

Adam J. Martin. An Analysis of Query Types for Prescription Drug Information. A Master's Paper for the M.S. in I.S. degree. November, 2010. 41 pages. Advisor: Robert Losee

Much is known about the extent to which people use the Internet to search for prescription drug information as well as the usefulness of printed drug information. Little is known about the types of drug information that are queried. This study uses query data from the most relied on search engine to determine the types of drug information that the U.S. population seeks online. It finds that searches for prescription drug information most often fall under the types *General or Synonym*, *Drug Substitute or Comparison*, and *Adverse Effects*, and that few queries fall under the types *Mechanism of Action*, *Administration*, and *Fertility, Pregnancy, or Lactation*. This study also provides evidence to support the conclusion that the U.S. population searches for prescription drug information in the context of information about other drugs and that usefulness criteria for online drug information should take this into account.

Headings:

Prescription Drugs

Information Seeking Behavior

Usefulness of Consumer Medication Information

Infodemiology

Taxonomy

Insights for Search

AN ANALYSIS OF QUERY TYPES FOR PRESCRIPTION DRUG INFORMATION

by
Adam J. Martin

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Information Science.

Chapel Hill, North Carolina

November 2010

Approved by

Robert Losee

TABLE OF CONTENTS

Introduction.....3

Literature Review.....5

 Quality of Online Health Information.....5

 Usefulness of Prescription Drug Information.....6

 Prescription Drug Information Seeking Behavior.....8

 Infodemiology.....11

Methodology.....14

 Google Insights for Search and Operational Definitions.....14

 Given Query Selection.....18

 Prescription Drug Query Taxonomy and Coding Top Search Queries.....19

Results.....25

Discussion.....27

 Findings Concerning Prescription Drug Information Seeking Behavior.....27

 Findings Concerning the Usefulness of Prescription Drug Information.....29

 Findings Concerning Infodemiology.....31

 Limitations and Suggestions for Future Work.....31

Conclusions.....33

References.....35

Appendix.....40

INTRODUCTION

As access to the Internet grows, more people are seeking information about medications online. Research conducted for the Pew Internet & American Life Project reveals that a full third of Americans now go online for information about drugs – a climb from 19% in 2002 (Fox, Jones 2009). This trend bears the potential for health improvements as well as dangerous outcomes and has fueled a substantial body of research. To date, much work has been conducted on the quality of drug information supplied via the Internet, but surprisingly little is known on the nature of the information demanded by the public.

Eysenbach (2002) has examined Internet-related phenomena from an epidemiological perspective. In one study (Eysenbach, 2006), he introduced a technique for measuring information demand that proved capable of predicting important health outcomes, e.g., influenza epidemics. His ingenious technique involved the purchase of flu-related keywords from Google AdSense (Mountain View, CA), which granted him aggregated information on the prevalence and geographic location of queries involving these terms. In 2008, Google introduced Insights for Search, an analytics tool that makes keyword data publically available.

This study will rely on Insights for Search data to gain an understanding of drug information demand. Due to the dearth of information on this subject, this study will be exploratory in nature and will seek to provide descriptive statistics on the type of

information that people in the United States search for with respect to the most highly prescribed drugs.

LITERATURE REVIEW

Quality of Online Health Information

Initially, much of the attention given to online health information was focused on the quality of the information available on health and drug Web sites. Health care professionals were concerned by the fact that it had become easier than ever to disseminate erroneous and dangerous information. Many sought to equip the public with criteria for evaluating Web sites.

The title of an early article by Silberg, Lundberg, and Musacchio (1997) warns, “Let the Reader and Viewer Beware” and introduces metadata-based criteria for the evaluation of medical information on the Internet. A study published in the *Journal of the American Medical Association* motivates a meta-analysis of quality evaluation tools as follows:

The rapid growth of the Internet has triggered an information revolution of unprecedented magnitude. Despite its obvious benefits, the increase in the availability of information could also result in many potentially harmful effects on both consumers and health professionals who do not use it appropriately. (Jadad & Gagliardi, 1998, p. 611)

Kim, Eng, Deering, and Maxfield (1999) published a similar review of quality evaluation criteria. Shon and Musen (1999) also relied on metadata as criteria for evaluating the quality of information of medical information in general. Concerning prescription drug information specifically, Hatfield, May, and Markoff (1999) used metadata and other criteria to evaluate the information on four Web sites.

Quality evaluation research quickly proliferated, and a systematic review of the literature was soon conducted. Eysenbach, Powell, Kuss, and Sa (2002) reviewed 79 empirical studies. Quality evaluation research was used to make trustworthy Web sites more easily identifiable. The HONcode (Health on the Net Foundation, Geneva, Switzerland) certification system is a prominent example of efforts to guide consumer health information seeking behavior towards high quality information Web sites.

Usefulness of Prescription Drug Information

The quality of prescription drug information can be considered from two perspectives – that of the experts evaluating the supply of information, and that of the consumers who are demanding it. Eysenbach points out the importance of identifying health information areas “where there is a knowledge translation gap between best evidence (what some experts know) and practice (what most people do or believe), as well as markers for ‘high-quality’ information” (Eysenbach, 2002, p. 763). Indeed, many quality evaluation criteria comprehended the need for considering the consumer’s perspective. For instance, criteria often took into account the usability of a Web site by measuring things like average loading time. Some quality evaluation research incorporated criteria relating to the usefulness of health information.

Criteria concerning the usefulness of drug information sometimes drew on federal guidelines concerning the written information provided to consumers at pharmacies (Thompson & Graydon, 2009). Typically, Consumer Medication Information (CMI) in the format of leaflets is provided to consumers upon receipt of their prescriptions. Guidelines for the usefulness of CMI were made available in an Action

Plan which identified “steps for assessing, evaluating, and revising these criteria, component, and format suggestions as additional information is gathered through consumer testing and other appropriate means” (Department of Health and Human Services, December 1996, p. 20). Wolf, Davis, Shrank, Neuberger, and Parker (2006) examined 40 FDA approved medication guides in order to assess their readability. Kim (2005) measured Perceived Usefulness of CMI in patients at three rheumatology/pain clinics in order to develop models capable of predicting CMI use. Winterstein, Linden, Lee, Fernandez, and Kimberlin (2010) evaluated CMI leaflets distributed by pharmacies based on federal criteria of usefulness.

Similar studies have been conducted outside of the United States. Newby, Drew, and Henry (2001) used a telephone survey and follow-up interviews of a random sample of Australians from the Hunter region to investigate drug information seeking as well as satisfaction and understanding of received information. In another study, a tool “for measuring consumers’ perceptions of the comprehensibility, utility, and design quality of written medicine information, was tested in Australian consumers” (Koo, Krass, and Aslani, 2007, p. 951).

