more clicks and may appear higher in the organic ranks and therefore are easier to find after the ban.

To study consumer search and click activity for pharmacy websites, we use data from the comScore Search Planner suite from September 2008 to September 2012, which provides click-through data generated by queries on the major search engines. We start with a large set of queries and gradually narrow down to a subset of health-related queries that contribute to most of the website traffic through search engines. Our main empirical strategy employs a synthetic control method that explores the differential changes in click patterns between the affected pharmacy websites and the control websites before and after the sponsored search ban and the Google-DOJ settlement. Synthetic controls are constructed using clicks on

---

[6]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.

[7]While most of our analysis focuses on tier-B and tier-C pharmacies, we also consider the effects on tier-A pharmacies in some specifications.

health-related non-pharmacy websites following the same set of queries used to generate our treated group: clicks on pharmacy websites.

Data analysis suggests that the banned pharmacies experience a reduction in the number of total clicks through search engines after the ban, but the effect is heterogeneous. In particular, tier-B websites experience a statistically significant increase in organic clicks after the sponsored search ban, partially offsetting the reduction in sponsored clicks. The pattern of increased organic clicks continues after the Google-DOJ settlement. In contrast, tier-C websites experience much less substitution to organic clicks after the ban, and the effect is not statistically significant. After the Google-DOJ settlement, tier-C websites suffer a decline in the probability of receiving any positive organic clicks in a given month. We also explore the differences between tier-C and tier-B website substitution conditioning on a website's organic click volume before the ban, its reliance on sponsored clicks before the ban, and consumers' search intention as reflected in specific pharmacy or drug queries. Overall, we conclude that the ban has increased search cost for tier-B websites, but at least some consumers overcome the search cost by switching from sponsored to organic links. This also suggests that sponsored advertising is effective for these websites although some of the sponsored clicks are merely replacing clicks that would have otherwise been generated from organic links. While we do not observe the quality of drugs offered by each pharmacy, the ban may have also increased health or safety concerns for tier-C websites, which may explain why there is little substitution to organic links for these websites.

The paper proceeds as follows. In section 2, we provide background on the online market for prescription drugs, the role of search engines as gateways to this market, as well as changes to Google's policy regarding sponsored search ads from online pharmacies. We describe the data and present descriptive statistics in section 3. Our econometric framework using synthetic controls and a two-part difference in differences model is presented in section 4, while results for both the average treatment effects and heterogeneous effects are shown in section 5. Section 6 concludes.

2 Background

2.1 The Online Market for Prescription Drugs

A literature review by Orizio et al. (2011) found that the percent of the 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 large, given the size of the U.S. prescription drug market.[8] According to Skinner (2006), sales to U.S. consumers from 278 Canadian or seemingly-Canadian pharmacies reached CDN$507 million in the 12-month period ending June 2005.[9] More than half of the sales were of top-selling brand-name prescription drugs consumed primarily by seniors.

One major concern associated with purchasing from online pharmacies 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.[10] 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."[11] Therefore, a consumer who purchases a drug from a foreign pharmacy for personal use faces some uncertainty regarding the likely reaction by the FDA.

---

[8] According to IMS, prescription drug sales in the U.S. totaled $310 billion in 2015, an increase of 8.5% over 2014.

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

[10] In particular, the NABP study reviewed 7,430 internet pharmacies as of December 2010 and found 96% of them operating out of compliance with U.S. state and federal laws and/or NABP patient safety and pharmacy practice standards. Among these non-NABP-recommended pharmacies, 34% had server locations in a foreign country, 27% had a physical address out of U.S., 56% did not provide any physical address, 84% did not require a valid prescription, 62% issued prescriptions via online consultation, 50% offered foreign or non-FDA-approved drugs, 83% did not offer medical consultation, and 16% did not have secure websites.

[11] See http://www.fda.gov/ICECI/ComplianceManuals/RegulatoryProceduresManual/ucm179266.htm.

To address the safety concerns, the FDA publicizes anecdotes about unsafe pharmaceuticals on the internet and warns consumers about rogue websites (which could be foreign or domestic). They also advise consumers to avoid any foreign websites and only make online purchases from the U.S. websites certified by the NABP. The NABP certification ensures that U.S. websites comply with laws in both the state of their business operation and the states to which they ship. As of February 29, 2012, the NABP certified 30 online pharmacies, 12 of which are run by large pharmacy benefits management 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, was endorsed by the NABP to screen pharmacy websites after the Google ban. As of November 2016, LegitScript monitors over 80,000 internet pharmacies and estimates that between 30,000 and 35,000 are actively selling prescription drugs at any one time. Among active websites, 96% do not satisfy LegitScript's certification criterion and therefore are not operating lawfully.[12] 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 endorses the use of LegitScript by domain name registrars to assist in identifying illegally operating websites, throughout this paper we treat LegitScript the same as NABP and label websites certified by either agency as NABP-certified.

The other two private certifiers – PharmacyChecker.com and the Canadian International Pharmacy Association (CIPA) – are fundamentally different from NABP and LegitScript. CIPA is a trade association of Canadian pharmacies and only certifies Canadian websites that comply with Canadian laws, while PharmacyChecker covers U.S., Canada, and many other countries. Similar to NABP, PharmacyChecker also charges fees for an approved website to

---

[12]LegitScript's certification criterion includes a valid license with local U.S. jurisdictions, a valid registration with the U.S. Drug Enforcement Administration (DEA) if dispensing controlled substances, valid contact information, a valid domain name registration, a valid prescription, only dispensing FDA approved drugs, and protecting user privacy according to the HIPAA Privacy Rule (45 CRF 164).

be listed on PharmacyChecker.com beyond a short period of initial approval. As of March 9, 2012, PharmacyChecker has approved 73 foreign websites and 51 U.S. websites.[13] 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.

Among the four certification agencies, PharmacyChecker is the only one that provides head-to-head drug price comparison across online pharmacies. As detailed below, Google used to contract with PharmacyChecker to filter websites listed in its sponsored search results, but switched to NABP/LegitScript after it agreed to ban non-NABP-certified pharmacies in February 2010.

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 tier-B samples were 49.2% cheaper than tier-A samples after controlling for other factors. In comparison, samples from tier-C websites were 54.8% cheaper than those from tier-A websites, but eight tier-C samples failed the authenticity test, compared to zero failures in the tier-A and tier-B samples.[14]

Consistent with Bate et al. (2013), other studies also suggest that a lower price for brand-name prescription drugs is an important incentive for U.S. consumers to shop online. According to 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

---

[13]Certification requirements on PharmacyChecker include that any approved website has a valid pharmacy license from its local pharmacy board, requires a prescription for U.S. 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.

[14]The price difference was mostly driven by non-Viagra drugs. There was no significant price difference across tiers for Viagra, but all the tier-C failures were Viagra.

brand-name drugs were on average 43% below U.S. prices for the same drugs.[15] Quon et al. (2005) compared 12 Canadian internet pharmacies with three major online U.S. drug chain pharmacies and found that Americans can save an average of 24% on the 44 most-commonly purchased brand-name medications from Canada.[16]

Fox (2004) reported that 75% of drugs bought online were for chronic conditions, while the remainder were for other purposes, such as weight loss and sexual performance substances. Consistently, Skinner (2006) found resemblance between the top five therapeutic categories used by U.S. seniors and the top five therapeutic categories in the cross-border online sales from Canada to the U.S. This suggests that seniors are an important source of demand for Canadian pharmacies.

Bate et al. (2013) report the results of a survey of RxRights members about their purchases of prescription drugs. The results show that of 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.[17] Surprisingly, most respondents had medical insurance and/or some prescription drug coverage, and the percentage of being insured was not lower among online shoppers.

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 lowest price. Consistently, most online shoppers restrict themselves to

---

[15]This number has been 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 U.S. This explains why most cross-border sales from Canada to the U.S. are brand-name drugs.

[16]The large price difference between the U.S. 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 and federal legislators in favor of legalizing or facilitating the cross-border drug trade with Canada. According to Skinner (2006), the number of state and federal bills on this topic increased from three in 2002 to 84 in 2005.

[17]Because RxRights is a non-profit organization that tracks the cost of prescription drugs, their members are likely more price sensitive than the general population.

one primary website, sometimes with supplements from other websites. This suggests that many consumers, especially those searching for an affordable long-term supply of chronic drugs, are aware of credentialing agencies for foreign online pharmacies and use them as one way to discern the quality of online pharmacies. This behavior is consistent with a perceived quality difference between tier-B and tier-C websites, which motivates us to examine whether the ban on non-tier-A websites from sponsored listings has a differential effect on tier-B and tier-C websites.

2.2   Search Engine Policy Regarding Online Pharmacies

As summarized in Table 1, Google formerly contracted with PharmacyChecker to ensure that every pharmacy listed in Google's sponsored search results was 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. Shortly after (November 2009), the FDA issued 22 warning letters to website operators. 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."[18]

In response to the highlighted concern for drug safety, on February 9, 2010, Google announced two changes regarding its pharmacy advertising policy. The first change was to only accept ads from U.S. pharmacy websites that were certified by the NABP and from Canadian websites that were certified by CIPA. The second change was that the NABP-certified websites could only target their ads to Google users in the U.S. and the CIPA-certified websites could only target Google users in Canada. The new policy was only applicable to U.S. and Canada. Two months later (April 21, 2010), LegitScript announced assistance to Google

---

[18]The report is available here: http://www.legitscript.com/download/BingRxReport.pdf.

in implementing Google's internet pharmacy advertising policy in place of PharmacyChecker. On June 10, 2010, both Microsoft and Yahoo! started to require NABP certification for online pharmacy advertisers. Finally, on August 24, 2011, the settlement between Google and the DOJ was announced and Google acknowledged "that it improperly assisted Canadian online pharmacy advertisers to run advertisements that targeted the United States ..."[19]

Figures 1 and 2 present screen shots of a Google search results page for the cholesterol drug, Lipitor, before and after the ban. Before the ban, on the left hand side are organic links including brand-name websites such as Lipitor.com and information-oriented websites such as wikipedia.org. At the top of the whole page is a sponsored link for Lipitor.com from the brand's U.S. manufacturer (Pfizer). On the right hand side are other sponsored links, of which the top two are clearly foreign pharmacies (canadapharmacy.com and canadadrugpharmacy.com). After the ban, there are no sponsored links on the page except for Lipitor.com at the top. The void of sponsored links on the right hand side is filled by a drug fact label for Lipitor with links to official information about the drug's side effects, warnings, and user guidance from the National Library of Medicine.[20]

In light of these events, we define three regimes for our empirical analysis as shown in Table 2. Regime zero refers to a 17-month period from September 2008 to January 2010, prior to Google adopting the ban. Regime one ranges from March 2010 to July 2011, covering the period after the Google ban, but before the Google-DOJ settlement. Finally, the 13-month post-settlement period from September 2011 to September 2012 is referred to as Regime two. 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).[21]

---

[19]See http://www.justice.gov/opa/pr/2011/August/11-dag-1078.html.

[20]The drug fact label first appeared on June 22, 2010 under a partnership between Google and the National Institute of Health (NIH) and probably has diverted some click traffic following drug name queries after the ban.

[21]Because we define regimes one and two by time only, the difference between regimes one and two could be driven by a general trend or heightened consumer awareness. To address this concern, we count the number of searchers per month for queries related to pharmacy certification. This number fluctuates month to month, but starts to show an upward trend in the middle of 2011. We regress the log of this count on

In our main analysis, we consider consumer behavior on all search engines. This is both because other search engines adopted a similar policy change soon after the Google ban and because the Google-DOJ settlement was an effective warning to all search engines. As a robustness check, in Appendix Tables B2 and B3 we report the results of our main specification using only Google's search engine.

2.3 Organic and Sponsored Search

Internet search engines, such as Google, are one avenue consumers use to reach online pharmacies. Upon submitting a query, a user is presented with organic and sponsored results, in which the ranking of the latter is based, in part, on bids placed by the websites' owners.[22] While a website may increase its sponsored rank (and placement on the results page) by increasing its bid, better placement in the organic results can be achieved using search engine optimization (SEO) techniques, such as editing the website content to make it more relevant to the query. Many studies including Baye et al. (2009) have shown that click rates fall dramatically with lower placement on the results page and very few searches give rise to clicks beyond the first page of results.

The ban on non-NABP-certified pharmacies from search engines' sponsored results may be less effective if links to those same pharmacies also appear in the organic results. A rising literature has shown that sponsored links account for 15% of all clicks (Jansen and Sprink 2009), consumers have a preference against sponsored links (Jansen and Resnick 2006), consumers appreciate sponsored links as advertisements if they are relevant (Jansen, Brown and Resnick 2007), and organic and sponsored links from the same website of a national retailer are complements in consumer clicks (Yang and Ghose 2010). Two studies released by Google painted a somewhat different picture. Chan, et al. (2012) found that 81% of

---

regime and quarter indicators (to control for seasonality). The coefficient on the regime two indicator is 1.18 (stdev=0.39, p-value<0.01). In comparison, the coefficient on the regime one indicator is 0.27 (stdev=0.37, p-value=0.47). This suggests that the average monthly count of searchers that query pharmacy certification terms has increased significantly in regime two, but not in regime one.

[22]If the user clicks on a sponsored link, the link owner pays the search engine the next highest bid or the reserve price if there are no other bids.

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 the organic results. Chan, et al. (2011) examined 446 incidences between October 2010 to March 2011 where advertisers temporarily paused their sponsored ads to determine their effectiveness. 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.

In contrast, Blake, Nosko and Tadelis (2015) run a series of controlled field experiments in cooperation with eBay, Inc. and found strong substitution between organic and sponsored listings. In particular, when eBay paused its sponsored ad for the keyword "eBay," consumers simply substituted to eBay's (unpaid) organic link. Sponsored ads for non-branded keywords were also found to be ineffective for users that had experience with the product, while the authors found that these ads positively affected new user registrations and purchases. While we cannot measure user experience with our data, our results suggest that the willingness of users to substitute between organic and sponsored links depends, in part, on the certification status of the destination website.

Above all, the existing literature suggests that whether organic and sponsored results are substitutes or complements depends on the organic rank of the website. In our context, if a non-NABP-certified pharmacy appears high on the first page of organic results, the ban on its sponsored listing may redirect consumer clicks to its organic link. However, for many non-NABP-certified pharmacies that do not appear highly ranked in the organic results, the ban on their appearance in sponsored listings could be an effective tool to minimize consumer clicks on those websites. Our data do not contain the organic rank of each pharmacy website, but the organic click traffic before the ban is higher for tier-B than for tier-C websites, which implies that the ban on sponsored listings could generate a differential effect on tier-B and tier-C websites.

Our context also differs from the literature in terms of the incentives for organic substitution. In the literature, a sponsored ad pause is often driven by an internal decision from the advertising website, which does not necessarily carry any explicit message about the quality of the website. However, in our context, the ban on sponsored ads is imposed by search engines due to safety and legal concerns. This action alone may change consumers' willingness to access the organic link of a banned website as well as the website's incentive to manage its organic link in response to the ban. These two incentives may generate another differential effect on tier-B and tier-C websites if consumers believe that tier-B websites are more trustworthy than tier-C websites or if tier-C websites relied more on sponsored ads before the ban and therefore became more inclined to improve their organic rank after the ban.

2.4   Other Related Literature

Our study is indirectly related to two other literatures. At first glance, the sponsored search ban is a form of a minimum quality standard. 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 assume that the standard is well enforced in reality. This assumption does not hold for online pharmacies: after the ban, consumers can still access non-NABP-certified pharmacies through organic results. In addition, other channels exist for consumers to gather safety information about online pharmacies, such as consumer experience, word of mouth, and alternative certification agencies.

Our study also contributes to the literature on the willingness of consumers to switch to alternative information channels when one channel is removed. Goldfarb and Tucker (2011) consider how offline advertising bans are affected by the simultaneous exposure of online ads and show that government bans of offline ads are less effective in the presence of online ads. Chiou and Tucker (2011) show that, following the removal of search advertising for certain pharmaceutical drugs, consumers searching for information on medical conditions

13

were less likely to click on the drug company's website, but more likely to click on non-FDA regulated websites such as Canadian pharmacies.[23] Similar to Chiou and Tucker (2011), we study how the ban on sponsored search advertisements for online pharmacies affects consumer search. This could cause consumers to avoid certain pharmacies due to safety concerns or to substitute between different-tiered pharmacies. Our data are unique in that it allows us to analyze these effects through the ban's impact on both sponsored and organic clicks.

