MINING USER-GENERATED DATA TO EXAMINE ONLINE INFORMATION SEEKING FOR REPRODUCTIVE AND SEXUAL HEALTH

A Dissertation
Presented to
The Academic Faculty

By

Alicia L. Nobles

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
Department of Systems and Information Engineering

University of Virginia

September 2018

MINING USER-GENERATED DATA TO EXAMINE ONLINE INFORMATION SEEKING FOR REPRODUCTIVE AND SEXUAL HEALTH

Approved by:

Dr. Laura E. Barnes, Advisor
Department of Systems and
Information Engineering
University of Virginia

Dr. Matthew S. Gerber
Department of Systems and
Information Engineering
University of Virginia

Dr. William T. Scherer
Department of Systems and
Information Engineering
University of Virginia

Dr. Jessica Keim-Malpass
School of Nursing
University of Virginia

Dr. Michael Smith
Department of Systems and
Information Engineering
University of Virginia

Date Approved: September 7, 2018

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

With the ubiquity of the Internet, health information that previously could only be obtained from a healthcare provider is now easily accessible to consumers. Indeed, many people turn to the Internet for health information despite their preference to see a healthcare provider. Infodemiology leverages the digital footprints left behind from online health information seeking to derive insights into human behavior. By applying computational approaches to passively-collected, user-generated data, a number of useful insights have been generated for public health issues spanning including chronic diseases, infectious diseases, behavioral health, and mental health.

However, few infodemiology studies have examined reproductive and sexual health. This area is particularly ripe for exploration as the sensitivity of these health concerns often drive people to trade accuracy (by obtaining information from a trusted provider) in lieu of the anonymity of the Internet. The main contribution of this dissertation is to fill this existing gap in infodemiology literature by contributing two case studies of online information seeking for reproductive and sexual health.

Prior to delving into the case studies, I provide a general backdrop for online information seeking and existing models of health information seeking that conceptualize health information seeking behaviors and the search process. Following this background, I present the first case studies examining online search engine queries for contraceptive methods during a period of heightened uncertainty for U.S. healthcare reform. The second case study examines online information seeking on the subreddit r/STD for recent information seeking

related to sexually transmitted infections (STIs). Both of these case studies are extremely timely given the focus on the repeal of the Affordable Care Act and record high rates of STIs in the U.S., respectively.

Specifically, I point to some of the most interesting and important results of these case studies. First, I identified unprecedented information seeking for intrauterine devices (IUDs) in response to uncertainty regarding the contraceptive mandate. This is critical given many public health leaders have speculated about the potential implications of the recent election and ours is the first study to link the change in administration to potential health outcomes using data. Second, I identified salient information needs of a previously unstudied platform, the subreddit r/STD. This community primarily seeks information about the human papillovirus (HPV), herpes simplex virus (HSV), transmission and risk of STIs during oral sex, and medical testing window periods. In the process, I identified and examined sharing of online resources, which is understudied in public health. Finally, I find evidence supporting experience and salience (external and internal uncertainty) as motivators linking these health information searches.

Both of these studies have implications for medical decision making and can support recent calls in the public health community for data-driven approaches to determine and tailor health communication.

CHAPTER 1

INTRODUCTION

1.1 Motivation

Infodemiology, an emerging field, utilizes digital data trails left behind from when people use the Internet to search and communicate health-related information to generate useful insights for and inform public health and policy [1]. Computational approaches applied to these passively-collected, user-generated data have provided insights spanning infectious and chronic diseases, behavioral health, and mental health. To name a few examples, search engine queries have been used to examine pro- and anti-vaccination information seeking [2], Reddit has been used to examine self-disclosure of mental illness [3], and symptom-related search engine queries have been used to predict pancreatic cancer [4].

Why would an individual turn to the Internet for reproductive or sexual health information? As Friedman et al. (2016) stated "Despite the ubiquity of sex in the media, a culture of silence surrounds sexual health in the United States . . . [5]." The sensitivity of reproductive and sexual health drives people to often trade accuracy in lieu of the anonymity of the Internet. Indeed, many people first turn to the Internet for information about reproductive and sexual health despite their preference to first speak with a healthcare provider because of this exact reason. This makes reproductive and sexual health information seeking par-

ticularly ripe for infodemiology.

Few infodemiology studies have examined reproductive and sexual health with the majority focusing on human immunodeficiency virus (HIV) and the human papillomavirus (HPV) vaccination using Twitter. Both of these are serious public health issues; however, research has overlooked more general public health issues such as the ones examined in this dissertation. The aim of this dissertation is to demonstrate the value of automated mining passively-collected, user-generated online data for insights into recent reproductive and sexual health information seeking. More specifically, I present two case studies where I apply existing strategies to mine insights into information seeking behavior for contraceptives and sexually transmitted infections (STIs).

The first case study in this dissertation examines online search engine queries for contraceptive methods during periods where there is heightened uncertainty of a repeal of the Affordable Care Act (ACA). The second case study examines online information seeking on the subreddit r/STD to examine recent information needs related to STIs. These case studies are timely given the increased emphasis on a full or partial repeal of the ACA under a Republican-controlled presidency and Congress and the record-high incidence rate of STIs in U.S., respectively. I take an atheoretical approach to both case studies, however, the results are compared to current models of health information seeking for a qualitative understanding of the motivators contributing to the information search. In doing so, the main contribution of this dissertation is extracting recent information needs for reproductive and sexual health (two understudied applications) using novel online data sources (Google search queries and Reddit) thereby filling a gap in existing literature.

By leveraging user-generated online data to listen to the public, we discover the public's, an often not involved stakeholder in health communication, potential information needs. Our current methods to examine health information seeking behavior and identify health information needs are costly, resource-intensive, and lag in time relying on surveys, focus groups, or in-depth interviews. The approaches demonstrated in this dissertation examine health information seeking behavior and salient information needs using publicly available data, are minimally intrusive, and can be done in near real-time complementing more traditional approaches. In the future these approaches could be used to inform health communicators (healthcare providers, public health experts, and medical librarians), another stakeholder and the decision makers, of platform-specific community health information needs and trends. Understanding the needs of online communities is critical for effective tailoring of health communication. As an HIV digital strategist recently stated "As communicators, we need to do more than understand which channels we can use to reach our audiences. We need to understand why individuals visit different social media sites and the unique characteristics of each platform, including: what content performs bests ... and how we can best contribute to the conversation. To understand these channels, we need to listen - listen to our audiences, and listen to the data [6]."

To summarize, this dissertation has the following contributions. First, I identified and addressed a literature gap related to reproductive and sexual health in the field of infodemiology. Second, I identified online information seeking behavior in response to changing healthcare policy. More specifically, I identified unprecedented information seeking for intrauterine devices (IUDs) in response to increased uncertainty in U.S. healthcare policy

reform. Third, I identified salient information needs related of a previously unstudied online community. More specifically, I identified salient information needs related to sexual health and STIs on the subreddit r/STD. In doing so, I also identified online resources that are being shared by consumers, which has been understudied in public health. Finally, both case studies provide supporting evidence that uncertainty is a motivator of health information searches linking the studies.

1.2 Dissertation Overview

This dissertation is organized into the following five chapters.

Chapter 1: Introduction provides a high-level motivation behind my dissertation research.

Chapter 2: Background and Literature Review offers an overview of online health information seeking, existing models of health information seeking, and concludes with a discussion of the existing infodemiology literature. In this chapter, I discuss a common vocabulary for the field that I will adhere to for the remainder of this dissertation.

Chapter 3: Online Information Seeking for Reproductive Health contributes the findings from the first of the two case studies demonstrating the application of computational approaches to passively-collected, user-generated online data to identify trending information seeking for contraceptives. Specifically, this chapter explores increases in information seeking for IUDs in response to heightened uncertainty surrounding a repeal of the ACA.

I find evidence of a positive relationship between increased information seeking for IUDs and events of heightened uncertainty. Additionally, this increased information seeking is elevated across all states irrelevant of the state's electoral vote in the 2016 U.S. presidential election.

Chapter 4: Online Information Seeking for Sexual Health contributes the findings from the second of the two case studies demonstrating insights from data-driven computational approaches to passively-collected, user-generated online data to identify the salient information needs of r/STD. Specifically, I find evidence that this community's most salient information needs are related to HPV, herpes simplex virus (HSV), transmission and risk of STIs during oral sex, and medical testing window periods. To ensure validity of the computational approaches, I compare the results of topic modeling to a qualitative analysis of a subsample of the data. Finally, I include an analysis of online resource sharing, which is rare in public health.

Chapter 5: Conclusions and Future Work concludes this dissertation by summarizing takeaways as well as pointing out implications and future directions. In particular, I discuss implications to tailor health communications tying the results of the dissertation to the original motivation of the case studies. In closing, I describe opportunities for future work including moving beyond characterization of information seeking to automated content recognition and identification of misinformation to further inform future health communication.

CHAPTER 2

BACKGROUND AND LITERATURE REVIEW

This chapter reviews the academic literature in the field of infodemiology (information epidemiology; also referred to as digital epidemiology) that uses passively-collected, user-generated data (like the data used in this dissertation). In doing so, I motivate the main contribution of this dissertation: that monitoring of reproductive and sexual health has been understudied in the larger infodemiology literature. To support this literature review, I first provide a brief introduction to online information and existing models of health information seeking.

The intention of this chapter is to provide a brief overview of online health information seeking, models of health information seeking, and applications of infodemiology rather than a comprehensive overview. When available, I point to more comprehensive reviews of the literature. Brief overviews of the literature specifically related to reproductive and sexual health are provided in the motivation sections of Chapters 3 and 4, respectively.

2.1 Online Health Information Seeking

The Internet has "created an avalanche of easily accessible information" [7]. Seventy-four percent of U.S. adults use the Internet and up to 80% of users use the it as a source for

health information [8]. Furthermore, with the ubiquity of smartphones, consumers can now access health information anytime, anywhere with limited effort that previously could only be obtained by consulting a healthcare professional [9, 10]. In 2012 alone, approximately 52% of smartphone owners report using their phone to search for health information [11].

There are a number of reasons why one may turn to the Internet for health information. First, information is available immediately or near real-time. Users may gather information or receive support allowing them to evaluate their need to see a healthcare provider, provide reassurance during the intermediate time while waiting on an interaction with a provider, or simply allowing users to use the information in lieu of seeing a provider. Second, with the advent and rise of Web 2.0 in the past decade, the Internet is no longer limited to passive browsing. There is a plethora of new sources on the Internet (e.g., online communities, social networking sites, and social Q&A) allowing people to engage with each other sharing personal experiences, sharing knowledge, increasing awareness, expressing emotions, and exchanging advice [10, 12, 13, 14, 15]. Third, individuals may find comfort in the anonymity of the Internet and online communities making them ideal sources of information for stigmatized or sensitive topics such as sexual and reproductive health [16, 17, 13]. Finally, the Internet may be an attractive source of information for those facing barriers to access (such as individuals lacking health insurance in the U.S.). The decision to use an online resource for health information is often not rational, but instead guided the cognitive effort to understand the resource and perception of social risk [13]. As such, individuals have formed online communities and often rely on the health information presented there to guide decision-making for seeking additional health care [18].

Online health information has far-reaching effects on public health by influencing health-care decisions and outcomes [15, 19]. There is increasing evidence of people using online resources for health information and social support. The effects of information presented in public online communities likely extends much further than just its active participants because a significant portion of searches for health information consist of passive monitoring behaviors (casually browsing websites or social media) [13]. It is imperative that we (researchers and health communication experts) understand how people use the Internet to seek health information and, subsequently, how accessed information impacts decision making.

2.2 Models of Health Information Seeking

Although information seeking spans many fields, a coherent approach to information seeking in general has yet to be developed [20]. Most models of health behavior include information seeking as a static variable and do not explicitly consider the dynamic information seeking process in-depth. Similarly, most models of communication do not capture the power of health-related concerns to motivate information seeking, hence the need for models that explicitly attempt to capture the health information seeking process [20]. More comprehensive overviews of human information interaction [21], health behavior [22, 23], and health information seeking [20] are provided in the aforementioned citations. This section reviews several commonly used, existing models for health information seeking including the Health Information Acquisition Model (HIAM), Expanded Conceptual Model of Information Seeking (ECMISB), Model of Risk Information Seeking and Processing

(RISP), and Comprehensive Model of Information Seeking (CMIS). For each of these models, I provide an overview of its origination, application, usage, description, and limitations. Prior to discussion of these models, below I discuss health information seeking in general.

This dissertation focuses on active information seeking (often shortened to just health information seeking) or those with an expressed intent to seek health information. That said, information seeking behavior also includes passive receipt of information and avoidance (although not explored in this dissertation).

Health information seeking behavior has been commonly studied under three contexts: (1) coping with a health-related problem, (2) its relationship to medical decision making, and (3) behavior change and preventative behavior [24]. Health information seeking is often perceived as a problem-focused coping strategy with some arguing that it may also serve as an emotion-focused coping strategy. Those that argue in favor of it being viewed as an emotion-focused coping strategy frequently cite that participating in health information seeking changes the relationship between an individual and a stressor, potentially reducing negative emotions associated with uncertainty [24]. According to Kuhlthau's "uncertainty principle", uncertainty is a cognitive state that increases anxiety and reduces confidence [25]. In light of Web 2.0 that allows users to interact with each other online, I would argue that health information seeking may serve as both a problem-focused coping strategy and emotion-focused coping strategy (or either). Ultimately, acquiring new health information may result in changes to health behavior by modifying an individual's beliefs and attitudes, informing individuals about risk, and providing resources to support behavioral changes [24]. Health information seeking has also been conceptualized as a process to

obtain adequate information needed to make a medical decision.

Most studies focused on health information seeking behavior, an ill-defined concept itself [24], have not incorporated existing models. Additionally, the basis of existing health information seeking models is on individual research participants; however, infodemiology studies are observational by nature not involving individual research participants [26] making operationalization of some components of the existing models difficult.

Health Information Acquisition Model

The HIAM was developed to model cancer information seeking based on descriptive analysis of responses from callers who sought information from the Cancer Information Service hotline [27]. The HIAM has been used for question development and qualitative analysis of focus group interviews that explored how new Hispanic mothers use health information sources to inform decision-making during the first 1000 days of their child's life [28] and a survey that explored how public library users in Botswana seek health information [29].

As shown in Figure 2.1, the HIAM is a sequential flow model which outlines the actions in the information seeking process. Feedback loops allow for cycling through sequential steps during the information seeking process based on the information found thus far in the search process.

The model is initialized by an internal or external stimulus, which prompts the user to evaluate if their current knowledge is sufficient. If the user decides that their current knowledge is sufficient, a search process is not initiated. Otherwise, a search process is

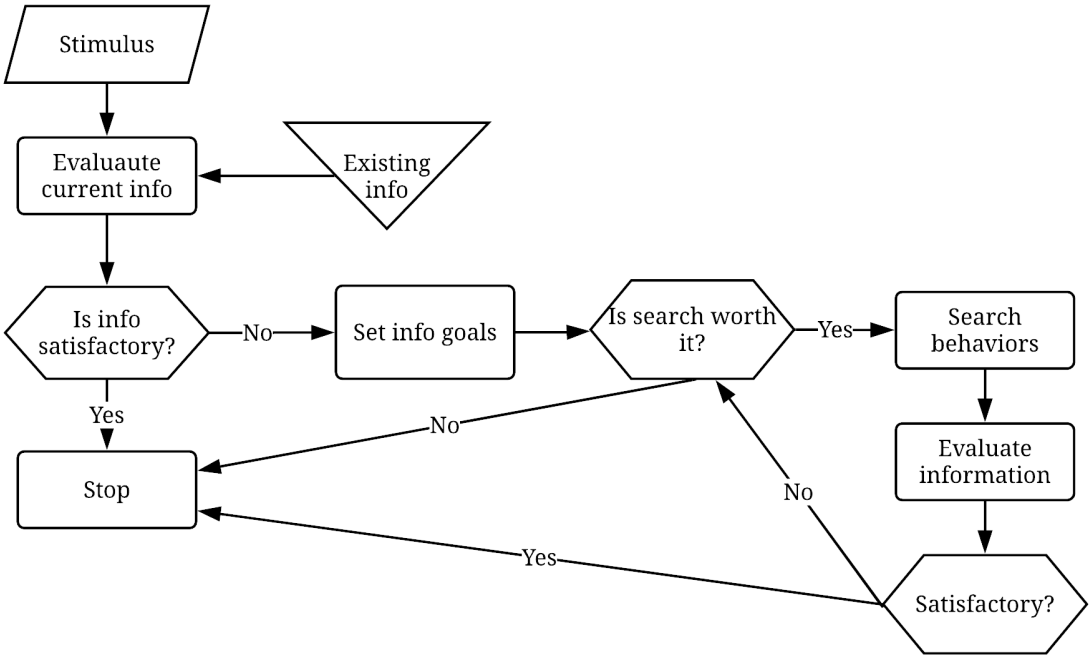


Figure 2.1: Health Information Acquisition Model (HIAM)

initiated by the user. The first step in the process is to evaluate the objective of the search and if the search effort is worth the benefit of retrieving additional information. If the user decides that the cost of the search outweighs the potential benefits, the search process does not continue. However, if the user decides that the benefits of a search outweigh the cost, then the user engages in a search behavior and evaluates the information obtained. If the information gained is sufficient, the search ends; otherwise the search cycles through the steps again until either additional searches are not worth it or sufficient information is obtained.

Although the HIAM was explicitly developed to provide guidance to planners interested in disseminating information to the public there are a number of limitations. First, it assumes that only professional resources are consulted during the information search and does not capture the complexity of the information field or varying attributes of different

carriers. Second, it assumes that all searches are active searches and does not account for the passive receipt of information. Third, information seeking is complex with a number of feedback loops that do not always follow a sequential pattern. As Johnson (2012) [20] states, "these models offer rational approaches to irrational phenomena." Finally, the model does not account for contextual or personal factors that may underlay the search behavior.

