Proceedings Article

Pick Your Poison: Pricing and Inventories at Unlicensed Online Pharmacies

NEKTARIOS LEONTIADIS, Carnegie Mellon EPP/CyLab
TYLER MOORE, Southern Methodist University
NICOLAS CHRISTIN, Carnegie Mellon INI/CyLab

Electronic commerce has transformed how goods are supplied to consumers, but has also exposed weaknesses in supply regulations of certain goods, such as alcohol, weapons or prescription drugs. While licensed pharmacies have tread carefully with online sales, many enterprising operators have been selling pharmaceuticals without a license for years. Despite facing considerable adversity, unlicensed online pharmacies have managed not only to survive, but even to generate considerable revenue. In this paper, we attempt 1) to understand the economic reasons for their success, while facing stiff competition from both legal and illegal alternatives, and 2) to identify characteristics of their supply chains that could be used to disrupt illicit sales. We collected six months' worth of inventory and pricing data from 265 online pharmacies that advertise through search-engine poisoning. We compare this to data from Silk Road, an anonymous online marketplace, and from *familymeds.com*, a licensed online pharmacy. We discover that instead of directly competing with licensed pharmacies, unlicensed pharmacies often sell drugs that licensed pharmacies do not or cannot sell. Furthermore, unlicensed pharmacies are not only cheaper overall, but they also offer volume discounts. Clustering analysis of inventories reveals that only a few suppliers appear to cater for most unlicensed pharmacies, which suggests that cutting them off could disrupt unlicensed sales. Cross-validating our data with inventories from a random sample of 265 different pharmacies deemed "not recommended" by the National Association of Boards of Pharmacy shows that our results are consistent across different types of questionable vendors.

Categories and Subject Descriptors: K.4.1 [**Public Policy Issues**]: Abuse and crime involving computers

General Terms: Measurement, Security, Economics

Additional Key Words and Phrases: Online crime, pharmacies, web frauds, search-engine poisoning

1. INTRODUCTION

The advent of electronic commerce has created new challenges in regulating the distribution of certain goods, such as alcohol, weapons or prescription drugs. Online pharmacies in particular need to meet a number of licensing requirements before they can operate legally in a large number of countries. Because these requirements can be quite stringent, many entrepreneurs decide to forgo them and operate unlicensed online pharmacies instead.

Unlicensed online pharmacies are often the butt of jokes regarding the types of products they sell, such as erectile dysfunction treatments. They additionally face several hurdles designed to stymie their success. First, unlicensed pharmacies encounter considerable scrutiny when advertising online, leading many operators to employ questionable techniques (e.g., spam, search-engine poisoning), which are likely a lot less effective at bringing customers than legitimate advertising channels (e.g., Google AdWords). Second, the payment processors on which they rely to complete transactions may be pressured into cutting off service [McCoy et al. 2012a]. Third, unlicensed pharmacies face stiff competition both from established pharmacy stores and, more recently, from anonymous online

This work was partially funded by the Department of Homeland Security (DHS) Science and Technology Directorate, Cyber Security Division (DHS S&T/CSD) Broad Agency Announcement 11.02, the Government of Australia and SPAWAR Systems Center Pacific via contract number N66001-13-C-0131; and by the National Science Foundation under ITR award CCF-0424422 (TRUST) and SaTC award CNS-1223762. This paper represents the position of the authors and not that of the aforementioned agencies.

EC'13, June 16–20, 2013, Philadelphia, USA.

marketplaces [Christin 2013]. Yet despite these obstacles, many unlicensed online pharmacies have managed to not simply survive, but actually thrive economically, with reported annual revenues between US $12.8M and $67.7M for large pharmaceutical networks [McCoy et al. 2012b].

This paper attempts to answer two related —and complementary— questions. First, we seek to understand why the unlicensed online pharmacy business remains viable despite such adversity. Second, we investigate whether certain characteristics of the unlicensed online pharmacy ecosystem could be used to disrupt illicit sales.

Our hypothesis is that customer demand is the main driver behind the continued success of unlicensed online pharmacies. Economics tell us that customer demand should be reflected by the supply – i.e., the goods provided by these pharmacies. Thus, different from most related work, e.g., [Leontiadis et al. 2011; Levchenko et al. 2011; McCoy et al. 2012a; McCoy et al. 2012b; Kanich et al. 2008], we elect to study inventories and prices, rather than advertising techniques, payment systems, or affiliate network structure.

More specifically, we collect and analyze six months worth of inventories and prices at 265 unlicensed online pharmacies that have been advertising through search-engine poisoning [Leontiadis et al. 2011]. We compare these inventories and prices with those of a different group of 265 pharmacies simply characterized as "not recommended" by the National Association of Boards of Pharmacy (NABP). We also compare unlicensed pharmacy inventories with the inventory of a licensed pharmacy (*familymeds.com*), and with goods that can be found on Silk Road, a prominent anonymous online marketplace with a focus on illicit drugs [Christin 2013].

Our measurements lead us to discover that unlicensed online pharmacies provide inventories that are markedly different from what a typical licensed pharmacy like *familymeds.com* offers. For instance, drugs treating chronic medical conditions such as cardiac and psychiatric disorders are found disproportionately often at unlicensed pharmacies, while cancer medications are under-represented. Price-wise, unlicensed pharmacies are, as expected, cheaper than *familymeds.com* overall. More interestingly, we find that the largest discounts are offered for fake generics – i.e., non-existent generic versions of a branded drug, which are probably counterfeits. In addition, while most legitimate pharmacies do not offer volume discounts, unlicensed pharmacies do.

While unlicensed pharmacies provide strong economic incentives for customers to patronize them, clustering unlicensed pharmacies by their inventories reveals a potential weakness in their business model: More than half of the pharmacies we surveyed belong to one of only eight inventory clusters. In turn, this tells us that most of the unlicensed pharmacy ecosystem relies on a small number of suppliers, which makes these suppliers a natural target for law enforcement intervention. These results are independent of the data source used – both the pharmacies on the NABP blacklist and those advertising through search-engine poisoning exhibit similar properties.

In the rest of this paper, we start by contrasting the different types of online pharmacies and discuss some of the advertising techniques employed by unlicensed pharmacies in Section 2. We describe our data collection methodology in Section 3, analyze inventories in Section 4, and examine pricing strategies in Section 5. We compare our contributions with related work in Section 6, and draw conclusions in Section 7.

2. BACKGROUND

Categorizing online pharmacies as either "legitimate" or "rogue" oversimplifies the diversity of the market. In this section, we first give a high-level overview of the licensing requirements and accreditation programs that exist to help assess the legitimacy of online pharmacies. We then discuss advertising techniques, which may themselves be a good indicator of whether an online pharmacy is engaging in questionable business or not. We also briefly introduce emerging anonymous online marketplaces, which are at the far end of the legitimacy spectrum.

2.1. Accreditation and reputation

The first distinction one can make is between *licensed* and *unlicensed* pharmacies. Licensed pharmacies are either online front-ends to brick-and-mortar stores with a valid pharmacy license, or online pharmacies that obtain prescriptions only through third-party pharmacies with verified licenses. Licensing requirements themselves vary from country to country, or even from state to state in the case of the United States. Thus, an online pharmacy may have a perfectly valid license in Barbados, but would not necessarily be licensed to sell drugs in the United States.

Because online pharmacies may ship drugs across jurisdictions, accreditation and reputation programs have been developed to assist consumers in making informed choices. For instance, the National Association of Boards of Pharmacies (NABP) is a professional association whose members are boards of pharmacies from across North America, Australia and New Zealand. Since 1999 the NABP has established the Verified Internet Pharmacy Practice Sites (VIPPS) program,[1] which provides accreditation, for a fee, to law-abiding online pharmacies. In turn, VIPPS accredited pharmacies carry a special logo on their website, which can be used by visitors to verify the legitimacy of the online pharmacy. Additionally, NABP provides an extensive list of "not recommended" online pharmacies,[2] which fail to demonstrate that they abide to the law of their jurisdiction. Likewise, LegitScript[3] is an online service that provides a list of law-abiding pharmacies. LegitScript is backed by the NABP, and is reportedly used by Google and Microsoft to determine whether pharmacies are legitimate or not. Many other online verification programs do exist. Their stringency varies considerably, ranging from requiring valid pharmacy licenses in the US or Canada (e.g., `pharmacychecker.com`) to merely relying upon reputation forums (e.g., `pharmacyreviewer.com`). Because of the large number of online pharmacies, many pharmacies are neither accredited or licensed, nor blacklisted. For instance, `eupillz.com` does not appear, at the time of this writing, in any of the aforementioned databases.