3 Data Description and Summary Statistics

Our primary data source is comScore.[24] ComScore tracks the online activity of about one million households in the U.S. It then extrapolates the observed activity in the households it tracks using various demographic weights, which represents the aggregate activity of all U.S. internet users. ComScore data have been used to study internet search behavior by a number of economists including Chen and Waldfogel (2006), Chiou and Tucker (2011), and George and Hogendorn (2013).

3.1 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 searches searchers, and clicks in a given month. In addition, clicks are broken down into organic versus paid and by destination URL. We observe 49 months of data from September 2008 to September 2012. At times, due to small sampling of some queries, click activity is censored when comScore is unable to reliably

---

[23]In addition, Manchanda, et al (2006) focus only on online banner advertisements and show that ads increase the probability of purchase for current customers. Chatterjee et al (2003) find that repeated exposure to banner ads reduces click probabilities especially for consumers with a higher innate tendency to click on ads.

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

extrapolate the observed activity to the whole population. 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 are some small amount of clicks on the website during that month, but extrapolation to the population would not produce a reliable estimate of the exact number of clicks. The synthetic control method is defined for continuous variables, so our main analysis using synthetic controls treats censored clicks as zero. However, in a robustness check, we use a two-part model to separately explore the probability that a website receives non-censored clicks and the probability that a website receives any (including censored) clicks.

Figure 3 shows an example of the term destination report for Lipitor in January 2012. The report lists the total clicks, divided between organic and paid, following queries for Lipitor in January 2012. By selecting "match all forms," the 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 U.S. on all websites following the query. In addition, the clicks are broken down by specific entity.[25] 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. While comScore also tracks how many clicks a website receives via direct navigation, bookmarks, etc., the comScore suite we use only allows us to observe click traffic originating from search engines.

3.2 Query List and Website Classification

To determine the clicks received by a given website, a list of queries must be submitted to comScore to extract website clicks from queries. To create a list of drug and pharmacy related query terms, we use several resources. The first is a list of brand names from the FDA's Orange

---

[25]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.

Book of all approved drugs. The second resource is a list of drug manufacturers from Kantar Media. 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.[26] Using our list of drug names, we download from comScore the top pharmacy website names in comScore's "Pharmacy" category.[27] This list, plus any pharmacy names that we can find on any of the four certifying websites, gives us a preliminary list of pharmacy websites. To address the possibility that searchers may reach pharmacy 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.[28] Together with all keywords that the FDA bids on sponsored ads, it allows us to identify a list of keywords from which the majority of traffic on online pharmacies are generated.

The combination of all these sources led to over 8,000 queries. We ran this list of queries through comScore twice in 2009 and 2012 and restricted our sample to queries that accounted for the top 90% of clicks in either window. We also complement the list with drugs that were advertised (in the Kantar media data) and/or prescribed by a physician from 2006-2009.[29] This left us with 690 queries. Lastly, because comScore reports the clicks from queries that are an exact match with the term and variations of the term, for example, both "canada pharmacy" and "canada online pharmacy" are counted as a search for "canada pharmacy", we only use queries that are not variations of another to avoid double counting. This further

---

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

[27]A website may have multiple classifications, but we classify any website with this classification as a pharmacy.

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

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

restricts our sample to 528 queries. Each query was then submitted to comScore and monthly reports from each search engine were downloaded for our analysis.

These 528 queries are then classified into different query types (see Table 3), mainly drug queries, pharmacy queries, and other health queries. Among pharmacy queries, we classify them according to whether they are exact names of the pharmacies, whether they are used to search for pharmacy certifiers, and whether it is used to search for discount pharmacies. In particular, among the pharmacy names queries, we separate them into the pharmacy's tier (tier-A, tier-B, or tier-C). Queries that target foreign pharmacies or pharmacies that sell cheap or discount drugs are classified as discount pharmacy search terms.[30] And the remaining pharmacy related queries are classified as general pharmacy terms.

Table 3 shows the total query count in each category of queries. Among the 528 queries, drug and pharmacy queries are more likely to lead to online pharmacy websites. In regime zero, drug and pharmacy queries account for 97.7%, 96.7%, and 97.4% of the clicks on tier-A, tier-B, and tier-C websites respectively. In comparison, drug and pharmacy queries account for 54% of clicks on non-pharmacy health-related websites, while the remaining 46% originate from other health queries. In the main analysis, when we count website clicks, we condition on clicks following only drug and pharmacy queries because these clicks are generated by consumers whose likely intention it is to search for drugs or pharmacies. 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.[31]

All empirical results presented below pool data from all five search engines. For robustness, we also estimate our main specification using clicks only on Google. Results are presented in the Appendix.

---

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

[31]The seasonality mainly occurs for clicks on tier-A pharmacies, which are not included in the main specification. As a robustness check of the two-part model that includes tier-A pharmacies, we control for seasonality.

3.3 Summary Statistics

Almost all search engine clicks on pharmacy websites originate from drug and pharmacy queries. Therefore, Table 4 focuses on drug and pharmacy queries only. It summarizes the number of searches for each type of drug and pharmacy queries and the number of clicks on pharmacy websites following the searches. The total pharmacy clicks to total searches ratio (column 3) shows how often a search leads to a click on a pharmacy website. This ratio reflects consumers' search intention and the ease of finding the website they intended to visit. The sponsored advertising ban makes tier-B and tier-C websites harder to find following a search and may also reduce consumers' interest in tier-B and C websites due to health safety concerns, both leading to a decline in the click-to-search ratio for queries targeting pharmacies as observed in the data. However, for a search on the exact pharmacy website name, the website is likely to be ranked high in the organic results and not affected by the sponsored ads ban so the click-to-search ratio only represents consumers' interest in reaching pharmacy websites after the ban. We see that the ratio stays the same for tier-B name queries, but declines for tier-C name queries.

Another statistic that shows the relevance of pharmacy websites and the difficulty in finding them is the percentage of clicks on pharmacy websites (column 4). Pharmacy clicks per search and the percentage of pharmacy clicks are highly correlated, as both depend on the relevance of pharmacy websites to the studied queries. In general, pharmacy queries lead to more clicks on pharmacy websites than drug queries. Among all types of queries, tier-B names generate the highest percentage of pharmacy clicks (93-98%), followed by tier-A names (78-81%), and discount pharmacy keywords (59-67%).[32] Tier-C names are associated with the lowest percentage of pharmacy clicks among all pharmacy name queries and this percentage drops sharply from 39.8% before the ban to 7.1% after the DOJ's settlement with Google. In contrast, the percentage of pharmacy clicks increases for tier-B names after the

---

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

ban.

The remaining columns of Table 4 report paid and organic clicks separately by pharmacy tier. After the ban, paid clicks drop to nearly zero for tier-B and tier-C websites as expected. Tier-C pharmacy websites still receive a few clicks after the ban because some label themselves as selling nutritional supplements rather than prescription drugs and are not perfectly screened by the ban. Organic clicks on tier-B and tier-C websites increase after the ban for almost all pharmacy and drug queries suggesting some substitution to organic clicks when sponsored links are no longer available.

To better illustrate the potential organic substitution after the ban, Table 5 summarizes the organic and paid click volumes by regime for each pharmacy tier. Clicks counts reported in this table are conditional on searches involving drug and pharmacy queries. Tier-B pharmacies were accessed mostly via paid clicks before the ban, with an average of 6,531 monthly paid clicks and 1,823 monthly organic clicks. The ban results in 100% loss in paid clicks over time, but part of the loss is offset by a large increase in organic clicks, suggesting that searchers are substituting toward organic links. For tier-C websites, the average number of paid clicks fell as expected. The average organic clicks rose in regime one, but the increase is small relative to the loss in paid clicks. Tier-C's organic clicks fell in regime two after the news coverage of the DOJ's settlement with Google, potentially suggesting a negative effect on tier-C websites associated with health concerns. Tier-A pharmacies are clearly very different from tiers B and C in absolute click volumes, which makes it difficult to compare the changes between tiers. The number of paid clicks on tier-A pharmacies grew steadily after the ban, while the number of organic clicks dropped slightly after the ban, followed by a substantial increase after the Google-DOJ settlement. While the average organic clicks across tier-A websites increased, the organic clicks for the median and 75th percentile website fell after the settlement, suggesting a distributional shift in organic clicks toward the largest websites.

The last three columns of Table 5 show the distribution of websites that have positive clicks in each regime. With the same set of queries in each regime, the number of online

19

pharmacy websites that are recorded as having any clicks is relatively stable for tier-A and tier-B pharmacies, but declines 48% for tier-C from 135 to 70. This decline suggests some tier-C websites may suffer, but this may benefit the remaining tier-C pharmacies if consumers with a preference for tier-C pharmacies continue to buy from them.

The differential changes in organic clicks on tier-B and tier-C websites after the ban can also be seen in Figure 5, where we plot the monthly number of paid and organic clicks for each pharmacy tier. Organic clicks on tier-B websites almost doubled after the ban, while organic clicks on tier-C websites experienced no increase and are shown to slowly decline after the ban.

4 Empirical Strategy

4.1 Control Groups

The changes in organic clicks on pharmacy websites after the sponsored search ban and the Google-DOJ settlement do not tell us the treatment effect of these events without comparing them to an appropriate counterfactual. Therefore, a key challenge of our study is to find a set of control group websites that exhibit a similar trend to the foreign pharmacy websites, but are immune to the events that affected online pharmacies. An ideal control group should reflect the mindset of the same consumers that click on tier-B and tier-C websites, follow the same trend as tier-B and tier-C sites before the intervention (ban), but consist of search and click activities that are not affected by the ban and the Google-DOJ settlement. Non-health websites may satisfy the last condition, but consumers interested in pharmacies may be very different from consumers interested in food, clothing, toys, electronics, or other non-health merchandise.[33]

This concern motivated us to focus on health-related websites as potential controls. The

---

[33]We compared the trends in search for drug and pharmacy-related keywords to the Google Domestic Search Index for consumer product categories, such as auto and computer/electronic, and we find that the time trends are very different.

health-related websites we use to construct the control group consist mainly of health and pharmaceutical information websites (such as WebMD), health organization websites (such as the American Heart Association), and pharmaceutical manufacturers(such as Pfizer). More specifically, we construct the control group using only clicks following drug and pharmacy queries, which is the same set of queries we use to calculate the clicks on pharmacy websites, as described above. Drug and pharmacy queries generate around 97% of the pharmacy website traffic we track before the intervention and 53% of the traffic on health-related non-pharmacy websites. Conditioning on the same set of drug and pharmacy queries, we argue that the clicks on health information websites and pharmacies reflect the same underlying trend in consumers' intention in seeking information. The control group is valid if the probability of consumers seeking pharmacies to purchase drugs relative to the probability of consumers seeking treatment information conditional on the drug and pharmacy search stays the same in the absence of the events affecting foreign pharmacies.

The assumption that clicks on the control group websites are not affected by the ban and the Google-DOJ settlement is not directly testable, but we offer empirical evidence to show that this is likely to be the case. To test this, we calculate the paid and organic clicks following general health queries that focus on health conditions. These queries lead to a very small fraction of the click traffic on pharmacy websites (less than 3% both before and after the intervention), and therefore the intervention that affect online pharmacies are unlikely to affect click traffic generated from general health queries. Taking the control websites' paid and organic clicks from general health queries as the benchmark, we test whether the interventions differentially affect the clicks following drug and pharmacy queries into the same control websites. We find that there is no differential change in the proportion of clicks generated from the two sets of queries for the control websites, which suggests that the interventions do not contaminate clicks on the control websites. The tests and discussions are included in Appendix A.

Furthermore, we examine whether the fraction of traffic generated by drug and pharmacy

21

queries follows similar trends for treated and control websites. The changes in this fraction tell us whether consumers have changed the set of queries they use to reach the intended websites. For health-related non-pharmacy websites, the fraction of organic clicks from drug and pharmacy queries rises from 53% in regime zero to 56% in regime one and two. For tier-B pharmacies, the fraction increases from 96% in regime zero to 99% in regime one and two. For tier-C pharmacies, the fraction is 98%, 99%, and 96% in the three regimes, respectively. The changes in the fraction are similar among the treated and control website, although there is slight decline in the fraction of tier-C clicks following drug and pharmacy queries in regime two.

One approach to determining the effects of the paid search ban and announcement is to estimate a difference-in-differences model using the full set of control websites. This requires that the treated and control websites follow a similar trend in the pre-ban period. In Table 6, we show that this is likely the case on average: we are unable to find a statistically significant difference in the time trends between the treated and control groups in the pre-ban period. However, for a given website, the level of clicks per month is highly variable over time and the click levels are heterogeneous across websites within the treated and control groups. Therefore, due to the concern that the parallel trend assumption may not hold separately for each individual treated website, we use the synthetic control method that selects and weights the websites in the pool of controls so that the resulting weighted average synthetic control follows the same level and trend as the treated website. We will describe the synthetic control method in detail in the next subsection. The advantage of the synthetic control method is that it does not rely on the assumption that the trend of each individual website in the control pool matches the treated website. The method is valid so long as the synthetic control website matches the treated website closely and the interventions do not affect the pool of controls.

4.2 Synthetic Control Method

The synthetic control method is developed in Abadie and Gardeazabal (2003) and Abadie, Diamond, and Hainmueller (2010). The method proposes a systematic approach to construct a synthetic control for each treated unit by selecting and weighting units from the pool of controls so that the weighted average of the selected control units matches the treated website's trend and level in the pre-ban period. The selected control units and their weights are kept fixed, and the weighted average outcome after the intervention is compared to the treated unit to derive the treatment effect. Abadie, Diamond, and Hainmueller (2010) use the method in a setting in which a single treated unit is presented, while the method is also used in a setting with multiple treatment units, such as Acemoglu et al. (2016). Athey and Imbens (2016) and Doudchenko and Imbens (2016) further discuss recent advances and advantages of the method. In this section, we discuss the way we construct the synthetic controls, the estimation and inference method, and also several robustness checks.

As discussed in Abadie, Diamond, and Hainmueller (2010), the key assumption required for the validity of the synthetic control method is that the units in the donating pool of controls are not affected by the treatment intervention. We discussed the empirical evidence that supports this assumption in the previous section. Moreover, the method has restrictions on the weights assigned, and requires that the synthetic control can match the level and trend in the treated units in the pre-ban period. The weights are chosen to minimize the difference in outcomes between the synthetic and treated websites in the pre-ban period.

Let $y_{\phi t}$ denote the clicks on treated website ϕ in period t and $Y_t^0 = [y_{1t}^0, ..., y_{Jt}^0]'$ denotes the vector of control website clicks. The vector of weights, $W = [w_1, ..., w_J]'$, needs to satisfy the conditions $w_j \geq 0$ and $\sum_{j=1}^J w_j = 1$. The synthetic control for website ϕ is denoted $\tilde{y}_{\phi t} = \sum w_j y_{jt}^0 = W'Y^0$. Denote $g(y) = [g_1, ..., g_K]'$ as functions of the outcome clicks. W is found by solving the following problem:

$$\min_W \{(g(y_\phi) - g(W'Y^0))'V(g(y_\phi) - g(W'Y^0))\}, \tag{1}$$

23

where $y_\phi = [y_{\phi 1}, ..y_{\phi T_0}]'$, $Y^0 = [Y_1^0, ..., Y_{T_0}^0]'$, T_0 is the last month before the policy intervention, and V is a positive definite weighting matrix. The objective function represents the distance between the synthetic control and the treated websites. $g(\cdot)$ is a vector of weighted averages of monthly organic clicks. We use weighted average rather than $g_k(y_t) = y_t$ for all $t \leq T_0$ to avoid over fitting. We have tried different sets of $g(\cdot)$ and checked robustness in terms of match quality and estimates. While synthetic matching papers usually set V equal to the identity matrix (for example in Acemoglu et al. (2016)), we follow the method discussed in Abadie, Diamond, and Hainmueller (2010) and choose V to minimize the mean squared prediction error of the outcomes in the pre-ban period.