Expanded Conceptual Model of Information Seeking Behaviors

The ECMISB was developed to model information seeking for breast cancer, derived from patients' experiences, and empirically tested with survey data related to breast cancer and diabetes [30, 31, 32]. As shown in Figure 2.2, the ECMISB outlines that major components, inspired by the Health Belief Model (HBM), that are considered during an individual's search process without specifying the relationships between the components. The major components include contextual (e.g., health status) and personal (e.g., demographics) attributes that influence information seeking behavior and outcomes.

The ECMISB acknowledges the difference between individuals who are actively seeking information versus those who passively come across information through their daily activities. Both categories of information seeking have varying levels of access to and receipt of information as well as the influence of the newly found information on decision making. Ultimately, this results in a patient outcome, which are influenced by the HBM. Finally, The ECMISB is considered one of the few viable models of health information seeking behavior by the National Cancer Institute [31].

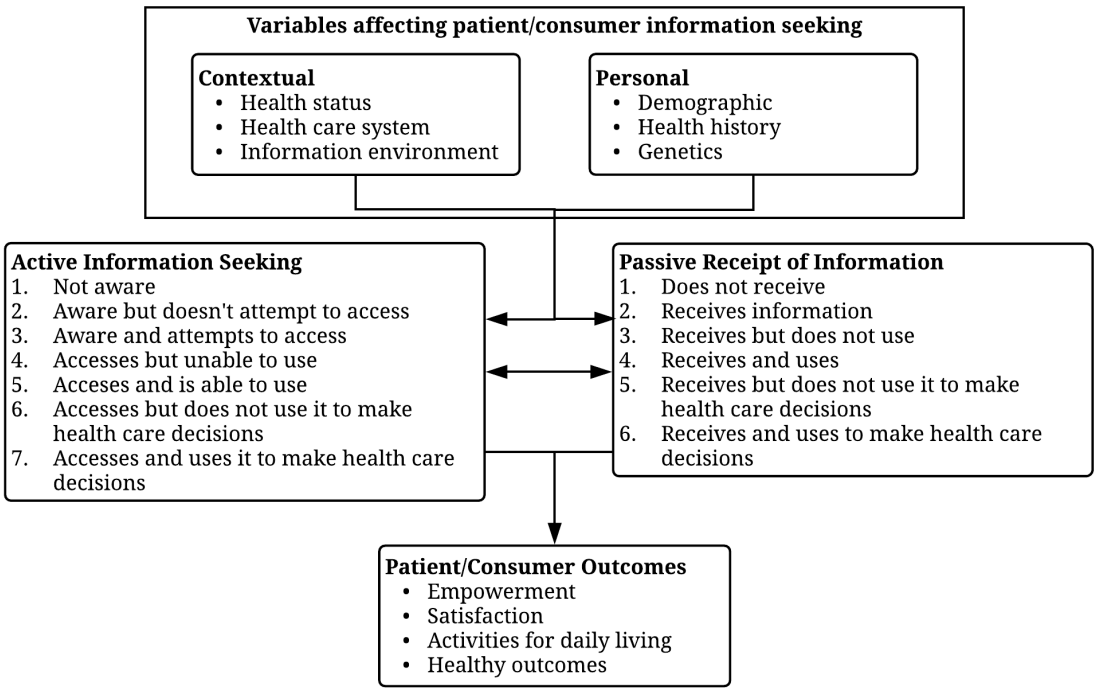


Figure 2.2: Expanded Conceptual Model of Information Seeking Behaviors (ECMISB)

Model of Risk Information Seeking and Processing

The RISP model is primarily based on the Heuristic Systematic Model of Information Processing (HSM) [33] and the Theory of Planned Behavior [34]. It has been empirically tested with survey data related to environmental and health risks [35, 36, 37, 38, 39, 40]. Several studies related to health information seeking have been guided by the RISP. For example, examining survey data for influenza vaccine information seeking of health care workers [41], survey data for H1N1 flu vaccination information seeking of college students [42], and survey data for food-related risk information seeking of consumer [43].

As shown in Figure 2.3, the RISP model specifies that information insufficiency, relevant channel beliefs, and perceived information gathering capacity influence an individuals'

motivation to seek and ability to process information related to risks [35].

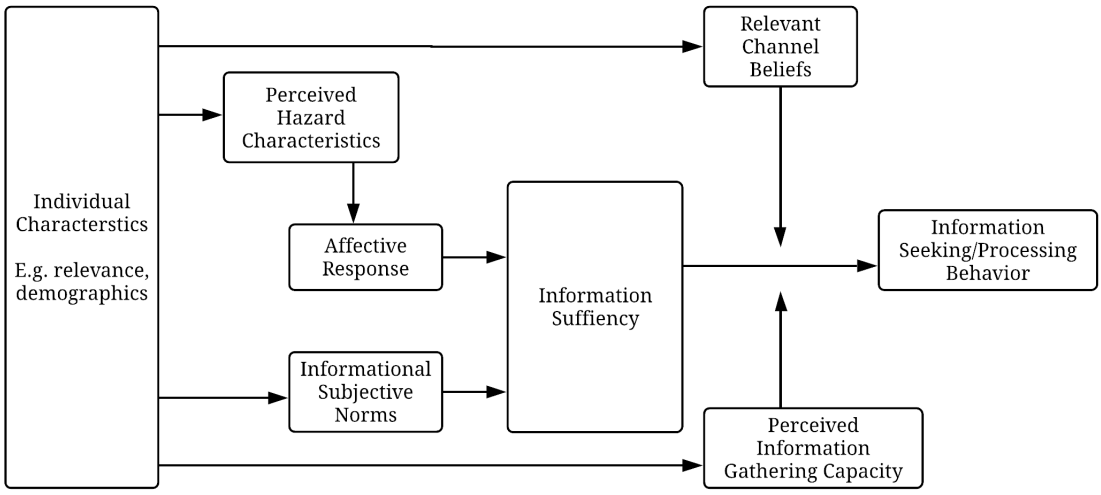


Figure 2.3: Model of Risk Information Seeking and Processing (RISP)

Information insufficiency is the difference between the amount of knowledge an individual currently has and their perception of how much knowledge is needed to handle a given risk. In other words, each individual defines how much information they believe is necessary to handle with a given risk. Individuals are more motivated to fill a perceived gap in knowledge if they sense a greater information insufficiency [44]. The concept of information insufficiency is rooted in the HSM.

The perception of an individual's information insufficiency is influenced by individual characteristics, perceived hazard characteristics, affective responses, and informational subjective norms. Individual characteristics include attributes such as demographics. Perceived hazard characteristics refer to the perceptions of a given risk and have been examined under three contexts in the literature (risk perception, institutional trust, and self-efficacy) [37]. Affective responses include emotions such as worry or anxiety that influence an individual's perception of information insufficiency. Informational subjective norms refers to

the individual's perception of knowledge held by others in their social network. In other words, an individual may perceive a societal pressure to increase their knowledge.

Relevant channel beliefs refers to the beliefs an individual has about a particular information carrier. For example, if the source of information is trustworthy or useful. Perceived information gathering capacity refers to an individual's perception of their self-efficacy to gather information.

Comprehensive Model of Information Seeking

The CMIS was developed as a model for cancer information seeking. It is based on the HBM, uses and gratification research, and Media Exposure Appraisal model [45]. As shown in Figure 2.4, the CMIS is a causal process model with three major classes of variables in the information seeking process: antecedents, information carrier factors, and information seeking actions. Similar to the ECMISB, the CMIS is also considered one of the few viable models of health information seeking behavior by the National Cancer Institute [31] and is also classified as an uncertainty management theory [46].

The antecedents (demographics, experience, salience, and beliefs) in the CMIS are based on antecedents identified by health behavior models (HBM and Transtheoretical Model) [47] and refer to background factors that predispose an individual to seek information from particular information carriers. Both demographics (age, gender, ethnicity, education, and socioeconomic status) and personal experience influence an individual's choice of their preferred information carrier. Experience refers to an individual's proximity

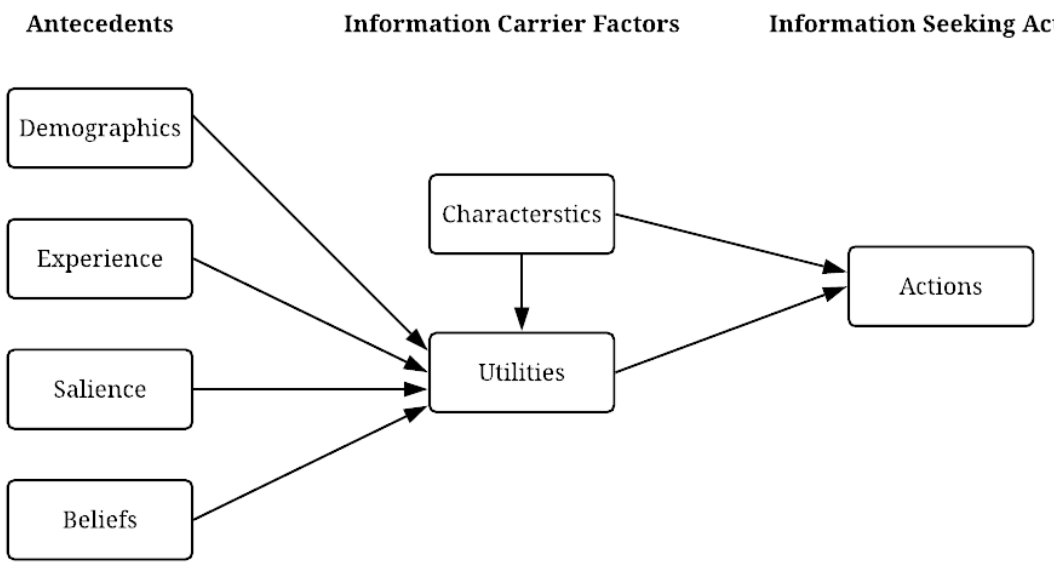


Figure 2.4: Comprehensive Model of Information Seeking (CMIS)

to disease and can often be a cue to action that triggers information seeking. There are four stages of proximity to disease along a continuum of health statuses. First, the causal stage is a general lack of concern about a disease. Any information gathered in this stage is passive (not an intentional search). Second, the purposive-placid stage is characterized by a lack of urgency, but general interest in health consciousness. This stage is focused on preventing disease or promoting health. Third, the purposive-clustered stage is characterized by confrontation with a particular disease and an individual is focused on information that will directly address a particular disease. Finally, the fourth stage is the directed stage which follows a confirmatory diagnosis and an individual focuses on finding information about coping with their diagnosis.

The two personal relevance factors, salience and beliefs, are the primary determinants influencing an individual's perceived knowledge gap (information insufficiency) that motivates information seeking [20]. Salience refers to an individual's belief about the relevance

of the health information to them [47]. An example of salience is the perception of risk to an individual's health. Individuals are motivated to seek information when the health information becomes salient (or relevant) to them. Beliefs refer to an individual's beliefs about the severity of the disease, their level of control over a situation, and their self-efficacy as well as cultural and religious beliefs that shape health behavior [48].

Information carrier factors, based on a model of Media Exposure and Appraisal [49], focus on the characteristics and usefulness of the sources selected by individuals. When selecting a source of information (information carrier), individuals often favor ease of accessibility over the authoritativeness of the source [13, 47, 50]. Sources that mimic face-to-face interaction may be preferable as individuals strongly prefer to receive information from other people [50].

Information seeking actions refer to actively, purposefully searching for information. The process an individual takes in their information search is likely to be influenced by antecedents and information carrier factors described above.

The model has been empirically tested primarily with cancer information seeking. Some examples studies guided by the CMIS include examining computer-assisted telephone interviewing data for information seeking in magazines for mammography screening [50], survey data for information seeking by cancer survivors [51], survey data for cancer information scanning [52], survey data for cancer information seeking for Korean Americans [53], and focus group and telephone interviews for physical activity information seeking for individuals diagnosed with multiple sclerosis [54].

A limitation of the CMIS is there no feedback loops, which overlooks the dynamic process of information seeking [47, 24]. Additionally, stages and actions in the process are not identified and there is no explicit treatment of outcomes [47]. Although context is important, it has not been explicitly defined in the CMIS, and there is a dearth of literature considering both context and information seeking together [47, 20].

The CMIS can relate to traditional health behavior models in four ways [20]. First, the CMIS can interact with health behavior models. That is, variables that determine health behaviors also determine information seeking. Second, the CMIS can be parallel to a health behavior model. That is, an individual may be gathering information, but not acting or may passively acquire information that will be a basis for information seeking later. Third, the CMIS can act as an input into a health behavior model. For example, finding information may be a cue to action. Fourth, the CMIS may be an outcome of a health behavior model. For example, an individual may determine no further action is required following a search concluding an end process for a health behavior model.

2.3 Monitoring User-Generated Data from Online Health Information Seeking

Public health organizations systematically collect data using environmental monitoring, surveys, notifications, and registries to monitor public health issues [55]. However, this traditional surveillance is limited and often time- and resource-intensive. Examination of online user-generated content, which is cost-effective and in near real-time, may hold promise as a supplement to traditional sources of data.

Online user-generated content can be collected both actively (individuals consent to extract their online data) and passively (usage of publicly available online data) data collection. Both types of data can be analyzed using both qualitative and automated computational techniques. This dissertation focuses on passively collected data related to reproductive and sexual health and application of automated computational methods to examine online information seeking. Similarly, the following review focuses on computational approaches applied to passively collected online user-generated content for infectious disease, chronic disease, behavioral health, and mental health. More comprehensive reviews of infodemiology are presented by Paul and Dredze (2017) [56] and Zeraatkar and Ahmadi (2018) [26]. This review draws upon the aforementioned overviews as well as recently published studies not included in those references.

The majority of studies that examine passively collected, online user-generated content are descriptive in nature followed by surveillance and prediction of human behavior. The most widely used data sources are search engine queries (Google Trends, Bing, or Yahoo) and microblogs (Twitter) [26, 57]. Beyond search engine queries and tweets, other sources of online user-generated content include blogs (Wordpress, Tumblr), other microblogs (Weibo), social network platforms (Facebook, LinkedIn), media sharing platforms (YouTube, Instagram), general purpose sharing platforms (Reddit), online review websites (Yelp), patient community websites (PatientsLikeMe), and crowd-sourcing services (Amazons Mechanical Turk). These platforms host a wide variety of content (text, images, video), metadata (time, location), and social network information (relationships between users).

Methodologies for online user-generated content range from content analysis to identify prominent themes of discussion (using supervised learning), trend inference to identify interest or activity across time and locations (using regression or autoregression methods), individual analysis to predict or learn predictive attributes about human behavior (using supervised learning), and techniques to validate quantitative findings (using comparison to external data, out-of-sample validation techniques, or qualitative analysis using smaller samples of data to provide a richer understanding). The majority of studies are not compared to a validated output; however, approximately one-third of studies using Google Trends do compare their results to a validated output from another source [26]. Finally, most of these studies are conducted in the U.S.

2.3.1 Infectious and Chronic Diseases

The most well-known and widely published studies using online user-generated content to supplement traditional surveillance focus on influenza-like illness (ILI) [56, 26]. These studies have employed a variety of data sources to monitor ILI including search engine queries [58], Twitter [59], Wikipedia page views [60], and restaurant reservation cancellations on OpenTable [61].

Beyond ILI, there have been a number of studies focused on digital surveillance of a assortment of infectious diseases including dengue fever using search engine queries and Twitter [62, 63]; Lyme disease using search engine queries [64]; norovirus using search engine queries [65]; Escherichia coli using Twitter [66]; malaria using search engine queries

[67]; bubonic plague in Madagascar using Twitter [68]; infectious intestinal disease using Twitter [69]; fear in response to the first imported case of Ebola in the U. S. using search engine queries and Twitter [70]; public awareness and response during the Zika virus using Twitter [71]; and conjunctivitis using search engine queries and Twitter [72].

There have been relatively few studies using passively collected data related to STIs. Young et al. (2014) found a significant positive relationship between geolocated HIV-related tweets and county-level HIV prevalence [73]. Using the annotated dataset, Young et al. (2017) built a classifier that distinguished HIV-related tweets (tweets containing content related to sexual risk or HIV-related substance abuse) from non-HIV-related tweets exemplifying that machine learning can identify HIV-related tweets comparable to a domain expert's ability to identify HIV-related tweets [74]. Using the same dataset, Young et al. (2018) found geolocated tweets containing sexual risk-related content was predictive of an annual increase in syphilis incidence [75]. Ireland et al. (2015) examined if words indicating future orientation and risky or safe leisure activities in geolocated tweets were correlated with county-level HIV prevalence [76]. The researchers found that risky counties (i.e., with high population density and access to risky activities) with relatively low rates of HIV were associated with tweets containing future-oriented language. The researchers hypothesize that future-oriented thinking may buffer HIV risk in vulnerable communities because people who are more future-oriented tend to engage in less risky sexual behavior. Oh and Park (2013) used a dictionary-based approach in SPSS Modeler Text Analytics to identify the prominent themes in health questions related to sexually transmitted diseases (STDs) on Yahoo! Answers finding the top five prominent themes to be STDs, herpes,

sex, help, and HIV [77]. Zheluk et al. (2013) found a strong correlation between HIV and AIDS search engine queries and HIV prevalence in Russia. Similarly, Domnich et al. (2014) found that a high positive correlation between HIV and syphilis incidence rates and search engine queries in Russia [78]. Finally, Breen et al. (2016) examined tweets to identify trending discussions about Pre-Exposure Prophylaxis (PrEP) for prevention of HIV [79].