2.2. Advertising techniques

Another indicator of the potential legitimacy of an online pharmacy is the type of advertising techniques it employs. Licensed, accredited pharmacies can purchase Google AdWords for instance, while unlicensed pharmacies have been barred from doing so since 2003. Thus, some unlicensed pharmacies resort to illicit advertising techniques. Email spam [Kanich et al. 2008] is perhaps the best known, but blog and forum spam, as well as search-engine poisoning, are increasingly gaining prominence [Leontiadis et al. 2011]. Because this latter form of advertising involves active compromise of unsuspecting Internet hosts, which is a criminal offense in many countries, we can almost assuredly categorize the pharmacies resorting to search-engine poisoning as unlicensed.

(Traditional) search-redirection attacks. Over the past couple of years, search-engine poisoning to advertise online pharmacies has increasingly relied on search-redirection attacks [Leontiadis et al. 2011; Lu et al. 2011]. In a search-redirection attack, the attacker first compromises a legitimate, unrelated website (e.g., a subdomain of `cmu.edu`). Then, depending on the type of visitor, the compromised webserver responds differently. If the site is visited by a search-engine crawler, the webserver responds with a page filled with links to other compromised webservers and a large number of drug names. This allows the compromised pages to enhance their rankings in search-engine results in response to drug queries. If the site is visited by somebody who came from a search engine looking for drugs, as evidenced by the presence of drug-related terms in the HTTP referrer field, the compromised website automatically redirects traffic to a different site under the attacker's control; after a succession of such redirects, the visitor is finally taken to an online pharmacy. Visitors that did not come to the compromised site looking for drugs are simply shown the original content, allowing these compromises to remain unnoticed for long periods of time [Leontiadis et al. 2011].

Novel variants of search-redirection attacks. We have identified two novel variants of the search-redirection attack, showing that the attackers adapt to defensive strategies. In response to attempts

---

[1] `http://vipps.nabp.net/`.
[2] `http://www.nabp.net/programs/consumer-protection/buying-medicine-online/not-recommended-sites/`.
[3] `http://www.legitscript.com/`.

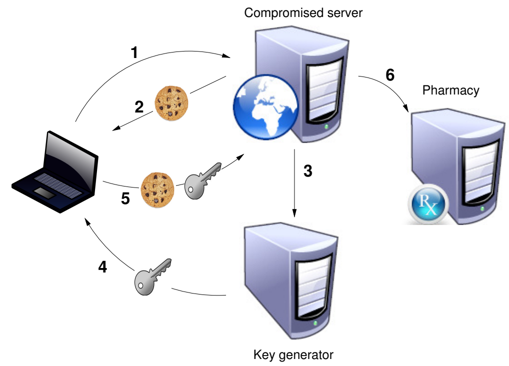


Fig. 1. Novel search-redirection attack, with a countermeasure against automated identification.

to partially anonymize the contents of the HTTP referrer field (e.g., by not showing query terms in increasingly popular HTTPS search-engine queries), attackers have started placing simple pharmacy storefronts *within* the compromised domains, and display them if they notice the traffic is coming from the search engine, regardless of the type of query being made. These storefronts typically consist of a few pictures with links; clicking on any of these links redirects the visitor to a pharmacy.

The second variant is slightly more complex, and is outlined in Figure 1. Upon connection to a compromised site (step 1), the visiting client receives a cookie (step 2), and is simultaneously redirected to a key generator site (step 3) which simply passes back a response key to the client (step 4), and redirects the client back to the compromised servers (step 5). The visiting client produces both the cookie received earlier and the response key, which triggers the compromised server to display a pharmacy storefront as in the previous variant. Clicking on any link takes the client to an actual pharmacy store (step 6). From a user standpoint, there is no difference between this attack and the previously described attack; from the attacker's standpoint, however, the use of cookies makes this type of attack significantly more difficult to detect by automated crawlers, which tend not to keep any state. Indeed if the cookie is not produced, an empty page, rather than a storefront, is displayed.

2.3. The emergence of anonymous online marketplaces

In addition to targeting consumers that use mainstream web services such as search engines, some enterprising drug peddlers have turned to anonymous online marketplaces to sell their products. Thanks to significant recent efforts to improve usability, Tor [Dingledine et al. 2004] has made anonymous Internet access widely available to even computer novices. In addition to providing anonymity for end users, Tor also supports anonymity for websites in the form of "hidden services," which are essentially webservers whose IP address is concealed. Coupled with the recent emergence of Bitcoin [Nakamoto 2008], a peer-to-peer distributed currency without any central governing authority, a number of hidden services have proliferated that sell contraband or illicit items [Christin 2013]. Different from unlicensed pharmacies, these anonymous online marketplaces do not make any claims of legitimacy: Users know that they are purchasing contraband items from an online "black market."

Perhaps the best known of those black markets is Silk Road, which primarily focuses on narcotics and prescription drugs, and has an estimated total yearly revenue of approximately $15 million [Christin 2013]. The Silk Road market operators do not themselves sell any goods, but instead provide an anonymous online forum for sellers and buyers to engage in transactions. As such, it is not a "pharmacy" per se so much as a middleman bringing together vendors of pharmaceutical goods (among others) with prospective customers. Of course, on Silk Road and other anonymous online marketplaces, no prescription or verification of any kind is required to make a purchase.

3. MEASUREMENT METHODOLOGY

We next discuss how we collected inventory and pricing data. We first explain how we selected pharmacy sites, before describing how we extracted inventories from each pharmacy.

3.1. Selecting and parsing pharmacies

We gathered data from four groups of pharmacies: 265 pharmacies that have been advertising using one of the variants of the search-redirection attacks discussed in Section 2, another 265 pharmacies that are listed as "not recommended" by the National Association of Boards of Pharmacies (NABP), 708 distinct vendors on Silk Road, and the licensed pharmacy *familymeds.com*.

Search-redirection advertised pharmacies. We identified pharmacies advertising through search-redirection attacks by adapting the crawler used in [Leontiadis et al. 2011], to analyze results from both Bing and Google. That crawler simply follows chains of HTTP 302 redirects found in response to a random subset of 218 drug-related queries until it reaches a final site, which it labels as an online pharmacy. As discussed in [Leontiadis et al. 2011], this simple heuristic is surprisingly accurate at identifying online pharmacies. We enhanced the crawler to reach pharmacies advertised using the novel attacks described in Section 2.2.

We then scraped all the candidate pharmaceutical sites our crawler identified. Starting on April 3rd, 2012, we attempted to scrape all the candidate pharmaceutical sites our crawler had identified until then; many of these domains had been taken offline, which is not overly surprising given the relatively short life span of online pharmacies. Then, between April 3rd, 2012 and October 16th, 2012, we scraped all candidate pharmaceutical sites at the time our crawler detected them.

We used `wget` to scrape the content of the candidate pharmacy domains, posing as a recent Windows Firefox client. We used random delays between different web page accesses in the same domain to avoid detection. As previously observed [Leontiadis et al. 2011], operators actively monitor visitor connections and respond to abnormal activity. Thus, we anonymized our traffic using Tor, changing Tor circuits every 15 minutes to evade IP blacklisting. Traffic anonymization came at the price of longer latencies – depending on the size of each pharmaceutical domain, the scraping process took from 4 to 12 hours to complete. As a result, we scraped each pharmacy only once.

From a set of 583 candidate pharmacies, we removed false positives (non-pharmaceutical sites), parked domains, and pharmacies for which we could not easily retrieve inventories. We subsequently obtained complete inventories for a total of 265 online pharmacies that advertise through variants of search-redirection attacks. By a slight abuse of terminology, we will refer to this set of pharmacies as the *unlicensed pharmacy set*.

NABP's "not recommended" pharmacies. We complement the unlicensed pharmacy set by a random sample of pharmacies labeled as "not recommended" by the NABP. There are 9 679 such such pharmacies. The details of how the NABP has assembled this list are unclear, but only 60 domains from the unlicensed pharmacy set are among the 9 679 "not recommended" pharmacies. This shows that the NABP is applying a set of criteria very different from ours to identify unlicensed pharmacies. Therefore, we used a random sample of those pharmacies solely to validate that our results can be generalized to a larger class of unlicensed pharmacies.

Out of the 9 679 pharmacies in the list, after excluding pharmacies in the unlicensed pharmacy set we draw a random sample of 265 domain names. We scrape these pharmacies to acquire their inventory as described above. Scraping took place between October 30th, 2012 and November 4th, 2012. We will denote this set of pharmacies as the *blacklisted pharmacy set*.

familymeds.com. Finding licensed online pharmacies for which we can collect inventory information was a surprisingly difficult task. The vast majority of popular online pharmacies we examined require valid sensitive private information, such as prescription insurance contract numbers, before granting access to their inventories and pricing information.