We then compute the treatment effect, $\lambda_{\phi s}$, on website ϕ in period s after the intervention as:

$$\lambda_{\phi s} = y_{\phi s} - \tilde{y}_{\phi s} = y_{\phi s} - W'Y_s^0. \tag{2}$$

We focus on the treatment effects of the ban in regime one and the effects of the DOJ-Google announcement in regime two. Averaging over the S months in regime r, we calculate the treatment effect for website ϕ as:

$$\bar{\lambda}_{\phi r} = \frac{1}{S} \sum_{s \in r} (y_{\phi s} - \tilde{y}_{\phi s}). \tag{3}$$

The treated websites may differ significantly in the number of clicks they receive and how those clicks are divided between organic and paid links before the ban. A website that receives relatively more organic clicks before the ban will experience a smaller treatment effect. Alternatively, a website that relied more on paid links before the ban will experience a larger treatment effect and the magnitude of the effect will depend on size of the website (measured by the total clicks received before the ban). Because of this potential heterogeneity in the effect sizes, we present both the average and median treatment effects in the results. More specifically, we calculate the mean and median of $\bar{\lambda}_{\phi r}$ across websites in each pharmacy tier.

The heterogeneity also motivates us to normalize the treatment effects using each website's pre-treatment click volume.

The average monthly clicks on the treated tier-B and tier-C websites and the synthetic controls we construct for them are shown in Figure 6. The figure shows that the synthetic control websites match the treated websites closely before the sponsored search ban in February 2010. The weights assigned to the control units to construct the synthetic control can be zero. As in Abadie, Diamond, and Hainmueller (2010), most of the treated websites in our data are matched with only a few control websites that receive positive weight. We find that 73% of tier-B websites and 55% of tier-C websites have a synthetic control constructed from only five or fewer control websites.[34]

The p-value of the estimate is derived using exact inference motivated by placebo tests. The idea of a placebo test is to derive the treatment effects for websites unaffected by the intervention and test whether these effects are statistically different from the treatment effect estimated for the treated unit. Using the exact inference method, we randomly draw a set of placebo websites from the control group, the number of which is the same as the number of treated websites. Then we compute the treatment effects using the synthetic control method described above. For every draw of placebo websites, we compute the average treatment effect. By repeating this 5,000 times, we get a distribution of placebo treatment effects. The p-value is then derived from the distribution of placebo effects. For statistics other than the average treatment effect, we calculate the statistics in each of the 5,000 draws of the placebo sample, from which we obtain the distribution of the statistics. We compare the statistics we obtain for the true treated sample to the distribution to determine the p-value. In Figure 7, we illustrate the distribution of placebo treatment effects by showing 25 random draws of the

---

[34]There are 5,129 websites in the pool of control websites, however, the support of many control websites does not overlap with the treated websites and many have trends far different from the treated website. Before we estimate the weights using equation (1), we perform two initial steps to ease the computational burden. First, we eliminate all websites with no overlapping support with the treated websites. Second, we calculate the mean square differences between the treated website and each of the control websites in regime zero and keep the 80 websites with the smallest difference. Using this restricted pool, we then perform the minimization in equation (1). For robustness, we vary the criterion in the second step to the top 100 and top 120 websites, and we find the treatment effect is not sensitive to increasing the size of the pool.

placebo websites.

We conduct several main robustness checks for the synthetic matching method. The synthetic control estimates depend on finding the weighting matrix, W, to construct the synthetic control units. In the first robustness check, we check the robustness of the synthetic control by using different sets of $g(\cdot)$ in the objective function. In the main specification, we match the quarterly averages of organic clicks, while in alternative specifications, we vary the periods over which we calculate the averages. The resulting pre-ban matching quality is similar using different objective functions. In the second robustness check, we conduct the exact inference placebo test on the "treatment effect" before the intervention, and we do not find any statistical significance. In the third robustness check, we conduct a placebo test in the temporal dimension. We set a placebo intervention date that is earlier than the true intervention date and determine whether we find a treatment effect during the placebo treatment period. This test validates that the synthetic control is able to mimic the treated units when extrapolated to the pre-ban sample. We set the placebo ban to two months earlier than the true search engine ban and estimate the treatment effect on the last two months before the intervention. We find no statistically significant differences in the outcomes of the treatment and synthetic control websites in the placebo treatment period. We also estimate the model via ordinary least squares (OLS) using the full set of control websites as detailed in the next section. As shown in Panel C of Table 7, the OLS estimates are of similar magnitude to the synthetic control method.

Although the synthetic control websites match the pre-ban outcomes of the treated websites better than the full set of control websites, the method has two limitations. First, the website traffic is characterized by censored clicks in some months followed by spikes in other months. We are interested in disentangling the aggregate effects of the ban and the DOJ settlement into the effects on the probability that pharmacies receive *any* clicks (the extensive margin) and the increase in click traffic, conditional that the website attracts positive traffic (the intensive margin). To examine these two effects separately requires that

the synthetic controls mimic the pulsing behavior of the treated website traffic. However, the synthetic control method is developed to examine a continuous outcome and does not match the discrete jumps in click traffic. Second, in order for a treated website to be matched, we need units in the donating pool that cover the same support of the outcome as the treated website. However, the click traffic for some tier-A websites is so large that we are unable to find any control websites that have a similar level of clicks. In the next section, we detail a two-part difference-in-differences model that can address both of these limitations.

4.3 Two-Part Model in a Difference-in-Differences Framework

Complementary to the synthetic control method, we estimate a two-part model in the difference-in-differences (DID) framework using the full control sample. The distribution of clicks per website-month is characterized by a spike at zero and a bell-shaped positive distribution skewed to the right. To capture this data pattern, we assume that the clicks are generated from the following two-part model.[35] The first part of the model estimates the extensive margin (whether a website receives any positive clicks in a month), which we fit using a logit model:

$$
\begin{aligned}
Prob(Y_{it} > 0) \quad = \quad & F\Big(\alpha + \sum_{k \in \{B,C\}} \beta_k Tier_k + \sum_{r=1}^{2} \gamma_r Regime_r \\
& + \sum_{k \in \{B,C\}} \sum_{r=1}^{2} \theta_{kr} Tier_k Regime_r\Big),
\end{aligned}
\tag{4}
$$

where Y_{it} denotes the clicks website i receives in month t, $Tier_k$ is an indicator for the type of pharmacy, $Regime_r$ is an indicator for different regimes, and $F(\cdot)$ is the logistic function. The second part of the model estimates the intensive margin (the number of clicks a website

---

[35]Two-part models used to analyze such data patterns are well discussed in the health literature when examining health expenditures (see Mullahy (1998) and Manning (1998)).

receives, conditional on receiving positive clicks), which we estimate via OLS:

$$
\begin{aligned}
(Ln(Y_{it})|Y_{it} > 0) \;=\; & \delta_i + \sum_{r=1}^{2} \lambda_r * Regime_r \\
& + \sum_{k \in \{B,C\}} \sum_{r=1}^{2} \vartheta_{kr} * Tier_k * Regime_r + \epsilon_{it},
\end{aligned} \tag{5}
$$

where δ_i denotes website fixed effects.[36] Because website fixed effects absorb the tier dummies, $Tier_k$ only appears in the interaction term with $Regime_r$. The treatment effects of the two interventions (ban and announcement) are captured by the coefficients θ_{kr} and ϑ_{kr}.

Due to censoring, we cannot observe the exact number of clicks when the number of clicks is small and cannot be precisely measured, but we can estimate the probability of censoring. In the total clicks specification, we separately estimate the logit model using the probability of non-censored positive clicks and the probability of any positive clicks (including censored clicks) as an outcome.[37]

5   Results

5.1   Average Effects

The time trends of the treated websites and the synthetic controls are plotted in Figure 6. After the ban in February 2010, the organic clicks on tier-B websites increase quickly compared to the synthetic controls and the organic clicks on tier-C websites follow similar trends as their synthetic controls. After the media coverage of the DOJ settlement with Google in August 2011, organic clicks on tier-B websites continue to be higher than their synthetic controls while organic clicks on tier-C websites decline relative to the clicks on the synthetic controls. The inference method we use is illustrated in Figure 7. We compare the

---

[36]We use the log of clicks on the left-hand side because the distribution of clicks appears log-normal with a long right tail. The log transformation results in an approximately normal distribution.

[37]Besides the two-part model, we also consider two alternative models, Tobit and Heckman selection models. We argue that the two-part model is the most appropriate to study the censored click problem in our setting. Results are available upon requests.

treatment effects of the true treated group to those of the placebo treated group.[38] The bold lines represent the true treatment effects by month for tier-B and tier-C pharmacies and the gray lines represent the effects estimated for different placebo samples. In the figure, we see that the tier-B true effects are above the placebo effects unlike the tier-C effects. Therefore, based only on this illustration, we expect tier-B effects to be statistically significant while the tier-C effects to be insignificant.

The patterns shown in Figure 6 and Figure 7 are consistent with the average effects estimated in the first three columns of Panel A in Table 7. Organic clicks on tier-B websites increase by 5,779 in regime one and by 5,561 in regime two compared with the control group; both statistically significant. Organic clicks on tier-C websites increase by 422 in regime one and decrease by 387 in regime two; in contract to tier-B, both estimates of tier-C are not statistically significant. The difference between the two regimes in the rate of substitution to organic clicks on tier-C websites suggests that the media coverage of the DOJ settlement with Google may have increased the consumer awareness about the safety of purchasing prescription drugs online.

The effects are measured as the change in the absolute number of organic clicks, so the effects can be especially large for websites with higher click traffic or websites that relied more on paid clicks before the ban. Due to the log-normal distribution of clicks and the potential influence of the large websites on the mean measure, we also present the median treatment effects in columns (4)-(6). The average effects are much larger than the median effects, suggesting the influence of the websites with much larger effect sizes. The median effect for tier-B websites in regime one and two are 954 and 1,553 respectively (both statistically significant), while the median effect for tier-C websites are both negative and not statistically significant.

To put the effects measures in clicks into perspective, we compare each treated website's change in organic clicks to the website's total clicks in regime zero, before the search engine

---

[38]Although we generate 5,000 placebo draws to derive the inference, we only plot a random subset of placebo draws for illustration.

29

ban.[39] Panel B shows the change in organic clicks relative to the website's level of total clicks before the ban. The median effects in columns (4)-(6) show a significant increase in tier-B organic clicks of 44% and 36% relative to the pre-ban level of clicks and the effects for tier-C websites in both regimes are negative and insignificant. Because websites differ in the fraction of clicks that comes from paid links before the ban and the fraction affects the size of the organic click effect, we will analyze the change in organic clicks conditional on websites with similar fractions of paid clicks in the next section.

In order to verify the robustness of using the full control sample in the estimation, we estimate a similar specification using OLS on the full control sample. To ensure comparability, the OLS analysis treats censored clicks as zero, as in the synthetic control method. The results are shown in Panel C. The estimates using the full control sample are similar to the estimates using the synthetic control sample, confirming the robustness of conducting analysis on the full control sample. As discussed in section 4.3, when analyzing the change in organic clicks on the extensive and intensive margins using the two-part model, we use the full control sample.

The synthetic control estimates show us the change in the average number of clicks, but not the mechanics of whether the change comes from the website's probability of receiving organic clicks (the extensive margin) or the number of clicks conditional on receiving non-censored clicks (the intensive margin). As discussed above, the synthetic controls do not match the extensive margin of the treated websites well, so we use the full control sample to estimate the two-part model. The first three columns of Table 8 focus on total clicks and the last two focus on organic clicks. Within total clicks, column (1) examines whether website i received any clicks (including censored clicks) in month t, column (2) examines whether website i received any uncensored clicks in month t, and column (3) examines the change in the total number of clicks conditional on a website receiving non-censored clicks in the month. Columns (4)

---

[39]In theory, we can compare the change to the website's paid clicks and organic clicks in regime zero separately. However, because the websites differ in the fraction of clicks coming from paid and organic, neither is a good base to represent the pre-ban size of a website.

and (5) are similar to columns (2) and (3), but for organic clicks only.

The comparison between the tier-B and tier-C estimates in columns (1) and (2) tells us the change in the probability that online pharmacies receive a positive number of clicks or a positive but censored number of clicks after the two events. The probability that tier-B websites receive non-censored clicks does not change after the sponsored search ban, while the probability that tier-B websites receive any clicks (both censored and uncensored) increases by 5% and 8% in regimes one and two respectively. Possibly due to the inability of large tier-B websites to advertise, this suggests that smaller tier-B websites had a greater chance to receive clicks and therefore, an average tier-B website is more likely to be clicked after the interventions. However, tier-C websites have a smaller probability of receiving non-censored clicks (as well as any clicks at all) after the interventions. Column (3) shows a greater percentage reduction in total clicks for tier-B websites compared to tier-C websites conditional on the website receiving any noncensored clicks. The number of observations drops significantly in this regression because click traffic on many websites is close to zero and therefore censored.

Focusing on organic clicks, the last two columns of Table 8 indicate strong organic substitution for tier-B websites: on average, the websites experienced an 81% increase in organic clicks in regime one and an 87% increase in regime two compared to the control group if the websites continue to receive positive number of clicks.[40] Combined with the decrease in total clicks, this suggests that the loss in paid clicks on tier-B websites was only partially offset by organic substitution. Total clicks still suffer significantly compared to the baseline group as an average tier-B website received 66% of its clicks from sponsored advertising before the ban. Moreover, the size of organic substitution effect is similar across the two regimes. In contrast, at the intensive margin, tier-C websites experienced a 5.3% increase in organic clicks in regime one and a 8.7% decrease in clicks in regime two and both effects are not statistically significant. Furthermore, the probability of tier-C websites receiving

---

[40]The two effects are not statistically different from each other.

31

non-censored organic clicks decreases by 3.2% after the media coverage of the Google-DOJ settlement. 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 websites. The cost of searching, and potentially safety concerns, results in almost no organic substitution for tier-C websites.

While we could not match the level of clicks for tier-A websites using synthetic controls, we are able to include the tier-A pharmacies in the analysis using the two-part model. Tier-A websites are likely to benefit in both regimes one and two as the total traffic on tier-B and tier-C online pharmacies decline. The increase in both organic and paid clicks for tier-A pharmacies is shown in Table 5. The average number of paid clicks increased from 46,398 to 73,812 from regime zero to regime two and organic clicks increased from 554,389 to 700,187. However, in the two-part model results, we only found a small increase in the probability of tier-A websites receiving any (including censored) clicks (4% and 7% in regimes ones and two, respectively), but a decrease in the log level of clicks in the two regimes. To reconcile the increase in the number of clicks and the decrease in the log level of clicks, we checked whether the change in clicks was heterogeneous across different size tier-A pharmacies.

In Appendix B, Table B1, we show that the total number of clicks on the four largest tier-A websites (cvs.com, walgreens.com, walmart.com, target.com) increased by an average of 2.3 million clicks per month from regime zero to regime two. The average monthly organic clicks on these four websites increased by 2 million clicks. For the remaining 43 tier-A websites, the average monthly total and organic clicks decreased by 22,151 and 24,586 between regimes zero and two. Although the increase in absolute click is large for large tier-A websites from regime 0 to regime 2, the increase is relatively small in log scale, smaller than the increase of 0.6 in log scale for the control group. These factors contribute to the negative tier-A effects at the intensive margin measured in log level. This comparison shows that although the smaller tier-A websites experience losses in clicks (potentially due to competition with the larger tier-A pharmacies or reasons unrelated to the ban), the magnitudes are much smaller

than the increases for the largest tier-A websites. Overall, tier-A pharmacies have not been negatively affected by the ban.

We have estimated auxiliary models to assess the robustness of the organic substitution results. Although treatment and control groups follow a similar linear trend before the ban, it is possible that the use of regime indicators is not sufficient to account for the temporal changes common to all groups. To address this, we add a linear trend in the main specification and rerun the analysis for organic clicks. Results are similar in magnitude and significance. We also check for the impact of seasonality by including indicators for the holiday months of November and December for all websites. Neither of these specifications impacted the qualitative results.[41]

Because the sponsored search ban on tier-B and tier-C pharmacies was imperfect (as shown in Figure 5), we also conducted robustness checks on the starting date of regime one in two ways. First, we used a new regime one 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 test by adding a hypothetical 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 treatment period before the actual ban. In the first strategy, we also divided regime one into two halves corresponding to before and after September 2010. We find similar coefficients for these two periods, except that the reduction in total clicks on tier-C websites at the extensive margin is deepened in the second half of regime one.