Compared to infectious diseases, there are fewer studies examining chronic diseases. Studies of chronic diseases have focused on prediction of pancreatic cancer diagnosis using search engine queries [4], content-analysis of pain-related tweets [80], prediction of asthma prevalence using tweets [81], content analysis of diabetes-related tweets [82], content analysis of migraine-related tweets [83], information seeking for movement disorders using search engine queries [84], information seeking for epilepsy using Wikipedia visits [85], information seeking for multiple sclerosis using search engine queries [86], and forecasting incidence of dementia-related outpatient visits using search engine queries [87]. Table 2.1 presents an overview of the platforms and applications for infectious and chronic disease.

2.3.2 Behavioral and Mental Health

Behavioral health focuses on how people's thoughts and behaviors influence their well-being. Again, traditional sources of data, such as physician visits and national surveys [56], are often time- and resource-intensive. Additionally, they rely on self-reported behaviors which are prone to bias. Computational approaches using passively collected online user-

Domain	Topics (Analysis)	Platforms
Infectious Disease	ILI (surveillance)	search engine queries, Twitter, Wikipedia, OpenTable
	dengue (surveillance)	search engine queries, Twitter
	Lyme disease (surveillance)	search engine queries
	norovirus (surveillance)	search engine queries
	Escherichia coli (surveillance)	Twitter
	malaria (surveillance)	search engine queries
	bubonic plague (surveillance)	Twitter
	infectious instestinal disease (surveillance)	Twitter
	Ebola virus (public perception)	search engine queries, Twitter
	Zika virus (spatiotemporal and content analysis)	Twitter
	conjunctivitis (surveillance)	search engine queries, Twitter
	syphilis (surveillance)	search engine queries, Twitter
	STDs (content analysis)	Yahoo! Answers
	HIV (surveillance)	search engine queries, Twitter
	HIV-related (sexual or substance abuse risk) content (classification)	Twitter
Chronic Disease	pancreatic cancer (predictive modeling)	search engine queries
	pain (content analysis)	Twitter
	asthma (surveillance)	Twitter
	diabetes (content analysis)	Twitter
	migraine (content analysis)	Twitter
	movement disorders (online information seeking)	search engine queries
	epilepsy (online information seeking)	Wikipedia
	multiple sclerosis (online information seeking)	search engine queries
	dimentia-related outpatient visits (prediction)	search engine queries

Table 2.1: Previous works focused on infectious and chronic disease using computational approaches on passively collected online user-generated content.

generated content have been applied to a number of behavioral health issues including diet, physical activity, substance abuse, vaccination, public awareness, and reproductive health. Below is an overview of some of the studies in these areas.

Dietary habits have been examined by analyzing the content of tweets and Instagram posts that mentioned food consumption [88, 89]. Studies have focused on analyzing the content of tweets that contain mentions of physical activity to inform future health marketing interventions [90]. Other studies have examined the content of online weight loss and pro-eating disorders communities hosted on a variety of platforms (discussion forums, blogs, and Reddit) [91, 92, 93].

A number of studies have focused on examining substance abuse using passively-collected, user-generated data. Studies have examined tobacco use (e.g., interest in e-cigarettes using search engine queries [94]), alcohol use (e.g., alcohol abuse using Twitter [95]), marijuana use (e.g., online sales using search engine queries [96]), use of synthetic cannaboids (e.g., tracking interest using search engine queries [97]), responses to policy changes (e.g., examine behavioral responses to Russian policy changes for codeine-containing medication using search engine queries [98]), use of over-the-counter, prescription, and illicit drugs (e.g., examining attitudes towards illicit and prescription drugs using Twitter [99], and the connection between prescription drug abuse and online illicit pharmacies using Twitter [100]).

Several studies have examined vaccinations including reasons for vaccination refusal using tweets [101], opinions about the HPV vaccination using tweets [102], interest in vac-

cinations using search engine queries [103], and interest in measles, mumps, and rubella vaccines using search engine queries [2]. A number of studies have used computational approaches on passively collected online user-generated content to examine public awareness on a variety of issues due to celebrity disclosure including HIV using search engine queries [104], suicide following celebrity suicide using Reddit [105], and cancer information seeking using search engine queries [106].

A descriptive study examined user characteristics and content analysis of general sexual health queries for males and females using the ChaCha question and answer platform [107]. Finally, one study used tweets to investigate gender and ideological differences differences between constituents in states that adopt anti- versus pro-abortion policy [108].

Public health organizations also systematically collect data from a number of surveys to monitor population mental health, however these surveys do not focus specifically on mental health [109]. To supplement traditional sources of data, there have been a number of studies focused on mental health using online user-generated content.

Studies examining depression have used a variety of social media platforms including Twitter, Reddit, Facebook, and search engine logs. For example, researchers have leveraged behavioral clues in tweets postings to predict the risk of depression before onset [110]. Reddit has been used to examine the role of self-disclosure and social support for several mental health issues including depression [3]. Facebook has been used to examine disclosures of depression by college students [111]. Search engine queries have been examined to determine seasonality of depression [112, 113]. There have been a number of studies in

suicide research. De Choudhury et al. (2016) [114] compared individuals who transitioned from posting on online mental health subreddits to a suicide watch subreddit to individuals who only posted on mental health subreddits using features of language providing insight into language markers indicating suicidal ideation. Braithwaite et al. (2016) [115] compared tweets of suicidal to non-suicidal participants, labeled according to a screening tool for suicidal symptoms, using a model with language features as variables which provided evidence that short messages can provide sufficient information to differentiate suicidal from non-suicidal individuals.

Other mental health issues have also been examined including classifying deviant eating disorder behavior using Instagram posts [116], mental health impacts during the Great Recession using search engine queries [117], postpartum changes following in emotion and behavior using tweets and Facebook [118, 119], and stress following gun violence near college campuses using Reddit [120]. Table 2.2 presents an overview of the platforms and applications for behavioral and mental health.

Domain	Topics (Analysis)	Platforms
Behavioral Health	nutrition (content analysis)	Twitter, Instagram
	physical activity (content analysis)	Twitter
	weight loss and eating disorder (content analysis)	discussion forums, blogs, Instagram, Reddit
	e-cigarettes (surveillance)	search engine queries
	online marijuana sales (suveillance)	search engine queries
	alcohol abuse (surveillance and content analysis)	Twitter
	synthetic cannaboids (surveillance)	search engine queries
	illicit/prescription drugs (content analysis)	Twitter
	online illicit pharmacies (content analysis)	Twitter
	reaction to drug policy (surveillance)	search engine queries
	vaccine refusal (content analysis)	Twitter
	opinions about HPV vaccination (content analysis)	Twitter
	interest in vaccines (surveillance)	search engine queries
	MMR vaccines (surveillance)	search engine queries
	HIV testing following celebrity disclosure (surveillance)	search engine queries, sales of OraQuick tests
	suicide following celebrity suicide (surveillance and content analysis)	Reddit
	cancer information seeking following celebrity disclosure (surveillance)	search engine queries
	pro- and anti-abortion conversation (content analysis)	Twitter
	sexual health (content analysis)	ChaCha
	pre-exposure prophylaxis (PrEP) (content analysis)	Twittter
Mental Health	depression risk (predictive modeling)	Twitter
	self-disclosure and social support for mental health issues (content analysis)	Reddit
	self-disclosure of depression (explanatory modeling)	Facebook
	seasonality of depression (surveillance)	search engine queries
	suicide risk (predictive modeling and causal inference)	Reddit
	suicide risk (predictive modeling)	Twitter
	eating disorders (classification)	Instagram
	mental health during the Great Recession (surveillance)	search engine queries
	post-partum changes in emotion and behavior (predictive modeling)	Twitter, Facebook
	stress following on-campus gun violence (predictive modeling and content analysis)	Reddit

Table 2.2: Previous works focused on behavioral and mental health using computational approaches on passively collected online user-generated content.

CHAPTER 3

ONLINE INFORMATION SEEKING FOR REPRODUCTIVE HEALTH

3.1 Motivation

More than half (51%) of pregnancies in the U.S. are unintended [121] resulting in an estimated $21 billion in total public expenditures annually [122]. Beyond the financial burden on the health care system, an unintended pregnancy places a long-term financial burden on the mother and family [123]. Increased access to contraception following legalization of oral contraceptives (birth control pills) has been associated with lower rates of poverty, higher rates of work-force participation, and higher wages for women [124, 125, 126, 127]. All said, prevention of unintended pregnancies is not only beneficial to the mother and family directly involved, but also to maintain public expenditures on health care costs and vital to a healthy economy.

In comparison to unintended pregnancies, contraceptives are less costly and generally seen as a cost-saving measure by insurers [124]. Contraception is one of the most pervasive medical interventions [124] with almost all sexually-active women (99%) in the U.S. having used at least one contraceptive method during their lifetime [128]. To further put this into perspective, 39 million women in the U.S. use contraception of which 30 million use a method more effective than condoms (78% of contraception users) and 4 million use an

IUD (10% of contraception users), a form of long-acting reversible contraceptive (LARC). Oral contraceptives and IUDs are the most popular forms of contraception in the U.S. Less than 1.5% of contraception users use another form of long-acting reversible contraceptive [129, 130, 131, 132].

The IUD is cost-effective, has high patient satisfaction, and is one of the most effective contraceptive methods available with a failure rate less than 0.8% with typical use (compared to 9% for oral contraceptives) [133]. Despite the appeal of IUDs, the utilization rate remains low because barriers to utilization including clinician knowledge, lack of same day insertion, low patient awareness, and high upfront costs [134, 135, 136]. The prohibitively high upfront cost, which can exceed $1,000, has been a major determent for higher IUD utilization rates [124, 137, 138, 129, 134]. Costs associated with an IUD include consultation, insertion, removal, and checking placement [129]. Although IUDs have a high upfront cost, they last multiple years whereas oral contraceptives cost between $160 to $600 annually, often making IUDs a more cost-effective option long-term [139]. As a matter of fact, in the Contraceptive CHOICE Project, one of the largest prospective cohort studies examining women seeking reversible contraception, when the barriers of cost, access, and knowledge were addressed, women selected long acting reversible contraceptives (including the IUD) at a 75% uptake rate [140].

With the exception of religious exemptions and grandfathered plans, the ACA contraceptive mandate, which went into place in August 2012, required all private insurance plans (from both employers and the Marketplace) to offer coverage of prescription contraceptives with no cost sharing by January 2013 [124, 141]. In May 2015, the Depart-

ments of Labor, Health and Human Services and the Treasury issued guidelines stating that health insurance must offer at least one of the eighteen FDA-approved contraceptive methods [142]. The mandate eliminated the cost barrier for most contraceptives, especially for those with high upfront costs like the IUD. Many women saw a substantial decrease in out-of-pocket spending for most contraceptive methods following the implementation of the mandate [124]. These decreases are likely to be more noticeable for women with private health insurance [124]. Before the mandate, contraceptive expenses represented a significant amount, 30%-44%, of a woman's out-of-pocket expenses for health care [124] resulting in an estimated $1.4 billion of annual savings for oral contraceptives alone [124].

The ACA "has experienced among the most controversial and conflictual implementations of any social welfare policy law in American history [143]." The bill faced an uphill battle starting during the legislative process. Congress approved the ACA without the support of any Republican votes; only Democrats voted in support of passage of the bill and it was unanimously opposed by all Republicans [144]. Since it's passage, the implementation of the ACA has remained highly politicized. There have been significant financial investments to convince the public of the demerits of the ACA, which has mounted public opposition against the bill [145, 143, 144]. Additionally, one of the primary campaign messages of Republicans has been opposition against the ACA. By the 2016 U.S. presidential election, the House, controlled by a Republican majority since early 2011, had voted to repeal ACA more than 50 times [146]. With divisive public opinion about the merits of the ACA, especially the contentious contraceptive mandate, the 2016 U.S. presidential and congressional election was viewed as a "mortal challenge [145]" to the ACA with Repub-

lican leaders vowing extensive repeal of the ACA.

Continued access to contraceptives has been thrown into question [147, 148] with the heightened focus on a repeal of ACA following the 2016 election of a Republican president and Republican majorities in the House and Senate.

3.2 Research Questions and Hypotheses

The aim of this case study is to examine how the public is responding to a potential repeal of the contraceptive mandate by examining online information seeking for IUDs during a one-year period post-election. More specifically, I examine the following research questions:

- **Research Question 1**: Is there a relationship between online information seeking for IUDs and heightened focus on the repeal of the ACA?
  - **Hypothesis 1.1**: Observed queries (information seeking) for IUDs are greater than the forecasted counterfactual scenario following the 2016 U.S. election.
  - **Hypothesis 1.2**: Observed queries (information seeking) for IUDs during time periods corresponding with increased uncertainty of the repeal of the Affordable Care Act (election of Republican-controlled administration, inauguration of Republican-controlled administration, and successful House vote to repeal ACA) are greater than the forecasted counterfactual scenario.
- **Research Question 2**: Is there a relationship between online information seeking and state-level electoral votes for the 2016 presidential election?

– For the second research question, a one-sided a priori hypothesis is not included because on one hand, we expect that queries will not be elevated in comparison to the counterfactual scenario in states that voted Republican because the repeal of the ACA is primarily a Republican campaign message. However, on the other hand, we expect that queries will be elevated across all states (regardless of the 2016 electoral vote) because access to birth control impacts women across the political spectrum.

3.3 Background

3.3.1 Internet as a Source of Information for Reproductive Health

The Internet is a significant resource of information about contraception, pregnancy, and childbirth during a woman's reproductive years [149]. In a self-administered survey of 3,181 young reproductive-aged women, 48% of women (59% of the Internet users in the survey) indicated that they use the Internet to look for information about contraception (oral contraceptives, contraceptive injection, IUDs) [150]. Conception-related searches were the second most common type of reproductive health-related search after pregnancy. Similarly, in a study focused on sexual health information seeking of college students, more than half of the students indicated they had searched the Internet for information about preventing pregnancy (53%) and contraceptives (50%) [151].

3.4 Data

Because little or no traditional data is contemporaneously available at the time of the study, we turned to Internet searches to understand public engagement with contraceptive options [152]. The fraction of Google searches (google.com/trends) emerging from the United States from January 2004 through October 2017 for the three most popular reversible contraceptive methods was monitored. This included oral contraceptives (including all queries with "birth" and "control" and "pill(s)," or "oral" and "contraceptive(s)/tion," e.g., "birth control pill," "get oral contraceptives," or "best birth control pills"), intrauterine devices ("IUD" or IUD brands "ParaGard", "Mirena", "Skyla", "Liletta", and "Kyleena") and condoms ("condom(s)").

3.5 Methods

3.5.1 RQ1: Information Seeking during Heightened Focus on Health Policy

Monthly search volumes for oral contraceptives, IUDs, and condoms after the presidential election (November 2016 to October 2017) were compared against expected search volumes generated from AutoRegressive Integrated Moving Average (ARIMA) forecasts [153] using all prior searches to project a synthetic counterfactual [152]) with prediction intervals. The residuals were checked to ensure no evidence of autocorrelation using the Ljung-Box statistic (all p-values were greater than 0.05 indicating no remaining significant autocorrelation) and visual inspection of the autocorrelation function (ACF) and partial au-

tocorrelation function (PACF). Bootstrap confidence-intervals of the cumulative difference between observed and expected search volumes were calculated. Raw search counts were inferred from comScore estimates (comscore.com).

3.5.2 RQ2: Information Seeking across Political Ideology

The analysis outlined in the section above was repeated for the IUD searches using state-level query data. A Welch t-test was used to determine if there were statistically significant differences in cumulative searches for IUDs between states that voted for the Republican candidate (Trump) versus the Democratic candidate (Clinton).

3.6 Results

3.6.1 RQ1: Information Seeking during Heightened Focus on Health Policy

As shown in Figure 3.1, IUD searches have reached all-time highs since the presidential election. Searches were cumulatively 15% (95%CI: [10,20]) higher than expected following the election (from November 2016 through October 2017), reflecting 10-21 million more searches than expected. Conversely, searches for oral contraceptives confirmed with expected volumes (0%; 95%CI: [-2,1]). Although not covered by the ACA, searches for condoms, were lower than expected on average (-4%; 95%CI: [-5, -2]). Observed search volumes for both oral contraceptives and condoms were within the 95% prediction intervals (in other words, expected search volumes based on searches pre-election). In contrast,

searches for IUDs eclipsed the forecasted 95% prediction-interval for six of the twelve months (November 2016, January 2017, February 2017, May 2017, June 2017, and July 2017) and reached a record high in May 2017, with 8.3 million total searches that month, when an ACA repeal advanced in the House. These dates correspond with the time periods of elevated uncertainty of the fate of the ACA including the election, inauguration, and successful House vote to repeal the ACA.

3.6.2 RQ2: Information Seeking across Political Ideology

As shown in Figure 3.2, cumulative differences IUD searches were statistically significantly higher than expected in all states, except NV (2%; 95%CI: [-4, 10]), ranging from 9% greater in in FL, IN, TN, and WA to 24% in MO. These increases were consistent across states that voted Democratic (Clinton) or Republican (Trump) (Welch t-test= 0.60, p = 0.548). Table 3.1 presents the model, mean, and 95% CIs for each state. Appendix A shows the temporal changes for each state.