Thus, instead of registering for any of these domains, we opted to use *familymeds.com*, a VIPPS-accredited pharmacy based in Connecticut, which makes its entire inventory and prices freely available on its website, as our source of legitimate prices and inventories. We will subsequently show that *familymeds.com* has both good inventory coverage, and that its prices are relatively close, albeit slightly cheaper, than those at five major pharmacy chains.

Silk Road. Finally, we collect data from Silk Road, an online anonymous marketplace. As part of a related study [Christin 2013], we obtained the entire inventory of 24 385 items available on Silk Road between February 3, 2012 and July 24, 2012. We immediately discarded all Schedule I drugs (e.g., marijuana), which constitute the majority of Silk Road's inventory, as they are never sold in pharmacies (licensed or unlicensed). Then, we matched each item against a comprehensive list of drug names provided by [U.S. Food and Drug Administration 2010]. Excluding items for which no match was found narrowed down the list to 5 511 items, which were offered by 708 different vendors. After additional inspection, we discarded a number of items that either were completely irrelevant (e.g., books about drugs wrongfully listed under "prescription drugs"), or did not have all the information we need (e.g., missing dosage or number of units sold), which further reduced this list to 4 208 unique items. Even though vendors are different entities, we consider Silk Road as a unique "pharmacy" in the rest of this paper, because we conjecture that differences from one vendor to the next are small in comparison to the differences between Silk Road-like markets and online pharmacies.

3.2. Extracting inventories

Once we have the webpages of interest, we need to extract the inventory data. We wrote a generic HTML parser to accomplish this task. More importantly, building inventories requires us to identify what constitutes a "drug," and associating it the right data. Defining the notion of drug is not as simple as it sounds: is a drug defined by its brand name, or by its active ingredient? Should we include dosage in the definition, considering that, at different doses, a medication might shift from over-the-counter to prescription only?

[Aizcorbe and Nestoriak 2010] have previously discussed sets of features that, taken together, could adequately describe and track a drug. Following their lead, we decided to collect as much information as possible regarding a given "drug." Specifically, we gather the following 5-tuples for each medication: (1) *drug name* (e.g., "Viagra"), (2) *active ingredient(s)* (e.g., Sildenafil), (3) *dosage* (e.g., 10mg, 10mcg, 10%), (4) Number and type of units (e.g., 10 tablets, 1 bottle, 2 vials), which we will collectively refer to as *unit*, (5) and the *type of drug* (i.e., generic vs. brand). We then associate each of these tuples with a *price* (e.g., 10.83) and a *currency* (e.g, USD, GBP, Euro).

For each pharmacy page we scraped, we identified the main drug advertised using a list of known prescription drugs [U.S. Food and Drug Administration 2010], computing the term frequency-inverse document frequency (TF-IDF) score [Salton and McGill 1986] and picking the drug name with the highest score. In a few cases, we were able to determine the name of the main drug simply by looking at the HTML file name.

We used the same method to determine the type of drug (brand or generic). These terms are often used ambiguously, or even deceptively, by unlicensed pharmacies. For example, many online pharmacies advertise "generic Viagra." However, a generic can only be produced and traded when the associate intellectual property rights have expired, or in jurisdictions where the intellectual property rights do not apply. In the case of Viagra the relevant patent is still in effect, which means that "generic Viagra" does not legally exist in most countries. Whether this means the product sold is counterfeit medication, or simply mislabeled, is unclear without purchasing and analyzing the drug.

Using the displayed drug names, we identify the active ingredients by querying the RxNorm[4] database of normalized names for clinical drugs [Liu et al. 2005]. Unlicensed pharmacies often sell drugs that are either not licensed in the US (e.g., Silagra, Kamagra) or are simply counterfeit combinations of existing drugs (e.g., Super Hard ON). Such drugs do not have any associated ingredient in the RxNorm database, and we exclude the 119 701 such tuples from our our analysis.

[4] `http://www.nlm.nih.gov/research/umls/rxnorm/`

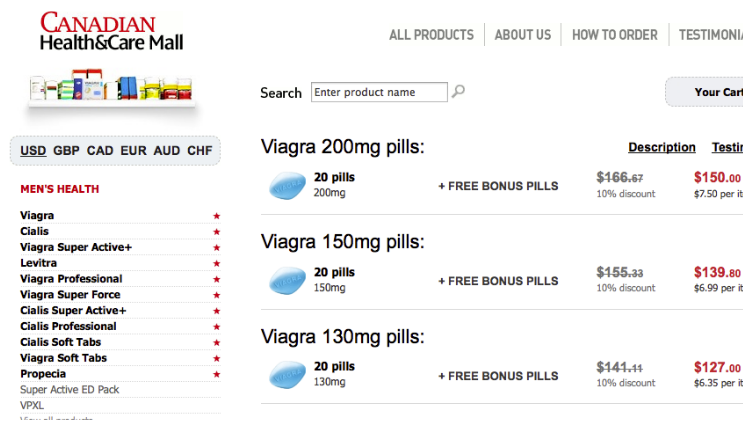


Fig. 2. Example of multiple drug names, dosages, currencies and prices presented within a single page (`rxcaredesign.com`). From this page, our parser produces three separate inventory entries.

Table I. Summary data for all four data sources. Each line presents the number of pharmacies, drug names, ingredients associated with the drugs, tuples denoting drug combinations, size of inventories, and diseases associated with the available drugs for each data source. In the case of Silk Road, we show the number of different “vendors” rather than a number of pharmacies.

Data Source	# Pharm.	Inventory size (median / mean)	Diseases targeted	# Drug names Sched.	Narc.	All	# Ing.	# Records
unlicensed pharmacies	265	157 / 170	652	42	9	1 000	557	1 022 635
familymeds.com	1	697	616	4	0	657	500	7 277
Silk Road	708	272	335	69	12	237	183	4 208
blacklisted pharmacies	265	64 / 107	726	51	7	1 283	774	417 467
Total	532		755	90	15	1 611	939	1 451 587

We then collect pricing information for each tuple. Figure 2 shows a typical example of how pricing information associated with a drug is presented. In this figure, our parser would produce three separate inventory entries. For instance, the entry corresponding to the first row in the figure would be “Viagra, Sildenafil, 200mg, 20 pills, brand, US $150.”

3.3. Collecting supplemental data

We complement our inventory entries by gathering supplemental drug attributes from several different sources.

Scheduled drugs and narcotics. We collect information related to the schedule and narcotic status of each drug.[5] The schedule classification was established as part of the Controlled Substances Act in 1970 and includes five ordered classes of drugs. Drugs are assigned to any of the schedules based on their potential for abuse and addiction. Schedule I drugs (e.g., marijuana) have the highest potential for abuse and are not deemed to have any acceptable medical use in the United States, while Schedule V drugs (e.g., Robitussin) have the lowest potential for abuse of all schedules.

Diseases treated. We use the National Drug File - Reference Terminology (NDF-RT) [Lincoln et al. 2004; Brown et al. 2004] to associate active ingredients with the diseases they treat or prevent. We use this information to group similar drugs together, and identify groups that are more or less likely to appear in a specific class of pharmacies. In addition, this information enables us to correlate pricing strategies with specific types of drugs.

[5] `http://www.deadiversion.usdoj.gov/schedules/orangebook/c_cs_alpha.pdf`.

WebMD drug classification. We supplement the NDF-RT information by data collected from WebMD.[6] WebMD groups drugs into 100 categories of medical conditions that the drugs are designed to treat, such as “Acne” or “Headache.” We extracted the drug names associated with each condition. We also used WebMD to get an idea of the drug popularity, by extracting the 180 drug names classified by WebMD as “top drugs,” which were selected “according to the number of searches submitted on WebMD for each individual drug.”

FDA drug shortage list. The FDA tracks when drugs are in currently in short supply.[7] We gathered the list of 110 drug ingredients listed as in shortage, to check their availability at unlicensed pharmacies, *familymeds.com* and Silk Road.[8] This list includes the National Drug Code (NDC) identifiers of the drugs, which are directly associated with specific combinations of drug names and dosages. We used the RxTerms[9] database [Fung et al. 2008] to decode the collected NDCs into information compatible with our drug data.

We combine the inventory information given by the 5-tuples discussed earlier, with this supplemental information, and create separate records in our database for each drug so observed.