5.2 Heterogeneous Effects

The previous section showed the effects of the paid search ban and the announcement of the settlement on different pharmacy website tiers using both the synthetic control method and a two-part model in the difference-in-differences framework analyzing the intensive and

---

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

extensive margins separately. To better understand these differences, we explore three types of heterogeneous effects.

While we find that both tier-B and tier-C pharmacy websites experience a significant reduction in total clicks in the two-part model, the mechanisms are different: the reduction for tier-B websites concentrate on the intensive margin and is partially offset by organic substitution. In comparison, the reduction for tier-C websites is driven by both extensive and intensive margins for both total and organic clicks. One explanation for the tier-B and tier-C comparison is that the search engine ban and the Google-DOJ settlement have heightened the health concerns of uncertified pharmacies and therefore consumers may perceive tier-C as having higher health risks than tier-B websites. However, another possible explanation is that tier-C websites are ranked low in the organic search results and consumers have difficulty finding them after the ban on paid clicks. As shown in Table 5, tier-C websites have a lower number of organic clicks before the paid search ban compared to tier-B websites.

To distinguish these two explanations, we first check whether consumers can differentiate tier-B and tier-C websites when searching for online pharmacies. If consumers cannot distinguish tier-B from tier-C websites, it is unreasonable to argue that tier-B and tier-C websites generate different health concerns. One way to determine the certification tier of a pharmacy is searching for pharmacy certifiers on the internet. Therefore, we aggregate the total number of searches using pharmacy-certifier-related queries. As reported in Table 9, LegitScript's monthly searches increased from 18 in regime zero to 278 in regime one and 1,275 in regime two. Similarly, the monthly search for "NABP" and "VIPPS" (the certification program of NABP) have more than tripled from 6,895 in regime zero to 27,098 in regime two. Searches for queries containing "pharmacy check", "pharmacy rating" and "pharmacychecker" increased from 731 in regime zero to 1,685 in regime one and 4,127 in regime two. The only exception is that queries for the Canadian certifier, "CIPA," have fallen over time, possibly because one of the first organic results following the query "CIPA" is the Children's Internet Protection Act rather than the Canadian Internet Pharmacy Association. These patterns are

34

consistent with the hypothesis that at least some consumers are either aware or concerned with the certification differences between online pharmacies and actively search more for certification status after regime zero.

To further uncover the mechanisms that drive the differential changes in tier-B and tier-C websites, we compare clicks on tier-B and tier-C websites using subsets of the data based on the organic visibility of the website, the fraction of a website's clicks which are paid, and clicks from queries that may imply a stronger intention to reach foreign pharmacies.

In the first set of heterogeneous results, we classify websites by their total organic clicks in the pre-ban period. We identify the common support of organic clicks for the control group, tier-B, and tier-C websites and divide websites into three bins based on total organic clicks in the first six months of the data.[42] Bin one includes websites with 150 to 400 monthly organic clicks, bin two includes websites with 700 to 1,750 monthly organic clicks, and bin three includes websites with 2,000 to 4,000 monthly organic clicks. The bin cutoffs are unevenly spaced because the number of tier-B and tier-C websites drops significantly in bins with a larger number of clicks. We choose these cutoffs to ensure there is no statistical difference in the monthly organic clicks in the pre-ban period among the treated and control websites, and there is a large enough sample in each bin.

Tier-B websites receive more clicks than tier-C websites on average. If the differences between tier-B and tier-C are driven by the potential organic rank a website can obtain after the ban, comparing websites with similar organic clicks volumes before the ban should remove the differential effect of the ban on tier-B and tier-C websites. The bin-by-bin results in Table 10 do not support this prediction. In Table 10, we present the synthetic control results in Panels A and B, and the two-part model results in Panel C. To show more robust and conservative estimates, we present the median synthetic control estimates in Table 10, while reporting the synthetic control mean effects as a robustness check in Appendix Table B4. We observe an increase in organic clicks for tier-B websites in all three bins. The increases

---

[42]Tier-A websites are excluded because they are typically much larger in the absolute click volume than tier-B and tier-C websites.

in tier-B organic clicks are strongest for the smallest websites (in bin one).[43] In contrast, organic clicks on tier-C websites generally decline in both regimes and the decline measured in clicks is largest (and statistically significant) for tier-C websites in bin two.

These patterns are confirmed in the two-part model. Tier-B pharmacies in bin one experience a large increase at both the extensive and intensive margin. The increases in organic clicks for other tier-B pharmacies mainly occurs at the intensive margin rather than the extensive margin. Tier-C pharmacies generally show a decline at both the intensive and extensive margins with the only exception being an increase at the intensive margin for pharmacies in bin two. The contrast between tier-B and tier-C pharmacies conditional on similar levels of organic clicks suggests that there are reasons other than the organic ranking that drive the tier-B and tier-C differences in organic substitution.[44]

Our second check of heterogeneous effects focuses on websites that derive a high fraction of their traffic from sponsored links before the ban. One argument for the smaller increase in organic clicks for tier-C pharmacies is that the tier-C pharmacies have a lower fraction of paid clicks (47%) than tier-B pharmacies before the paid search ban, and hence we would expect less organic click substitution. This motivated us to analyze the tier-B and tier-C pharmacies with the same fraction of paid clicks before the ban. Moreover, the reliance on sponsored links depends on a website's bidding strategy for sponsored links and SEO strategies. This choice, in turn, sheds light on a website's comparative advantage in optimizing its paid and organic ranking. A website with most of its clicks originating from sponsored links may have a low organic rank and may find it harder to boost its organic rank after the ban. As a result, it may experience less organic substitution after the ban.

We use the fraction of clicks coming from paid search clicks, $FracSponsored_{initial}$, to measure the reliance on sponsored search. A greater value of $FracSponsored_{initial}$ implies a

---

[43]We only have two tier-B websites in bin one. One of them, "canadapharmacyonline.com" experienced a significant surge in organic clicks that contribute to the large estimates in bin one. Note that the small number of treated units in bin one does not affect our ability to derive the inference.

[44]For robustness, we also estimate the two-part model in Table 8 conditional on the same volume of total clicks on tier-B and tier-C pharmacies and the results do not change qualitatively.

greater negative shock from the ban as the loss of all paid clicks represents a larger fraction of the website's total clicks. Whether that shock translates into a larger or smaller organic substitution effect depends on several factors. On the one hand, the organic substitution effect can be larger for websites with a higher $FracSponsored_{initial}$ because a larger fraction of searchers are forced to seek alternative links to click, and the website's organic link is one such alternative. On the other hand, foreign pharmacy websites with higher $FracSponsored_{initial}$ previously relied more on traffic from sponsored links and may not excel at promoting their organic links using SEO. However, at the same time, they may have stronger incentives to engage in SEO in order to obtain greater organic substitution after the ban. We explore this heterogeneity by conditioning on websites with a different value of $FracSponsored_{initial}$.

In the raw data, tier-B websites have a higher $FracSponsored_{initial}$ than tier-C websites (see Table 5). The average fraction of paid clicks across websites in regime zero is 66% for tier-B websites and 27% for tier-C websites. Because very few tier-B websites have $FracSponsored_{initial}$ below 45%, we focus on two subsamples, one with $FracSponsored_{initial}$ between 45% and 80% (medium fraction) and the other above 80% (high fraction). The results are shown in Table 11. Panels A and B show the median synthetic control estimates and Panel C shows the estimates from the two-part model.[45] Panels A and B show that organic clicks on tier-B websites increase in both subsamples with a larger increase for websites with a medium fraction of paid clicks before the ban. When normalizing by the level of pre-ban total clicks, tier-C websites with a high fraction of paid clicks experience a negative and significant decrease in organic clicks in both regimes, while the effects on medium fraction tier-C websites are insignificant.

The estimates from the two-part model show that the increase in organic clicks for tier-B websites mainly occurs at the intensive margin with similar estimates for both medium and high fractions. The probability that tier-C pharmacies receive any clicks declines and the effect is stronger for the medium fraction websites. Again, the tier-B and tier-C differences,

---

[45] As a robustness check, we report the synthetic control mean effects in Appendix Table B5.

conditional on a similar initial fraction of paid clicks, suggest that there might be reasons beyond consumer search costs that drives the different organic substitution effect for tier-B and tier-C websites.

The third type of heterogeneous effect relates to a searcher's willingness to substitute sponsored ads with organic links, as it determines how much time and effort the searcher will spend looking for the organic link of a banned website. This willingness, in turn, depends on the searcher's expected benefit and cost of using a tier-B or tier-C website. The ban and the Google-DOJ settlement may have raised the perceived health, legal and technical cost of using a foreign pharmacy website, but searchers that expect overwhelming cost-saving or privacy benefits from these websites may continue to search for these websites. In contrast, those that do not expect high enough benefits from these websites may be persuaded by the warning message to avoid such websites.

This logic motivates us to look for heterogeneity in searchers' expected benefits from foreign pharmacy websites. One direct way is to examine clicks conditional on pharmacy queries that target discount pharmacies or tier-B and tier-C pharmacy names. Another approach is to examine differences across drug types. The existing literature suggests that cost saving and privacy are the most cited reasons for using online/foreign pharmacies before the ban. Thus consumers that target chronic or lifestyle drugs may expect higher benefits from the banned websites and are more willing to continue using their organic links after the ban.[46] These arguments suggest that we should analyze queries about chronic/lifestyle drugs separately from other drug queries.

Among the different pharmacy queries, we classify discount pharmacy searches and tier-B and tier-C name searches as targeted pharmacy searches and the other general pharmacy queries as non-targeted pharmacy searches. In terms of drugs queries, we define a drug query

---

[46]Following the Oxford English dictionary, we define a lifestyle drug as "a drug prescribed to treat a condition that is not necessarily serious or life-threatening but that has a significant impact on the quality of life." See http://www.oed.com/view/Entry/108129. In addition, one medical article, Gilbert et al. (2000), in the *British Medical Journal*, 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." Viagra is a prominent example of a lifestyle drug.

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 drug query is defined as non-chronic if the average prescription frequency is below 3.5 per patient per year. In total, we have identified 73 chronic drug queries.[47] In addition, 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 U.S. government (23 queries).[48]

Table 12 compares how organic clicks into tier-B and tier-C websites change after the ban conditional on whether the clicks originated with a targeted pharmacy query.[49] We present the median effects based on the synthetic controls in Panels A and B and results from the two-part model in Panel C. As a robustness check, we report the synthetic control mean effects in Appendix Table B6. Tier-B websites experience increases in organic clicks through both targeted and non-targeted query searches. As expected, the increase is much larger following queries that target discount pharmacies. For tier-C pharmacies, the reduction in organic clicks is smaller following queries that target discount pharmacies. The results of the two-part model are generally consistent and show that following queries that target discount pharmacies, the positive effect is larger on the intensive margin for tier-B pharmacies, while the negative effect is smaller (and insignificant) for tier-C pharmacies.

Conditional on different drug queries, Table 13 shows the change in organic clicks following queries for chronic drugs and lifestyle drugs, respectively. As before, we report median effects from the synthetic control method and marginal effects from the two-part model. The synthetic control mean effects are reported in Appendix Table B7 as robustness check. Tier-B websites experience a modest increase in the number of clicks following chronic drug queries. The number of organic clicks following queries for lifestyle drugs decline for both tier-B and

---

[47]Appendix Table C1 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.

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

[49]For the estimates based on synthetic controls, we reconstruct the control pool and the synthetic control websites based on clicks following each subset of queries.

tier-C pharmacies. When benchmarked by the total number of pre-ban clicks, the increase in organic clicks on tier-B pharmacies following chronic drug queries accounts for less than 10% of the total clicks in regime zero. Comparing these numbers to the increase in pharmacy queries, it suggest that even though consumers searching for the discount and lifestyle drug queries have a higher incentive to purchase from foreign pharmacies, they have an easier time finding tier-B pharmacies through pharmacy queries than through drug queries after the ban. Organic clicks on tier-C pharmacies decline following both types of drug queries. It again shows that consumers respond differently to online pharmacies based on their certification status.

To summarize, the heterogeneous effects presented above suggest that the differential organic substitution for tier-B websites is not just driven by the fact that tier-B websites are on average larger and relied more on sponsored links than tier-C websites. Rather, the ban and the Google-DOJ settlement increased searches for online pharmacy certification and tier-B websites enjoyed positive organic substitution even when compared with tier-C websites of similar click volume or those that have a similar fraction of sponsored clicks before the ban. Moreover, tier-B websites experience positive organic substitution through queries that target discount pharmacies and drugs that treat chronic conditions while tier-C websites experience no organic substitution. The organic substitution for tier-B websites is strongest through queries targeting discount pharmacies.

6   Conclusion

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

The websites certified by non-NABP agencies, referred to as tier-B websites, experience a reduction in total clicks, though some of their lost paid clicks are replaced by organic clicks.

The organic substitution effect does not change significantly after the Google-DOJ settlement. In contrast, pharmacies not certified by any of the four major certification agencies are less likely to receive any clicks at all, and the substitution to organic links after the ban is much weaker.

Overall, we conclude that the ban has increased search cost for tier-B websites, but at least some consumers overcome the search cost by switching from paid to organic links. In addition to search cost, while we do not observe the quality of drugs purchased on pharmacies in each tier, our results suggest that the ban may have increased health concerns for tier-C websites and discouraged consumers from reaching them via organic links. It is also possible that tier-C websites are buried deeper in organic results than tier-B websites, and the extra obscurity adds difficulty for consumers to switch to organic links for tier-C websites. However, this explanation cannot fully explain the differential organic substitution effect for tier-B websites. After the ban and the Google-DOJ settlement, consumers searched more for online pharmacy certification and tier-B websites enjoyed positive organic substitution even compared to websites with similar organic visibility. The organic substitution for tier-B websites is strongest following queries that target discount pharmacies.

Our study is limited in that we only observe clicks via search engines. Due to the lack of individual click-through data, we do not know whether consumers switch between drug, pharmacy and other queries after the ban. Nor do we know whether the banned pharmacies have engineered their organic results or the NABP-certified pharmacies have increased their prices or changed their advertising strategy after the ban. Finally, because clicks on tier-A pharmacies tend to be orders of magnitude larger than other pharmacies, we are prevented from fully analyzing the effects on tier-A pharmacies using synthetic controls. These questions warrant further study.

References

1. Abadie, A., A. Diamond and J. Hainmueller (2010): "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association*, 105: 493-505.

2. Abadie, A. and J. Gardeazabal (2003): "The Economic Costs of Conflict: A Case Study of the Basque Country." *American Economic Review*, 93(1): 113-132.

3. Acemoglu, D., S. Johnson, A. Kermani, J. Kwak, and T. Mitton (2016): "The value of connections in turbulent times: Evidence from the United States." *Journal of Financial Economics*, 121(2):368 - 391.

4. Athey, S. and G. Imbens (2016): "Recursive Partitioning for Heterogeneous Causal Effects." Proceedings of the National Academy of Science, 113(27): 7353-7360.

5. Bate, R., G. Z. Jin and A. Mathur (2013): "In Whom We Trust: The Role of Certification Agencies in Online Drug Market", *The B.E. Journal of Economics Analysis and Policy.* Contribution Tier, Volume 14, Issue 1, Pages 111-150, ISSN (Online) 1935-1682, ISSN (Print) 2194-6108, DOI: 10.1515/bejeap-2013-0085.

6. Baye, M. R., J. R. J. Gatti, P. Kattuman, and J. Morgan, J. (2009): "Clicks, Discontinuities, and Firm Demand Online." *Journal of Economics & Management Strategy*, 18(4): 935-975.

7. Blake, T., C. Nosko and S. Tadelis (2015): "Consumer Heterogeneity and Paid Search Effectiveness: A Large Scale Field Experiment", *Econometrica*, 83: 155-174.

8. Chan, D. X., D. Kumar, S. Ma, and J. Koehler (2012): "Impact of Ranking Of Organic Search Results On The Incrementality of Search Ads" available at http://research.google.com/pubs/pub37731.html.