3.7 Discussion

Americans are considering IUDs in record numbers, while interest in other forms of birth control have remained stable (oral contraceptives) or declined (condoms). It is unclear if queries corresponded with utilization as claims data are not available. However, as the vice president of research at athenahealth stated “It certainly looks like some women are concerned that full coverage for contraceptive services will be more expensive for them, and so

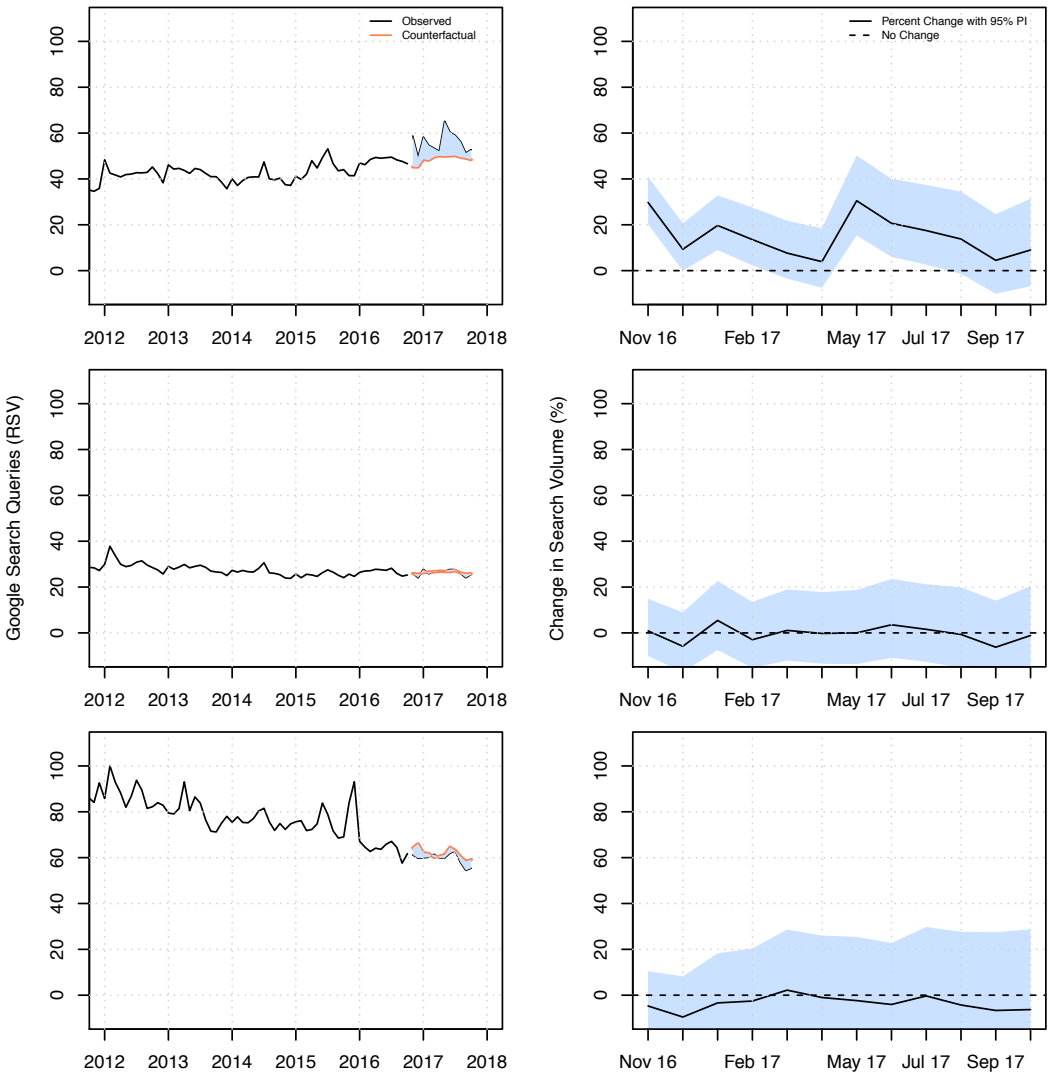


Figure 3.1: Relative search volume (RSV) and percent change in RSV for Google queries for one-year following the 2016 U.S. election. Row 1 = Queries for intrauterine devices (IUDs). Row 2 = Queries for birth control pills. Row 3 = Queries for condoms.

are getting IUDs without cost while they still can [139].” Indeed women view LARCs as very effective, but not as affordable as oral contraceptives and are often inhibited by financial constraints in contraceptive decision making [137]. According to athenahealth, there was a 19% increase in IUD management or insertion-related visits between October and December 2016 [139] and 21% increase between October 2016 and January 2017 (three

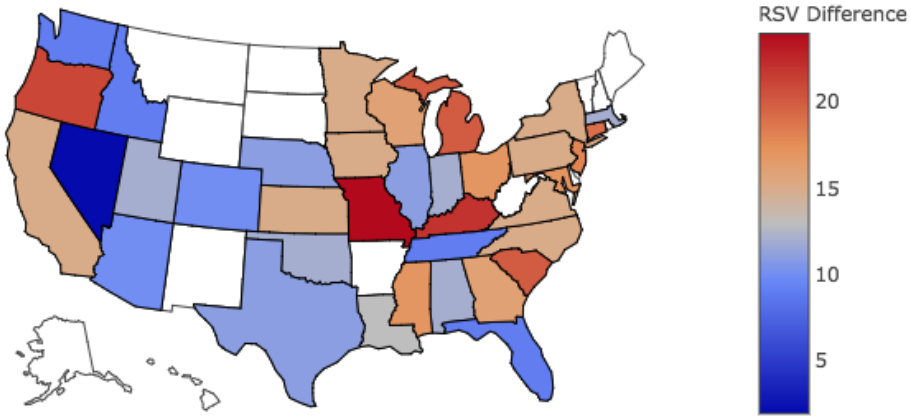


Figure 3.2: Within-state percent changes in relative search volume (RSV) for intrauterine device (IUD) queries for one-year following the 2016 U.S. election.

months post-election) in comparison to the same time period a year prior [154], which is consistent with our results. This analysis was based on electronic health records database of 1 million patient visits across a network of 85,000 providers. The IUD-related visits were primarily for commercially-insured patients, who are most at risk for losing the no cost-sharing contraceptive coverage if the ACA or contraceptive mandate was overturned, while rates remained constant for Medicaid patients [139]. Additionally, similar to our analysis, athenahealth reported an increase in IUD-related visits in counties regardless of voting Democratic or Republican in the 2016 presidential election [139].

According to athenahealth, as of May 2017, demand for IUDs has reduced to nearly pre-election levels and researchers speculated this indicated a decrease in anxiety related

to losing the no cost-sharing contraceptive coverage or that concerned women had already went through with an IUD insertion [155]. However, researchers warned that demand is dynamic and could increase again based on further uncertainty around healthcare policy shifts [155]. According to the chair of the American College of Obstetricians and Gynecologists working group on long-acting reversible contraception, "This is about people's fears about the contraceptive mandate going away [154]," who also indicated that patients were increasingly seeking information about LARCs.

Searches predict other behaviors, such as medical testing [104], and at minimum show the change in administration has coincided with elevated engagement with IUDs. The etiology of this shift is less clear, but in part may be because IUDs with a lifespan of three to ten years will confer continued protection even after an ACA repeal thereby providing a medical hedge against any potential ACA repeal.

3.7.1 Health Information Seeking Models

This case study used an atheoretical approach to examine online information seeking for IUDs in response to heightened uncertainty surrounding a repeal of the ACA. We compare our results to the health information seeking models presented in Chapter 2, which were primarily built to analyze individual-level survey or interview data with structured questions. More specifically, below we discuss (1) the limitations of the RISP (see Figure 2.3), HIAM (see Figure 2.1), ECMISB (see Figure 2.2) for the passively collected search engine query data, (2) how the CMIS (see Figure 2.4) could potentially be operationalized for

future studies with search engine queries or similar data, and (3) the suitability of current health information seeking models for aggregate search engine queries.

Key components (individual characteristics, individual beliefs, and perceived information gathering capacity) related to an individual's health information search according to the RISP would be difficult to operationalize using passively collected data from search engine queries. Similarly, the HIAM focuses on a sequential search process, which is difficult to elicit from the passively-collected, aggregate search engine query data. Operationalizing the HIAM model would be more suitable for other types of passively collected social data such as individual-level, temporal search engine query data. Finally, the ECMISB integrates contextual and personal demographics into the model, which are difficult to elicit from passively collected search engine query data. Search engine queries are aggregated, therefore we cannot infer the contextual variables that affected the individual information search. Operationlizing the ECMISB may be better suited for survey or interview data, where questions could elicit the contextual and personal demographics.

According to the CMIS, information seeking actions are active, purposeful searching for information. Unfortunately, we can not determine the process that an individual in the aggregated sample takes in their information search. However, we find evidence that information seeking for IUDs is positively related to the degree of uncertainty about healthcare policy reform. Uncertainty occurs "when details of the situation are ambiguous, complex, unpredictable, or probabilistic; when information is unavailable or inconsistent; and when people feel insecure in their own state of knowledge or the state of knowledge in general. [156]" Contributing to uncertainty is the hostile political environment [145] intensifying

the politicization of healthcare reform [144]. Most of the existing models of health information seeking focus on rational contexts (thinking, reflecting), however, research has often found that health information seeking is irrational reflecting emotions and subjectivity [20]. Healthcare providers have confirmed that there is indeed "financial uncertainty for women as healthcare policy shifts [139]." Periods of heightened focus on reform are likely to create heightened uncertainty because of the unpredictability and complexity of the outcomes of reform [157]. Consistent with other studies, we find evidence that higher rates of information seeking are related to heightened uncertainty [158] supporting Kuhlthau's "uncertainty principle", which posits that uncertainty increases anxiety and reduces confidence [25]. Although we cannot know the context behind the individuals in the aggregate search engine queries, we find evidence that with each event perceived as potential threat to the ACA.

Online information seeking may be seen as a way to take control during these periods of heightened uncertainty [149]. Self-efficacy, a important predictor of information seeking [159], is the "belief in one's capabilities to organize and execute the course of action required to produce given attainment [160]." Health self-efficacy are "individuals' beliefs about their ability to manage their health [157, 161]." It's likely that the individuals searching feel a potential reform is salient to them. Finally, the antecedent of beliefs or their level of control over a situation is evidenced by the information search combined with the analysis conducted by athenahealth indicating increases in IUD-related visits.

As mentioned in the limitations, some antecedents, such as demographics, cannot be inferred from the aggregate data sample. Additionally, experience is the proximity of the

individuals to disease. In this scenario, searchers are not motivated by disease, but are potentially motivated to prevent unintended pregnancy without incurring higher costs of prevention.

Another input into information seeking actions, as identified by the CMIS, is the characteristics of the information carrier. Unfortunately, there is not a way to identify this for each individual. However, many individuals trust search engines, such as Google, and often initiate their search using one [162].

Ultimately, new information acquired from the search may modify an individual's beliefs, attitudes, and decisions to seek treatment, inform about risk, and provide resources to support behavioral changes [24]. In this sense, the information seeking may act as an input into the HBM.

Computational analysis can identify a high-level view of the relationship of information seeking and events of perceived uncertainty. However, current health information seeking models were primarily developed for analysis of survey or interview data. In comparison to health information seeking field, infodemiology is a relatively new field. Health information seeking models that account for passively collected data have yet to be developed, but development of such models, specific to the today's rapidly changing information environment, would be beneficial for standardization and comparison across platforms.

3.7.2 Research and Practical Implications

The benefits of IUD usage extend beyond cost-savings. For example, greater uptake of IUDs and implants has preceded fewer births and abortions [163, 164, 165]. The American College of Obstetricians and Gynecologists suggests that any "repeal and replace" effort should continue coverage of no cost-sharing for FDA-approved contraceptives [154]. Research has indicated that women with interest in the IUD are significantly more likely to have one if the incurred cost is less than $50 [134]. Those who may be most adversely impacted are individuals who had high out-of-pocket expenses for contraceptives prior to the ACA since this group had the highest demand for IUDs after implementation of the ACA [166].

Given the benefits of IUD usage and increased information seeking, future work should focus on the quality of information shared in online resources. Specifically, the availability of factual information since people like to obtain their information social networks. Negative narratives and misconceptions can discourage women from utilizing IUDs [167, 168, 169]. In one study, few websites offered current information about LARCs [170]. This is especially relevant for IUDs because of persistent misinformation stemming from the Dalkon incident. The faulty Dalkon Shield IUD in the 1970s and 1980s resulted in over 200,000 gynecological complications and 18 deaths which damaged the reputation of IUDs and contributed to fear and decline in popularity in the U.S. [171]. Exploring womens' concerns and beliefs about the IUD, as presented in social media, could help inform effective health communication. Additionally, because of the partisan nature of healthcare

reform, the divisive opinions of political entities could shape the public's perception and understanding of the contraceptives and the contraceptive mandate [157].

Beyond reproductive health-specific implications, traditional surveillance data such as health insurance claims or large, nationally-representative surveys are costly, resource-intensive, and often significantly lag in time. For example, the next National Survey of Family Growth [172], which inquires about contraceptive methods, covers a four year period from 2015 to 2019 (unfortunately it does not inquire when women change contraceptive methods [129]). As another example, a license for two seats of one-year of full access to health insurance claims data is $30,000 [173]. Monitoring search engine queries may be used to inform health communicators of emerging information needs as well as gauge real-time public. Public opinion can shape legislative activity and decisions made by health insurance organizations [157]. Future efforts should focus on validation of the search engine query data using claims or survey data as well as building monitoring similar to the platform to identify mental health trends [174]. Knowing of emerging health concerns will enable health communicators to provide information, so the public can make an informed decision, instead of a decision out of fear [148].

3.7.3 Limitations

Google Trends aggregate search engine queries only track the segment of the population that use this service [26], therefore we do not claim to model demand for all individuals in the U.S. We see this analysis as a step towards understanding health information needs and

shifts in public reactions towards healthcare policy in near real-time. Currently, external validation is not available, however, as mentioned earlier, our results do correlate with an analysis of visits records in an electronic health records database presented by athenahealth. Another limitation of aggregate search engine queries is the lack of detailed information of the individuals conducting the queries, so we can not ensure results are representative of the U.S. population nor track individual-level characteristics.

State	Model	Mean	95% CI	Election Result
AL	ARIMA(0,1,1)(2,0,0)[12] with drift	12	[4, 21]	Trump
AZ	ARIMA(0,1,1)(2,0,1)[12] with drift	10	[5, 17]	Trump
CA	ARIMA(2,1,0)(2,0,0)[12]	15	[10, 21]	Clinton
CO	ARIMA(1,1,1)(2,0,0)[12] with drift	10	[6, 15]	Clinton
CT	ARIMA(0,1,1)(1,0,0)[12] with drift	19	[14, 24]	Clinton
DC	ARIMA(0,1,1) with drift	11	[4, 19]	Clinton
FL	ARIMA(2,1,1)(1,0,0)[12] with drift	9	[2, 16]	Trump
GA	ARIMA(0,1,1)(1,0,0)[12]	16	[10, 24]	Trump
IA	ARIMA(1,1,1)(2,0,0)[12] with drift	15	[12, 20]	Trump
ID	ARIMA(2,1,1)(2,0,0)[12] with drift	9	[6,13]	Trump
IL	ARIMA(0,1,1)(0,0,1)[12] with drift	11	[6, 16]	Clinton
IN	ARIMA(0,1,1)(2,0,0)[12] with drift	12	[5, 19]	Trump
KS	ARIMA(0,1,1)(1,0,2)[12]	15	[7, 25]	Trump
KY	ARIMA(1,1,1)(2,0,0)[12]	22	[16, 30]	Trump
LA	ARIMA(0,1,1)(2,0,0)[12] with drift	13	[5, 22]	Trump
MA	ARIMA(0,1,2)(2,0,0)[12] with drift	12	[7, 17]	Clinton
MD	ARIMA(0,1,1)(1,0,0)[12] with drift	17	[11, 23]	Clinton
MI	ARIMA(0,1,2)(0,0,1)[12] with drift	20	[14, 27]	Trump
MN	ARIMA(1,1,2)(1,0,0)[12] with drift	15	[9, 22]	Clinton
MO	ARIMA(1,1,1)(2,0,0)[12]	24	[18, 30]	Trump
MS	ARIMA(5,1,1)(2,0,1)[12]	17	[8, 29]	Trump
NC	ARIMA(1,1,1)(2,0,0)[12] with drift	15	[11, 19]	Trump
NE	ARIMA(4,1,0)(2,0,1)[12] with drift	11	[4, 19]	Trump
NJ	ARIMA(2,1,1)(2,0,0)[12]	18	[10, 26]	Clinton
NV	ARIMA(3,1,1)(2,0,1)[12] with drift	2	[-4, 10]	Clinton
NY	ARIMA(2,1,1)(1,0,2)[12]	15	[9, 22]	Clinton
OH	ARIMA(2,1,3)(2,0,0)[12] with drift	17	[12, 23]	Trump
OK	ARIMA(3,1,3)(2,0,1)[12]	12	[4, 20]	Trump
OR	ARIMA(0,1,1)	21	[16, 26]	Clinton
PA	ARIMA(1,1,1)(2,0,2)[12] with drift	15	[10, 22]	Trump
SC	ARIMA(0,1,1)(2,0,1)[12]	20	[11, 30]	Trump
TN	ARIMA(0,1,2)(2,0,0)[12] with drift	9	[5, 14]	Trump
TX	ARIMA(0,1,1)(2,0,1)[12]	11	[6, 16]	Trump
UT	ARIMA(1,1,1)(0,0,1)[12]	12	[7, 18]	Trump
VA	ARIMA(1,1,1)(1,0,0)[12]	15	[10, 21]	Clinton
WA	ARIMA(3,1,1)(2,0,0)[12] with drift	9	[5, 13]	Clinton
WI	ARIMA(2,1,1)(2,0,0)[12]	16	[10, 21]	Trump

Table 3.1: State-Level AutoRegressive Integrated Moving Average (ARIMA) Models.

CHAPTER 4

ONLINE INFORMATION SEEKING FOR SEXUAL HEALTH

4.1 Motivation

Reportable STDs (chlamydia, gonorrhea, and syphilis) hit a record high of 2.1 million new diagnoses in the U.S. in 2016 [175]. STDs are a substantial public health burden with approximately 20 million new STDs annually resulting in $16 billion of annual healthcare costs in the U.S. [175]. With many people undiagnosed, this may be an underestimate of the incidence and cost. Beyond economic consequences, STDs can result in emotional distress, infertility, create inflammatory conditions, and increase the risk of acquiring HIV [176, 177].