Astrom et al. (2000) used both quantitative and qualitative methods to develop a tool for measuring patients’ drug information preferences. They relied on data on demographic characteristics, an intrinsic desire for information scores, and open-ended questions. Open-ended questions included: “What kind of information about your medicines do you want?” (p. 161). However, only a few example responses are provided in the Results section. The authors’ introductory remarks provide motivation for the present paper:

To be satisfied one should have been given the right amount of information and also find the information useful. Satisfaction with information is a subjective measurement and does not say anything about quality of information. The key questions explored in this study are ‘what?’ drug information the patient desires, ‘how?’ it should be presented and ‘by whom?’ it should be given. (p. 159)

These studies all testify to the importance of the consumer’s circumstances with respect to prescription drug information. Many also highlight the usefulness of the information as an important criterion. However, none of these studies provide usefulness criteria that are based on the types of prescription drug information that consumers actually seek. The fact that the term “information prescription” (Leisey & Shipman, 2007) has currency in this literature is a reflection of a general tendency to impose information needs on consumers based on what experts believe they should be exposed to. This is certainly an important approach – perhaps even the *more important* approach – yet it is not so important that the alternative approach taken in this paper should not be adopted. Indeed, these approaches can be complementary.

Prescription Drug Information Seeking Behavior

Many studies investigate prescription drug information seeking behavior. In addition to the PEW survey mentioned in the Introduction section, there are two large-sample, quantitative studies. Baker, Wagner, Singer, and Bundorf (2003) sought “to measure the extent of Internet use for health care among a representative sample of the US population” (p. 2400). This study found that a weighted 33% of respondents (unweighted n=3,670) had used the Internet or e-mail to learn more about a drug. Another large-sample (n=1,084) study found that “the Internet tops doctors as the go-to resource for information about health- and wellness-related topics. Whereas 55 percent of online

adults ask their physician for health- and wellness-related information, 59 percent resort to Internet-based resources” (iCrossing, 2008, p. 4). This study posed the question: “If you had to determine what sources contribute to your decision whether or not to take a prescription medication, how important would a suggestion from each of the following sources be to you?”(iCrossing, 2008, p. 7). 11% of respondents identified Internet resources as “Extremely important” and 39% chose “Very important.”

A survey using a quota sample drawn from an outpatient pharmacy of the National University Hospital in Singapore was used to evaluate patient needs and sources of drug information (Ho, Ko, and Tan, 2009). Among participants who had ever used the Internet (n=201), 53% reported having used the Internet as a drug information source. This study also found that “the type of DI that respondents usually wanted involved adverse effects (72.6%), dosing (54.7%), indication (54.2%), herb– drug or drug– drug interactions (38.8%), and mechanism of action (25.9%). Relatively few respondents were interested in DI about the use of devices such as inhalers (13.4%) and whether the drug could be used during pregnancy or breastfeeding (10.9%)” (p. 734). The current paper will also investigate the type of drug information sought by patients, but will not use a survey method and will focus on the U.S. population, rather than a population of mostly Chinese nationals in Singapore (p. 733).

Some research has focused on developing models that predict health information seeking behavior. Weaver et al. (2009) developed logistic regression models that show “home computer ownership, online time per week, and health care system use are all positively linked with [internet medical information]-seeking behavior” (p. 714). Koo, Krass, and Aslani (2006) used a cross-sectional questionnaire study to investigate the

influence of patient factors, e.g., health literacy and coping strategy, on seeking and reading written medicine information.

Qualitative designs have also been used to investigate drug information seeking behavior. Peterson, Aslani, Stud, and Williams (2003) used focus groups to assess how consumers searched for and appraised medicines on the Internet. They found that participants “searched for information on medicines using a range of search techniques from simple 1-word searches and advanced techniques to suboptimal techniques (Discussion section, para. 4). They also found that “many participants searched for information on a medicine by typing the brand name into a search engine” (Discussion section, para. 6). This finding informs the present paper, which will examine query data related to brand names rather than active ingredients, i.e., generic names.

Another qualitative study used human-computer interactions approaches – namely, naturalistic observation in a usability laboratory and interviews – to assess the health information retrieval and appraisal techniques of consumers (Eysenbach & Kohler, 2002). This work relied on a database of anonymized queries to an “ask doctor” Web site. Zeng et al. (2004) also conducted an interview and observation study and “found that many consumers were unable to find satisfactory information when performing a specific query” (p. 45). For this study, the consumers were asked to state a health information need as a goal prior to searching.

An article by Williams, Dennis, and Nicholas (2005) investigated what users were looking for on Drugscope via an open question regarding users’ last site visit. This work also examined information needs with page visit frequency data and found information

needs that related to euphoria-seeking behavior and its dangers as well as legislation and policy questions.

The literature on prescription drug information seeking behavior would benefit from a study that addresses the types of drug information that are searched for in the United States.

Infodemiology

In a seminal article, Eysenbach (2002) introduces the term “infodemiology” as follows:

A new research discipline and methodology has emerged—the study of the determinants and distribution of health information and misinformation—which may be useful in guiding health professionals and patients to quality health information on the Internet. Information epidemiology, or infodemiology, identifies areas where there is a knowledge translation gap between best evidence (what some experts know) and practice (what most people do or believe), as well as markers for “high-quality” information. (p. 763).

One application of infodemiology is to complement surveys. Eysenbach and Kohler (2003) “think that direct analysis of searches elicit a much more accurate picture of what people are doing and looking for on the web than for example survey data such as the Pew Internet Survey, which currently dominate the literature” (p. 229). In order to address the “surprising dearth of evidence on what consumers are searching for on the web and how consumers do it” (p. 225) in regard to health information, Eysenbach and Kohler conduct a study “aimed to determine the actual prevalence of health-related searches on the web by analyzing search terms entered by people into popular search engines and to make some preliminary attempts in qualitatively describing and classifying these searches” (p. 225). Likewise, the present paper will attempt to address

the “surprising dearth of evidence on what consumers are searching for on the web,” but in the health information sub-domain of prescription drug information.

There is evidence that the concept of infodemiology has taken hold. In the *Journal Epidemiology*, Lee relates that “the basic premise of infodemiology is that certain information patterns on the Internet may be caused by, or may cause, population-health patterns” (Lee, 2010, p. 761). Studies that rely on this premise are beginning to proliferate – especially since the launches of Google Trends and Google Insights for Search.

Work (Ginsberg et al., 2009) published by researchers affiliated with Google and the Centers for Disease Control and Prevention (CDC) built on the Influenza surveillance study mentioned in the Introduction section. This culminated in a means “to estimate consistently the current [Influenza-Like Illness] percentage 1-2 weeks ahead of the publication of reports by the CDC’s US Influenza Sentinel Provider Surveillance Network” (p. 1013). Other researchers “found Google Trends to approximate certain trends previously identified in the epidemiology of Lyme disease” (Seifter, Schwarzwald, Geis, and Aucott, 2010, p. 135).