9. Chan, D. X., Y. Yuan, J. Koehler, and D. Kumar (2011): "Incremental Clicks Impact Of Search Advertising", available at http://static.googleusercontent.com/media/research.google.com/en/us/pubs/archive/37161.pdf.

10. Chatterjee, P., D. L. Hoffman, and T. P. Novak (2003): "Modeling the clickstream: Implications for web-based advertising efforts" *Marketing Science*, 22(4): 520:541.

11. Chen, L. and J. Waldfogel (2006): "Does Information Undermine Brand? Information Intermediary Use and Preference for Branded Web Retailers." *Journal of Industrial Economics.*

12. Chiou, L. and C. Tucker (2016): "How Does Pharmaceutical Advertising Affect Consumer Search?" Available at SSRN: http://ssrn.com/abstract=1542934.

13. Doudchenko, N. and G. Imbens (2016): "Balancing, Regression, Difference-In-Differences and Synthetic Control Methods: A Synthesis." NBER Working Paper No. 22791. Available at: http://www.nber.org/papers/w22791.

14. Fox, S. (2004): "Prescription Drugs Online" Pew Internet & American Life Project. Available at http://www.pewinternet.org/~/media//Files/Reports/2004/PIP_Prescription_Drugs_Online.pdf.pdf.

15. George, L. and C. Hogendorn (2013): "Local News Online: Aggregators, Geo-Targeting and the Market for Local News." Available at SSRN: https://papers.ssrn.com/sol3/papers2.cfm?abstract_id=2357586.

16. Gilbert, D., T. Walley, and B. New (2000): "Lifestyle Medicines" *British Medical Journal*, 321:1341.

17. Goldfarb, A. and C. Tucker (2011): "Advertising Bans and the Substitutability of Online and Offline Advertising" *Journal of Marketing Research*, 48(2): 207-227.

18. Gurau C. (2005): "Pharmaceutical marketing on the internet: Marketing techniques and customer profile" *Journal of Consumer Marketing*, 22(7):421.

19. Jansen, B. J. and M. Resnick (2006): "An examination of searchers' perceptions of non-sponsored and sponsored links during ecommerce Web searching" *Journal of the American Association for Information Science and Technology*, 57: 1949-1961.

20. Jansen, B. J., A. Brown and M. Resnick (2007): "Factors relating to the decision to click on a sponsored link" *Decision Support System*, 44: 46-59.

21. Jansen, B. J. and A. Spink (2009): "Investigating customer click through behaviour with integrated sponsored and nonsponsored results" *International Journal of Internet Marketing and Advertising*, 5(1/2): 74-94.

22. Manchanda, P., J.-P. Dube, K. Y. Goh, and P. K. Chintagunta (2006): "The effect of banner advertising on internet purchasing" *Journal of Marketing Research*, 43(1).

23. Manning, W. G. (1998): "The logged dependent variable, heteroscedasticity, and the retransformation problem" *Journal of Health Economics*, 17(3).

24. Mullahy, J. (1998): "Much Ado About Two: Reconsidering Retransformation and the Two-Part Model in Health Econometrics." *Journal of Health Economics*, 17(3).

25. NABP (2011): "Internet Drug Outlet Identification Program Progress Report for State and Federal Regulators: January 2011" available at `http://www.nabp.net/news/assets/InternetReport1-11.pdf`.

26. Orizio, G., A. Merla, P. J. Schulz, and U. Gelatti (2011): "Quality of Online Pharmacies and Websites Selling Prescription Drugs: A Systematic Review" *Journal of Medical Internet Research*, 13(3).

27. Quon, B. S.; R. Firszt, and M. J. Eisenberg (2005): "A comparison of brand-name drug prices between Canadian-based Internet pharmacies and major U.S. drug chain

pharmacies." *Annals of Internal Medicine*, 143(6):397-403.

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

29. Skinner, B. J. (2006): "Price Controls, Patents, and Cross-Border Internet Pharmacies Risks to Canada's Drug Supply and International Trading Relations" The Fraser Institute, Critical Issues Bulletin 2006. Available at `https://www.fraserinstitute.org/studies/price-controls-patents-and-cross-border-internet-pharmacies-risks-to-canadas-drug-supply`.

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

Figures and Tables

Figure 1: Google Search Screenshot, Before the Ban
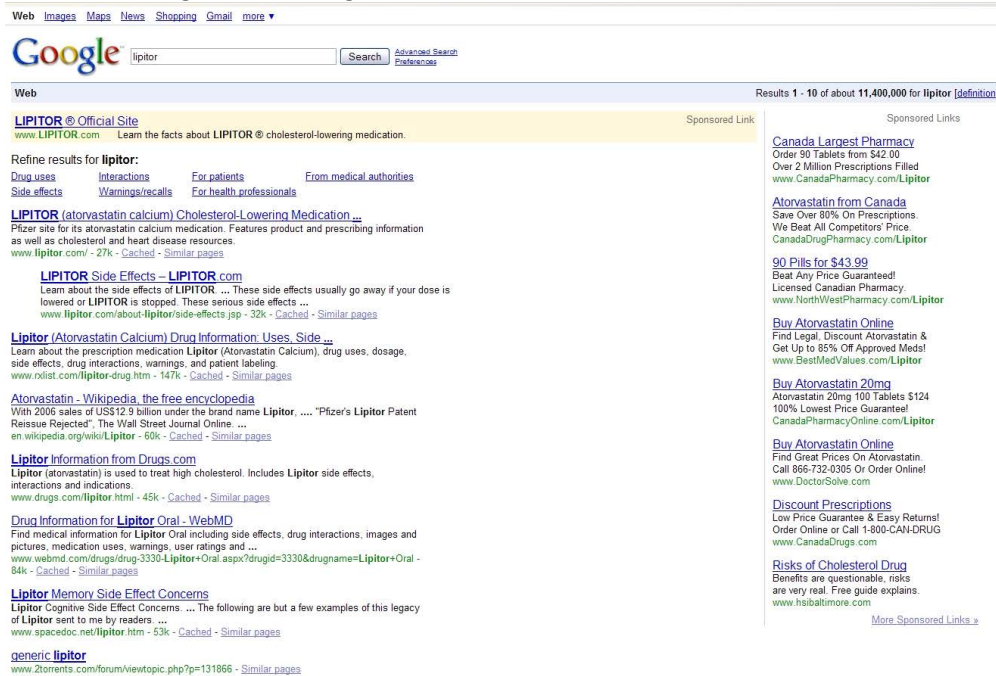
Figure 2: Google Search Screenshot, After the Ban
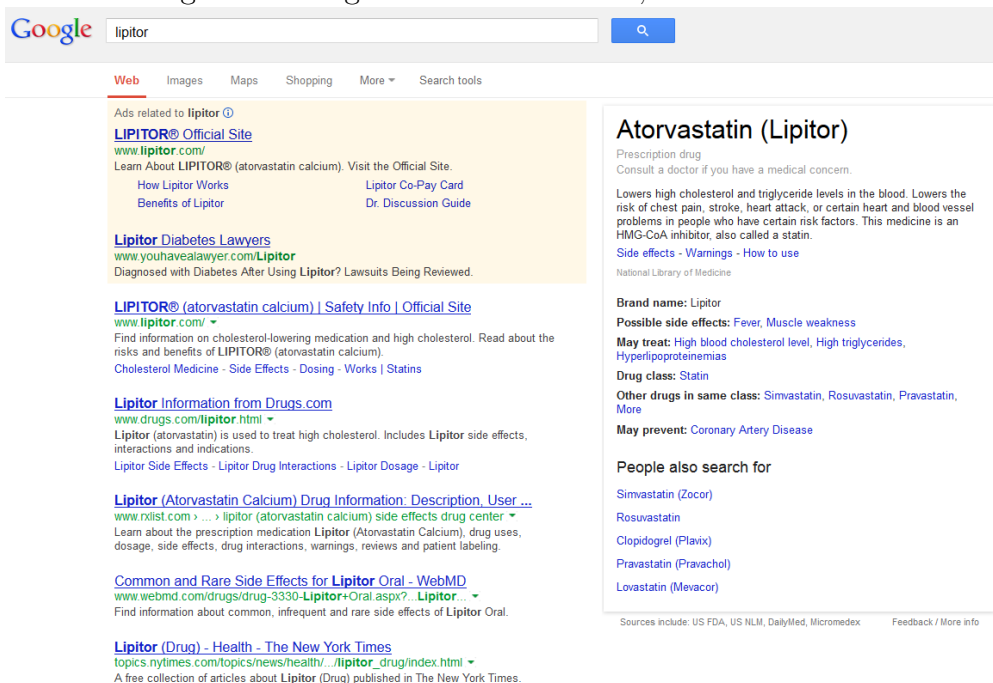
Figure 3: Example ComScore Data

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	…

Figure 4: Searchers and Searches by Broad Query Type
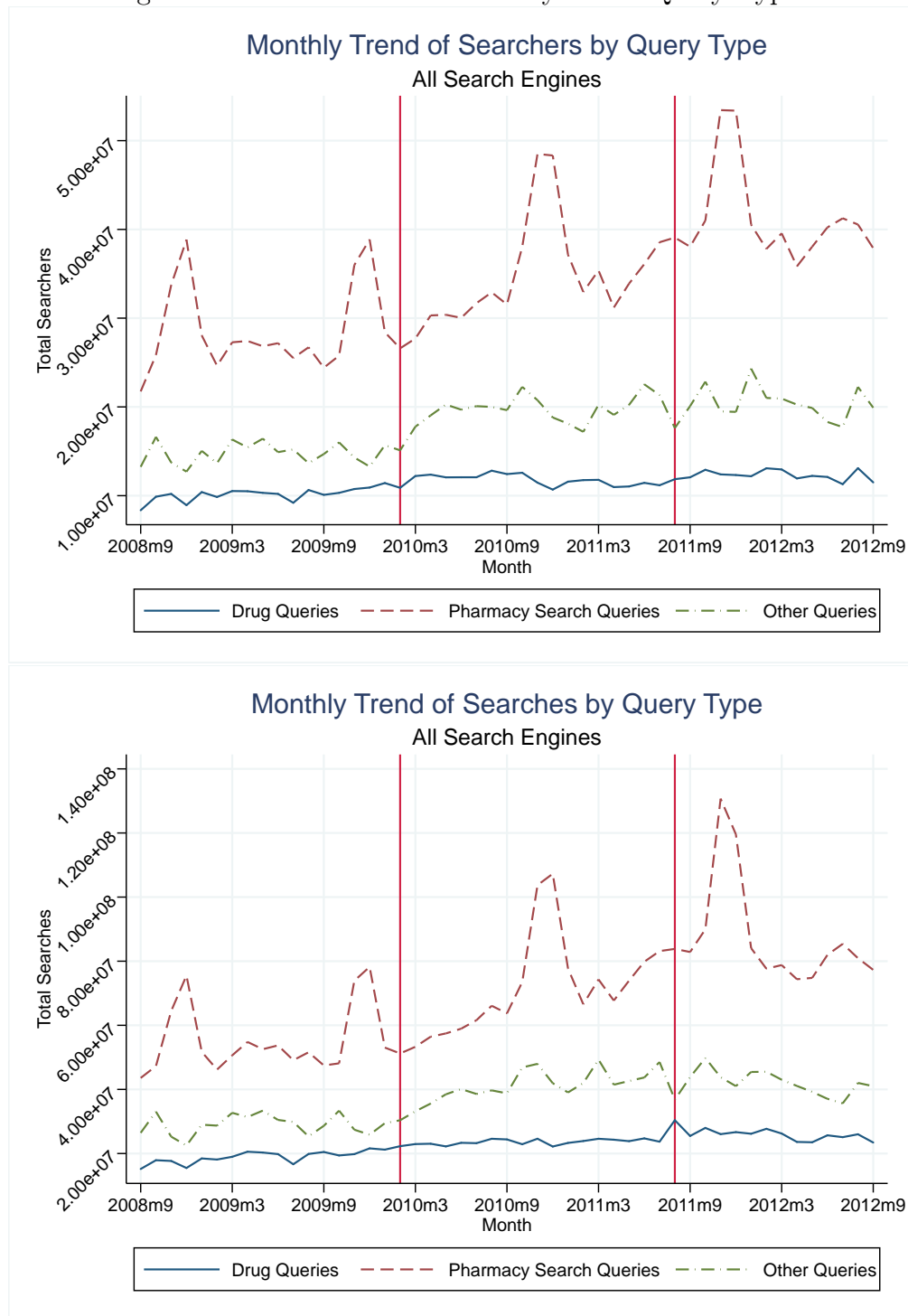
Notes: The top figure shows the total number of searchers of each query type in each month. The bottom figure shows the total number of searches of each query type in each month.

Figure 5: Clicks On Pharmacy and Control Websites
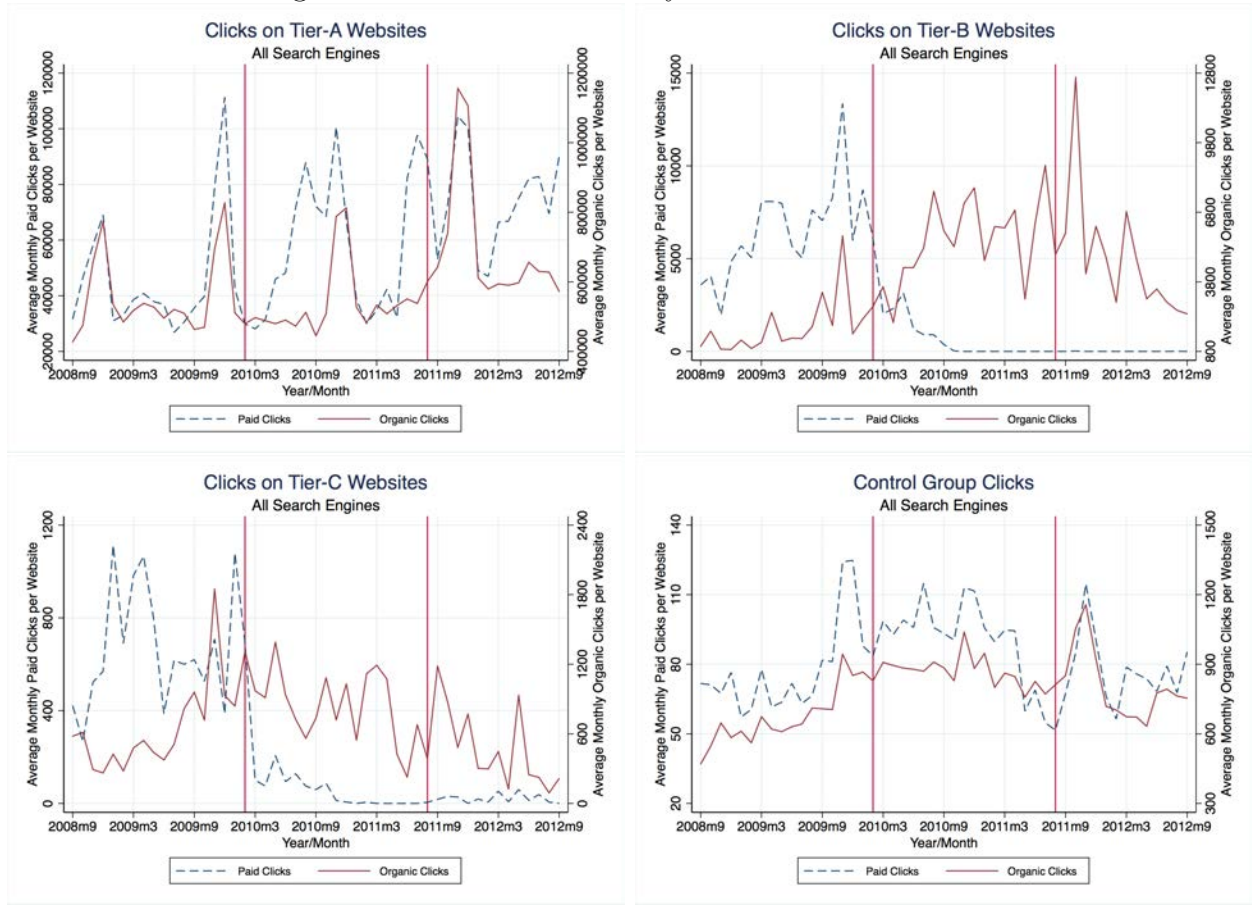
Notes: 1. The figures show the total monthly paid and organic clicks of each tier of online pharmacy website. 2. For non-pharmacy websites, we calculate total clicks by aggregating click traffic from all types of health queries excluding drug and pharmacy queries. For pharmacy websites, we calculate total clicks by aggregating click traffic from all types of health queries. 3. If the ban on sponsored links has been perfectly implemented, we should observe zero paid clicks for tier-B and tier-C websites in regimes 1 and 2. However, because screening is imperfect, we still observe a small volume of paid clicks on these websites.