Despite the preference to obtain health information from a health care provider, sensitive issues such as sexual health, often drive people to consider less trustworthy sources of information readily available on the Internet. Beyond privacy concerns, there may be other barriers for people who choose to access the Internet as their first source health information including, but not limited to, lack of access to a healthcare provider (rural locations), lack of health insurance, and financial barriers (a financial barrier may still exist even if an individual has health insurance). As a result, the accessibility of the Internet may be attractive for individuals who are concerned they may have transmitted or acquired a STD as well as

to seek information about risk and treatment.

The advent of Web 2.0 has generated a plethora of new sources of information available on the Internet (e.g., online communities, social networking sites, and social Q&A) that extend beyond passive viewing of sites. These sites encourage people to engage with each other collectively contributing knowledge and providing social support [13].

Reddit is a social news sharing and discussion site that is organized into subreddits that focus on specific topics. As of December 2017, there were 1,194,569 subreddits [178]. Reddit is becoming an increasingly popular source of information [179]. According to Alexa Internet, Reddit is currently the third most popular site in the U.S. following Google (first) and Youtube (second) based on a combined measure of page views and unique site users [180]. To put that into perspective, Facebook is ranked fourth, Amazon is ranked fifth, and Wikiipedia is ranked sixth.

r/STD is a subreddit that has existed for seven years with the self-described goal of "...help calm the anxiety that comes with a potential STD infection through education, awareness, and prevention techniques [181]." A brief description provided by the site follows: "Feeling paranoid about that one night stand? Condom broke? Got some scary blood results back? Need help understanding what your labs are saying? Share your stories, concerns and questions here. Anything and everything STD related [181]."

Online health information has far-reaching effects on public health by influencing health-care decisions and outcomes [15, 19]. Furthermore, there is increasing evidence of people using online resources not only for health information, but also social support. From an or-

ganizational perspective, there is increasing interest in using social media to provide health communication to the public [12]. For example, r/STD was one of 25 health-related subreddits recently targeted for sharing health information sources [182]. Additionally, STDs are highly stigmatized and, despite the health care community's robust knowledge about these conditions, misinformation and misunderstanding is pervasive among the public. Despite increases in STDs, public health programs for STDs have received fiscal cuts in recent years [175], potentially exacerbating existing barriers and increasing the need for innovative strategies for public outreach. It is imperative that we understand who, why, and how people use social media for health information so that online resources can be monitored to ensure that health communication can be accurately tailored and distributed to the appropriate audiences [10].

4.2 Research Questions

Our goal was to characterize the community and content of information shared on r/STD. As such, the work was motivated by three research questions:

- **RQ1**: What are the general patterns of user engagement?
- **RQ2**: What is the content of the information exchange?
- **RQ3**: What are the psycholinguistic properties that original posters (OPs) articulate?

To answer these questions, we conducted a mixed methods analysis using one-year (prior to this case study) of content from r/STD. Computational analysis of all original

posts (OPs) and comments was conducted to examine content and qualitative content analysis was conducted on a subsample of data to ensure the computational analysis captured the dynamics of the community during the study period. Note, we use a common convention on Reddit of referring to both the initial post and initial poster as OPs. The findings discussed in this chapter reveal that r/STD largely serves as a platform to crowdsource a diagnosis as both a first opinion prior to interacting with a healthcare provider and a second opinion to follow-up to an interaction with a healthcare provider. Beyond crowdsourcing a diagnosis, users often articulate concerns regarding test window periods and transmission and risk of STDs during oral sex. This study addresses a gap in the literature about online health information seeking for STDs. A key strength is the characterization of a previously unstudied, but active, online community. Reddit, specifically the r/STD subreddit, offers the opportunity for researchers to better understand the methods in which users interact, discuss, and share health information revealing patterns and content that could be used to tailor health interventions specific to this platform in the future. Finally, although this study took an inductive, atheoretical approach to our analysis, we discuss how existing health information seeking models could be operationalized for studying the r/STD community.

4.3 Background

4.3.1 Internet as a Source of Information for Sexual Health

In a semi-structured interview of adolescent males, all participants indicated that sexual health is an "embarrassing and stigmatized" topic and this affected their information seek-

ing behavior. The participants preferred to "seek information in private" using the Internet although they acknowledged the potential trade-off for privacy in lieu of accuracy [183]. Other studies have also indicated that the "Internet is a leading source of sexual health information" for young adults because of the sensitivity of the subject [151, 184, 185, 186]. For example, adolescents turn to the Internet for sexual health information that they fear parents and healthcare providers may be judgmental about, including STDs. Anonymity is of particular concern for this age group and they prefer to use online resources that do not require personally identifying information [187]. Furthermore, the lesbian, bisexual, gay, transgender, queer/questioning [LBGTQ] community seek sexual health information online at a significantly higher rate compared to heterosexual individuals [185, 188] potentially because of the additional challenges they face with their sexuality [189]. In a qualitative interview of 32 LGBT people aged 16-24, all participants endorsed using the Internet to seek information about STDs or HIV [189]. This behavior is not limited to digital natives (young adults raised with the Internet). Middle-age and older adults also believe sexual health is a sensitive issue and use the Internet as a source of information, but at lower rates than young adults [190, 184].

4.3.2 Reddit as a Source of Health Information and Social Support

What makes Reddit an attractive source of information despite a large number of credible online resources with information about STDs? Selecting a source for health information is commonly based on the perception of social risk and cognitive effort to understand content, sometimes resulting in irrational choices for information [13]. Reddit has number of dis-

tinguishing features, as identified by the CMIS [13, 191], including accessibility (free and publicly available), quality (individuals crowdsource and evaluate information, although the quality of information is unknown), usability (passive viewing of content without an account), and interactivity (individuals interact with each other by creating an account). This makes Reddit an ideal source for seeking health information, especially for stigmatized topics, and an increasing number of researchers are examining health-related subreddits for stigmatized conditions [3, 114, 192]. Additionally, Reddit simulates interactive health communication providing human-to-human interaction facilitated by the Internet [20].

It is imperative for researchers and public health experts to understand cultural norms and characterize the information exchanged on online communities such as Reddit. An increasing number of researchers are examining health-related subreddits for actionable insights for mental health, weight loss, and monitoring adverse drug reactions. For example, researchers have examined mental health discourse on Reddit [3], language attributes indicating a shift in the user's posting on a mental health to the suicide watch subreddit [114], shifts in posting activity and content following celebrity suicides [105], helpful comments from a suicide prevention perspective on a suicide watch subreddit [193], similarities of content among mental health subreddits [192], the relationship between social support and weight loss [109], stress of college students following gun violence on campus [194], and adverse drug reactions for psychiatric medications [195].

Reddit is an ideal platform to examine health information seeking for STDs given that the demographics of Reddit and incident STD infections align with more than half of Reddit users are under the age of 25 (50.5%) [196] and half of the estimated 20 million new

STD diagnoses in the U.S. are among individuals aged 15 to 24 [175]. r/STD is an active community with 2,500 subscribers, albeit likely many more readers that are not subscribers. To the best of our knowledge, this is the first study to examine sexual health information seeking on Reddit.

4.4 Data

The Python Reddit API wrapper [197] was used to collect OPs, comments, and associated metadata from r/STD for the one-year period between August 1, 2016 and July 31, 2017 from Reddit's official API [198]. A total of 1,802 OPs and 5,948 associated comments created by 2,083 unique users were collected. In accordance with Reddit tradition, an OP is the initial post and comments are the responses to the OP. For each OP and comment, we collected the text and metadata including the date, parent ID of the OP, permanent URL, author, score, number of upvotes, and number of downvotes. The title and number of associated comments was collected for each OP. Note that OPs can contain both a title and follow-up text. Twenty-three percent (22.5%) of OPs contain only titles. The titles and text of each OP were combined to obtain the total text associated with each OP. Based on a qualitative review, titles are typically either openers or briefly summarize the intention of the OP. Comments do not have a title and only contain a text field.

4.5 Methods

4.5.1 RQ1: User Engagement

Activity and Participation

Descriptive statistics including the number of OPs and comments, number of comments per an OP, length of the OPs and comments, unique number of contributors and activity, response time to OPs, usage of the built-in scoring system, and daily activity were examined to determine characteristics of the community.

Throwaway Accounts

A common practice on Reddit is the usage of a throwaway accounts to maintain anonymity. These accounts often contain the word "throw" in the username (e.g., "throwdissaaway"). We used a regular expression search to identify accounts that contained the word "throw" signifying that is was specifically created with the intention of maintaining anonymity. This approach cannot identify all accounts that were specifically created with the intention of anonymity, however, this approach ensures that only throwaway accounts were identified.

4.5.2 RQ2: Content of Information Exchange (Quantitative Analysis)

For computational content analysis, we examined the prevalence of explicit mentions of STDs, top unigrams, online resources that were shared, and topic modeling. Text analysis

was conducted using scikit-learn [199], pandas [200], Natural Language Toolkit (NLTK) [201], and gensim [202].

Prevalence of STDs

Regular expressions were used to find keywords related to STDs present in the OPs and comments. An OP or comment was labeled with the STD if it specifically mentions a keyword. All OPs, comments, and keywords were stemmed using NLTK's snowball stemmer to reduce words to their root form (e.g., syphilis to syphili) prior to the regular expression search. "AIDS" was corrected using a regex search for "aids" because the stemmed version "aid" produced false positives. The OP or comment was labeled with the STD type if it contained any of the associated keywords. For example, the STD type "total HSV" includes text that contained the stemmed variants of "HSV" OR "herpes" OR "cold sore." We then further classified the STDs into three categories consistent with the CDC: (1) acute, reportable STDs (i.e., chlamydia, gonorrhea, syphilis) (2) intermediate, non-reportable STDs (i.e., HSV, HPV, and molluscum contagiosum), and (3) bloodborne chronic, reportable STDs (HIV, acquired immunodeficiency syndrome [AIDS], and hepatitis). Table 4.3 presents the STD categories and associated keywords. These categories are not mutually exclusive. For example, if an OP mentioned chlamydia and HSV, then it was classified as both an acute, reportable STD and intermediate, non-reportable STD.

Top Unigrams

To identify the frequency of unigrams, the text was pre-processed by removing special characters, removing common stop words, stemming words, and joining frequently adjacent co-located unigrams into phrases (e.g., cold_sores) using phrase modeling [203]. The unigrams were then sorted by term frequency and term frequency-inverse document frequency (tf-idf) to obtain the top 50 unigrams in OPs and comments.

Topic Modeling

Unsupervised topic modeling was used to discover latent themes present in the OPs and comments based on the unigrams. From the unigrams, we created matrices of term frequency and term frequency-inverse document frequency (tf-idf) to input into topic models using Latent Dirichlet Allocation (LDA) and non-negative matrix factorization (NMF), respectively. Because both LDA and NMF require a pre-specified number of topics, we iterated through models the number of topics varying from 10 to 40 by increments of five. The latent topics were then manually labeled and the model that produced the most coherent, representative topics was selected. The same process was used for the comments with the exception that all comments associated with an OP were aggregated to maintain the conversational nature of commenters responding to an OP.

Although LDA is arguably the most widely used topic modeling technique, we found LDA to produce substantially less coherent topics. For example, for the OPs, LDA produced one coherent topic compared to 21 coherent topics from NMF for a 25-topic model.

We suspect this is because the OPs and comments contain similar language and NMF produces higher topic coherence for niche content [204]. This finding is in line with other studies examining topic coherence of NMF compared to LDA [205, 206, 204, 207]. Unlike LDA which returns normalized topic probabilities for each document (i.e., topic probabilities sum to one for each document), NMF does not return normalized topic probabilities [208, 209].

Online Resources

Online resources were extracted from the text using the package URLExtract [210] by identifying top-level domains as defined by the Internet Assigned Numbers Authority (iana.org), and extracting all characters to each side of the top-level domain until a stop character (white space, comma, quotation marks) is encountered. A top-level domain is the rightmost label of a URL (e.g., the domain name www.example.com belongs to the top-level domain 'com'). After domain names were extracted from the full URLs, each site was visited and qualitatively labeled with the type of online resource.

4.5.3 RQ2: Content of Information Exchange (Qualitative Analysis)

A 1-in-20 systematic sampling approach was used to identify a subsample of threads or conversations (OPs with associated comments) resulting in 91 OPs with 289 associated comments (4.18 ± 3.08 comments per a OP). An inductive coding process was used for thematic analysis of the threads. One researcher (AN) identified preliminary themes, which

was developed into an initial codebook. AN conducted multiple iterations of independent coding until thematic saturation was achieved and formalized a codebook. Another researcher (JK) coded a subsample of the data (random twenty percent sample of the threads) using the formalized codebook, discussed any points of disagreement, and made refinements to the coding scheme achieving an inter-rater reliability of 0.81 (percent agreement). Related codes were grouped into seven high-level categories representative of prominent themes. Topics that emerged from the qualitative analysis were compared to topics that emerged from the computational analysis (topic modeling) to identify if topic convergence existed between the quantitative and qualitative approaches. Below is a discussion of the findings and each conversation (or thread) is indicated by the unique id assigned to the OP.

4.5.4 RQ3: Psycholinguistic Properties

Each OP was scored using the 2015 version of Linguistic Inquiry Word Count (LIWC) [211]. LIWC is a language analysis software that evaluates and outputs scores for linguistic and psychological features of language. For our analysis, we focused on the following categories: affect, biological processes, cognitive processes, past focus, present focus, future focus, and time. The score for each OP is expressed as the percentage of words belonging to a given category (e.g., score of 5.0 for 'affect' indicates 5% of words in the OP belonged to the 'affect' category) [211]. The means and bootstrapped 95% confidence intervals of the seven psycholinguistic categories were calculated and compared to examine the proportion of psycholingustic content present in the OPs.

The cognitive processes category includes words that reflect cognition such as “notice”, “why”, “curio*”, and “how”. The biological processes category includes words that reflect language about the body, health, and sexuality such as “condom”, “heal”, “penis*”, and “gay.” Past, present, and future focus categories reflect time orientation (e.g., “were” for past focus) and include variations in inflection to modify the tense of a verb (e.g., “appeared” to “appear”). The time category includes words referring to a specific or relative time (e.g., “monday” and “earlier”). The affect category includes words that reflect the overall emotional tone of the OP (e.g., positive emotion, fear, and anxiety) such as “scared”, “anxious”, “asham*”, “happy”, and “trust.”

4.6 Results

4.6.1 RQ1: User Engagement

Activity and Participants

Tables 4.1 and 4.2 present descriptive statistics of the OPs and comments contained in the dataset. In general, the r/STD community averaged 5 $\pm$ 2 OPs and 16 $\pm$ 10 comments per a day during the study period. The majority of activity for both original posting and commenting on r/STD occurred on weekdays rather than the weekend, as shown in Figure 4.1.

The median response time to comments was 7.50 hours (min = 3 seconds; max = 178 days). The mean number of OPs per poster was 1.15 $\pm$ 0.59 indicating that most posters were one-time users of r/STD. The mean number of comments per commenter was 3.98 $\pm$ 16.10.

No. of OPs	1802
Mean OPs per day	5 ± 2
No. of unique original posters	1557
Mean length of OP titles	9 ± 7
Mean length of OP text	146 ± 138
No. of OPs with titles only	406 (22.5%)

Table 4.1: Descriptive statistics of original posts (OPs).

No. of comments	5948
Mean comments per day	16 ± 10
No. of unique commenters	1387
Mean length of comments	38 ± 58
Mean comments per OP	3 ± 3

Table 4.2: Descriptive statistics of comments. OP = original post.

The average user activity for contributing to both OPs and comments was 3.51 ± 13.27 unique contributions. Forty-eight percent (n = 861) of original posters also contributed to comments, however, we did not distinguish if these comments were limited to the thread that they initiated. Surprisingly, the community rarely used the built-in scoring system available in Reddit for this study period. The mean score of OP and comments were 1.39 ± 0.95 and 1.32 ± 0.90, respectively.

Throwaway Accounts

In general, 14% of all users specifically used throwaway accounts with 17% of the 1557 unique OPs using a throwaway account and 12% of the 1387 unique commenters using a throwaway account. We did not distinguish if these commenters were actually commenting on the OP that they initiated.

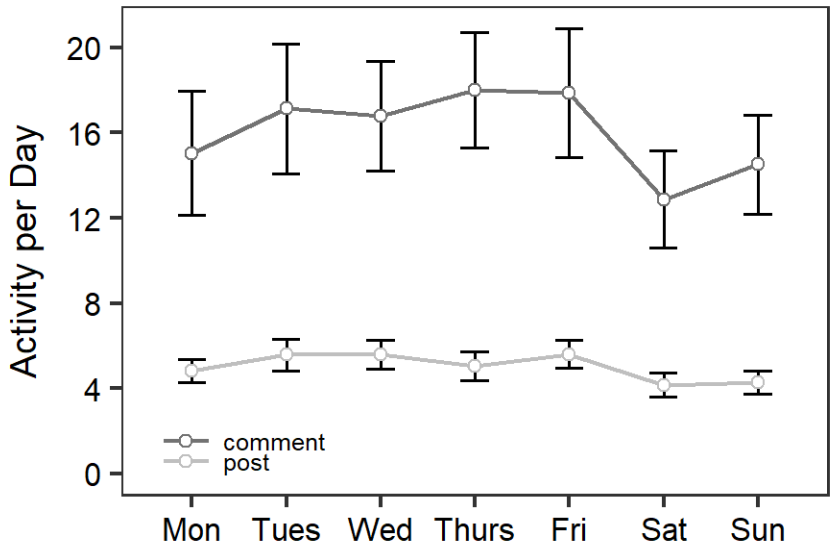


Figure 4.1: Mean number of submissions per day for original posts and comments in the dataset.