4. INVENTORY ANALYSIS

We next present an analysis of the inventory data we gathered. In this section, we focus on item availability, rather than prices. We start with an overview of the data we have, before discussing the granularity which we will use to define “drugs.” We then compare the availability of different drug classes (scheduled drugs, for instance) across pharmacies, and we specify main types of medical conditions targeted by the unlicensed pharmacy set. Last, we perform clustering analysis on the available inventories, to identify a common pattern in the suppliers of online pharmacies.

4.1. Drug availability by pharmacy type

Table I presents a breakdown of the collected data from the unlicensed pharmacy set, the blacklisted pharmacy set, *familymeds.com* and Silk Road. We collected a total of 1 451 587 distinct (drug name, active ingredient, dosage, unit) records. These records contain 1 611 different drug names.

Drug availability. Both unlicensed pharmacies and blacklisted pharmacies exhibit the largest number of different drugs being sold, but the total number of different actual active ingredients are similar to those available on *familymeds.com*. A possible explanation is that, compared to licensed pharmacies, unlicensed pharmacies try to offer a wide variety of drug names to attract a wider range of customers. In addition, unlicensed pharmacies also target markets outside the United States, where same active ingredients often carry different market names. For instance, generic variants of Tylenol (acetaminophen) in the United States are sold as “paracetamol” in the United Kingdom. This seems to be confirmed by the fact that there are between 4.4 and 4.7 different drug names listed per disease/condition treated in the unlicensed and blacklisted pharmacies, compared to 3.4 different drug names associated with a given condition in *familymeds.com*, and 2.7 in Silk Road.

Scheduled drugs. 90 of the 1 611 drug names we found are listed under Schedules II to V, including 15 drugs categorized as narcotics. However, no pharmacy, licensed or not, sells Schedule I drugs. The licensed pharmacy *familymeds.com* does not sell any narcotics and lists only four scheduled drugs in its inventory. Both blacklisted pharmacies and unlicensed pharmacies, on the other hand, appear to sell more scheduled drugs and narcotics, and both sets appear relatively similar to each other. Silk Road tops the list in both scheduled drugs and narcotics, even after having purposefully removed all Schedule I drugs from the analysis.

[6] `http://www.webmd.com/drugs/index-drugs.aspx`.

[7] `http://www.fda.gov/Drugs/DrugSafety/DrugShortages/ucm050792.htm`.

[8] We do not consider the blacklisted pharmacies for this analysis, as we show that this dataset is very similar to the unlicensed pharmacy set.

[9] `https://wwwcf.nlm.nih.gov/umlslicense/rxtermApp/rxTerm.cfm`

Table II. Ingredients of scheduled drugs, narcotics, drugs in shortage, and top drugs at *familymeds.com*,unlicensed pharmacies and Silk Road. Numbers in bold represent statistically significant difference ($p < 0.0001$ from χ^2 tests) between two types of pharmacies for a given type of ingredient.

Type of ingredient	#Ing.	In unlic. pharm. (all)		In Silk Road		In unlic. pharm. (median)		In *familymeds.com*		#Unlic. pharm. with ingred.
In shortage	150	**75**	**(50%)**	**21**	**(14%)**	**8**	**(5.3%)**	**32**	**(21.3%)**	265
Top WebMD Drugs	283	**255**	**(90.1%)**	**93**	**(32.9%)**	**57**	**(20.1%)**	**146**	**(51.6%)**	265
Narcotics	166	10	(6.0%)	11	(6.6%)	2	(1.2%)	0	(0%)	8
Schedule (all)	486	33	(6.8%)	44	(9.1%)	1	(0.2%)	3	(0.6%)	63
Schedule II	93	10	(10.8%)	15	(16.1%)	2	(2.2%)	0	(0%)	8
Schedule III	116	9	(7.8%)	7	(6.0%)	1	(0.9%)	1	(0.9%)	46
Schedule IV	135	14	(10.4%)	21	(15.6%)	2	(1.5%)	2	(1.5%)	28

Table III. Similarities in drugs sold across different data sources. Pharmacies may sell the same drugs, but it is less common to sell the same drug and dosage, and rarer still to sell the same drug, dosage and number of pills.

Drug name	Dosage *example tuple*	Units	Unlicensed pharmacies # matches	# pharmacies	% records
∩ ***familymeds.com***					
Viagra			391	260	54.6%
Viagra	*100mg*		318	247	25.6%
Viagra	*100mg*	*30 pills*	299	243	15.4%
∩ **Silk Road**					
Viagra			164	261	32.1%
Viagra	*100mg*		138	257	11.1%
Viagra	*100mg*	*30 pills*	62	250	1.2%
∩ ***familymeds.com*** ∩ **Silk Road**					
Viagra			85	256	25.2%
Viagra	*100mg*		69	245	7.4%
Viagra	*100mg*	*30 pills*	26	234	0.6%

Comparison of drug availability by ingredient type. Beyond the absolute numbers of drugs for sale at different types of pharmacies, we are also interested in studying how comprehensive the inventories are in terms of active drug *ingredients* they provide.

Table II reports on the prevalence of different categories of drug ingredients on unlicensed pharmacies, *familymeds.com* and Silk Road. The blacklisted pharmacies have similar characteristics as the unlicensed pharmacies, and are therefore omitted from the table for simplicity. In addition to the schedule and narcotics categories mentioned above, the table also shows the availability of popular drugs and those currently in shortage. For example, 75 of the 150 drug ingredients currently in shortage are for sale at one or more unlicensed pharmacies. While this is much higher than the 32 shortage ingredients for sale at *familymeds.com*, it would be wrong to conclude that there is better availability at unlicensed pharmacies than at licensed ones, since we are comparing the inventories of 265 pharmacies to just one. A fairer comparison is between *familymeds.com* and the median number of shortage ingredients offered by unlicensed pharmacies (8).

Hence, to compare proportions between *familymeds.com* and unlicensed pharmacies using a χ^2 test, we compare median values for unlicensed pharmacies to the observed value at *familymeds.com*. We conclude that this difference in proportions (8 / 150 versus 32 / 150) is statistically significant ($p < .0001$). By contrast, when comparing inventories on the Silk Road to unlicensed pharmacies, it is better to compare the complete inventory for unlicensed pharmacies since both rely on many sellers. In the case of shortages, unlicensed pharmacies offer much greater coverage than do sellers on Silk Road (50% vs. 14%).

4.2. Product overlap between different types of pharmacies

We next investigate the extent to which products offered by different types of pharmacies overlap. Recall from Section 3, that a drug is fully described by five features: active ingredient(s), name, dosage, units, and whether the drug is a brand or a generic. As such the definition of "overlap" in inventory is actually dependent on the level of granularity we choose to define what a "drug" is. Table III shows the effect of choosing a specific level of granularity to look for matches across pharmacies. The left-hand side of the table displays the set of pharmacies and drug features we are using to look for matches within the unlicensed pharmacy set. The right hand-side of the table indicates the number of matches, number of pharmacies that contain a match, and overall fraction of records for which a match is found. For example, the first row describe matches when we only use the drug name for comparison, ignoring other attributes. An example would be to simply search matches for "Viagra," ignoring differences in dosages and units. Matching by drug names only, we find that there are 391 drugs sold both by *familymeds.com* and unlicensed pharmacies; we are able to find a match in 260 of the unlicensed pharmacies. These matches correspond to 54.6% of the 1 022 635 records (drug/price combination) we collected from the unlicensed pharmacy set.

Obviously, the more features we use to identify matching drugs, the fewer records we have available to draw conclusions from. On the other hand, these finer records are of better quality, since we know that we are comparing similar items. A particularly interesting result in Table III is that, regardless of the level of granularity considered, inventories in unlicensed pharmacies and *familymeds.com* are considerably different.

This shows that one of the ways unlicensed pharmacies compete with legitimate pharmacies is by offering different items. The fact that a large number of unlicensed pharmacies coincide with licensed pharmacies indicates that unlicensed pharmacies collectively offer a larger inventory than we can find at *familymeds.com*. This finding is confirmed by what we observe when looking at Silk Road. Silk Road, as described above, has a much richer inventory in controlled substances than both unlicensed pharmacies and *familymeds.com*. In other words, a key lesson from Table III is that, rather than purely competing on substitutes with legitimate pharmacies, unlicensed pharmacies and anonymous online market vendors are providing complementary inventories.