Figure 6: Average Monthly Organic Clicks: Treated Vs. Synthetic Controls
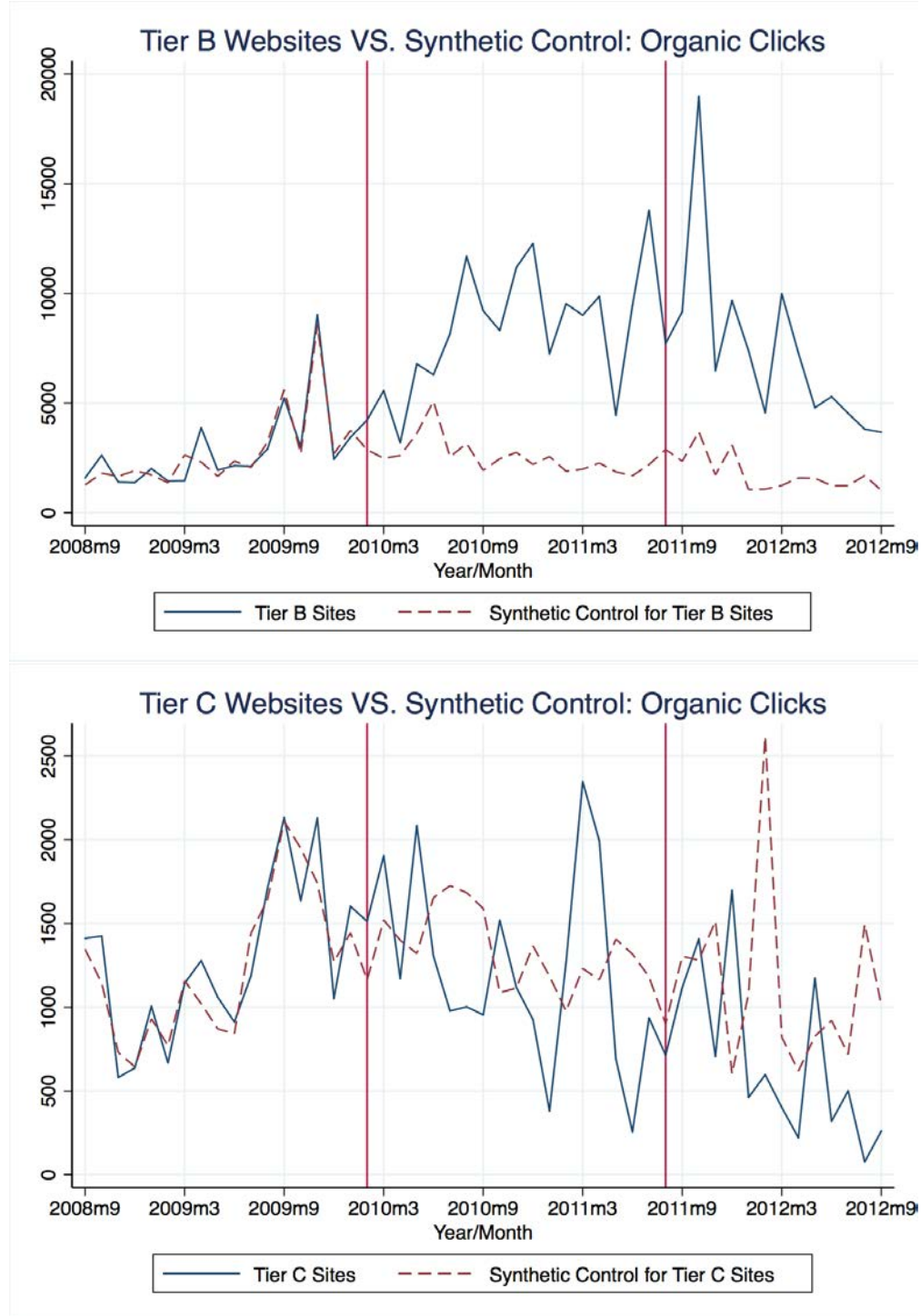
Notes: The top figure shows the average monthly organic clicks on tier-B websites and their synthetic control websites. The bottom figure shows the average monthly organic clicks on tier-C websites and their synthetic control websites.

Figure 7: Treatmemnt Effects: True Treated Websites vs. Placebos
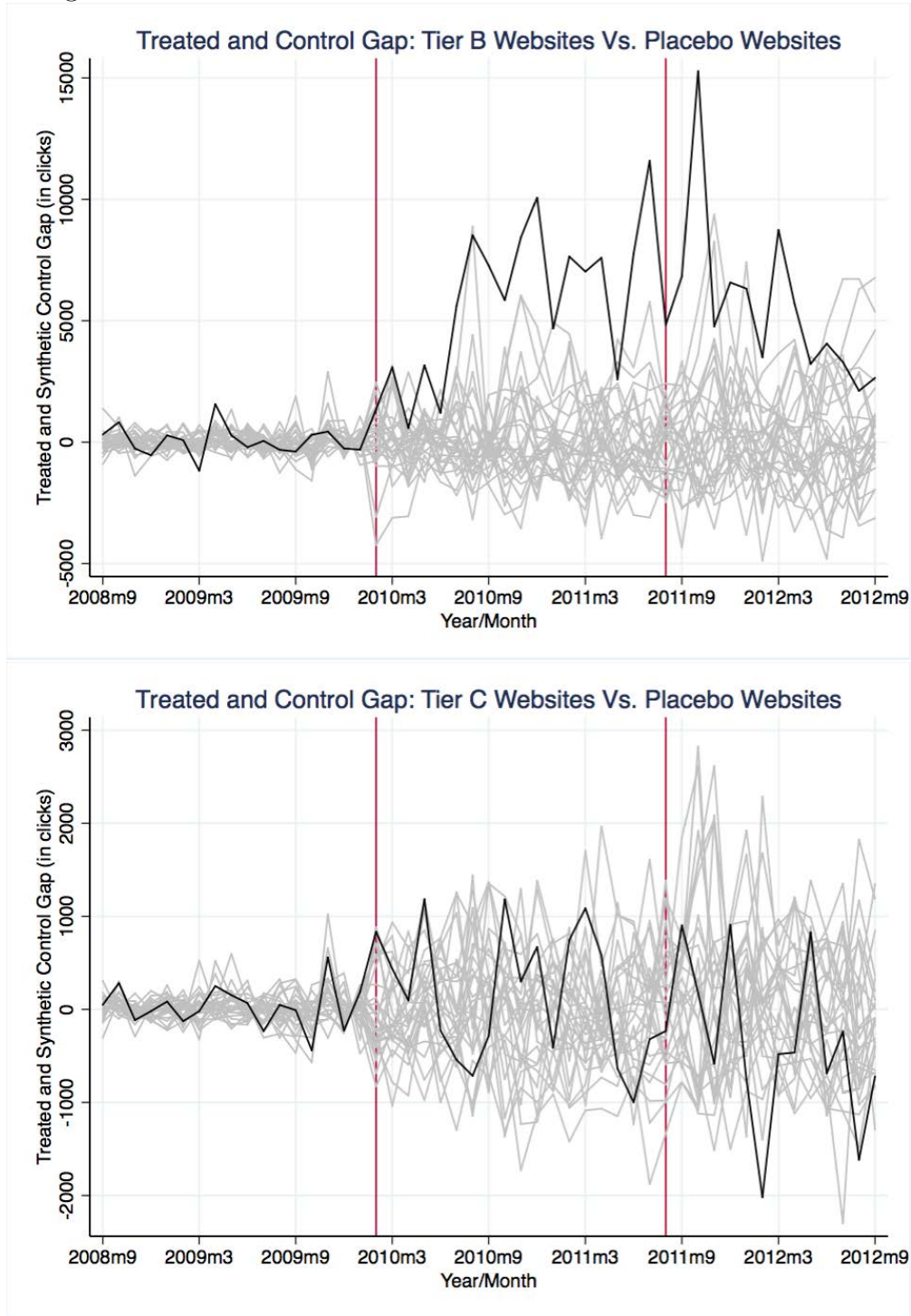
Notes: The p-value of the estimate is derived using exact inference motivated by placebo tests. For each draw of placebo websites, we randomly draw a set of websites from the control pool, the number of which is the same as the number of tier-B and tier-C treated websites. We then compute the treatment effects using the synthetic control method described above. For every draw of placebo websites, we compute the average treatment effect. By repeating this 5,000 times, we get a distribution of placebo treatment effects. The p-value is then derived from the distribution of placebo effects. The figure illustrates the distribution of the placebo treatment effects by showing 25 random draws of placebo.The bold line represents the effects for the true treated sample, and the gray lines represent the effects for each placebo sample.

Table 1: 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 U.S. 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 U.S. 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 U.S. consumers.
August 24, 2011	DOJ announced its settlement with Google

Table 2: 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

Table 3: 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 websites	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		

Table 4: Query Statistics: Overall Number of Searches and Clicks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
					Paid Clicks			Organic Clicks		
Query Type	Reg	Total Searches*	PharmClicks/ Searches	Pharmacy Clicks %	Tier-A	Tier-B	Tier-C	Tier-A	Tier-B	Tier-C
Pharmacy Queries										
General Pharmacy Search	0	832.6	0.10	27.9%	94,325	20,843	6,692	306,419	6,312	13,792
	1	1,156.6	0.08	39.7%	72,707	2,483	1,390	259,706	16,445	18,972
	2	1,208.7	0.06	21.0%	88,117	0	222	268,329	10,373	17,160
Discount Pharmacy Search	0	9.0	0.39	66.5%	932	5,889	776	3,673	2,900	3,815
	1	11.8	0.33	58.5%	1,825	815	19	3,097	10,353	5,184
	2	11.7	0.26	62.4%	1,512	1	0	3,571	10,370	3,166
Tier-A Pharmacy Names	0	5,546.1	0.50	80.6%	230,232	71	20	2,883,102	55	183
	1	7,167.0	0.51	78.2%	283,555	0	0	2,794,803	105	217
	2	8,853.2	0.45	78.8%	380,141	0	0	3,793,243	794	568
Tier-B Pharmacy Names	0	2.4	0.50	92.9%	632	366	98	2,088	652	96
	1	4.7	0.53	93.0%	721	64	0	1,695	3,319	0
	2	3.9	0.50	97.9%	958	0	0	740	3,543	0
Tier-C Pharmacy Names	0	1.4	0.47	39.8%	0	0	160	0	0	250
	1	0.6	0.48	31.4%	0	0	104	113	0	684
	2	0.6	0.00	7.1%	0	0	0	0	0	15
Certifier Search	0	2.8	1.17	6.5%	59	0	0	77	0	0
	1	2.2	0.01	1.3%	0	0	0	44	0	0
	2	4.1	0.04	1.5%	109	0	0	0	0	0
Drug Queries	0	71.9	0.14	22.1%	273	1,039	1,092	6,348	63	578
	1	89.9	0.02	2.6%	329	238	121	1,750	535	1,439
	2	97.6	0.03	3.5%	559	2	111	2,171	713	1,344

* in thousands

Notes: 1 All statistics in this table are averages across queries within each *query type × month* and the statistics related to clicks are conditional on queries that resulted in positive clicks on any pharmacy website. Column (2) "Total Searches" counts the average monthly searches per query. Column (3) "PharmClicks/Search" is the ratio of average monthly clicks on pharmacy website to the number of searches conditional on a query. Column (4) "%Pharmacy Clicks" is the average percentage of total clicks landing on pharmacy websites conditional on each query search. Column (5)-(7) for paid clicks and column (8)-(10) for organic clicks show the number of average monthly clicks that land on each tier of pharmacy websites. 2 The large number of searches of tier-A websites is due to the fact that tier-A includes retailer stores such as Walmart and Target.

Table 5: Pharmacy and Non-Pharmacy Website Statistics

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
		Mean		StdDev		Median		75th percentile		N	N	N
	Regime	paid	organic	paid	organic	paid	organic	paid	organic	Any Clicks	(Paid>0)	(Organic>0)
TierA	0	46,398	554,389	149,166	2,274,422	1	1,247	1,178	10,721	47	23	36
	1	57,602	539,122	225,225	2,277,205	0	1,487	269	13,682	40	18	36
	2	73,812	700,187	247,523	3,096,116	0	711	1,211	8,331	39	17	30
TierB	0	6,531	1,823	9,669	3,300	2,105	379	8,432	2,265	25	17	17
	1	639	5,642	1,090	10,328	0	1,110	1,088	3,906	24	12	22
	2	2	4,824	8	7,190	0	1,104	0	5,867	22	2	16
TierC	0	697	670	2,912	1,629	0	14	0	336	135	28	74
	1	52	880	281	3,208	0	0	0	185	102	13	54
	2	22	497	258	2,008	0	0	0	39	70	1	35
Non-pharmacy	0	80	694	1,642	16,390	0	0	0	7	20,226	1,210	5,129
	1	95	895	2,904	23,164	0	0	0	0	10,606	878	3,958
	2	80	832	1,770	20,654	0	0	0	0	7,353	621	2,561

Notes: 1. The click counts in the table are based on the month×website level and the statistics are calculated for each website type×regime. We keep a balanced sample of 47 tier-A websites, 25 tier-B websites, 135 tier-C websites, and 20,226 non-pharmacy websites. We excluded websites that have zero total clicks in the pre-ban period. 2. We define websites with "any clicks" as websites having received either censored or positive clicks from the set of queries in our data. The two columns report the number of websites in each regime that have any positive (non-censored) paid clicks, and have any positive (non-censored) organic clicks.

55

Table 6: Test of Pre-Trends in Regime Zero

	(1)	(2)
	$I(OrgClicks > 0)$	$ln(OrgClicks)$
t	0.0334***	0.0251***
	(0.00148)	(0.00214)
Tier-A$\times t$	-0.00266	-0.0196
	(0.0134)	(0.0126)
Tier-B$\times t$	-0.0295	0.0348
	(0.0238)	(0.0282)
Tier-C$\times t$	-0.0188	0.00567
	(0.0143)	(0.0161)
Tier-A	2.716***	–
	(0.248)	
Tier-B	2.417***	–
	(0.337)	
Tier-C	1.176***	–
	(0.158)	
Constant	-2.668***	7.475***
	(0.0214)	(0.00153)
Observations	347,361	23,577
FE		website

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

Notes: 1. The regression only includes pre-ban observations. 2. t corresponds to the month in the data and takes values between -8 and 8. 3. We exclude the indicator variables for the control group, non-pharmacy websites. 4. The dependent variable in column (1) is if a website has any *non-censored* positive organic clicks in a given month, and the dependent variable in column (2) is the website's actual number of non-censored organic clicks. 5. Standard errors are clustered at the website level.

Table 7: Synthetic Control Estimates: Effects on Pharmacy Organic Clicks

	(1)	(2)	(3)	(4)
	Mean		**Median**	
	Regime 1	*Regime 2*	*Regime 1*	*Regime 2*
Panel A. $\Delta = Y_1 - Y_0$				
Tier B	5778.7**	5561.3*	954.3**	1552.7***
p-value	0.010	0.071	0.019	0.004
Tier C	421.96	-387.31	-41.09	-69.04
p-value	0.238	0.673	0.633	0.855
Panel B. Δ/*Pre-intervention Total Clicks*				
Tier B	83.8%	67.5%	44.1%*	35.6%**
p-value	0.186	0.169	0.065	0.043
Tier C	57.6%	-0.5%	-7.2%	-8.1%
p-value	0.236	0.328	0.237	0.202
Panel C. OLS Estimates using Full Sample				
Tier B	5655.9***	4404.5***		
std.err.	(2205.1)	(1361.3)		
Tier C	54.05	-231.4**		
std.err.	(159.4)	(111.4)		

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. The table reports synthetic matching estimates of the effects of the two interventions on the organic clicks of tier-B and tier-C pharmacies. 2. The organic clicks are measured at the website×month level. *Panel A* reports the difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-intervention total clicks, *Panels B* reports the difference as a percentage of each website's pre-intervention total clicks. 3. Columns (1)-(2) report the mean effects while Columns (3)-(4) report the median effects. 4. As a comparison, *Panel C* presents the OLS estimates of the treatment effect using the full set of websites in the control sample.