Breyer and Eisenberg (2010) assert that “online search volume can provide useful information for epidemiologic study and medical research” (p. 585). This study used Insights for Search and tested the hypothesis “that chronic noninfectious diseases with known variations in seasonal incidence (such as diabetes mellitus, blood pressure, myocardial infarction, and nephrolithiasis) would show seasonal variations in number of searches” (p. 584).

To date, only one study has used Google query data to investigate prescription

drug questions. Schuster, Rogers, and McMahon (2010) examined certain Google search queries related to statins using Trends and Insights for Search. They conclude that “Internet search engine queries for drug information exhibit temporal and geographic patterns of healthcare utilization...search engine query data may prove helpful in providing payers and policy makers with a new window into healthcare utilization in our communities” (p. 218).

METHODOLOGY

This section will describe the methodology that this paper uses to address the question of the type of information that people in the United States search for with respect to the most highly prescribed drugs. It will begin by laying out the Insights for Search data and operational definitions, will proceed to explain choices made in selecting queries corresponding to the most highly prescribed drugs, and will end by describing the drug information taxonomy and coder arrangement.

Google Insights for Search and Operational Definitions

Alexa (Alexa Internet, Inc., San Francisco, CA) ranks Google as the number one Web site according to traffic volume both overall and in the United States. The Online Health Search 2006 survey conducted for the PEW Internet and American Life Project reports that “66% of health seekers say their last query began at a general search engine like Google or Yahoo” (Fox, 2006, p. 5). This makes Google’s Insights for Search analytic tool a relevant source of data for consumer information seeking behavior related to health and drug information.

For a given query, Insights for Search offers data sets corresponding to four main categories: interest over time, regional interest, top searches, and rising searches. In this paper, “given query” is defined as the query entered into – given to – Insights for Search and for which Insights for Search provides data. Consider the below screenshots

corresponding to the given query “vicodin,” the most dispensed drug of 2009.

The screenshot shows the 'Insights for Search' interface. It is divided into three main sections: 'Compare by', 'Search terms', and 'Filter'.
 - 'Compare by': Contains three radio buttons: 'Search terms' (selected), 'Locations', and 'Time Ranges'.
 - 'Search terms': Contains a text input field with 'vicodin' entered, a tip 'Tip: Use quotation marks to match an exact phrase. ("table tennis")', and a '+ Add search term' link.
 - 'Filter': Contains several dropdown menus: 'Web Search' (set to 'Web Search'), 'United State' (set to 'United State'), 'All subregion' (set to 'All subregion'), 'All metros' (set to 'All metros'), '2004 - present' (set to '2004 - present'), and 'Health' (set to 'Health'). A 'Search' button is located at the bottom right of the filter section.

Image 1: Entering the given query “vicodin” into Insights for Search. Regional, time period, and category filters are located on the right side of the interface.

<http://www.google.com/insights/search/#cat=45&q=vicodin&geo=US&cmpt=q>

Insights for Search results are specific to time ranges and geographic regions that researchers can adjust. This paper will restrict the geographic region to the United States. Data is available from January 2004 to the present. In most cases the entire 2004-present time period will be selected. However, in some cases the time period will be restricted in an attempt to control for a confounding factor – the original FDA approval of a drug. Pilot analysis suggested that including query data from time periods between 2004 and one year following the FDA approval date for a drug affected results. Since not all drugs to be evaluated had an approval date within the 2004-present time period, date ranges for drugs that did have an approval date between 2004 and the present were restricted to one year following the drug’s FDA approval date to the present. Researchers can also limit query data by category. This paper will restrict all query data by the Health category in order to remove data not associated with health information seeking behavior. Pilot analysis suggested that this will remove some results that are associated with other categories, especially the Industries category.

Top searches

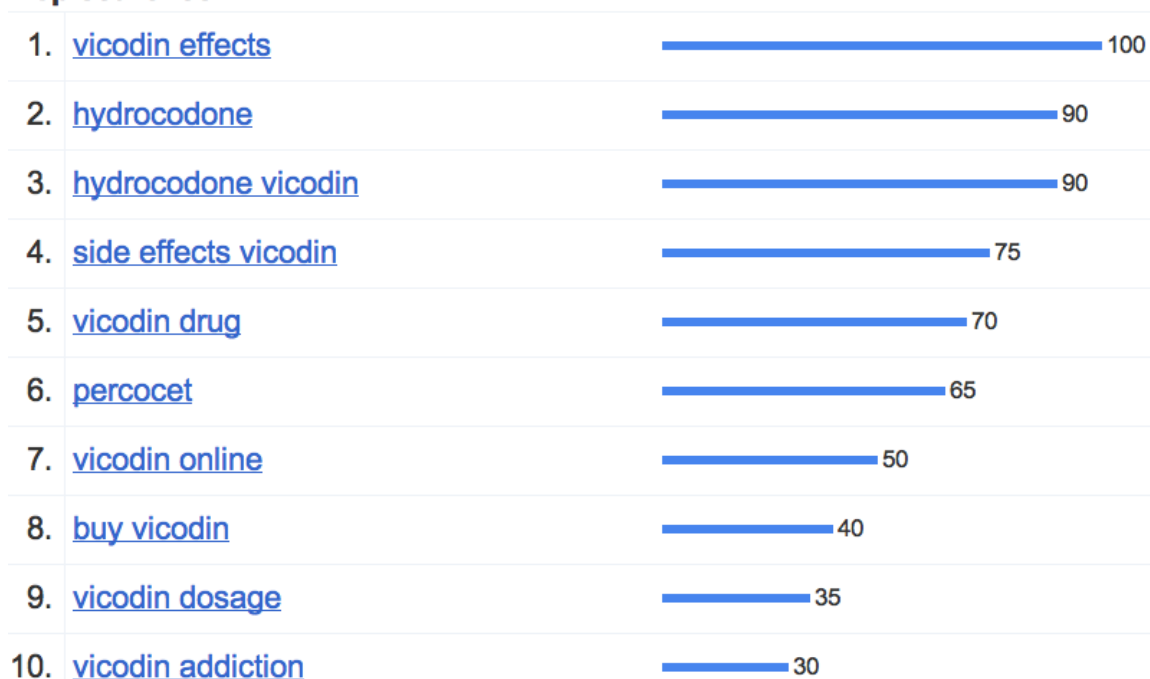


Image 2: Insights for Search top searches results for the given query “vicodin.” “Side effects vicodin” is the fourth top search query; it has a normalized, scaled relative frequency of 75.