4.6.2 RQ2: Content of Information Exchange (Quantitative Analysis)

Prevalence of STDs

As shown in Table 4.3, the majority of OPs focused on the intermediate, non-reportable STDs (n = 816, 45%) followed by acute, reportable STDs (n = 250, 14%) and chronic, bloodborne STDs (n = 209; 12%). The prevalence of explicit STD mentions in the OPs aligns with the prevalence of STDs in the general population [175]. That is, HPV and HSV are the most common STDs and, of the reported notifiable STDs, the most common, in order of incidence, were chlamydia, gonorrhea, and syphilis. There were minimal OPs related to trichomoniasis, lice, pelvic inflammatory disease, lymphogranuloma venereum, bacterial vaginosis, and mycoplasma genitalium (less than ten posts for each STD).

STD	OPs	Comments
HSV (herpes | hsv | cold sore)	556 (30.85%)	1084 (18.22%)
HPV (wart | papillomavirus | hpv | cancer)	301 (16.70%)	702 (11.80%)
Acute reportable (chlamydia | gonorrhea | syphilis)	250 (13.87%)	302 (5.08%)
Intermediate non-reportable (HSV | HPV | MC)	816 (45.28%)	1733 (29.14%)
Chronic bloodborne (HIV | AIDS | hepatitis)	209 (11.60%)	305 (5.13%)

Table 4.3: Prevalence of STDs in original posts (OPs) and comments. AIDS = acquired immune deficiency syndrome. HIV = human immunodeficiency virus. HPV = human papillomavirus. HSV = herpes simplex virus. MC = molluscum contagiosum.

Top Unigrams

Figures 4.2 and 4.3 present the top unigrams by term frequency and tf-idf for OPs and comments. The unigrams were organized into categories focusing on symptoms, health care, time-related, body, STDs, sex, and action words that imply information seeking.

Symptoms. Symptom words in the OPs included symptom, bump, pain, red, little, and itch. Symptom words in the comments included symptom and bump. OPs were more symptom-centric describing symptoms in great detail whereas comments tend to confirm if the symptoms are indeed representative of a STD.

Health care. Health care words in both the OPs and comments were test and doctor. OPs and comments tend to focus on testing for STDs, reflecting on a diagnosis, or seeking a second opinion. For example, one OP:

> . . . I went to the doctor the day that I had developed [a lesion] and she thought it was for sure an STD. She swabbed it for herpes, as well as [testing for HIV, syphilis, chlamydia, and gonorrhea]. They all came back negative . . .

The poster then discusses paranoia, multiple rounds of STD testing despite repeated negative results, and asks readers if they recommend "settling down" or additional testing. An example of a comment follows:

> . . . The fear that you have herpes and may infect another, is definitely a big enough fear to psychologically mess with your erections. I don't disagree that that might be in your head. You've had two doctors say you don't have herpes.

Time-Related. Temporal words in the OPs included time, start, day, week, month, and years. Similarly, temporal words in comments included time, day, week, month, and year. OPs tend to focus on describing a sexual encounter and subsequent symptoms. For example,

> Do I have HIV? I lost my virginity to a girl I met online about 2 months ago. A few days after that my penis started burning during urination and also when I am not urinating . . .

Whereas comments tend to focus on describing what symptoms to look for or reiterating what the poster stated. For example,

> You may or may not experience fever or rash, if you did it would typically come before an outbreak. Usually around 5-10 days but this is different from person to person. Yes, they usually start to heal pretty quickly. Again, everyone is different though. You want to get checked out before they scab over (usually day two or three of an outbreak).

Body. Body words in the OPs included area, penis, and skin. Body words in comments included oral and skin. OPs tend to focus on symptoms in relation to an area of their body (e.g., "Painless bump on pubic area for month") whereas comments tend to focus on describing symptoms to provide information. For example,

> If you've had shingles in the same area, I'd venture to say shingles. It could be herpetic but only way to know is to swab. Shingles and herpes are in the same family of viruses!

STDs. STD words in the OPs included herpes, STD, HPV, HIV, wart, and chlamydia. STD words in the comments included herpes, HPV, wart, STD, HIV, HSV, and virus.

Information Seeking. Information seeking refers to the action words that users employ when trying to understand their health concern. Action words such as look, think, know, notice, and worry were commonly used to describe the behaviors associated with making symptom comparisons and locating adequate condition-specific information.

Topic Modeling

OPs primarily focused on crowdsourcing information on transmission [a], symptoms, testing [b], and treatment of STDs [c]; expressing concern or fear over a potential diagnosis [d]; and seeking advice on how to speak to a partner [e]. Many OPs focused on crowdsourcing a diagnosis with posts titled such as "What is this?" and provided an image or description of their symptoms to aid commenters in diagnosis. Several OPs focused on fears of being

misdiagnosed by their healthcare provider [f] and were crowdsourcing a second opinion. These themes are expressed in the topic modeling. Twenty-three out of the 25 topics were coherent enough to identify a theme, as shown in Table 4.4.

[a] "Can someone explain [if] Herpes simples virus can be transmitted in this situation?"

[b] "HIV RNA test at 11 days, OraQuick at 43 days, both negative. Am I in the clear?"

[c] "Can you order the treatment for Chlamydia online?"

[d] "...I spend my days googling first symptoms and I think I'm starting to become hypochondriac ..."

[d] "I am worried to the point of having anxiety attacks, please tell me what you think"

[e] "Found out I have an STI, and nervous about telling my new partner. ...I know I should tell him, but I'm scared this will be the end of us. I am so ashamed :("

[f] "Misdiagnosed? Herpes or Molluscum? ...My doctor told me I had genital herpes after a visual inspection. No testing was done to verify I have HSV ..."

Topic Label	Unigrams	Example Original Post Title
Intermediate STDs	outbreak, virus, ingrown_hair, pimple	“Herpes or infected follicles?”
HPV	hpv, high_risk, wart, vaccine, cancer	”Does this look like HPV?”
Symptoms	fordyce_spots, sti, std, symptoms, blister	“Fordyce spots, ppp, or std???”
Symptoms	bumps, small, itchy, look_like, clear	“Should I be worried about these bumps?”
Crowdsourcing	help_identify, concerned, opinion, worse, idea	“Could anyone help me identify this?”
Information seeking	help, reddit, figure, doctor, scared	“Please help! Kinda scared”
HIV	hiv, risk, condom, exposure, oral_sex	“Risk of getting HIV from bleeding gums”
Chlamydia	chlamydia, treatment, tested_positive, doxycycline, azithromycin	“Can doxycycline cure chlamydia that I just got yesterday?”
HPV	wart, tyginta, acv, treatment, look_like	“Can I share Tyginta with my boyfriend even if he hasn’t got warts yet?”
Symptoms	penis, redness, shaft, dry, worry	“Bumps on Penis, Really worried”
HPV	genital_warts, look_like, terrified, tyginta, treatment	“Genital warts? 7 months pregnant - never had an std before - terrified.”
Testing	test, results, got_tested, negative, positive	“Can the results of the OraQuick test be trusted, especially if they are negative?”
Solicited sex	worried, prostitute, changes, girlfriend, sti	“I had unprotected sex with a prostitute. Should I be worried?”
Symptoms	bump, small, ingrown_hair, pimple, doesn_hurt	“Have this bump for almost 3 months now. Should I be alarmed?”
Risk	unprotected_sex, years, partners, months, tested	“Is there a place to get tested asap?”
MC	molluscum, vinegar, cured, spread, wondering	“Molluscum? Other? Help please”
Information seeking	sex, unprotected_oral, safe, worry, used_condom	“Freaking out! Had sex with a stranger.”
Crowdsourcing	got, noticed, painful, need_help, possible_std	“Help with a possible STD.”
Crowdsourcing	ideas, months, week, noticing, years_ago	“Had for over a week now...Any ideas?”
HSV	hsv, genital, oral, testing, contract	“Contracting HSV-2 indirectly through group sex?”
Symptoms	rash, red, itchy, inner_thigh, help_identify	“Single rash on pubic area.”
Crowdsourcing	opinion, thought, tested, seeing, going	“NSFW - Your opinion on what’s going on”

Table 4.4: Topics present in original posts. HPV = human papillomavirus. HSV = herpes simplex virus. MC = molluscum contagiosum.

Although the comments shared common themes with the OPs (see Table 4.5), the context of the comments differ. Comments primarily focused on sharing additional information about symptoms, testing, and treatment. If the poster was seeking a diagnosis, comments focus on providing a diagnosis. Some diagnoses by commenters sound definitive [g] compared to some that offer an opinion, but urge the poster to see a healthcare provider [h]. Some comments reassure the poster to not be anxious [i], however, not all comments are supportive of the poster's emotions [j]. Comments also focus on sharing their personal experiences of fears related to potential diagnosis of STDs as well as confirmed diagnoses of an STD. Finally, some comments share misunderstandings [k]. In contrast to the OPs, the themes in the comments are less coherent with only 16 coherent themes out of the 25 topics, as shown in Table 4.5.

> [g] "This is not an STD it's actually very common for men to have these tiny bumps on their penis. Consider it a form of genital acne that's harmless."

> [h] "It looks like Molluscum to me, but could also be warts. It's best to have a doctor look at it for confirmation . . . "

> [i] "Well, I'm no expert but based on the information you have provided here it doesn't sound like you have anything to worry about."

> [j] "It's a bump. If anything, it looks like Molluscum. And if you're tripping so much, maybe you should see a doctor instead of reassurance-seeking on Reddit."

Topic Label	Unigrams	Example Comment
HPV & MC	dermatologist, molluscum, genital_warts, warts, treat	"Those are warts for sure, you need to go to a Dermatologist."
HSV	herpes, outbreak, look_like, cold_sores, swab_test	"Looks like Herpes. Go to the doctors to get it swabbed for type confirmation."
HPV	hpv, cancer, vaccine, genital_warts, strains	"herpes and hpv are spread from skin to skin contact"
Skin irritation	yeast_infection, fungal_infection, looks_like, ingrown_hair, pimple	"Looks like an ingrown hair."
Skin irritation	fordyce_spots, papules, normal, probably, appear	"Fordyce spots, normal"
Risk	hiv, risk, stds, oral_sex, encounter	"...HIV can be transmitted though unprotected oral, anal, and vaginal sex. Bearing this in mind, the chances of transmitting HIV through oral sex is very unlikely ..."
Testing	test, weeks, months, blood_test, negative	"...a test at four weeks after last potential exposure is 95% accurate, and [for an accuracy of] 100%, you have to wait three months ..."
HSV	hsv, oral, hsv1, cold_sores, hsv2, genital	"I think if you already have hsv1, it gives you some immunity to hsv2 ..."
MC	molluscum, dermatologist, bumps, treatment, touch	"Could be Molluscum"
Testing	tested, stds, clinic, sure curable	"Be upfront, tell her what's happening with you and say for her own sake she should also get tested!"
Provider referral	doctor, wait, diagnosis, checked, clinic	"Go to the doctor"
Provider referral	checked, second_opinion, treated, looked, urologist	"You went to a dermatologist, did they not have any other idea aside from steroids? Have you sought a second opinion?"
Acute, reportable STDs	chlamydia, gonorrhea, infection, antibiotics, bacteria	"Chlamydia is extremely rare in the oral cavity ..."
HPV	appear, diagnosis, dermatologist, look_like, wart	"I thought there were no tests for HPV in men ...get the wart removed, ask your gf to get checked ...She could get the Gardasil vaccine and if the wart area is covered by a condom, use condoms until the end of the vaccine course which is I think 6 months!"
Symptoms	symptoms, anxiety, checked, develop, common	"...I think you should test, but genital focused anxiety can also make you think there is a problem. I'm not saying that you don't have something, but you haven't given any context ..."
Diagnosis	doesn_look, looks, does_look, like, good	"Doesn't look like herpes to me. Could be eczema."

Table 4.5: Topics present in comments. HPV = human papillomavirus. HSV = herpes simplex virus. MC = molluscum contagiosum.

No. of URLs	No. of Original Posts
1	344
2	68
3	31
4	6
5	4
6	2
7	1
8	1
13	1
14	1

Table 4.6: Number of online resources shared in original posts.

No. of URLs	No. of Comments
1	198
2	24
3	2
4	2
5	2
6	2
8	2
11	1

Table 4.7: Number of online resources shared in comments.

[k] "...Saliva kills the HIV virus..."

Online Resources

Twenty-five percent (25.3%; n = 457) of the OPs shared an online resource. Table 4.6 presents the number of online resources (URLs) shared per a OP totaling 669 online resources (URLs) shared among 457 OPs. Table B.1 presents the number of OPs that shared the top level domains of a URL and the type of oneline resource. The overwhelming majority of OPs that shared a URL linked to an image hosting service, where the poster referred to an image showing symptomology (93.0%; n = 425). OPs also contained links to social Q&A sites (2.4%; n = 11), informational sites (1.8%; n = 8), trusted information sites (government/university) sites (1.3%; n = 6), online STD testing (1.3% n = 6), herpes dating sites (0.7%; n = 3), information about the state of herpes vaccination (0.4%; n = 2), expert Q&A (0.4%; n = 2), professional education (0.4%; n = 2). Links to sites about a physical testing center, support group, news, online store, and alternative medicine were shared by

one OP each. Figure 4.4 presents the type of online resources shared in posts.

Approximately four percent (3.9%; n = 233) of the comments shared an online resource (URL). Table 4.7 presents the number of online resources (URLs) shared per a comment totaling 313 online resources (URLs) shared in the comments. Table B.2 presents the number of comments that shared the top level domains of an online resource (URL) and type of site. Again, the majority of comments that shared a URL linked to an image hosting service, where the comment referred to an image showing symptomology (36.5%; n = 85). Comments also contained links to informational sites (17.2%, n = 40), trusted information sites (government/university/WHO) (10.3%, n = 24), social Q&A (9.4%; n = 22), online testing (4.3%; n = 10), expert Q&A (3.4%; n = 8), online stores (2.6%; n = 6), research articles (2.1%; n = 5), health news (2.1%; n = 5), dating sites (2.1%; n = 5), health care centers or clinics (1.7%; n = 4), video hosting service (1.7%; n = 4), support sites (1.3%; n = 3), wiki (1.3%; n = 3), professional organizations/education (0.9%; n = 2), lifestyle websites (0.9%; n = 2), alternative medicine (0.4%; n = 1), and herpes vaccination (0.4%; n = 1). Figure 4.5 presents the type of online resources shared in comments. Appendix B presents the top level domains of the online resources that were shared.

4.6.3 RQ2: Content of Information Exchange (Qualitative Analysis)

We categorized the content of the conversations into seven broad categories (in order of discussion): (1) information seeking, (2) interaction with health care, (3) providing information, (4) decision making, (5) misinformation, (6) STD types, and (7) emotional content.

The broad categories as well as sub-codes are shown in Appendix C.

4.6.3.1 Information Seeking

The majority of OPs (n = 88; 97%) focused on seeking information about STDs. The type of information OPs asked about included crowdsourcing a diagnosis (n = 45; 51%), risk and transmission (n = 21; 24%), providers and testing (n = 15; 17%), treatment (n = 8; 9%), symptoms of STDs (n = 6; 7%), dating and disclosure of STD status (n = 3; 3%), asking others to share their personal experiences (n = 3; 3%), and general information about STDs (n = 2; 2%). Only one OP in subsample inquired about preventative measures (n = 1; 1%); specifically a male inquiring about how long before the HPV vaccination was effective.

Crowdsourcing a Diagnosis. The most common inquiry was related to a crowdsourced diagnosis (n = 45). For example, OP 44: "Do these look like genital warts?" and OP 76: "Whats this rash on my face?" Seventy-three percent (n = 33) of the OPs that focused on a crowdsourced diagnosis described their physical symptoms and progression of the symptoms to aid the commenters in diagnosing if they have a STD.

> "Does this look like an std? It started out small then about two weeks later it's a little bigger. No itch or burn can't even feel it's there." (OP 88)

> "Does this look like herpes? I had protected sex one month ago and today I noticed this bump around the base on my penis. It's not a blister or a sore, and it doesn't hurt. I'm planning on going to the doctor on Monday but my anxiety

is killing me right now." (OP 26)

Transmission and Risk. The second most common inquiry was related to transmission and risk of STDs (n = 21). These OPs focused on the mechanisms for infection, general probability of infection, and risk from encounters. The OPs explicitly mentioned chlamydia, gonorrhea, syphilis, HPV, HSV, nongonococcal urethritis, hepatitis, and HIV with most inquiries about transmission and risk being HIV-related (n = 8). Another common concern was transmission and risk from oral sex (n = 12).

> "For the sake of argument, just assume the female had gonorrhea. Is it possible to approximate the probability of contracting oral (pharyngeal) gonorrhea after one instance on cunnilingus (e.g. 50%, 1 in 10, etc)?" (OP 56)

> "Had unprotected oral... Just gave a urine and blood sample today, but want to know what you all think. I'm really only worried about herpes due to HSV1 (don't know my status) and HPV, as I'm aware the bacterial infections are curable and the chance that I got hep-C or HIV is infinitesimally small. From what I've gathered by reading posts on medhelp by Dr. Hunter Handsfield, the chance that I got either genital herpes or HPV (or gono or NGU, even) is exceedingly small (1 in 1000 he said)." (OP 84)

Half of the OPs (n = 11) focused on the odds of being infected with a STD or HIV after only one unprotected sexual encounter. For example, OP 10 inquired about the risk of being infected with HIV from a male-to-male oral sex encounter because he had reoccurring yeast

infections and read online that this could be a symptom of a weakened immune system compromised by HIV.

Other concerns of the OPs included: (1) the odds of transmitting high-risk HPV to a long-term partner during male-to-female oral sex because of her recent diagnosis of high-risk HPV, (2) probability of transmitting STDs from non-sexual activities including black-splash from a urinal, other's saliva touching your face during talking, and blood-to-skin contact, (3) risk of HSV transmission if the partner is taking anti-viral medication to manage HSV, and (4) general risk associated with oral and vaginal sex with sex workers (prostitute, escort).