Table 8: Two-Part Model Estimates: Clicks on Pharmacy Websites Compared to Control Websites

	(1)	(2)	(3)	(4)	(5)
	I(Any Click>0)	I(Total Click>0)	ln(Total Clicks)	I(Organic Clicks>0)	ln(Organic Clicks)
Tier-A×Regime1	0.332***	0.155*	-0.0927	0.153*	-0.0948
	(0.0985)	(0.0875)	(0.110)	(0.0888)	(0.114)
Tier-B×Regime1	0.398*	-0.157	-0.670***	0.123	0.807***
	(0.229)	(0.212)	(0.222)	(0.197)	(0.198)
Tier-C×Regime1	-0.146	-0.373**	-0.205	-0.211	0.0533
	(0.113)	(0.156)	(0.163)	(0.145)	(0.145)
Tier-A×Regime2	0.517***	0.290**	-0.278**	0.287**	-0.280**
	(0.136)	(0.122)	(0.136)	(0.123)	(0.127)
Tier-B×Regime2	0.614***	0.205	-0.810***	0.525***	0.872***
	(0.205)	(0.202)	(0.212)	(0.182)	(0.204)
Tier-C×Regime2	-0.426***	-0.578**	-0.334*	-0.410**	-0.0866
	(0.159)	(0.227)	(0.189)	(0.209)	(0.177)
Tier-A	1.973***	2.670***		2.704***	
	(0.241)	(0.248)		(0.247)	
Tier-B	1.971***	2.688***		2.405***	
	(0.305)	(0.343)		(0.337)	
Tier-C	1.093***	1.335***		1.166***	
	(0.120)	(0.159)		(0.157)	
Regime1	-0.316***	-0.145***	0.244***	-0.133***	0.261***
	(0.00856)	(0.0130)	(0.0174)	(0.0130)	(0.0179)
Regime2	-0.699***	-0.526***	0.410***	-0.515***	0.395***
	(0.0121)	(0.0197)	(0.0261)	(0.0199)	(0.0272)
Constant	-1.416***	-2.608***	7.717***	-2.656***	7.594***
	(0.0106)	(0.0206)	(0.0103)	(0.0211)	(0.0107)
Marginal Effects (External Margin)					
Tier-A×Regime1	0.043***	0.008*		0.009*	
	(0.013)	(0.005)		(0.005)	
Tier-B×Regime1	0.052*	0.007		-0.009	
	(0.03)	(0.011)		(0.012)	
Tier-C×Regime1	-0.019	-0.011**		-0.021	
	(0.015)	(0.008)		(0.009)	
Tier-A×Regime2	0.067***	0.015**		0.016**	
	(0.018)	(0.007)		(0.007)	
Tier-B×Regime2	0.08***	0.028		0.011***	
	(0.027)	(0.01)		(0.011)	
Tier-C×Regime2	-0.056***	-0.022***		-0.032**	
	(0.021)	(0.011)		(0.013)	
Observations	960,351	960,351	57,817	960,351	55,557
FE	-	-	Website	-	Website

Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.
Notes: 1. We exclude the indicator variables for control group, non-pharmacy websites, and pre-treatment period, regime zero. 2. Each observation is at a website×month level and we exclude the month in which the treatment occurred. 3. The dependent variable in column (1) is whether a website has any clicks, paid or organic, *including censored clicks*, in a given month. Dependent variables in columns (2) and (4) are whether a website has any *non-censored* positive total or organic clicks in a given month, respectively. The dependent variables in columns (3) and (5) are the number of non-censored positive total and organic clicks, respectively. 4. The regressions in columns (3) and (5) are conditional website/months in which the clicks are positive and non-censored. 5. Standard errors are clustered at the website level.

Table 9: Search Trends for Pharmacy Certifiers

Query	Regime 0	Regime 1	Regime 2
check pharmacist license	0	0	309
cipa	18,955	6,852	4,052
legitscript	18	278	1,275
nabp	6,369	10,121	23,996
national pharmacy certification	6	15	0
pharmacy check	226	727	1,686
pharmacy ratings	174	327	391
pharmacychecker	332	631	1,742
vipps	490	1,651	3,102

Notes: This table shows the monthly level of searches for all online pharmacy certifier related queries.

Table 10: Organic Click Analysis Grouped by a Website's Organic Click Volume in Regime Zero

	(1)	(2)	(3)	(4)	(5)	(6)
	Bin 1		**Bin 2**		**Bin 3**	
	Regime1	*Regime2*	*Regime1*	*Regime2*	*Regime1*	*Regime2*
Panel A. $\Delta = Y_1 - Y_0$ (Median)						
Tier-B	9467.9***	9202.5***	788.3**	1064.3**	-23.1	6656.3***
p-value	0.000	0.000	0.049	0.023	0.700	0.001
Tier-C	-17.7	-10.2	-1007.5**	-1177.9**	-829.4	-1020.2
p-value	0.282	0.330	0.020	0.030	0.266	0.374
Panel B. Δ/Pre-intervention Total Clicks (Median)						
Tier-B	367.8%*	324.8%*	13.2%	30.1%	-0.2%	49.5%
p-value	0.083	0.090	0.249	0.133	0.298	0.106
Tier-C	-1.0%**	-0.6%**	-53.0%	-31.0%	-17.3%	-29.9%
p-value	0.040	0.034	0.238	0.379	0.466	0.410
Panel C. Two-Part Model						
Marginal Effect on Probability of Positive Clicks						
Tier-B	0.153*	0.32*	-0.136	0.074	-0.001	0.047
s.e.	(0.08)	0.066	(0.105)	(0.07)	(0.171)	(0.107)
Tier-C	-0.029	-0.117	-0.273*	-0.261	0.012	-0.135
s.e.	(0.059)	(0.076)	(0.078)	(0.233)	(0.132)	(0.145)
Marginal Effect on Ln(Clicks) (Conditional on Positive Clicks)						
Tier-B	2.356***	1.696	0.340***	0.664***	0.933***	1.194***
s.e.	(0.551)	(1.039)	(0.0923)	(0.253)	(0.310)	(0.227)
Tier-C	0.167	-0.335	-0.0408	0.859***	-0.119	-0.326
s.e.	(0.115)	(0.388)	(0.392)	(0.174)	(0.206)	(0.268)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We split the websites into bins according to their monthly organic clicks before the search engine ban. We adjust the bin size so that the monthly organic clicks in the pre-intervention period passes the balancing test. Bin 1 websites have 150 to 400 monthly organic clicks; bin 2 websites have 700 to 1750 monthly organic clicks, and bin 3 websites have 2000 to 4000 monthly organic clicks. 2. *Panels A and B* of the table report *median* synthetic matching estimates of the effects of the search engine ban and the DOJ's settlement with Google. *Panel A* reports the median difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-intervention total clicks, *Panel B* reports the difference as a percentage of each website's pre-intervention total clicks. *Panel C* reports the estimates using the full set of controls in the two-part model. 3. The organic clicks are measured at the website×month level.

Table 11: Organic Click Analysis Grouped by a Website's Fraction of Paid Clicks in Regime Zero

	(1)	(2)	(3)	(4)
	Medium Fraction		**High Fraction**	
	Regime1	*Regime2*	*Regime1*	*Regime2*
Panel A. $\Delta = Y_1 - Y_0$ (Median)				
Tier-B	1362.4**	4097.9***	619.5**	2445.5***
p-value	0.013	0.001	0.041	0.002
Tier-C	-864.4*	-1143.9*	-24.6	-43.4
p-value	0.073	0.058	0.717	0.878
Panel B. Δ/Pre-intervention Total Clicks (Median)				
Tier-B	44.1%**	37.7%**	7.1%	22.7%
p-value	0.041	0.039	0.220	0.123
Tier-C	-17.4%	-8.1%	-3.2%*	-4.3%*
p-value	0.481	0.282	0.070	0.032
Panel C. Two-Part Model				
Marginal Effect on Probability of Positive Clicks				
Tier-B	-0.102**	-0.016	0.031	0.046**
s.e.	(0.044)	(0.039)	(0.023)	(0.021)
Tier-C	-0.357***	-0.464***	-0.0434**	-0.068*
s.e.	(0.063)	(0.197)	(0.020)	(0.041)
Marginal Effect Ln(Click) (Conditional on Positive Clicks)				
Tier-B	0.518***	0.758***	0.612	0.649
s.e.	(0.220)	(0.291)	(0.502)	(0.511)
Tier-C	-1.246	1.866***	-0.0451	-0.012
s.e.	(0.839)	(0.139)	(0.407)	(0.398)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We split the websites into bins according to the fraction of paid clicks a website receives before the search engine ban. Medium fraction websites are those with a fraction of paid clicks ranging between 48% and 80%, and high fraction websites are those with a fraction of paid clicks above 80%. The fractions are chosen by examining the clustering of fractions for tier-B and tier-C websites. Tier-B websites rarely have the fraction of paid clicks following below 50%, so we are unable to do the comparison for websites with a low fraction of paid clicks. 2. *Panels A and B* report the *median* synthetic control estimates. *Panel C* reports the estimates using the full set of controls in the two-part model. 3. The organic clicks are measured at the website×month level. *Panel A* reports the median difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-ban total clicks, *Panel B* reports the difference as a percentage of each website's pre-ban total clicks.

Table 12: Organic Click Analysis Conditional on Heterogeneous Pharmacy Queries

	(1)	(2)	(3)	(4)
	From Targeted Queries		**From Non-targeted Queries**	
	Regime1	*Regime2*	*Regime1*	*Regime2*
Panel A. $\Delta = Y_1 - Y_0$ *(Median)*				
Tier-B	1025.2***	2168.6***	54.5***	-13.6
p-value	0.000	0.000	0.002	0.883
Tier-C	-78.1**	-36.1	-130.2***	-58.0**
p-value	0.012	0.149	0.000	0.026
Panel B. Δ*/Pre-intervention Total Clicks (Median)*				
Tier-B	30.6%	46.4%	7.0%	-2.4%
p-value	0.230	0.191	0.164	0.685
Tier-C	-35.3%*	-26.0%	-57.9%	-52.0%
p-value	0.084	0.153	0.182	0.813
Panel C. Two-Part Model				
Marginal Effect on Probability of Positive Clicks				
Tier-B	0.007	0.13**	-0.014	-0.067
s.e.	(0.038)	(0.041)	(0.038)	(0.057)
Tier-C	-0.011	-0.098	-0.043*	-0.148***
s.e.	(0.039)	(0.064)	(0.024)	(0.043)
Marginal Effect on Ln(Clicks) (Conditional on Positive Clicks)				
Tier-B	1.037***	0.938***	0.702***	0.88***
s.e.	(0.204)	(0.277)	(0.197)	(0.243)
Tier-C	-0.268	-0.633*	0.276	0.756***
s.e.	(0.243)	(0.331)	(0.207)	(0.298)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We examine clicks generated from different pharmacy queries. Columns (1) and (2) are results from pharmacy queries that target discount pharmacies. Such queries include the name of tier-B and tier-C websites and queries that imply a search for inexpensive drugs (e.g., cheap drugs online). Columns (3) and (4) are results from all other pharmacy queries. When constructing clicks in the control group sample, we only condition on clicks generated by the same set of targeted or non-targeted pharmacy queries. 2. *Panels A and B report the median synthetic control estimates. Panel C reports the estimates using the full set of controls in the two-part model.* 3. The organic clicks are measured at the website×month level. *Panel A reports the median difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-intervention total and paid clicks, Panel B reports the difference as a percentage of each website's pre-intervention total clicks.*

Table 13: Organic Click Analysis Conditional on Heterogeneous Drug Queries

	(1)	(2)	(3)	(4)
	From Chronic Drug Queries		**From Lifestyle Drug Queries**	
	Regime1	*Regime2*	*Regime1*	*Regime2*
Panel A. $\Delta = Y_1 - Y_0$ *(Median)*				
Tier-B	44.8***	105.5***	-49.6*	-16.1
p-value	0.002	0.000	0.097	0.384
Tier-C	-186.3***	-113.3***	-80.6***	-16.1
p-value	0.000	0.002	0.005	0.336
Panel B. Δ*/Pre-intervention Total Clicks (Median)*				
Tier-B	3.8%	6.6%*	-3.1%	-9.2%
p-value	0.145	0.086	0.674	0.693
Tier-C	-49.7%	-41.8%	-11.7%	-11.1%
p-value	0.149	0.204	0.465	0.445
Panel C. Two-Part Model				
Marginal Effect on Probability of Positive Clicks				
Tier-B	-0.082	0.026	-0.067***	0.016
s.e.	(0.05)	(0.074)	(0.024)	(0.053)
Tier-C	-0.058	-0.095	-0.015	-0.072
s.e.	(0.069)	(0.099)	(0.038)	(0.052)
Marginal Effect on Ln(Clicks) (Conditional on Positive Clicks)				
Tier-B	0.819**	1.199	0.526	0.648***
s.e.	(0.401)	(0.735)	(0.455)	(0.195)
Tier-C	0.333	-0.117	0.0527	-0.269
s.e.	(0.275)	(0.595)	(0.290)	(0.361)

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We examine clicks generated from different types of drug queries. Columns (1) and (2) examine organic clicks generated from chronic drug queries. Columns (3) and (4) examine organic clicks generated from lifestyle drug queries. When constructing clicks on the control group sample, we only condition on clicks generated from the same set of chronic or lifestyle drug queries. 2. *Panels A and B* report the median synthetic control estimates. *Panel C* reports the estimates using the full set of controls in the two-part model. 3. The organic clicks are measured at the website×month level. *Panel A* reports the median difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-intervention total clicks, *Panel B* reports the difference as a percentage of each website's pre-intervention total clicks.

Banning Foreign Pharmacies from Sponsored Search: The Online Consumer Response
Appendix

Matthew Chesnes

Federal Trade Commission
mchesnes@ftc.gov

Weijia (Daisy) Dai

Lehigh University
dai@lehigh.edu

Ginger Zhe Jin

Federal Trade Commission & University of Maryland
jin@econ.umd.edu

May 10, 2017

Appendix A: Empirical Evidence for the No Contamination Assumption

The key assumption of the synthetic control methodology is that clicks on the control group websites are not affected by the ban and the Google-DOJ settlement. In this appendix, we present evidence that supports this assumption.

Foreign pharmacies mainly rely on drug and pharmacy queries to generate paid clicks. In theory, as soon as the foreign pharmacies are banned from bidding for paid clicks, other websites that have content related to drug and pharmacy queries will have a higher chance to win the paid search auctions. Therefore, control websites may receive disproportionately more paid clicks from drug and pharmacy queries than from other queries after the ban.

To test this potential contamination, we calculate the control websites' paid clicks generated from general health queries. General health queries focus on health conditions and exclude drug and pharmacy queries. They lead to less than 3% of the paid and organic clicks on foreign pharmacies. In light of this, banning foreign pharmacies from bidding on paid clicks is likely to only affect paid clicks from drug and pharmacy queries and not those from general health queries. Taking paid clicks from general health queries as the baseline, we compare it to paid clicks from drug and pharmacy queries. If control websites do not receive differentially more paid clicks from drug and pharmacy queries than from general health queries after the ban, it is unlikely that the ban has affected paid clicks on control websites. Using the same logic, since general health queries also lead to less than 3% of the organic clicks on foreign pharmacies, the events that affect clicks on foreign pharmacies should not affect organic clicks generated from general health queries, hence we can use the organic clicks generated from general health queries as a baseline.