<http://www.google.com/insights/search/#cat=45&q=vicodin&geo=US&cmpt=q>

This paper will examine “top search queries.” In this paper, top search queries refer to the top searches query results generated by Insights for Search. This paper will measure the “normalized, scaled relative frequency” of top search queries. These normalized, scaled relative frequencies are also provided by Insights for Search as in Image 2 above. It is important to emphasize that top search queries always correspond to a given query. For instance, “percocet” could not be considered as a top search query independently of the given query “vicodin,” since it may be a top search query for given queries other than “vicodin.” In these cases it may have a normalized, scaled relative frequency value other than 65.

All Insights for Search data are relative rather than absolute. They are derived by computing “how many searches have been done for the terms you've entered, relative to

the total number of searches done on Google over time” (Google, 2010). That is, each data point represents the relative frequency of the given query. Data are also normalized and scaled from 0 to 100. Insights for Search determines top search results “by examining searches that have been conducted by a large group of users preceding the search term you’ve entered, as well as after” (Google, 2010).

This paper will perform further computations to this data to enable certain comparisons. Consider again the given query “vicodin” which yields 49 top search queries. “Side effects vicodin” is among these top search queries; it has a normalized, scaled relative frequency of 75. “Vicodin dosage” is also among the top search queries; it has a normalized, scaled relative frequency of 35. Without further computation it is possible to compare the “importance” of “side effects vicodin” to “vicodin dosage”: “side effects vicodin” is about twice as “important” as “vicodin dosage.” But comparisons using the normalized, scaled relative frequencies are only valid within top search results for a given query. In other words, without further computation, it is not possible to compare “side effects vicodin” as a top search for the given query “vicodin” to “lipitor side effects” as a top search query for “lipitor.”

Fortunately, the calculation that allows this comparison is straightforward. First, the normalized, scaled relative frequency for each top search query for a given query is divided by the sum of the normalized, scaled relative frequencies of all top search results for a given query. This quotient is then multiplied by 100. The result of this procedure is referred to in this paper as a “query weight.” Return to the example of the top search query “side effects vicodin” for the given query “vicodin.” “Side effects vicodin” has a normalized, scaled relative frequency of 75. Dividing this by the sum of the normalized,

scaled relative frequencies of all of the top search queries for the given query “vicodin,” 1380, and multiplying the result by 100 yields a query weight of 5.00. Using query weights will enable comparisons between the top search results of different given queries. For instance, the query weight for “vicodin side effects” might be compared to the query weight for “lipitor side effects” with respect to the given query “lipitor” which is 20.20. This means that “lipitor side effects” is a “more important” top search query with respect to the given query “lipitor” than is “side effects vicodin” with respect to the given query “vicodin.”

Although this paper seeks to make comparisons across given queries, it will not do so by simply comparing the query weights for top search queries. Doing so poses two main problems. The first reason involves the large number of total top search queries. One advantage of Insights for Search over Trends is that data sets are available in comma separated value data files. These files are not limited to ten top search queries as shown in Image 2. For instance, the data file for the given query “vicodin” includes 49 top search queries. This study will use all available top search queries, which will yield more comprehensive results. This may also yield more interesting results; the 45th top search query corresponding to the given query “vicodin” is “snort vicodin.” However, without an a priori rationale for limiting possible comparisons that result from using all top search queries, there would be too many potential comparisons.

More importantly, the question arises as to how to treat top search queries that involve similar terms. For instance, the top search query “vicodin side effects” is very similar to “side effects vicodin.” This provides motivation for adding query weights. If such query weights are not added, comparison to similar query weights for different

given queries might not be valid. E.g., if the given query “lipitor” only generated “side effects lipitor” and not “lipitor side effects.” On a broader level, this points to a need for the top search queries to be categorized, a process that will be described below.

After all of the top search queries have been fit to the below-described taxonomy, the query weights for each drug information type can be computed for each given query. Return to the example of the given query “vicodin,” which had three top search queries that were coded as falling under the *Adverse Effects* drug information type. These query weights, 7.60, 6.08, and 2.28 add to 15.96. This result, 15.96, will be referred to as “the type weight” of *Adverse Effects* for the given query, “vicodin.” Since all query weights for a given query sum to 100, these type weights are equivalent to the percentage of query weight for each drug information type. These type weights can be summed over all given queries; the *Adverse Effects* type weight for the given query “vicodin” can be added to that for “lipitor” and so forth. Sum type weights for each drug information type will then be divided by the total of the sum type weights for all drug information types and then multiplied by 100. This gives the mean type weights, which are equivalent to the percentage of the sum type weights for each drug information type.

Given Query Selection

Crucial methodological decisions had to be made regarding given query selection. Since this paper aims to explore prescription drug information seeking behavior, selecting from the universe of prescription drugs was deemed an appropriate starting point. There is a vast number of FDA approved prescription drugs, however. Hatfield et al. (1999) took the following approach: “the 30 prescription drugs dispensed in the highest volume

to ambulatory care patients were selected to represent those medications for which consumers would be seeking information. The drugs to be evaluated (appendix) were selected by using the ‘Top 200 Rx Drugs’ list for 1996” (p. 2308). This paper will follow this approach, but it will evaluate the top 50 drugs and will substitute the most current version of the Top 200 Rx Drugs list (Bartholow, 2010). This list is titled, “Top 200 Products in the US Market By Dispensed Prescriptions, 2009” and the copyright is owned by IMS Health.

Certain features of this list require decisions regarding given query selection. Consider the drug metformin hcl. Top search queries for the given queries "metformin hcl" and "metformin" differ widely. Considering counts alone: only 11 top search queries are generated for the former, while 50 are generated for the later. This paper always selects given queries without qualifiers. In this case, the given query “metformin” is selected. Consider Image 3 below, which demonstrates the merit of this approach to given query selection.

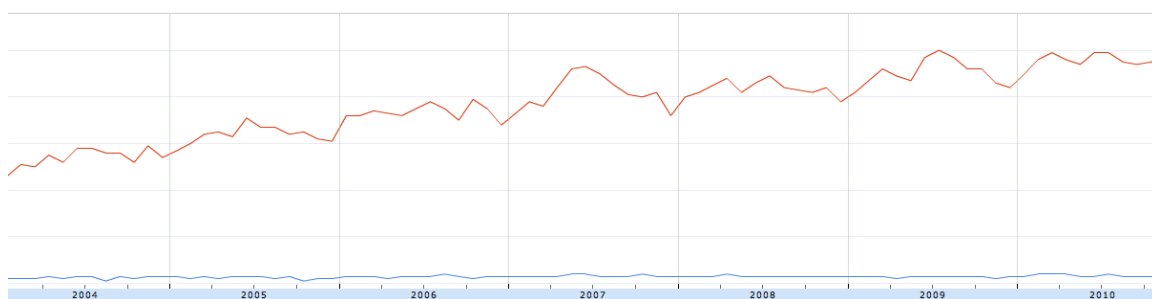


Image 3: Interest over time data for the given queries “metformin” and “metformin hcl.” “Metformin” (top series) maintains a higher normalized, scaled relative frequency than “metformin hcl” (bottom series).