4.3. Identifying drug conditions served by unlicensed pharmacies

In epidemiology, it is common to observe a disease and only afterwards identify risk factors that promoted transmission. Case-control studies are suited to this task [Schlesselman 1982], and we can use this method to identify which medical conditions are at greater "risk" of being served by unlicensed pharmacies. We use the data mapping drug ingredients to 100 medical conditions from WebMD to construct risk factors. We then check how many of these ingredients are offered at unlicensed pharmacies. For each condition and each drug, we compute the probability that the drug treats the condition and is available at an unlicensed pharmacy (p_{11}); treats the condition but is not available at an unlicensed pharmacy (p_{10}); does not treat the condition but is available at an unlicensed pharmacy (p_{01}); and does not treat the condition and is not available at an unlicensed pharmacy (p_{00}). We then compute an "odds ratio" for each category, given by $p_{11}p_{00}/p_{10}p_{01}$.

We calculate 95% confidence intervals for the odds ratio using the mid-p method. Any risk factor with lower 95% confidence bound greater than 1 is positively correlated with drugs appearing in unlicensed pharmacies. Similarly, any risk factor with upper 95% confidence bound less than 1 is negatively correlated with drugs appearing in unlicensed pharmacies.

Table IV lists the 36 conditions positively correlated with appearing in unlicensed pharmacies along with 7 negatively-correlated conditions. Due to space constraints, we omit the remaining 57 conditions whose differences are not statistically significant. We can see from the table that cardiac conditions, STDs and psychiatric conditions are among the meta-categories with multiple conditions positively associated with drug ingredients offered by unlicensed pharmacies. It makes sense that cardiac drugs would be featured prominently by unlicensed pharmacies, given their widespread use as ongoing maintenance medication and considerable expense. STDs and psychiatric disorders are also often chronic conditions, which require ongoing drug treatment and consequently, recurring expenses that many consumers would opt to reduce. Furthermore, some psychiatric drugs may be abused for recreational purposes, e.g., Xanax.

Table IV. Odds-ratios identifying the medical conditions that are over-represented or under-represented in the inventories of unlicensed pharmacies.

Condition	odds ratio	95% CI	p value	Meta-condition
Conditions with more drugs sold by unlicensed pharmacies				
Cold Sores	13.2	(2.4,332.7)	0.0015	
Sinus Infection	10.5	(2.8,73.9)	0.0002	Allergies
Stroke Prevention	7.7	(2.8,27.7)	0	Cardiac
Syphilis	7.3	(2.6,26.2)	0.0001	STD
Bipolar Disorder	6	(3.4,11.6)	0	Psychiatric
Tonsillitis	5.5	(2.3,15.5)	0.0001	
Strep Throat	5.3	(2.3,13.7)	0	
Chlamydia	5.1	(2.5,11.1)	0	STD
Acid Reflux	5.1	(2.3,12.4)	0	
Bronchitis	4.9	(2.4,10.7)	0	
Congestive Heart Failure	4.6	(2.9,7.6)	0	Cardiac
Heart Attack	4.3	(2.8,6.8)	0	Cardiac
Gonorrhea	4.2	(2.2,8.5)	0	STD
Gout	4.1	(1.3,15.6)	0.0155	
Depression	4	(2.4,7.1)	0	Psychiatric
Lyme Disease	3.7	(1.5,9.9)	0.0039	
Ulcer	3.7	(1.3,12.1)	0.0157	
Anxiety	3.5	(2.2,5.9)	0	Psychiatric
Fibromyalgia	3.5	(1.5,8.8)	0.0039	
High Blood Pressure	3.4	(2.4,4.7)	0	Cardiac
Ear Infection	3.4	(2.0,5.8)	0	
Stomach Flu	3.3	(1.1,11.0)	0.0312	
COPD	2.9	(1.6,5.4)	0.0004	
Dementia	2.9	(1.6,5.4)	0.0007	Psychiatric
Bursitis	2.9	(1.3,6.4)	0.0065	
Diabetes	2.7	(1.9,4.0)	0	
Asthma	2.6	(1.8,3.7)	0	
Tendonitis	2.6	(1.3,5.7)	0.0108	
Hives	2.3	(1.3,4.3)	0.0053	
Edema	2.3	(1.1,5.1)	0.0324	
Chest Pain	2.2	(1.3,3.6)	0.0038	Cardiac
High cholesterol	2.1	(1.1,3.8)	0.0227	Cardiac
Bladder Infection	2.1	(1.1,4.1)	0.0315	STD
Staph Infection	2	(1.3,3.1)	0.0012	
Pneumonia	1.7	(1.2,2.6)	0.0054	
Arthritis	1.6	(1.1,2.2)	0.0136	
Conditions with fewer drugs sold by unlicensed pharmacies				
Constipation	0.17	(0.01,0.88)	0.0316	
Lung Cancer	0.31	(0.11,0.68)	0.0026	Cancer
Endometriosis	0.34	(0.16,0.65)	0.0006	
Anemia	0.38	(0.23,0.60)	0	
Lymphoma	0.54	(0.36,0.79)	0.0012	Cancer
Leukemia	0.6	(0.38,0.92)	0.0179	Cancer
Psoriasis	0.66	(0.43,0.98)	0.0408	

By contrast, three of the seven conditions negatively associated with unlicensed pharmacies are forms of cancer. One possible explanation is that cancer medications are frequently administered by hospitals, and so consumers are less likely to fill prescriptions directly. Furthermore, many people might be willing to try an online pharmacy to treat chronic conditions such as diabetes and cardiac medication, but they would balk at doing so for drugs to treat cancer.

4.4. Identifying suppliers

We next turn to looking at similarities in inventories *within* the unlicensed pharmacy set. As has been described in previous papers [Leontiadis et al. 2011; Levchenko et al. 2011; McCoy et al. 2012a; McCoy et al. 2012b], unlicensed pharmacies often operate as parts of affiliate networks. That is, affiliates essentially set up storefronts, and are in charge of finding ways of bringing traffic to them. On the other hand, once a sale is completed, they are not actually involved in the shipping and delivery of the drugs. This task is handled by the affiliate network operators, who collect most of the sales revenues [McCoy et al. 2012b]. Prior work by [Levchenko et al. 2011] observed similarities in web pages from identical affiliate programs, but did not draw any conclusions regarding inventories. Whether or not different affiliate networks use different inventory suppliers, or whether a given affiliate network uses multiple suppliers is not entirely clear.

As in related work on malware classification [Jang et al. 2011] or webpage classification [Levchenko et al. 2011], we use the Jaccard distance to determine how (dis)similar two pharmacy inventories are. If A is the inventory of pharmacy $\mathcal{A}$ and B the inventory of pharmacy $\mathcal{B}$, the Jaccard distance J_δ between A and B is given by:

$$J_\delta(A,B) = 1 - J(A,B) = 1 - \frac{|A \cap B|}{|A \cup B|} \,.$$

If two pharmacies share the same exact inventories their Jaccard distance will be equal to 0, and if their inventories have nothing in common, then their distance is 1.

We plot a heat map of the Jaccard distances between all pharmacy pairs in Figure 4(a). The color of each point denotes the similarity in the inventories between a pair of pharmacies. Blue (distance 0) represents represents identical inventories, while brown (distance 1) represents no overlap. After reordering columns to pool together Jaccard distances that are close to each other, clusters of similar inventories appear quite clearly in the figure.

We can define two pharmacies as belonging to the same cluster if their Jaccard distance is below a threshold t. To recursively merge clusters we consider three alternatives:

Single linkage, where the distance between two clusters of pharmacies X and Y is defined as the distance of the two most similar members of the clusters. That is, the distance between two clusters is $J_\delta(X,Y) = \min_{x \in X, y \in Y}\{J_\delta(x,y)\}$, where x and y correspond to inventories of pharmacies in each cluster, respectively.

Complete linkage, where the distance between two clusters of pharmacies X and Y is defined as the distance of the two most dissimilar members of the clusters. That is, $J_\delta(X,Y) = \max_{x \in X, y \in Y}\{J_\delta(x,y)\}$.

Average linkage [Sokal 1958], where the distance between two clusters of pharmacies X and Y is defined as the average distance between all pairs of members in both clusters, that is $J_\delta(X,Y) = \frac{1}{|X|\cdot|Y|}\sum_{x\in X}\sum_{y \in Y} J_\delta(x,y)$.