Providers and Testing. The third most common inquiry was related to providers and testing (n = 15). Six OPs focused on if they should be tested for a STD. These OPs described a sexual encounter where either their partner disclosed their STD status (positive or negative) and inquired if they should be tested based on the information provided in the post. Four OPs focused on how, when, or where they should be tested. Some examples follow. OP 28, who has unprotected oral sex with a stranger, inquired about free testing sites because they were uninsured. OP 17 expressed concern about the anonymity of STD testing and inquired about "private" STD testing. OP 5 inquired about where rapid testing for STDs because they did not want to wait 10 days until their appointment with their general practitioner. Finally, OP 40 inquired "What is the most accurate/effective home HIV test kit I can get in the U.S.?"

Five OPs focused on interpreting a test result including defining what a "confirmatory

test" means, interpreting raw results for chlamydia and HSV tests, and the test window for HIV.

> "What are the odds of HIV getting past these tests? I had an rna pcr at 18 post exposure. A duo 4thgen at 29 days and a duo 4thgen at 43days. ... So based just on the tests all of which were negative. What's the odds that this turns positive at 3 months? ..." (OP 39)

> "Need help interpreting results...I got tested for chlamydia and the results are as follows: IgM <1:10 IgG <1:64 IgA <1:16. What does this mean?" (OP 47)

Treatment. The fourth most common inquiry was related to treatment (n = 8). OPs explicitly mentioned treatment for chlamydia, gonorrhea, HSV, and ureaplasma. Questions included the benefit of treating ureaplasma, the benefit of treating HSV, if one could receive treatment for HSV without a confirmatory test, and the effectiveness of specific medications to treat gonorrhea or manage HSV. For example, OP 25 was concerned gonorrhea could be the root cause of their ulcertative colitis and inquired if a z-pack would treat gonorrhea. OP 69 inquired if anyone had experience with "H-clr", a supplement marketed to strengthen the immune system against HSV outbreaks.

Symptomology. The fifth most common inquiry was related to symptoms of STDs (n = 6). The OPs explicitly mentioned HSV (oral and genital), syphilis, and trichomoniasis and as well as STDs in general. OPs focused on the "warning signs" of STDs, presentation of a syphilis sore, location of HSV sores during a primary infection, and the time window

between infection and presentation of HSV. The following is an excerpt from an OP which indicated the poster had received an uncertain diagnosis of HSV from their provider, described her symptoms in detail, and inquired if the symptoms were representative of HSV.

> "About a week ago I got a lump on my labia. At first it was a little painful (very similar feeling to acne on the face), and I assumed it was something like an ingrown hair. However although the pain faded, the lump remained there. . . . There is no itching whatsoever, and it's generally not painful. . . [The doctor] did a quick visual check and was like "ummm yeah probably herpes, maybe a bacterial infection, why don't you try taking medication for everything it could possibly be and I'm sure it'll go away?" He was pretty openly unsure what it was and said he did not have the equipment to do an actual herpes test, and apparently they are "not accurate anyway". What he ended up telling me to do is to take only herpes medication for a few days and if it goes away, that means I have herpes. Is this really a normal way to tell if you have herpes? . . . Does this sound like herpes? In terms of STDs, I was more worried about something like syphilis. . . " (OP 80)

Dating and Disclosure of STD Status. The sixth most common inquiry was related to symptoms of STDs (n = 3). All three OPs that mentioned dating were concerned about dating with HSV. Two OPs infected with HSV sought advice on how to disclose to a partner their STD status. One OP sought opinions on whether it is better to date someone who discloses their HSV status or someone who does not know.

> ”So it’s come to my understanding that if you’ve had more than a few partners, you’ve been easily exposed to hsv2. I’ve met a great lady who is hsv2+, disclosed, takes meds and has very few outbreaks... In this day and age in a casual sex environment. Would you feel safer having protected sex with someone who knew and took measures, or someone off a random dating site who didn’t know if she or he was +...” (OP 43)

4.6.3.2 Interaction with Health Care

One-third of the OPs (n = 30) indicate their interaction thus far with health care in their inquiry. Eleven OPs were seeking a crowdsourced diagnosis as a first opinion before seeing a provider. Six of which mention that they are currently waiting on a scheduled appointment. Most OPs were concerned about HSV or HPV, described their symptoms, and provided an image to aid the diagnosis. Only one OP was concerned about HIV and described symptoms.

> “...i have 3 different ”warts” in my intimate area. Now i am afraid that i have a high infectious diseas[e], What is your opinion about the warts? I have an appointment in 2 weeks at the doctor” (OP 61)

More than half of the OPs (n = 19) who mentioned their interaction with healthcare indicated they have already seen a provider or been tested for a STD. Seven of which disclosed a positive test result, five had a negative test result, five were waiting on test results, and two did not receive a diagnosis or testing during their interaction with a provider.

Of the OPs with positive test results, three OPs inquired about the effectiveness or necessity of treatment, two OPs sought a crowdsourced diagnosis, one OP indicated concern about other STDs, and one OP wanted information about how to move forward with their diagnosis. For example, OP 2 disclosed they were genital HSV-1 positive, thus far had only had symptoms during the primary infection, and inquired if their current skin irritation was indicative of another outbreak by providing an image of their symptoms. OP 83 inquired about the necessity of treatment for HSV-1.

> "Why bother get treatment for Herpes 1 if >50% has it? I noticed some slight skin rash and itchiness around my genitals, so I got my first STD test done through stdcheck.com recently. I learned that I have Herpes 1. Apparently ¿50% or even around 80% of the population has it, and they usually get it from childhood. So if that's the case, and I'm now in my 20s and didn't even know I had it until now, is there even any point to get any treatment for it? Is it worth paying the $95 that stdcheck.com charges for the consultation and prescription treatment?" (OP 83)

OP 37 inquired about the necessity of treatment for ureaplasma.

> "[My girlfriend has ureaplasma.] I also started doing research and found out it is not a very serious bacteria to have unless there are symptoms, to which my partner and I do not have. She started another round of antibiotics, and my doctor told me its not a big deal and does not want me to waste damaging the good bacteria with antibiotics to fight something that is not completely

understood when I have no symptoms or problems. My girlfriend is annoyed, though, that I did not insist to be treated. Has anyone had any experience with ureaplasma?" (OP 37)

Of the OPs with negative test results, three OPs sought reassurance that they indeed did not have a STD, one inquired about of the causes of nongonococcal urethritis, and the other about HIV test window periods.

The five OPs waiting on test results inquired about: (1) trichomoniasis symptoms, (2) the meaning of a "confirmatory test" (3) if they should see another provider for treatment of painful HSV sores rather than waiting on results, (4) a second opinion to diagnosis their "spots on penis" while waiting on the STD test results, and (5) a second opinion following their provider's diagnosis of folliculitis.

4.6.3.3 Providing Information

Comments primarily focus providing information in response to the OP's inquiry (n = 276; 96%). Comments focused on providing a diagnosis (n = 72; 26%), suggestions that the OP see a provider (n = 40; 14.5%), information about transmission and risk (n = 37; 13.4%), information about testing (n = 26; 9.4%), information about treatment (n = 20; 7.2%), information about symptoms of a STD (n = 15; %), information about prevention (n = 11; 5.4%), emotional support (n = 13; 4.7%), sharing a personal experience (n = 13; 4.7%), and suggestions about the type of provider to see (n = 10; 3.6%). Less than 10 comments focused on providing information about how to disclose STD status to a partner (n = 7),

combating stigma associated with a STD (n = 6), suggesting the the OP seek emotional support or may have a mental health issue (n = 4), or suggested posting on a more suitable subreddit (n = 2).

Diagnosis. The most common information provided in comments were a diagnosis in response to the OP inquiring if they have a STD. Most comments that provide a diagnosis suggest that the OP does not have a STD based on the description of their scenario (n = 41). Only three of the comments that suggest that the OP is STD-negative suggest follow-up with a healthcare provider. Below are examples of a few STD-negative diagnoses.

> "Looks okay. Just acne" (OP 16)

> "Yeah I think that's fungal or an allergy. Please go see the doctor asap, that doesn't look fun! . . ." (OP 45)

> "Best educated guess? Sebaceous glands." (OP 49)

The comment below responds to an OP's concern that they may have received a false negative HIV test result because of the test window period.

> "You are HIV Negative. After 28 days a 4th Gen test is considered conclusive. Only third generation tests are conclusive at 3 months. In real life almost everyone is positive by 6 weeks." (OP 13)

Finally, the comment below expresses support for the OP to schedule an appointment although they don't believe their symptoms are STD-related.

> "Don't feel nasty. Good job scheduling an appointment, always a good thing if you have concerns. Doesn't look like herpes to me." (OP 15)

Twenty-one comments provide a diagnosis suggesting that the OP does have a STD based on the OP's description. Eight of which suggest the OP should follow up with a healthcare provider. Below are examples of a comment in response to an OP inquiring about bumps on their genitals and an OP inquiring if they are experiencing another HSV outbreak.

> "Not herpes. Looks like molluscum. Do NOT open those, they're highly contagious. See a dermatologist to make sure." (OP 15)

> "Yes, it does look like another outbreak." (OP 2)

Ten comments did not definitely diagnosis the OP as STD-positive or negative. These comments suggest it could be a STD or a normal skin issues. Three of which suggest the OP should follow-up with a healthcare provider.

> "Some of them looks like warts, some like fordyce spots. Go to a Dermatologist." (OP 44)

Health Care Referral. The second most common information provided is suggesting to an OP that they should be seen, tested, or treated by a healthcare provider. For example, comment in response to OP 8: "Go to the clinic and get checked." Occasionally, the comment would also suggest a provider type.

> "That's almost definitely some sort of contact dermatitis. The uniformity of the "rash" indicates some sort of irritation or allergy as opposed to a sexually transmitted virus. I'd go to a walk-in clinic or some sort of non-emergency urgent care. You probably are just having an allergic reaction to something and need some sort of cortisone shot, etc." (OP 45)

Transmission and Risk. The third most commonly provided type of information focused on risk and transmission of STDs. Comments explicitly provided information about intermediate, reportable STDs (n = 26; 70%); acute, reportable STDs (n = 14; 46%); and chronic, reportable (n = 5; 14%). Again, comments could mention multiple STDs, so their focus on a particular STD group was not mutually exclusive.

The majority of comments (n = 23) focused on the general risk of acquiring a STD, including incidence and prevalence in the general population, and occasionally attempted to quantify risk. For example, the excerpt below from a comment discusses the general risk of acquiring several STD types from a single encounter involving oral sex.

> "... You can catch Gonorrhea from oral sex, but if you receive a blow job from someone who has Gonorrhea in the throat, you have a less likely chance of catching it than say vaginal or anal sex. Chlamydia is exceedingly rare in the throughly. You can't get Hepatitis from oral sex. You can't get HIV. Ureaplasma, Mycoplasma, trichomoniasis are not received from oral sex. So to say, you can pretty much get any STD from oral sex is false. On top of that, STDs, when present transmit inefficiently, especially via oral. ... The

odds of catching anything from any single exposure is quite minimal. With that, unprotected oral sex does have a risk for an STD. . . . [Oral sex] when compared to other forms of sex, [are] WAY more safe. Again not risk free, but no where near other forms of unprotected sex. . . . " (OP 28)

The excerpts below provide information about incidence or prevalence of a STD in the general population.

". . . In 2007, cervical cancer incidence in the United States was 7.9 per 100,000 women, with approximately 12,200 new cases reported. . . . About 13,000 women will be diagnosed with cervical cancer this year and it will actually kill about a third of them. . . . HPV causes about 40,000 cancers every year in the US." (OP 71)

" 1) Most people have some form of herpes. Likely you already have HSV1 or HSV2. 2) 1/6 guys in the US has HSV2, 1/4 women 3) Risk of female to male transmission over a year, without condoms, with valtrex, is 2%. that is for a YEAR of sex. pills do lower the risk by about 1/2, but they dont make it impossible." (OP 30)

Another common type of information provided was related to information about the modes of transmission for HSV, HPV, MC, gonorrhea, syphilis, and HIV (n = 14).

"In order for herpes to spread, it needs to have skin-to-skin contact with help of little friction for the cells to have microscopic cuts for the virus to easily

penetrate. Also, just because you had contact it doesn't necessarily mean that you're gonna get infected. . . ." (OP 54)

"It is more difficult usually than you touched for two seconds. Friction and micorabrasions from friction help spread stuff, but stuff can spread perfectly well without it, too. No one can really give you specific math." (OP 4)

The comment below focuses on the reduction in risk of transmitting HSV due to antiviral medication.

"There are two reasons to take antiviral medication if you have herpes: 1) It can cut down on symptoms, clearly if you have no symptoms or if they are not bothersome then you don't need to take antivirals 2) It can cut down on asymptotic shedding - when you are contagious with no symptoms - this is only relevant if you are getting intimate (kissing or performing oral are the two concerns) -having oral performed on you, or possibly genital to genital contact although transmission is rare this way - with someone who does not already have HSV 1" (OP 83)

Finally, the comment below focuses on transmission of HIV.

"Yes, HIV can be transmitted by vaginal fluid on to a broken pimple. But, the risk of that is super low. So low that it's not worth fretting about. But, the only thing you can do is wait and get tested to be sure." (OP 81)

Testing. The fourth most common type of information provided was related to testing. Approximately half of the 26 comments providing information about testing focused on the test window period for HSV and HIV. That is, the period of time between being infected and potentially receiving a false negative. The comments below share information about HSV and HIV testing in response to the OP's inquiries about HSV and HIV testing.

> "I usually read doctor Hunter Handsfield replies on a website called medhelp. He's a std expert and gives very helpful information about herpes. He also said that testing around 6-8 weeks will be around 80% accurate." (OP 30)

> " . . . After 28 days a 4th Gen test is considered conclusive. Only third generation tests are conclusive at 3 months. In real life almost everyone is positive by 6 weeks." (OP 13)

The next most common information provided about testing focused on testing for HPV. Many comments focused on explaining recommended screening methods for cervical cancer, that pap smears test for abnormal cells, and that direct testing for HPV is limited. For example, the comment below is in response to an OP who was recently diagnosed with high risk HPV and is seeking additional information about the diagnosis.

> ". . . A Pap smear is not a test for HPV. It's a test for cancer. They look at the cells scraped for your cervix and see if they are mutagenic. Again, if you're in the US, it is EXTREMELY unlikely that you were tested for HPV if you're under 30 and had a normal Pap. EXTREMELY! . . . Screening Methods for

> Average-Risk Asymptomatic Women. Age 21 to 29: Every 3 years with cytology (Pap testing), regardless of age of onset of sexual activity or other risk factors. Age 30 to 65: Every 5 years with HPV co-test (Pap + HPV test) OR every 3 years with cytology." (OP 71)

Several comments focused on where to purchase home HIV testing, online testing for STDs, and where to be seen for general STD testing (urgent care, walk-in clinics).

Finally, two comments focused on providing information about how to be tested for acute, reportable STDs. For example, the comment below is in response to an OP inquiring if z-packs (prescribed for an unrelated reason) will treat any chlamydia, gonorrhea, or syphilis.

> "Swab for gonorrhea/chlamydia. An infectious disease doctor or std clinic does that sort of thing all the time. No worries. Blood test for syphilis." (OP 25)

Treatment. The fifth most common type of information provided was related to treatment. Most comments focused on treatment of HSV and HPV. Comments related to HSV focused on the topical medications (lidocaine) to alleviate pain associated with herpetic lesions and antiviral medications to manage symptoms and reduce risk of transmission. Comments related to HPV focused on removal of visible genital warts and refraining from sexual activity until the skin healed. A few comments focused on specific antibiotic medications for acute, reportable STDs. For example in response to OP 58, which inquired about the

risk of chlamydia from a sexual encounter and treatment options,

> "We can test for chlamydia/gonorrhea via rectal, endocervical, throat, and urethral swabs and also via urine so if you performed oral or had anal intercourse, be honest so they can retrieve the proper samples. Treatment is super painless and they may even offer it to you before the labs come back. Lots of times it's just one pill (1 gram azithromycin) but there are alternatives if your allergic. Chlamydia can be sort of embarrassing, but be glad he told you about it so you can knock it out quickly and it won't cause long lasting issues. I'd just head to a PP clinic or STD clinic in your area and request a CT/GC workup. If it's cheap, you may as well go for a whole STD panel just for the peace of mind." (OP 58)

Symptoms. The sixth most common type of information provided was related to symptoms. Comments discussed when and where herpatic lesions appear during a primary infection, appearance of HPV (genital warts), how HPV may result in abnormal changes to cervical cells potentially leading to cervical cancer, and that gonorrhea and nongonococcal urethritis can present as asymptomatic. For example, comments in response to an OP inquiring about the presentation of HPV share how symptoms of HSV may present.

> "1) technically they can appear anywhere in the boxer region, but they often present where the infection entered. 2) Sometimes when experiencing a primary or subsequent outbreak flu like symptoms are part of it, swollen lymph

> nodes, body aches, etc. But most people who have the disease have no symptoms or symptoms so mild they are unnoticed or confused with something else. 3) without treatment they [c]an heal, yes, because your body has an immune response and will fight the virus and cause it to become dormant for a time. . . . " (OP 20)

Prevention. Only eleven comments focused on prevention of STDs making this the seventh most common type of information provided. All except one comment was related to HPV; the other comment was related to HSV. Many commenters suggested the HPV vaccination for prevention of HPV. For example, comment to OP 71, "The point I'm trying to make is that HPV is a major kiler worldwide [due to cervical cancer] and we were saved from it by the Pap smear until the vaccine came along in 2006! (OP 71)" The comment also suggested using condoms for prevention of HPV, however, they do not caveat that HPV can still be transmitted with condom usage. Finally, one comment focused on how to prevent transmission of HSV to partners

> "just tell your partners beforehand; take acyclovir when you have symptoms, and avoid sex when you have outbreaks . . . " (OP 54)

Sharing Personal Experience. The eighth most common type of information provided was related to sharing a personal experience by self-disclosing their experience with a specific STD or similar symptoms that turned out to not be a STD. For example, the comment below focuses on sharing their personal experience with cervical cancer in reference to an OP seeking information about the HPV vaccination.