<http://www.google.com/insights/search/#cat=45&q=metformin%20hcl%2Cmetformin%2C&geo=US&cmpt=q>

Another given query selection decision involves the fact that the drug list considers a generic drug made by a different manufacturer to be a distinct product. For example, simvastatin manufactured by Teva is ranked 14th and simvastatin manufactured

by Dr. Reddy's is ranked 26th. This paper does not consider a generic drug made by different manufacturers to be distinct drugs.

This raises the question of given queries for drugs for which there are no generic alternatives. These “branded” drugs might be referred to by their brand name or by the name of their active ingredient (which becomes the generic name after patent expiry). Peterson, Aslani, Stud, and Williams (2003) found that “many participants searched for information on a medicine by typing the brand name into a search engine” (Discussion section, para. 6). In accord with this, given queries for drugs for which patent expiry has not occurred will be the brand name of the drug. However, in the case of drugs for which patent expiry has occurred, this will not always be the case. For instance, for lisinopril the generic name will be selected because it had higher levels of normalized, scaled relative frequency than the brand name, Prinivil (see Image 4 below).

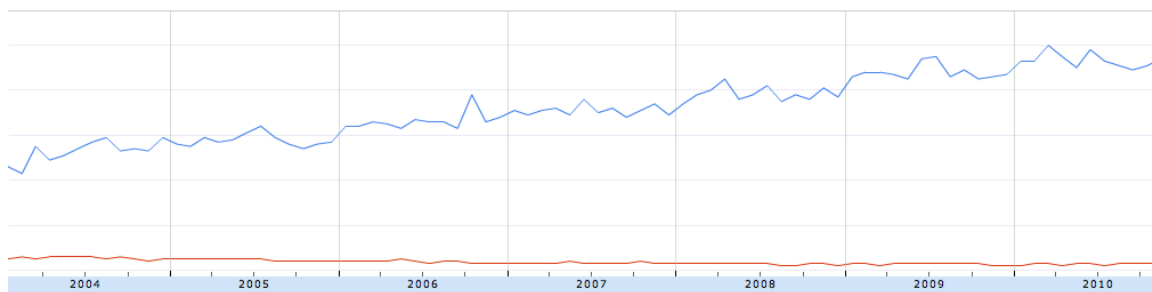


Image 4: Interest over time data for the given queries “lisinopril” and “prinivil.” The given query corresponding to the generic name, “lisinopril” (top series) maintains a higher normalized, scaled relative frequency than that corresponding to the brand name, “prinivil” (bottom series).

<http://www.google.com/insights/search/#cat=45&q=lisinopril%2Cprinivil&geo=US&cmpt=q>

In still other cases, the both the generic and the brand name were used in the given query. Consider the drug simvastatin (brand name, Zocor). Patent expiry for this drug occurred in June 2006. One year after this period, “zocor” had higher normalized, scaled relative frequencies than “simvastatin.” However, Image 5 shows that the reverse is true after

2008. Throughout the time period, both given queries had substantial search volumes. This problem was circumvented by using the Boolean OR operator, which is “+” in Insights for Search).

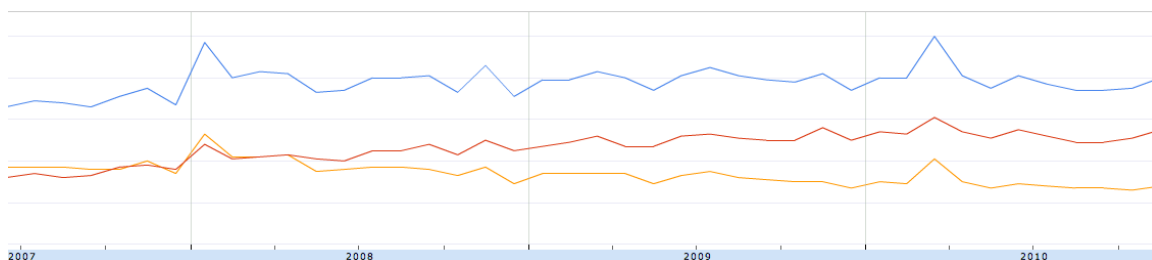


Image 5: Interest over time data for the given queries “simvastatin+zocor,” “simvastatin,” and “zocor.” “Simvastatin+zocor” (top series) captures search volume for “simvastatin” OR “zocor.” The given query corresponding to the generic name, “simvastatin” (from left, middle series) begins with a higher normalized, scaled relative frequency than that corresponding to the brand name, “zocor” (from left, bottom series), but this relationship does not hold over the entire time period.
<http://www.google.com/insights/search/#cat=45&q=simvastatin%2Bzocor%2Csimvastatin%2Czocor&geo=US&date=6%2F2007%2043m&cmpt=q>

In addition to the above reasoning, the final list of given queries was reviewed by a medical doctor for face validity. This final list is provided in the Appendix.

Prescription Drug Query Taxonomy and Coding Top Search Queries

For reasons outlined above, this paper will rely on a prescription drug query taxonomy. There is precedent for this in the health information seeking literature. Eysenbach and Kohler (2003) used coders to map health related search queries to a pre-existing taxonomy. However, “the Ely taxonomy – originally developed to classify physicians’ information needs – proved not to be very useful to code consumer questions” (p. 227). A study by Ho, Ko, and Tan (2009) described the following drug information types in relation to consumer Internet drug searching: adverse effects, dosing, indication, herb-drug or drug-drug interactions, mechanism of action, use of devices such as inhalers, and use during pregnancy. Pilot analysis shows that this

taxonomy would cover many of the top search queries, but that it needs to be modified to account for the differences between self-reported drug information needs and prescription drug search queries. The following taxonomy will be used in this study.

Query Type	Example Top Search Queries Forms
<i>Adverse Effects</i>	"[drug name] side effects," "[drug name] rash"
<i>Dosing</i>	"[drug name] 40 mg," "[drug name] daily"
<i>Indication</i>	"[drug name] [indication]"
<i>Drug-Drug or Herb-Drug Interactions</i>	"[drug name] interactions"
<i>Mechanism of Action</i>	"how [drug name] works"
<i>Fertility, Pregnancy, or Lactation</i>	"[drug name] pregnancy,"
<i>Administration</i>	"[drug name] cream," "[drug name] tabs"
<i>Drug Substitute or Comparison</i>	"[name of another drug in same class]," "[name of another drug with the indication of the given query drug]"
<i>General or Synonym</i>	"[drug name] medication," "[active ingredient name when generic not available]"
<i>Cost or Procurement</i>	"buy [drug name]," "[drug name] no prescription"
<i>Other</i>	"[drug name] recall"

Table 1: Drug query taxonomy.