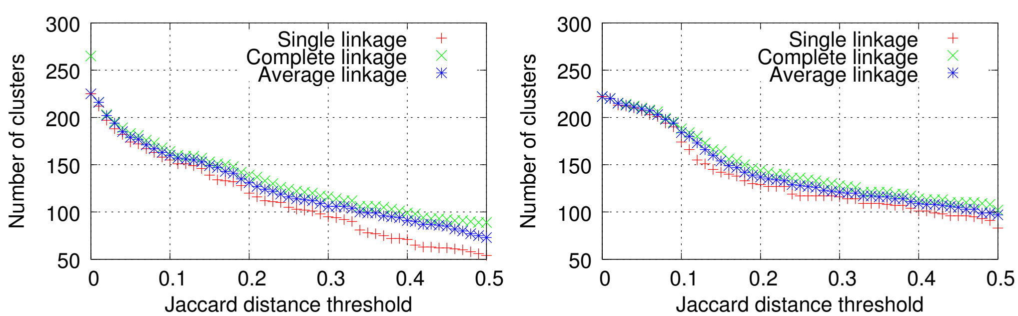
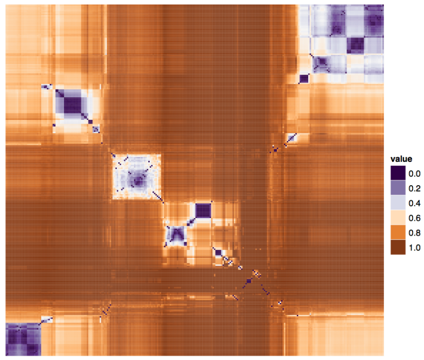
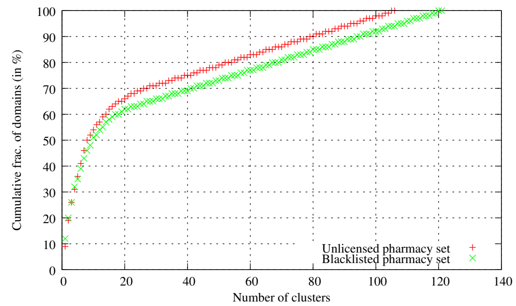


Fig. 3. Effect of different levels of distance threshold and different linkage criteria. The left plot left uses inventory data from the unlicensed pharmacy set. The one on the right uses inventory data from the blacklisted pharmacy set.

Figure 3 shows how many clusters are identified as a function of the distance threshold. The left plot corresponds to the unlicensed pharmacy set, while the right plot corresponds to the blacklisted pharmacy set. The lines correspond to the different linkage criteria. A good threshold value is empirically defined as a value for which the number of clusters remains constant even if we slightly adjust the threshold. Using average linkage, we find that $t = 0.31$ is a good choice for the threshold.



(a) Heat map of the Jaccard distances between all pairs of pharmacies in the unlicensed pharmacy set. After reordering pharmacies, we observe a number of clusters that appear to show similarities.



(b) Cumulative distribution of pharmacies as a function of the number of clusters considered (Using average-linkage, $t = 0.31$).

Fig. 4. Clustering the pharmacies based on their inventories.

Table V. Per-unit price discounts for different types of drugs. Positive differences mean that unlicensed pharmacies offer lower median prices. Differences are statistically significant according to Mann-Whitney U-test with $p << 0.0001$.

Price comparison	Types of drugs	Difference (median)	95% C.I.	Sig.?	# Records
familymeds.com – *unlicensed pharmacies*	All drugs	\$2.14	(\$2.12, \$2.17)	✓	171 098
	Fake generics	\$1.54	(\$1.49, \$1.58)	✓	41 669
	Drug in shortage	\$0.72	(\$0.59, \$0.85)	✓	3 966
	Popular drugs	\$0.36	(\$0.32, \$0.41)	✓	95 308
Silk Road – *unlicensed pharmacies*	All drugs	-\$0.46	(-\$0.54, -\$0.37)	✓	3 821

This value is incidentally very close to the value ($t = 0.35$) used by [Levchenko et al. 2011] in their related analysis. More interestingly, we find that $t = 0.31$ is an appropriate choice for both the unlicensed pharmacy set and the blacklisted pharmacy set.

In Figure 4(b) we plot the cumulative distribution of the pharmacies as a function of the number of clusters considered. Clusters are ranked by decreasing size. While we observe 82 singletons, the key finding is that, for unlicensed pharmacies, half of the pharmacies belong to one of eight clusters. We obtain similar results for blacklisted pharmacies (101 singletons, 9 clusters corresponding to 50% of all pharmacies), which is another piece of evidence that the unlicensed pharmacy set and the blacklisted pharmacy set have roughly similar properties.

In short, we do observe fairly large concentrations in similar inventories. This confirms that unlicensed pharmacies operate with a relatively small set of suppliers. From an intervention standpoint, this is good news: If the few factories supplying these drugs can be subject to more stringent controls, potential harm resulting from self-medication [Gelatti et al. 2013] could be significantly reduced.

5. PRICING STRATEGIES

We now turn our attention to the product prices offered by the sets of pharmacies we are studying. We measure the price variation from one type of pharmacy to another, and we look at factors that might be affecting it.

5.1. Pricing differences by seller and drug characteristics

Table V summarizes several price differences we examined. In the first test, we confirmed that prices are considerably cheaper at unlicensed pharmacies than at *familymeds.com*. For this test,

Table VI. Median unit prices and percentage discounts offered by *familymeds.com* and unlicensed pharmacies for 60-pill and 90-pill orders relative to the unit price of 30-pill orders. Differences are statistically significant according to Mann-Whitney U-test with $p << 0.0001$.

	30 pills	**60 pills**	discount				**90 pills**	discount			
	unit price	unit price	$	95% CI	Sig. diff.?	%	unit price	$	95% CI	Sig. diff.?	%
familymeds.com	$3.86	$3.86	$0.00	($0.00,$0.00)		0%	$3.86	$0.00	($0.00,$0.00)		0%
unlic. pharm.	$1.77	$1.60	$0.16	($0.15,$0.18)	✓	10.0%	$1.48	$0.27	($0.25,$0.29)	✓	16.9%

we compare the prices for drugs that are available at *familymeds.com* and an unlicensed pharmacy when a direct comparison is possible – that is, when both pharmacies sell a drug with the same name, dosage, and in the same number of units (e.g., a 10-pack of Lipitor 10mg pills costs $8.99 at `pills4everyone.com`, compared to $41.90 at *familymeds.com*). We normalize all prices to the per-unit price (e.g., the aforementioned pills cost $0.89 each at `pills4everyone.com`, a discount of $3.30 compared to the $4.19 at *familymeds.com*).

Overall, the median difference in per-unit prices between *familymeds.com* and unlicensed pharmacies is $2.14. This difference is statistically significant according to the Mann-Whitney U-test ($p \ll 0.0001$), while the 95% confidence interval is ($2.12, $2.17). Thus, we can safely conclude that unlicensed pharmacies are significantly cheaper than *familymeds.com*.

We confirmed this finding may generalize to other legitimate pharmacies by consulting GoodRx,[10] which provides pricing information across different pharmacies for specific drugs. We compared prices from *familymeds.com* with those offered at five major pharmacy chains (CVS, RiteAid, Target, Walgreens, and Wal-Mart), and observed that, for a sample of 433 drugs on which we could directly compare prices, *familymeds.com* is generally cheaper than the median price of these major chains, with a median price difference of $5.55. This difference is statistically significant (Mann-Whitney U, $p < 0.001$); the 95% confidence interval is ($2.34, $8.64). In other words, unlicensed pharmacies appear to be significantly cheaper than these major pharmacy chains as well.

We also want to find whether any other characteristics of the drugs on sale might influence the magnitude of the pricing advantage. To that end, we next study differences in the size of discount offered by unlicensed pharmacies relative to *familymeds.com*. One common deceptive tactic employed by unlicensed pharmacies is to offer "generic" versions of drugs where no such generic exists (e.g., because the patent is still in effect). We found around 42 000 such "fake generic" discrepancies in our dataset. The median per-unit price discount for fake generics is $3.14, compared to a $1.70 discount for other drugs not mislabeled as generic. The Mann-Whitney U-test estimates a median difference of $1.54 in the discount for fake generics. This suggests that deceiving customers with promises of branded generics can be financially enticing. We also find smaller, yet still statistically significant price discounts for drugs in shortage and those identified in WebMD as "top drugs." All these differences are significant for p-values much less than 0.0001.

How do prices compare between unlicensed pharmacies and drugs sold on Silk Road? While Silk Road has become notorious for selling narcotics even though other unlicensed pharmacies do not, sellers on Silk Road also offer non-narcotics for sale, many of which can also be bought from unlicensed pharmacies. Overall, drugs found on Silk Road are $0.46 cheaper per unit than their unlicensed counterparts. This is somewhat surprising, given that privacy-concerned customers drawn to Silk Road might have been expected to be willing to pay a premium for purchasing anonymously.