Table A1 presents the results of a simple OLS that tests whether the treatments (the ban and the DOJ settlement) differentially affect clicks from drug and pharmacy queries on control websites as compared to clicks from general health queries. Each observation in the regression is the monthly number of clicks on a control website generated from drug and

1

pharmacy queries or from general health queries. Column (1) focuses on paid clicks and column (2) focuses on organic clicks. The key coefficients are the interactions between the regime one and two indicators and the indicator of clicks originating from drug and pharmacy queries. None of the key coefficients are statistically different from zero. We repeat this analysis in the two-part model and obtain similar results in Table A2. In short, we do not find any evidence suggesting that paid clicks on control websites is contaminated by the treatment.

Furthermore, even after foreign pharmacies are banned from bidding on paid links, health information websites still face fierce competition from domestic pharmacies. As shown in column (1) of Table A3, the magnitude of paid clicks on control websites are very small before the intervention (80), relative to that of tier-A pharmacies (46,398), tier-B pharmacies (6,531) and tier-C pharmacies (697). After the intervention, control websites did gain some ground on paid clicks in absolute terms (from 80 to 95) but still far less than domestic pharmacies (from 46,398 to 57,602).

Table A1: Control Website Clicks from Drug and Pharmacy Queries Versus Clicks from Other Health Queries (OLS)

	(1)	(2)
	Paid Clicks	*Organic Clicks*
DrugPhQueries×Regime1	12.55	106.9
	(11.39)	(73.73)
DrugPhQueries×Regime2	-4.735	54.17
	(9.791)	(77.98)
Regime1	1.957	91.14***
	(3.582)	(31.46)
Regime2	3.507	81.97*
	(4.892)	(44.08)
DrugPhQueries	38.39***	107.6
	(12.52)	(177.6)
Constant	38.61***	562.1***
	(5.718)	(138.5)
Observations	2,055,452	2,055,452

Standard errors are clustered at the website×querytype level.

Table A2: Control Website Clicks from Drug and Pharmacy Queries Versus Clicks from Other Health Queries (Two-part Model)

	(1)	(2)	(3)	(4)
	I(Paid Clicks>0)	ln(Paid Clicks)	I(Organic Clicks>0)	ln(Organic Clicks)
DrugPhQueries×Regime1	0.0488	-0.0521	0.0238	0.0289
	(0.0461)	(0.0724)	(0.0201)	(0.0255)
DrugPhQueries×Regime2	0.0721	-0.102	0.0316	-0.0101
	(0.0675)	(0.105)	(0.0312)	(0.0395)
Regime1	-0.238***	0.0853	-0.135***	0.227***
	(0.035)	(0.0551)	(0.0155)	(0.0186)
Regime2	-0.539***	0.468***	-0.532***	0.403***
	(0.0531)	(0.0785)	(0.0243)	(0.0291)
DrugPhQueries	0.507***		0.522***	
	(0.072)		(0.0349)	
Constant	-4.777***	7.227***	-3.217***	7.592***
	(0.0576)	(0.0217)	(0.0278)	(0.00823)
Marginal Effects				
DrugPhQueries×Regime1	.0004		.00098	
	(.0004)		(.00068)	
DrugPhQueries×Regime2	.00065		.00130	
	(.00061)		(.00104)	
Observations	2,055,452	18,691	2,055,452	88,885
FE	-	Site×QueryType	-	Site×QueryType

Standard errors are clustered at the website×querytype level.

Table A3: The Number of Clicks from Drug and Pharmacy Queries

		Paid Clicks			Organic Clicks		
		(1)	(2)	(3)	(4)	(5)	(6)
	Regime	from drug & pharmacy queries	from other health queries	(1) / (2)	from drug & pharmacy queries	from other health queries	(4) / (5)
Control Sites	0	80	40	2.0	694	583	1.2
	1	95	41	2.3	895	670	1.3
	2	80	43	1.9	832	663	1.3
Tier A	0	46,398	2,713	17.1	554,389	11,642	47.6
	1	57,602	1,192	48.3	539,122	9,412	57.3
	2	73,812	1,996	37.0	700,187	10,867	64.4
Tier B	0	6,531	198	33.0	1,823	83	21.8
	1	639	17	38.0	5,642	45	124.0
	2	2	0	–	4,824	57	85.3
Tier C	0	697	21	32.7	670	15	46.2
	1	52	1	100.7	880	10	88.8
	2	22	0	–	497	23	21.5

Note : The table presents average monthly paid and organic clicks per website from two different types of queries.

There are two possibilities where the treatment could affect organic clicks on control websites. First, if control websites won more paid clicks after the treatment, they should have received a smaller number of organic clicks due to substitution. The second potential mechanism relates to the search engine algorithm. Google relies on click traffic to determine website relevance and organic rank so foreign pharmacies may be ranked lower after the intervention because they have lower overall traffic. However, they may also be ranked higher after the intervention because some have a higher volume of organic clicks. Foreign pharmacies may also try to engage more in search engine optimization in order to improve their organic ranking. These could have a crowd-in or crowd-out effect on the organic rank of informational websites. Therefore, informational websites could be ranked lower or higher in the organic search results. Both mechanisms predict that we should observe clicks from drug and pharmacy queries on control websites to change relative to clicks from general health queries. The prediction is not supported in the above tables. No significant change is

observed for organic clicks in the regression analysis (Table A1 and Table A2). Moreover, in the raw data shown in columns (4)-(5) of Table A3, we do not see any evidence that organic clicks from drug and pharmacy queries on control websites decrease after the ban, which provides further evidence against the hypothesis.

Appendix B: Additional Results

Table B1: Tier-A Click Changes

Regime	Tier-A (Small)			Tier-A (Large)		
	Paid Clicks	Organic clicks	Total Clicks	Paid Clicks	Organic clicks	Total Clicks
Absolute Scale						
0	4,298	44,378	48,676	498,973	6,037,001	6,535,974
1	4,361	18,873	23,234	629,943	6,131,801	6,761,744
2	6,733	19,792	26,525	794,909	8,014,433	8,809,342
Log Scale						
0	8.3	8.8	8.9	12.9	15.1	15.3
1	9.0	9.0	9.1	12.9	15.2	15.4
2	9.1	9.1	9.3	13.4	15.4	15.5
Control Websites						
Log Scale						
0	7.1	7.5	7.6			
1	7.3	7.8	7.9			
2	7.7	8.1	8.2			

Notes: 1 The table shows the average of website monthly clicks in absolute scale and in log scale. Large tier-A websites include Walmart, Target, CVS and Walgreen, and small tier-A websites include all other tier-A websites. 2 Although the increase in clicks measured in log scale is small for large tier-A websites, the increase in the absolute scale is large. 3 There is also an increase in the control website clicks in log scale, and the increase is larger than those of both small and large Tier-A websites. 4 The decrease in organic clicks and paid clicks for small tier-A websites show as an increase in log scale is due to that the decrease is driven by the larger websites in tier-A.

6

Table B2: Robustness Check on Table 7: Synthetic Control Estimate Using Only Google's Search Engine Data

	(1)	(2)	(3)	(4)
	Mean		**Median**	
	Regime 1	*Regime 2*	*Regime 1*	*Regime 2*
Panel A. $\Delta = Y_1 - Y_0$				
Tier B	4089.1***	3927.4***	1457.4***	2305.5***
p-value	0.000	0.004	0.000	0.000
Tier C	489.0***	-173.7	-15.6	-32.7
p-value	0.019	0.354	0.939	0.646
Panel B. Δ/*Pre-intervention Total Clicks*				
Tier B	250.7%*	222.6%*	32.6%**	61.4%**
p-value	0.057	0.069	0.033	0.011
Tier C	21.9%	-51.6%	-6.3%	-26.9%
p-value	0.267	0.439	0.909	0.892

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: This table replicates Table 7 in the paper using data only from Google's search engine.

Table B3: Robustness Check on Table 8: Estimate Effects on Pharmacy Organic Clicks Using Only Google's Search Engine Data

	(1) I(Any Click>0)	(2) I(Total Click>0)	(3) ln(Total Clicks)	(4) I(Organic Clicks>0)	(5) ln(Organic Clicks)
TierB×Regime1	0.00258	-0.577**	-0.246	0.0432	1.077***
	(0.238)	(0.272)	(0.291)	(0.261)	(0.169)
TierC×Regime1	-0.253*	-0.462**	-0.00565	-0.155	0.160
	(0.130)	(0.206)	(0.183)	(0.187)	(0.178)
TierB×Regime2	0.692***	0.268	-0.355	0.890***	0.918***
	(0.243)	(0.238)	(0.278)	(0.249)	(0.241)
TierC×Regime2	-0.589***	-0.668***	-0.169	-0.354	-0.00827
	(0.178)	(0.246)	(0.204)	(0.224)	(0.198)
TierB	2.080***	2.719***	.	2.134***	.
	(0.349)	(0.368)	.	(0.361)	.
TierC	1.045***	1.265***	.	0.988***	.
	(0.131)	(0.168)	.	(0.175)	.
Regime1	-0.308***	-0.142***	0.235***	-0.137***	0.239***
	(0.00925)	(0.0142)	(0.0189)	(0.0142)	(0.0196)
Regime2	-0.705***	-0.546***	0.422***	-0.544***	0.391***
	(0.0133)	(0.0219)	(0.0282)	(0.0222)	(0.0294)
Constant	-1.462***	-2.666***	7.591***	-2.705***	7.490***
	(0.0114)	(0.0225)	(0.0113)	(0.0230)	(0.0118)
Marginal Effects					
TierB×Regime1	0.0003	-0.0302**		0.0022	
	(0.0301)	(0.0143)		(0.0132)	
TierC×Regime1	-0.0319*	-0.0242***		-0.0079	
	(0.0164)	(0.0108)		(0.0095)	
TierB×Regime2	0.0874***	0.014		0.045***	
	(0.0307)	(0.0125)		(0.0126)	
TierC×Regime2	-0.0744***	-0.0349***		-0.0179	
	(0.0225)	(0.0129)		(0.0114)	
Observations	830,020	830,020	46,299	830,020	44,590
FE	-	-	Website	-	Website

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: This table replicates Table 8 in the paper using data only from Google's search engine.

Table B4: Robustness Check on Table 10: Synthetic Control Mean Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	Bin 1		**Bin 2**		**Bin 3**	
	Regime1	*Regime2*	*Regime1*	*Regime2*	*Regime1*	*Regime2*
Panel A. $\Delta = Y_1 - Y_0$ *(Mean)*						
Tier-B	9467.9***	9202.5***	717.4	1600.3*	5422.7***	6659.1***
p-value	0.000	0.000	0.141	0.074	0.002	0.004
Tier-C	144.7	50.3	-1101.4 *	-960.1	1530.0**	-871.1
p-value	0.128	0.316	0.059	0.181	0.028	0.288
Panel B. Δ/Pre-intervention Total Clicks (Mean)						
Tier-B	367.8%*	324.8%*	18.2%	25.9%	32.1%	55.2%
p-value	0.083	0.090	0.352	0.269	0.239	0.207
Tier-C	42.9%	5.0%	-67.1%	-50.0%	62.8%*	-28.0%
p-value	0.233	0.389	0.233	0.346	0.051	0.441

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We split the websites into bins according to their monthly organic clicks before the search engine ban. We adjust the bin size so that the monthly organic clicks in the pre-intervention period passes the balancing test. Bin 1 websites have 150 to 400 monthly organic clicks; bin 2 websites have 700 to 1750 monthly organic clicks, and bin 3 websites have 2000 to 4000 monthly organic clicks. 2. *Panels A-B* of the table report synthetic matching estimates of the effects of the search engine ban and the DOJ's settlement with Google. *A* reports the average difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-intervention period total clicks, *Panels B* reports the difference as a percentage of each website's pre-intervention (Regime zero) level of total clicks.

Table B5: Robustness Check on Table 11: Synthetic Control Mean Effects

	(1)	(2)	(3)	(4)
	Medium Fraction		**High Fraction**	
	Regime1	*Regime2*	*Regime1*	*Regime2*
Panel A. $\Delta = Y_1 - Y_0$				
Tier-B	6700.2***	6347.4**	6244.3***	6232.3***
p-value	0.000	0.010	0.000	0.002
Tier-C	-1107.9	-765.2	-127.5	-256.0*
	*			
p-value	0.064	0.189	0.169	0.091
Panel B. Δ/Pre-intervention Total Clicks				
Tier-B	40.3%	0.0%	68.1%	25.9%
p-value	0.194	0.770	0.190	0.269
Tier-C	-15.4%	0.2%	-7.8%	-50.0%
p-value	0.494	0.772	0.204	0.364

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We split the websites into bins according to the fraction of paid clicks a website receives before the search engine ban. Medium fraction websites are those with a fraction of paid clicks ranging between 48% and 80%, and the high fraction websites are those with a fraction of paid clicks above 80%. The fractions are chosen by examining the clustering of fractions for tier-B and tier-C websites. Tier-B websites rarely have the fraction of paid clicks following below 50%, so we are unable to do the comparison for websites with low fraction of paid clicks. 2. *Panels A-B* report synthetic matching estimates of the effects of search engines ban and the DOJ's settlement with Google. *Panel A* reports the median difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-intervention period total clicks, *Panels B* report the difference as a percentage of each website's pre-intervention (Regime 0) level of total clicks.

Table B6: Robustness Check on Table 12: Synthetic Control Mean Effects

	(1)	(2)	(3)	(4)
	From Targeted Queries		**From Non-targeted Queries**	
	Regime1	*Regime2*	*Regime1*	*Regime2*
Panel A. $\Delta = Y_1 - Y_0$ *(Mean)*				
Tier-B	4256.0***	3537.9***	304.7*	60.9
p-value	0.000	0.003	0.052	0.284
Tier-C	659.0**	-25.9	-224.0**	-337.4**
p-value	0.033	0.531	0.027	0.012
Panel B. Δ/*Pre-intervention Total Clicks (Mean)*				
Tier-B	35.0%	357.9%*	26.3%*	42.8%*
p-value	0.261	0.068	0.085	0.091
Tier-C	43.4%	-37.8%	-99.9%	-83.8%
p-value	0.211	0.362	0.157	0.184

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We examine clicks generated from different pharmacy queries. Column (1) and (2) are results from pharmacy queries that target discount pharmacies. Such queries include the name of tier-B and tier-C websites and queries that imply a search for inexpensive drugs (). Column (3) and (4) are results from all other pharmacy queries. When constructing clicks on the control group sample, we only condition on clicks generated by the same set of targeted and non-targeted pharmacy queries. 2. *Panels A and B* report the mean synthetic matching estimates of the effects of search engines ban and the DOJ's settlement with Google. *Panel A* reports the average difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-intervention period total clicks, *Panels B* report the difference as a percentage of each website's pre-intervention (Regime zero) total clicks.

Table B7: Robustness Check on Table 13: Synthetic Control Mean Effects

	(1)	(2)	(3)	(4)
	From Chronic Drug Queries		**From Lifestyle Drug Queries**	
	Regime1	*Regime2*	*Regime1*	*Regime2*
Panel A. $\Delta = Y_1 - Y_0$ *(Mean)*				
Tier-B	28.7	416.3*	-28.0	408.1
p-value	0.249	0.051	0.449	0.110
Tier-C	-835.6***	-1129.8***	-612.9***	-656.3*
p-value	0.007	0.009	0.007	0.059
Panel B. Δ/*Pre-intervention Total Clicks (Mean)*				
Tier-B	3.9%	16.5% **	-10.2%	3.1%*
p-value	0.148	0.026	0.671	0.094
Tier-C	-97.8%	-88.4%	-13.2%	-38.8%

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Notes: 1. We examine clicks generated from different types of drug queries. Column (1) and (2) examine organic click generated from chronic drug queries. Column (3) and (4) examine organic click generated from lifestyle drug queries. When constructing clicks on the control group sample, we only condition on clicks generated by the same set of targeted and non-targeted pharmacy queries. 2. *Panels A and B* report the mean synthetic matching estimates of the effects of search engines ban and the DOJ's settlement with Google. *Panel A* reports the average difference in organic clicks between the foreign pharmacy websites and their synthetic controls. To compare the effect sizes with the website's pre-intervention period total clicks, *Panels B* report the difference as a percentage of each website's pre-intervention (Regime zero) total clicks.

Appendix C: Examples of Top Chronic and Lifestyle Drug Queries

Table C1: Examples of Top Chronic and Lifestyle 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 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

* ED stands for erectile dysfunction.

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 websites 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.