Coding will be conducted by a primary coder, PC, who worked independently. PC is a licensed, practicing physician. Furthermore, as a primary care practitioner, PC is exposed to a wide variety of prescription drugs. Pilot analysis suggested that a skilled-coder arrangement might be necessary for reliability and efficiency. For instance, a coder lacking general awareness of prescription drugs would not know that the top search query "ppi" refers to the drug class "proton pump inhibitors." On the other hand, a medical doctor can be expected to recognize a wide variety of generic names, drugs within the same drug class as the given query drug, and drugs indicated for the same condition as the given query drug. PC was asked to consult the AHFS Drug Information database (American Society of Health-System Pharmacists, 2010) to resolve any questions. PC

did not require the use of this reference for coding.

RESULTS

Below, Table 2 shows the results for each given query type. The second column shows the sum of the weights for each given query type across all given queries. The third column shows the mean type weight across all given queries. Note that the mean type weight for each given query type is equivalent to the percentage of the sum type weight.

Given Query Type	Sum Type Weight	Mean Type Weight
<i>Administration</i>	27.21	0.54
<i>Adverse Effects</i>	834.45	16.69
<i>Cost or Procurement</i>	122.48	2.45
<i>Dosing</i>	438.56	8.77
<i>Drug Substitute or Comparison</i>	1156.50	23.13
<i>Drug-Drug or Drug-Herb Interaction</i>	132.59	2.65
<i>Fertility, Pregnancy, or Lactation</i>	28.89	0.58
<i>General or Synonym</i>	1599.64	31.99
<i>Indication</i>	336.18	6.72
<i>Mechanism of Action</i>	9.43	0.19
<i>Other</i>	314.07	6.28
Totals	5000	100

Table 2: Total Weight and Percent Total Weight by Given Query Type.

General or Synonym was the given query type with the highest mean type weight, followed by *Drug Substitute or Comparison*. These two given query types alone represent over half of all mean type weight (55.12). The only other given query type with a mean type weight greater than 10 was *Adverse Effects* (16.69).

Mechanism of Action was the given query type with the lowest mean type weight. This given query type had a mean type weight of less than 1 (.19). *Administration* and *Fertility, Pregnancy, or Lactation* also had mean type weights of less than 1 (.54 and .58, respectively).

By way of exploratory investigation, this study will briefly discuss the differences between drugs in the opioid pharmacological category and drugs that are not. In particular, the difference between opioids and non-opioids in terms of mean type weight for the *Cost or Procurement* drug information type will be investigated. Below, Table 3 shows results according to these categories.

	Cost or Procurement (sum type weight)	Not Cost or Procurement (sum type weight)
Opioids	36.68	4963.32
Non-opioids	85.79	4914.21
Total	122.48	9877.52

Table 3: Sum type weights for the Cost or Procurement type and Not Cost or Procurement types (i.e., all other drug information types) according to drugs in the opioid category and those in other categories.

This data lends itself to a Pearson's chi-squared test of the null hypothesis that drugs in the opioid class should be considered to belong to the same population as those that do not with respect to *Cost or Procurement*. The resulting test statistic is 20.1 and the probability that this null hypothesis is true ($p < .0001$) supports rejecting the null hypothesis.

DISCUSSION

The above results contribute to three bodies of literature, namely, those concerning prescription drug information seeking behavior, the usefulness of prescription drug information, and infodemiology. Results will be discussed in light of these bodies of literature. This section will conclude with a discussion of limitations and by highlighting some of the work that could be done to build on the present study.

Findings Concerning Prescription Drug Information Seeking Behavior

Hitherto, no drug information seeking behavior studies have described the types of drug information that people in the United States (US) search for. The PEW study (Fox, 2009) and the Baker, Wagner, Singer, and Bundorf study (2003) both restricted their surveys to the extent to which people in the US use the Internet for drug-related information needs. Neither asks the further question of what types of information are of interest. The Ho, Ko, and Tan study (2009) does pose this further question, but in a population of mostly Chinese nationals visiting the outpatient pharmacy of the National University Hospital in Singapore.

In addition, this paper has an important advantage over these studies, which all relied on surveys. As Eysenbach and Kohler (2003) point out: “direct analysis of searches elicits a much more accurate picture of what people are doing and looking for on the web than for example survey data such as the Pew Internet Survey, which currently

dominate the literature” (p. 229). They reason:

Not only is it difficult for people to recall in a survey which kind of information they retrieve on the web most frequently, the accuracy of survey data also suffers from a social desirability bias – rarely people will for example admit to be seeking pornographic material, although these kind of searches are apparently the most prevalent. (229)

These findings support this reasoning. For example, the 45th top search for the given query “vicodin” was “snort vicodin.” This may be less likely to have been self-reported than information needs that do not involve drug abuse.

On the other hand, surveys have merits that the present paper’s method lacks. For many top search queries, it is possible to make a valid inference from the query to a specific information need. For instance, the top search query “vicodin overdose” represents an information need having to do with vicodin overdoses. In other cases, such inferences are not possible. It is impossible to determine a specific information need from the top search query “hydrocodone,” which is the main active ingredient of vicodin. For this reason, this author suggests a weakening of Eysenbach and Kohler’s remarks regarding surveys. In assessing prescription drug information seeking behavior, surveys are complementary to methods that directly measure online search behavior.

The results show that a high percentage of top search queries fell under the *Drug Comparison* (mean type weight = 23.13) and *Drug-Drug or Drug-Herb Interaction* (mean type weight = 2.65) query types. These categories involving other drugs together account for over 25% of all mean type weight. This suggests that information seeking behavior is not focused on single prescription drugs in isolation. Rather, people in the US search for information about multiple drugs in succession. Consumers may find information that places prescription drugs in the context of other prescription drugs to be

helpful. In particular, prescription drug information Web sites which allow simultaneous searches of multiple drugs and which return juxtaposed results along with any drug-drug interactions may be desirable to consumers.

Findings Concerning the Usefulness of Prescription Drug Information

This paper also has findings that bear on the criteria used to assess the usefulness of online prescription drug information. The drug information types that people seek – and therefore to some extent find useful – may be different from those corresponding to Consumer Medication Information. Prima facie, it makes sense for consumers to have different information needs when turning to the Internet for prescription drug information vis-à-vis when reviewing the printed leaflets distributed by pharmacies with prescriptions. Consumers searching the Web for prescription drug information do not necessarily have a prescription; they may be querying in order to determine whether they would receive therapeutic benefit from prescription drugs. People in this position could be expected to have little interest in information having to do with administration and more interest in general questions, drug substitutes or comparisons, and adverse events. The mean type weights of the top search queries bear this out to some degree. The mean type weight for *Administration* was 0.54, while those of *General or Synonym, Drug Substitute or Comparison*, and *Adverse Events* were 31.99, 23.13, and 16.69, respectively. Of course, many people conducting Google searches for top 50 drugs will have been prescribed them, so it is not surprising that *Dosing* queries had a mean type weight of 8.77.