5.2. Volume discounts as competitive advantage

Another way for unlicensed pharmacies to entice prospective customers is to offer discounts when buying at higher volumes. We examined the prices of drugs offered in both *familymeds.com* and unlicensed pharmacies at the same dosage and number of units. Of the 171 098 matching tuples,

[10] `http://www.goodrx.com/`.

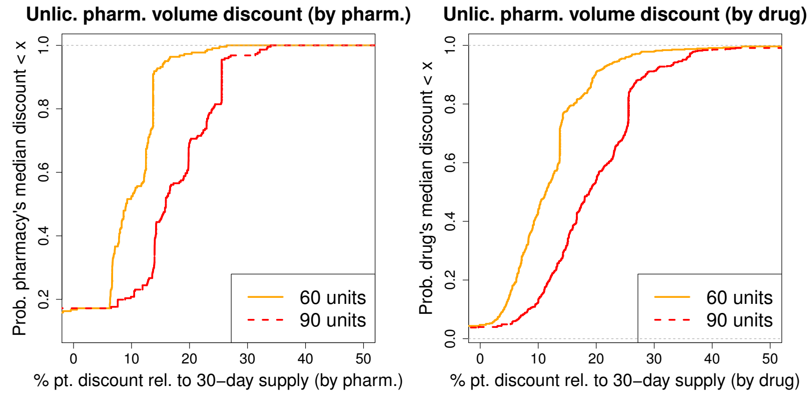


Fig. 5. Cumulative distribution functions of the median percentage-point price discount per pharmacy (left) and per drug (right).

156 136 (91%) offered 30, 60, or 90 pills. These drugs were offered by 221 unlicensed pharmacies, 83% of the total. We therefore focus our analysis on only these drugs.

A single unlicensed pharmacy can sell same combination of drug, dosage and units at several prices. This happens for two reasons. First, the drug may be sold in different currencies. Second, the pharmacy may sell multiple variants of the same drug (e.g., Super Viagra and Viagra Professional) at different prices. To simplify comparison, for every pharmacy and drug, dosage and unit combination, we compute the median of all per-unit prices. For example, `rx-pharm-shop.com` sells a 30-pack of Viagra 10mg in USD, GBP and EUR in different varieties, yielding 9 different prices. Its median per-unit price is \$2.74, falling to \$2.44 for 60-pack prices and \$2.06 for 90-pack prices. Overall, we observe median per-unit 30, 60, and 90-day prices for 20 124 distinct drug-dosage-pharmacy combinations.

We check for discounts in the per-unit prices of 60- and 90-unit supplies relative to the per-unit price of a 30-unit supply. Table VI presents our findings. First, we never observed a per-unit discount on drugs from *familymeds.com*. By contrast, the 221 unlicensed pharmacies offered a median discount of 10% for 60-day supplies, rising to a 16.8% discount for 90-day supplies. The unit price on unlicensed pharmacies falls from \$1.77 on 30-day supplies to \$1.60 for 60-day supplies and \$1.48 for 90-day supplies. These discounts are statistically significant according to the Mann-Whitney U-test for $p << 0.0001$.

To describe how discounts vary by pharmacy, the left plot in Figure 5 presents a CDF of the median percentage point discount offered by each unlicensed pharmacy. We can see that over 80% of pharmacies offer a discount of at least 7%. While volume discounts are the norm, around 15% of pharmacies actually charge more per-unit for larger volumes, which is surprising since a consumer could simply buy multiple 30-unit supplies instead. The discounts are consistently greater for 90-unit supplies than for 60-unit supplies. Finally, a few pharmacies offer very deep discounts at higher volume – around 5% of pharmacies offer median discounts exceeding 15% for 60-day supplies and 25% for 90-day supplies.

We also observe substantial variation in discounting according to the drug sold, as shown in Figure 5 (right). The median discount for drugs is 11.3% for 30-day supplies and 18.8% for 90-day supplies. However, the 10% most deeply-discounted drugs save at least 20% for 60-day supplies and 29% for 90-day supplies. We conclude that unlicensed pharmacies can use volume discounting as a way to attract prospective customers, particularly as the tactic may not be used widely by legitimate pharmacies. Indeed, prescriptions often limit the amount of drugs that can be purchased at once, making volume discounting less practical for legitimate pharmacies.

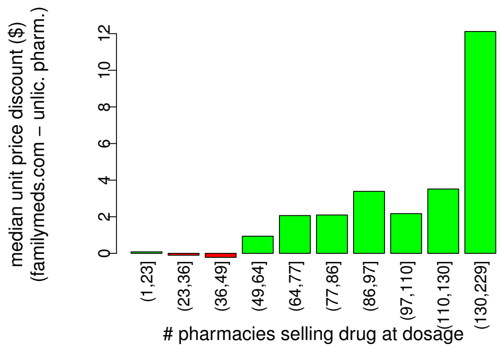


Fig. 6. Bar plot of the median unit price discount for drug-dosage combinations grouped in increasing number of unlicensed pharmacies selling the drug at the specified dosage.

5.3. How competition affects pricing

We have already seen that unlicensed pharmacies adjust prices strategically in order to attract customers, ranging from discounting volume sales to offering fake generics. They must also react to competition from other unlicensed pharmacies. Some common drugs are sold by nearly all the pharmacies we studied, while other more obscure drugs are sold by just a few. Microeconomic theory predicts that competition drives prices down; we now examine whether prices set by unlicensed pharmacies do in fact fall when competition among sellers is high and rise when competition is low.

To answer this question, we examine all prices for combinations of drugs and dosages. We normalize each price by the number of units sold at the drug's dosage and then compare the median normalized prices offered at *familymeds.com* and unlicensed pharmacies. We compute the price difference between *familymeds.com* and each unlicensed pharmacy selling a drug-dose combination. We then compute the median of this difference across all pharmacies selling that drug-dose combination. For example, the following 7 pharmacies sell Mirapex 1mg at these prices per pill:

Pharmacy	unit price (unlic. pharm.)	unit price (*familymeds.com*)	discount
`yourhealthylife.cc`	$4.37	$4.57	$0.20
`drugs-medshop.com`	$1.89	$4.57	$2.67
`pharmaluxe.com`	$4.14	$4.57	$0.43
`24medstore.com`	$4.00	$4.57	$0.56
`online-canadian-drugshop.com`	$4.37	$4.57	$0.20
`safetymedsonline.com`	$4.37	$4.57	$0.20
`7-rx.com`	$1.86	$4.57	$2.71
Median	-	-	$0.43

We then check whether the number of unlicensed pharmacies influences the median discount offered. We group the drug-dosage combinations into deciles according to how many pharmacies sell them. Figure 6 plots the median discount offered for each decile. We can see that less popular drugs offer a very small discount, and sometimes even charge slightly more than *familymeds.com* does. However, as more pharmacies sell the drug, competition drives pharmacies to sell at a higher discount relative to the price charged by *familymeds.com*. For example, the median discount for drugs sold by 87–97 pharmacies is $3.39. The discount rises to $12.12 for the 10% most popular drugs.

Discounts in the blacklisted pharmacies. We performed a similar analysis to check for a significant difference between the discounts in the two sets of unlicensed pharmacies. The result of Mann-Whitney U-test showed that the difference in the observed discounts between the unlicensed pharmacy set and the blacklisted pharmacy set are statistically insignificant. In other words, the observation of discounts for volume purchases is not limited to the main set of unlicensed pharmacy set, and is not caused by any measurement bias. On the contrary, the discounting phenomenon is characteristic of all unlicensed pharmacies.

6. RELATED WORK

Since the realization that online crime is financially motivated [Moore et al. 2009], a number of studies have emerged that focus not only on the technical details of illicit online activities, but on their financial aspects as well. For example, [Levchenko et al. 2011] studied the supply chain utilized to monetize spam, and [McCoy et al. 2012a] go into deeper detail, studying the payment networks utilized in this monetization chain. Other examples in the domain include [Moore et al. 2011] that examine the monetization of trending terms, as well as [Franklin et al. 2007] that measure the advertised prices of illicit commodities in online underground forums.

Our work is, on the one hand, related to measurement studies that focus on specific aspects of online markets to gain better insights over some observed behavior. In this category, we can relate to [Scott and Yelowitz 2009] that manually collect inventory information from a number of Internet-based stores, and analyze anomalies in the pricing of diamonds; and to the work of [Lin et al. 2006] that study the effect of reputation systems in online auction markets, by collecting auction information pertaining to specific categories of items.