This author recommends that future studies not exclusively rely on FDA

guidelines concerning CMI for criteria to evaluate the usefulness of prescription drug information online. Thompson and Graydon (2009) took this approach and used the following categories shown in Table 4 below.

Criterion	Description
1	Drug names, indications for use, and how to monitor for improvement
2	Contraindications and what to do if they apply
3	Specific directions on how to use and store the medication and overdose information
4	Specific precautions and warnings about the medication
5	Symptoms of serious or frequent possible adverse reactions and what to do
6	Certain general information including encouraging patients to communicate with health care professionals and disclaimer statements
7	Information that is scientifically accurate, unbiased in tone and content, and up to date
8	Information in an understandable and legible format that is readily comprehensible to consumers

Table 4: Thompson and Graydon's usefulness criteria for evaluating prescription drug information Web sites (2009, p. 42).

These usefulness criteria fail to include a criterion that corresponds to the *Drug Substitute or Comparison* type, which had the second-highest mean type weight at 23.13.

Given that the *General or Synonym* type could have captured some information needs related to drugs' mechanism of action, it is impossible to infer from the fact that *Mechanism of Action* had the lowest mean type weight (0.19) to the conclusion that consumers do not have information needs along these lines. It may be that consumers have difficulty formulating queries for this information type. Or, it may be that consumers are more interested in what medications do than how they work.

Qualitatively, this is supported by the present study, which uncovered many top search queries – some with high query weights – that asked about the *effects* of a particular drug.

Findings Concerning Infodemiology

Inasmuch as the literature of infodemiology overlaps with those of prescription drug information seeking behavior and drug information usefulness, the above findings are applicable to this literature. By giving an overview of the types of prescription drug information searched for in the US, this study also provides groundwork for later infodemiology studies. The following exploratory result is suggestive of the kind of studies that may follow this one. The top 50 drugs can be organized according to pharmacologic class. Opioids are one class that seemed to exhibit different mean type weights. In particular, it seemed that more queries fell under the *Cost or Procurement* category for drugs in the opioid category. Indeed, Pearson's chi-squared test supports rejecting the null hypothesis that drugs in the opioid category should be considered to belong to the same population as those that do not with respect to *Cost or Procurement* ($p < .0001$). This finding could be useful as a marker of prescription drug abuse. Leveraging Insights for Search's geographic analysis feature, this information could be used as a means of assessing the extent of prescription drug abuse at the state or local level.

Limitations and Suggestions for Future Work

The coder, PC, described occasional uncertainty in deciding between the *Drug-Drug or Drug-Herb* given query type and the *Drug Substitute or Comparison* type. Future studies should strive for two coders, which would allow inter-coder reliability assessment. Primary care physicians or pharmacists represent ideal candidates for skilled coders.

Also, given queries comprised the top 50 prescribed drugs, and these are not likely to be representative of all drugs. For instance, drugs indicated for rare conditions, “orphan drugs,” may be associated with different information needs and this may result in different mean type weights. Future studies might explore orphan drugs. Or, studies might rely on random sampling from a sampling frame including all FDA approved drugs.

This paper provides evidence in support of distinguishing between usefulness criteria of CMI (i.e., the printed leaflets distributed at pharmacies) and those pertaining to online prescription drug information. Future studies should not simply adopt FDA guidelines for the usefulness of CMI in order to evaluate the usefulness of prescription drug information on the Internet. This is because guidelines for CMI do not emphasize the importance of placing information for a particular drug in the context of other drug information.

By way of exploratory investigation, this study has also shown that drugs in the opioid category are characterized by different *Cost or Procurement* search behavior, which could be a useful marker of drug abuse. Future infodemiology studies can use the drug information taxonomy as well as the mean type weights presented in this study to investigate other relationships that have the potential to improve public health outcomes.

CONCLUSIONS

This study has provided descriptive statistics on the type of information that people in the United States search for with respect to the most highly prescribed drugs. Using data from Google's Insights for Search, it has shown that searches for prescription drug information most often fall under the types *General or Synonym*, *Drug Substitute or Comparison*, and *Adverse Effects*. Very few search queries fall under the types *Mechanism of Action*, *Administration*, and *Fertility, Pregnancy, or Lactation*.

This study has also shown that methodologies that exploit actual search behavior of large populations are important complements to large-scale surveys. In particular, this study was able to move beyond the large-sample surveys conducted for the PEW Internet & American Life Project (Fox, 2009) and Baker, Wagner, Singer, and Bundorf (2003). Unlike these studies, the findings of this paper are not attended by the bias that may result from self-reported survey responses.

This study also provides evidence for the fact that the US population searches for prescription drug information, not in isolation, but in the context of other drugs. This population may find useful Web sites that juxtapose information about prescription drugs alongside that of other drugs. Usefulness criteria specific to printed CMI should not be applied to online drug information Web sites without the addition of criteria that address the presence of contextual information about other drugs.

Finally, chi-squared tests showed that opioids differ from drugs not in the opioid

category in terms of *Cost or Procurement* search behavior. Future studies may establish this as a marker of drug abuse or may explore other categorical differences.

References

- Astrom, K., Carlsson, J., Bates, I., Webb, D. G., Duggan, C., Shanghani, P. & McRobbie, D. (2000). Desire for information about drugs: A multi-method study in general medical inpatients. *Pharmacy World & Science*, 22(4), 159-164.
- Baker, L., Wagner, T. H., Singer, S. & Bundorf, M. K. (2003). Use of the Internet and e-mail for health care information. *Journal of the American Medical Association*, 289(18), 240-246. doi: 10.1001/jama.289.18.2400
- Bartholow, M. (2010). Top 200 Prescription Drugs of 2009. *Pharmacy Times*. Retrieved from:
<http://www.pharmacytimes.com/issue/pharmacy/2010/May2010/RxFocusTopDrugs-0510>
- Breyer, B. N. & Eisenberg, M. L. (2010) Use of Google in study of noninfectious medical conditions. *Epidemiology*, 21(4), 584-585.
- Eysenbach, G. (2002). Infodemiology: The epidemiology of (mis)information. *The American Journal of Medicine*, 113, 763-765.
- Eysenbach, G. (2006). Infodemiology: Tracking flu-related searches on the Web for syndromic surveillance. *Proceedings of the American Medical Informatics Association Symposium*, 244-247.
- Eysenbach, G. & Kohler, C. (2002). How do consumers search for and appraise health information on the World Wide Web? Qualitative study using focus groups,

- usability tests, and in-depth interviews. *British Medical Journal*, 324, 573-577.
- Eysenbach, G. & Kohler, C. (2003). What is the prevalence of health-related searches on the World Wide Web? Qualitative and quantitative analysis of search engine queries on the Internet. *Proceedings of the American Medical Informatics Association Symposium*, 225-229.
- Eysenbach, G., Powell, J., Kuss, O. & Sa, E. (2002). Empirical studies assessing the quality of health information for consumers on the World Wide Web: A systematic review. *Journal of the American Medical Association*, 287(20), 2691-2700.
- Fox, S. (2006). *Online Health Search 2006*. Retrieved from PEW Internet and American Life Project website:
http://www.pewinternet.org/~media/Files/Reports/2006/PIP_Online_Health_2006.pdf
- Fox, S. & Jones, S. (2006). *The social life of health information*. Retrieved from PEW Internet and American Life Project website:
http://www.pewinternet.org/~media/Files/Reports/2009/PIP_Health_2009.pdf
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S. & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012-1014. doi: 10.1038/nature07634
- Google, Inc. (2010). *Insights for Search* [Online database]. Retrieved from:
<http://www.google.com/insights/search/#>
- Hatfield, C. L., May, S. K. & Markoff, J. S. (1999). Quality of consumer drug information provided by four Web sites. *American Journal of Health-System*

Pharmacists, 56, 2308-2311.

Ho, C., Ko, Y. & Tan, M. (2009). Patient needs and sources of drug information in Singapore: Is the Internet replacing former sources? *The Annals of*

Pharmacotherapy, 43, 732-739. doi: 10.1345/aph.1L580

iCrossing, a digital marketing company. (2008). How America searches: Health and wellness. Retrieved from:

<http://www.icrossing.com/articles/How%20America%20Searches%20-%20Health%20and%20Wellness.pdf>

Jadad, A. R. & Gagliardi, A. (1998). Rating health information on the Internet:

Navigating to knowledge or to Babel? *Journal of the American Medical Association*, 279(8), 611-614. doi: 10.1001/jama.279.8.611

Kim, P., Eng, T. R., Deering, M. J. & Maxfield, A. (1999). Published criteria for evaluating health related Web sites: Review. *British Medical Journal*, 318, 647-649.

Koo, M. M., Krass, I. & Aslani, P. (2006). Enhancing patient education about medicines: Factors influencing reading and seeking of written medicine information. *Health Expectations*, 9, 174-187.

Koo, M. M., Krass, I. & Aslani, P. (2007). Evaluation of written medicine information:

Validation of the consumer information rating form. *The Annals of Pharmacotherapy*, 41, 951-956. doi: 10.1345/aph.1K083

Lee, B. (2010). Epidemiologic research and Web 2.0 – the user-driven Web.

Epidemiology, 21(6), 760-763.

Leisey, M. R. & Shipman, J. P. (2007). Information prescriptions: A barrier to

fulfillment. *Journal of the Medical Library Association*, 95(4), 435-438. doi: 10.3163/1536-5050.95.4.435

Newby, D. A., Hill, S. R., Barker, B. J., Drew, A. K. & Henry, D. A. (2001). Drug information for consumers: Should it be disease or medication specific? Results of a community survey. *Australian and New Zealand Journal of Public Health*, 25(6), 564-570.

Peterson, G., Aslani, P., Cert, G., Stud, E. & Williams, K. A. (2003). How do consumers search for and appraise information on medicines on the Internet? A qualitative study using focus groups. *Journal of Medical Internet Research*, 5(4). doi: 10.2196/jmir.5.4.e33

Schuster, N. M., Rogers, M. A. M. & McMahon, L. F. (2010). Using search engine query data to track pharmaceutical utilization: A study of statins. *The American Journal of Managed Care*, 16(8), 215-219.

Seifter, A., Schwarzwalder, A., Geis, K. & Aucott, J. (2010). The utility of “Google Trends” for epidemiological research: Lyme disease as an example. *Geospatial Health*, 4(2), 135-137.

Shon, J. & Musen, M. A. (1999). The low availability of metadata elements for evaluating the quality of medical information on the World Wide Web. *Proceedings of the American Medical Informatics Association Symposium*, 945-949.

Silberg, W. M., Lundberg G. D. & Musacchio, R. A. (1997). Assessing, controlling, and assuring the quality of medical information on the Internet: Caveant lector et viewer – let the reader and viewer beware. *Journal of the American Medical*

Association, 277(15), 1244-1245.

Thompson, A. E. & Graydon, S. L. (2009). Patient-oriented Methotrexate information sites on the Internet: A review of completeness, accuracy, format, reliability, credibility, and readability. *The Journal of Rheumatology, 36(1), 41-49.* doi: 10.3899/jrheum.080430

U.S. Department of Health and Human Services. (1996, December). *Action plan for the provision of useful prescription medicine information.* Retrieved from: <http://www.fda.gov/downloads/AboutFDA/CentersOffices/CDER/ReportsBudgets/UCM163793.pdf>

Williams, P., Dennis, K. & Nicholas, D. (2005). An evaluation of the websites of charities and voluntary organisations providing support for young people: Case study: Drugscope. *Webology, 2(3).* Retrieved from: <http://www.webology.ir/2005/v2n3/a16.html>

Winterstein, A. G., Linden, S., Lee, A. E., Fernandez, E. M. & Kimberlin, C. L. (2010). Evaluation of consumer medication information dispensed in retail pharmacies. *Archives of Internal Medicine, 170(15), 1317-1324.*

APPENDIX

Rank (Number of Dispensations, 2009)	Given Query	Rank	Given Query
		25	ambien
		26	percocet
1	vicodin	27	cymbalta
2	lipitor	28	seroquel
3	synthroid	29	effexor
4	lisinopril	30	flomax
5	amoxicillin	31	xanax
6	nexium	32	trazodone
7	plavix	33	actos
8	toprol	34	fosamax
9	singulair	35	bactrim
10	lexapro	36	prevacid
11	proair	37	klonopin
12	simvastatin+zocor	38	tramadol
13	amlodipine+norvasc	39	levaquin
14	azithromycin	40	prozac
15	metformin	41	prednisone
16	metoprolol	42	prilosec
17	hydrochlorothiazide	43	atenolol
18	crestor	44	lantus
19	furosemide	45	augmentin
20	warfarin	46	trikor
21	advair	47	celebrex
22	ibuprofen	48	aricept
23	zoloft	49	vytorin
24	diovan	50	keflex

Appendix: Given queries for the top 50 dispensed prescription drugs of 2009.