On the other hand, this paper is closer is spirit to the study of illicit online markets. In particular, it builds on [Leontiadis et al. 2011], which describes search-redirection attacks and identifies concentration effects among participants. Similarly, [McCoy et al. 2012b] provide ground-truth data on the transactions information of three major pharmaceutical affiliate programs, totaling about $170 million/year. Our work also builds on the study of the Silk Road marketplace [Christin 2013], which shows overall revenue to be in the range of $15 million/year.

[Lee 2012] first employed the case-control method to cybercrime, identifying which academic departments were targeted most by spear-phishing emails laced with malware. We employed the case-control method in Section 4.3 to identify which medical conditions are targeted most by unlicensed pharmacies.

Compared to prior work, our study uniquely combines a variety of data sources to provide a comparative analysis of drug pricing and availability information across different shades of legitimacy.

7. CONCLUSIONS

Unlicensed pharmacies circumvent legal requirements put in place to protect consumers from physical harm. They operate, however, in the context of a broader ecosystem where consumers can choose among licensed pharmacies, unlicensed pharmacies and anonymous contraband marketplaces. Consequently, unauthorized pharmacies must offer a compelling reason for consumers to do business with them instead of more legitimate alternatives. One approach is for unlicensed operators to fake legitimacy through clever website design and deception. The web suffers from asymmetric information – it can be very hard for the average consumer to distinguish good websites from bad. Licensed pharmacies combat this with certification schemes such as VIPPS and LegitScript. Unfortunately, our findings suggest that seals cannot do the job on their own.

Unauthorized pharmacies are already competing hard by offering deep inventories and discount prices. Inventories at unlicensed pharmacies can rival those at licensed pharmacies, and can be more extensive for certain classes of drugs (e.g., scheduled drugs). We have shown evidence of sophistication in how prices are set by unlicensed pharmacies. While cheaper across the board than at a reference licensed pharmacy, unlicensed pharmacies also employ deceptive tactics such as fake generics to attract customers, in addition to more straightforward volume discounts.

Our study also identified possible intervention policies against unlicensed pharmacies. A somewhat typical approach is to devote resources to blacklisting unlicensed pharmacies. Unfortunately, blacklists only offer a partial solution, since switching domain names is easy, quick, and cheap. On the other hand, by clustering pharmacies by their inventories, we could identify a much smaller number of suppliers. Hence, tracking down the laboratories that provide unlicensed pharmacies with their goods appears a much more viable and cost-effective intervention strategy.

Finally, all of our results are essentially independent of the data source we used to identify unlicensed online pharmacies. This further suggests that, even though combating illicit advertising (spam, search-engine poisoning) has been the core focus of the computer security community, dealing with unlicensed online pharmacies in particular, and online crime in general, requires a more comprehensive economic treatment. We hope this paper will stimulate further work in that area.

REFERENCES

AIZCORBE, A. AND NESTORIAK, N. 2010. Price indexes for drugs: A review of the issues. Bur. of Econ. Analysis working papers.

BROWN, S., ELKIN, P., ROSENBLOOM, S., HUSSER, C., BAUER, B., LINCOLN, M., CARTER, J., ERLBAUM, M., AND TUTTLE, M. 2004. VA national drug file reference terminology: a cross-institutional content coverage study. *Medinfo 11,* Pt 1, 477–81.

CHRISTIN, N. 2013. Traveling the Silk Road: A measurement analysis of a large anonymous online marketplace. In *Proc. WWW'13*. Rio de Janeiro, Brazil. To appear. Preprint available as arXiv:1207.7139.

DINGLEDINE, R., MATHEWSON, N., AND SYVERSON, P. 2004. Tor: The second-generation onion router. In *Proc. USENIX Sec.'04*. San Diego, CA.

FRANKLIN, J., PAXSON, V., PERRIG, A., AND SAVAGE, S. 2007. An inquiry into the nature and causes of the wealth of internet miscreants. In *Proc. ACM CCS'07*. Alexandria, VA, 375–388.

FUNG, K., MCDONALD, C., AND BRAY, B. 2008. RxTerms–a drug interface terminology derived from RxNorm. In *AMIA Ann. Symp. Proc.*. Vol. 2008. American Medical Informatics Association, 227.

GELATTI, U., PEDRAZZANI, R., MARCANTONI, C., MASCARETTI, S., REPICE, C., FILIPPUCCI, L., ZERBINI, I., DAL GRANDE, M., ORIZIO, G., AND FERETTI, D. 2013. 'You've got m@il: Fluoxetine coming soon!': Accessibility and quality of a prescription drug sold on the web. *Int. J.of Drug Policy*.

JANG, J., BRUMLEY, D., AND VENKATARAMAN, S. 2011. Bitshred: Feature hashing malware for scalable triage and semantic analysis. In *Proc. ACM CCS'11*. Chicago, IL.

KANICH, C., KREIBICH, C., LEVCHENKO, K., ENRIGHT, B., VOELKER, G., PAXSON, V., AND SAVAGE, S. 2008. Spamalytics: An empirical analysis of spam marketing conversion. In *Proc. ACM CCS'08*. Alexandria, VA.

LEE, M. 2012. Who's next? identifying risks factors for subjects of targeted attacks. In *Proc. Virus Bull. Conf.*. 301–306.

LEONTIADIS, N., MOORE, T., AND CHRISTIN, N. 2011. Measuring and analyzing search-redirection attacks in the illicit online prescription drug trade. In *Proc. USENIX Sec.'11*. San Francisco, CA.

LEVCHENKO, K., CHACHRA, N., ENRIGHT, B., FELEGYHAZI, M., GRIER, C., HALVORSON, T., KANICH, C., KREIBICH, C., LIU, H., MCCOY, D., PITSILLIDIS, A., WEAVER, N., PAXSON, V., VOELKER, G., AND SAVAGE, S. 2011. Click trajectories: End-to-end analysis of the spam value chain. In *Proc. IEEE Sec. & Privacy'11*. Oakland, CA.

LIN, Z., LI, D., JANAMANCHI, B., AND HUANG, W. 2006. Reputation distribution and consumer-to-consumer online auction market structure: an exploratory study. *Decision Support Systems 41,* 2, 435–448.

LINCOLN, M., BROWN, S., NGUYENA, V., CROMWELLA, T., CARTER, J., ERLBAUM, M., AND TUTTLE, M. 2004. US department of veterans affairs enterprise reference terminology strategic overview. In *Proc. Medinfo 2004*. 391–395.

LIU, S., MA, W., MOORE, R., GANESAN, V., AND NELSON, S. 2005. RxNorm: prescription for electronic drug information exchange. *IT Professional 7,* 5, 17–23.

LU, L., PERDISCI, R., AND LEE, W. 2011. SURF: Detecting and measuring search poisoning. In *Proc. ACM CCS'11*. Chicago, IL.

MCCOY, D., DHARMDASANI, H., KREIBICH, C., VOELKER, G., AND SAVAGE, S. 2012a. Priceless: The role of payments in abuse-advertised goods. In *Proc. ACM CCS'12*. Raleigh, NC.

MCCOY, D., PITSILLIDIS, A., JORDAN, G., WEAVER, N., KREIBICH, C., KREBS, B., VOELKER, G., SAVAGE, S., AND LEVCHENKO, K. 2012b. Pharmaleaks: Understanding the business of online pharmaceutical affiliate programs. In *Proc. USENIX Sec.'12*. Bellevue, WA.

MOORE, T., CLAYTON, R., AND ANDERSON, R. 2009. The economics of online crime. *J. Econ. Persp. 23,* 3, 3–20.

MOORE, T., LEONTIADIS, N., AND CHRISTIN, N. 2011. Fashion crimes: Trending-term exploitation on the web. In *Proc. ACM CCS'11*. Chicago, IL.

NAKAMOTO, S. 2008. Bitcoin: a peer-to-peer electronic cash system. `http://bitcoin.org/bitcoin.pdf`.

SALTON, G. AND MCGILL, M. J. 1986. *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., New York, NY.

SCHLESSELMAN, J. 1982. *Case-control studies: design, conduct, analysis*. Oxford University Press, USA.

SCOTT, F. AND YELOWITZ, A. 2009. Pricing anomalies in the market for diamonds: evidence of conformist behavior. *Economic Inquiry 48,* 2, 353–368.

SOKAL, R. 1958. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull. 38*, 1409–1438.

SORENSEN, A. 2000. Equilibrium price dispersion in retail markets for prescription drugs. *J. Pol. Econ. 108,* 4, 833–850.

U.S. FOOD AND DRUG ADMINISTRATION. 2010. National drug code directory. `http://www.fda.gov/Drugs/InformationOnDrugs/ucm142438.htm`.