> "...If you never got the vaccine, you're still at significant risk and shouldn't miss a Pap. About 13,000 women will be diagnosed with cervical cancer this year and it will actually kill about a third of them. If you've never met one of these women, Hi! ;-) ..." (OP 71)

Emotional Support. Tying for eighth most common time type of information provided is emotional support. Most comments do not explicitly address the emotional component of the OPs' concerns. However, thirteen comments do directly make suggestions and provide emotional support. Comments employ several tactics including reassurance and suggests the OP reduce their stress or "relax." Offers of emotional support are not limited to concerns over a specific STD, but instead are related to OPs about HSV, HPV, and HIV. For example, below are two comments to an OP providing emotional support to someone panicking over the possibility of HIV from a sexual encounter and an OP who is concerned about having HSV because it is "nasty."

> "...Try to distract yourself from these thoughts. 222 am is not a time that's amenable to reasonable consideration of circumstances, and a reasonable consideration of circumstances would lead to the conclusion that you are fine. Things will look better in the morning." (OP 10)

> "Don't feel nasty. Good job scheduling an appointment, always a good thing if you have concerns. Doesn't look like herpes to me." (OP 15)

Disclosure. Of the few comments that discussed disclosure (n = 7), most suggested that

you should disclose your STD status, specifically for HPV and HSV, to a partner. Below are comments that support disclosing status.

> "Yes, you should tell people that you have been exposed to HPV in the past, so that they could get the vaccine. Also you should use a condom every time." (OP 32)

> "just tell your partners beforehand; take acyclovir when you have symptoms, and avoid sex when you have outbreaks. Herpes is just like common cold, but causes pimples on your genital or wherever you have it. It's all good :D FACT: Anywhere from 65-90% of the adult population has oral herpes and about 1 in 5 people over 12 have genital herpes." (OP 54)

The one comment that did not support disclosing based their argument on the prevalence of HPV.

> "IMO it's pretty stupid to tell to your new partners that you have HPV, i mean, everyone has it. My suggestion is that if you have a visible wart, you need to remove it, don't have sex till the warts are gone and the skin is healed." (OP 32)

Combating Stigma. Six comments focused on combating stigma against HSV. Commenters do so by mentioning statistics about the prevalence of HSV, downplaying the

symptoms of HSV, or disclosing their experience with HSV. Note all comments that focused on combating stigma of STDs were related to HSV. Comments to OP 54, OP 54, and OP 15 that focus on sharing statistics about the prevalence of HSV:

> "Genital herpes is caused by a herpes simplex virus (type 1 or type 2). Nearly everyone (70%) will catch at least one type, sooner or later. No biggie man. You should seriously chillaz tho, unless you wanna get unnecessary diseases that are cuased by stress and hurts your body more than herpes, which isnt a life threatening condition(aka painful pimples on your face or pubic area, or wherever)! . . . " (OP 54)

> ". . . FACT: Anywhere from 65-90% of the adult population has oral herpes and about 1 in 5 people over 12 have genital herpes." (OP 54)

> "This is about the 100th post I've come across describing the anxiety of having herpes as being "nasty", "disgusting", "gross"; it's honestly a bit of a piss off. I have herpes, many people on this site have herpes- we're not gross, it's a common virus that millions of people have, and there is nothing "nasty" about it." (OP 15)

Comments to OP 15 and OP 54 that focus on downplaying the symptoms of HSV follow.

> "Of course herpes is not fun, but it's not, in my opinion, gross or nasty. But

it is pretty benign. I mean besides having a few blisters a couple times a year what else does it really do? . . . ”

“ . . . Herpes is just like common cold, but causes pimples on your genital or wherever you have it.”

4.6.3.4 Decision Making

Only 13 OPs share their decisions about their health behavior in a follow-up comment. Seven OPs indicated that they intended to see a provider in a follow-up comment based on feedback from the commenters. Three of which indicated that they do not think it is an STD, but they intend to see a provider nonetheless. Five OPs indicated that they did see a provider and were not diagnosed with an STD. One OP indicated that they would not see a provider because of financial limitations.

4.6.3.5 Misinformation

Seventeen comments (6%) contained misleading information or misinformation. For this analysis, misdiagnosis was not considered misinformation. Comments presenting misleading or misinformation included recommendations for alternative treatment, and incorrect information about transmission, testing, and symptoms.

Seven comments in three threads contained misinformation about HPV. In response to OP 80 asking for information about why the HPV vaccination consists of three shots,

one comment implied that condoms would protect against HPV (`. . . use a condom [and] condom sense . . . ") and another comment suggested the side effects, including infertility, are worse than the benefits of the HPV vaccination. The comment did not caveat that condoms may not sufficiently protect against HPV (misleading). One comment in the thread focused on correcting the misinformation presented about the side effects of the HPV vaccination.

> ". . . has [been] through peer reviewed clinical trials and has shown to be effective against certain HPV strains as well as causing little to no long-term side effects . . . " (OP 80)

OP 71 focused on her recent diagnosis of a high-risk strain of HPV and inquired about how to prevent her male partner from being infected. Four comments presented misleading or misinformation about HPV. One comment stated `It's HPV, no symptoms - no problem. Do whatever you feel like doing." Another comment suggested to not be concerned because " . . . you can develop cancer with other things way worse than having sex." Other comments stated there was no way to test for HPV and that it is a temporary condition without acknowledging that HPV does not clear for all individuals. Three comments focused on correcting the misinformation presented in thread. For example, one commenter checked the history of comments left by the individual who stated "no symptoms - no problem" and identified that the individual frequently posts this and proceeded to correct information presented by other commenters. Finally, OP 90 sought a crowdsourced diagnosis for their symptoms. The OP was subsequently diagnosed with genital warts by the commenters.

One comment contained misinformation.

> "HPV has no way of being diagnosed and is usually asymptomatic. ... With this disease you're at risk for some cancers but like I said it is mostly asymptomatic and you will have little to no problems with it."

Four comments contained common misperceptions about HSV including that HSV-1 is limited to oral outbreaks (i.e., not genital) (OP 83, OP 28), HSV-1 always results in noticeable symptoms during the primary infection (OP 54), and HSV does not remain dormant for long periods of time (OP 54).

Three comments contained misinformation about transmission of STDs during oral sex. These comments ranged from suggesting that the OP not be concerned about risk of STD transmission during oral sex (OP 28) to suggesting no HIV risk resulting from oral sex (OP 28) to reassuring an OP that recently tested positive for a high-risk HPV strain that HPV cannot be transmitted to her male partner during oral sex (OP 71). One comment corrected the misinformation presented in the OP 28 thread stating "Some very bad advice in this thread. Unprotected oral sex is NOT "safe sex.".

Two comments contained misinformation about test window periods. In response to OP 9, which was concerned about the potential of an STD because of newly formed blisters on the male poster's genital area and sensitive lymph nodes after unprotected sex with a new partner, a comment responded that it takes 3 months for a conclusive HSV test. Although it can take up to several months for antibodies to develop, the comment doesn't indicate that the window period may be shorter based on each individual (misleading). In response to OP

13 seeking information about whether a negative fourth generation HIV antibody/antigen test at 74 days from exposure could be considered a true negative, a comment stated that you must wait 96 days for a conclusive result. The window period for fourth generation antibody/antigen test is one month.

Finally, one comment suggested "Chlamydia/gonorrhea can be cured by herbal medicine called diuretic and anti-inflammatory pill http://www.diureticspill.com/ efficiently." in response to an OP 9's concerns about transmission of chlamydia and gonorrhea.

In general, misinformation was limited in the content reviewed during qualitative analysis, however, it seems particularly prevalent in discussions about transmission of STDs during oral sex, test window periods, and general information about HSV and HPV. Sometimes the information was contested by other commenters who jumped in the conversation to correct misinformation, however this did not always occur.

4.6.3.6 STD Types

OPs frequently mentioned specific STDs (n = 70; 77%) that they were either concerned about, had previously been tested for, or were considering being tested for. Referrals to specific STDs, in order of decreasing prevalence, included non-reportable STDs (n = 37; 53%), acute, reportable STDs (n = 16; 23%), and chronic, reportable STDs (n = 13; 19%). Several OPs referred to other STDs including nongonococcal urethritis (n = 1), scabies (n = 1), trichomoniasis (n = 1), and ureaplasma (n = 1).

Acute Reportable STDs. Eight OPs referred to chlamydia, five OPs referred to gonorrhea,

and three OPs referred to syphilis. Concerns about chlamydia included how to interpret raw test results, uncertainty if treatment was effective, risk of reinfection while partner is being treated, and risk of infection from unprotected sex (anal, oral, vaginal). Concerns about gonorrhea included risk of infection from oral sex, uncertainty if treatment was effective, a crowdsourced diagnosis, and inquiring about types of medications for treatment. OPs that mentioned syphilis were seeking a crowdsourced diagnosis. OPs frequently described their symptoms in an attempt to aid the commenters in their response.

Intermediate Non-Reportable STDs. Of the intermediate, non-reportable STDs, nineteen OPs mentioned HSV, fifteen OPs mentioned HPV, and two OPs mentioned molluscum contagiosum.

By and far, when OPs mentioned a word related to HSV, it was was because they were crowdsourcing a diagnosis if their symptoms were indicative of HSV. Other concerns included the benefits of treatment, if a confirmatory test is necessary to obtain treatment, how to interpret raw test results, how to manage depression related to a diagnosis, dating, and risk and transmission from protected and unprotected sex (anal, oral, vaginal), mouth-to-mouth contact, and accidental contact with bodily fluids.

Similar to HSV, when OPs mentioned a word related to HPV, it was because they were crowdsourcing a diagnosis if their symptoms were indicative of genital warts. Other concerns included risk from oral sex, how to prevent transmission of a high-risk HPV strain, and whether one should disclose to future partners that a previous partner was infected with a high-risk strain of HPV. Only one OP focused on the HPV vaccination inquiring why it

is necessary to do the entire series.

OPs that explicitly mentioned molluscum contagiosum were crowdsourcing a diagnosis by providing a description and image of their symptoms.

Chronic Reportable STDs. Of the chronic reportable STDs, eleven OPs mentioned HIV and two OPs mentioned hepatitis. OPs that explicitly referred to HIV were primarily concerned about transmission, risk, and testing. OPs sought information about transmission of HIV from blood-to-skin contact during protected sex and bodily fluids encountered during non-sexual contact and oral sex. Information seeking for HIV testing focused on the window period during which an individual is infected but tests negative and reliability of home HIV testing. Several OPs explicitly mentioned they were more concerned about STDs and not HIV because did not seek information the risk of being infected by HIV is "infinitesimally small." Most users explicitly expressed anxiety when explicitly discussing HIV. These OPs primarily reflected fear rather than a rational concern about contracting HIV.

> "Now, my anxiety is kicking in. ASSUMING she's HIV positive, am I at great risk of getting it? I do not have any open wounds or sores. Also, I was wearing a condom throughout the whole thing." (OP 12)

4.6.3.7 Emotional Content

Twenty-two (26%) OPs contained content related to emotions. All of these OPs expressed anxiety with one OP also expressing depression. Some of the language used in the OPs characteristic of emotional content included words such as "worry", "concern", "beside

myself", "paranoid", "depressed", "scared", "afraid", "panic", "anxiety", "anxious", and "freak out." Most OPs reflecting emotional content also contained a mention of HIV (n = 10) followed by HSV (n = 9), and HPV (n = 4).

Four OPs were anxious about the reliability of their negative test results, especially about test window periods, for HIV, HSV, and STDs in general. For example, two OPs with negative HIV test results were concerned if they were outside of the test window period. Another example is OP 67, who tested twice for HSV IgG-class antibodies and, despite having no visual symptoms, was "paranoid" because his raw results had varied between the two tests. The OP inquired if he should pursue a Western blot test to further validate his negative test results. Finally, OP 77 had symptoms including"spots on penis," sore throat, and swollen lymph nodes. Although he and his partner recently tested negative for STDs, he wanted to know if anyone had a similar experience as he was "fairly concerned and decided to get tested this week to clarify things."

Two OPs with descriptions of symptoms similar to the common cold expressed concern about being infected with oral chlamydia or gonorrhea. For example, OP 9 was "so beside [him]self at the moment because he was "convinced that he definitely had something." OP 56 acknowledged his symptoms could be due to a common cold and he may be a "worried well," but would appreciate any information nonetheless .

Four OPs were anxious about high risk encounters such as unprotected sex, casual sex with strangers, and with a commercial sex worker. For example, OP 82 expressed concerned because they had their first oral sex encounter with a partner who had a rash,

which the partner informed were mosquito bites.

> "Please help me I had my first ever hook up and now that it's over I'm so paranoid. . . . We did oral for each other while I had a slight sore on my cheek. Will I be safe? Do you think I will contract aids/std from this person? Does this symptom correlate with aids or any stds? Better to [be] safe than sorry since what's done is done." (OP 82)

OP 87 shared their concern about an encounter with a sex worker.

> ". . . I gave oral sex to a prostitute before we got into the other stuff. She was on the higher end of the spectrum so she seemed clean, but I'm worrying nonetheless. I know HIV is almost impossible to get from oral, so I'm not too worried about that. I am worried about the other stuff though. How likely is it to get STDs through oral sex? And what are the warning signs I should look out for in the next few weeks?" (OP 87)

Two OPs were anxious because their partner disclosed their STD status or a new STD-related infection. OP 30 described having unprotected sex with a partner who retrospectively disclosed she was HSV-positive followed by unprotected sex with multiple partners afterwards. The OP inquired if the antiviral medications his partner was taking during their encounter would prevent transmission.

> "I may have contracted herpes - going to the doctor's tomorrow. Am I being paranoid? So I had sex with this girl who insisted on not using a condom

> . . . [She] disclosed that it was herpes, and that she has had it for a few years. She did not tell me before, during, or after (well yeah a month after) the intercourse. She had no visible signs of herpes, and she says that she doesn't infect others while she is under the pill treatment, but I am still scared as fuck. Can anyone confirm/deny some of this?" (OP 30)

OP 54 expressed extreme distress about cold sores.

> "Feeling depressed. . . . I know this is rambling, but how do you move forward? I'm basically thinking I can't date anymore. I'm now so paranoid about everything that I can't even think about sex. I'm afraid I'd give it to someone else. At least I've lost 12 pounds because I haven't been able to sleep for two weeks."

Three OPs were anxious about the possibility of transmission from unprotected oral sex encounters. For example, OP 10 was concerned that he was infected with HIV from an unprotected oral sex encounter with another male. The OP described uncontrollable anxiety that kept him up at night.

> "major HIV panic attack middle of the night . . . I know the odds are very small but I am panicking so much now I don't know what to do. I will call my doctor tomorrow morning but it's 2:22 AM right now and I cannot bear to last through the night like this. I don't know why it's so bad all of a sudden but everything I read seems to point towards the fact that getting hiv from oral sex is not

> impossible, and that my symptoms are hiv symptoms. ...I don't even know what I expect from this but the anxiety is unbearable. I cannot do anything to distract myself or fall asleep." (OP 10)

Three OPs described hypervigilence or hyperinformation seeking. For example, OP 27 inquired about the risk of being infected by HSV from someone's saliva during non-sexual interaction (specifically someone's saliva landing their lip while they are speaking). The OP disclosed that they have obsessive-compulsive disorder and that "these sort of things ruin me for days on end." OP 35 described themselves as an "anxious person" and shared that they had seen a doctor who diagnosed their symptoms as "idiopathic." The OP found it beneficial to write and share their "health anxieties" on the subreddit.

Four OPs mentioned emotional distress over symptoms that seem legitimately STD-related. For example, "I'm freaking out sos" (OP 36) and "my anxiety is killing me now" (OP 26).

4.6.4 Comparison of Qualitative and Quantitative Approaches

The themes uncovered by computational approaches (topic modeling) and qualitative approaches (open qualitative coding) are difficult to compare on a one-to-one basis because of the granularity of detail that each method provides. In general, qualitative analysis allows for a richer level of detail and deeper understanding of context than topic modeling, which identifies main themes in text. However, for comparative purposes, we compared the topics found using topic modeling to the broader themes found during qualitative analysis in

an attempt to quantify the percent overlap. Additionally, we describe how the topics align into the broader themes from qualitative analysis. We discuss reasons why discrepancies between the qualitative analysis and computational approaches may have occurred.

The results of the computational approach are largely consistent the results from qualitative analysis with 71% agreement in broader themes. Topics from the computational analysis overlapped with five out of the seven topics identified during qualitative analysis: emotional content, health care interaction, information seeking, providing information, and STDs. The chlamydia, HIV, HPV, HSV, intermediate STD, HPV, HIV, and molluscum contagiosum topics in the OPs and acute STD, HPV, HSV, intermediate STD, and molluscum contagiosum topics in the comments align with the the theme of STDs. The crowdsourcing, information seeking, sex with sex workers, symptoms, risk, and testing topics in the OPs align with the theme of information seeking. The diagnosis, provider referral, risk, skin irritation, symptoms, and testing topics in the comments align with the theme of providing information. Finally, there are elements of the health care interaction and emotional content themes in the crowdsourcing, information seeking, and testing topics (e.g., the words 'concerned', 'doctor', 'got_tested', 'results', and 'scared').

The two themes that were not found during topic modeling were decision making and misinformation. We hypothesize this is because (a) OPs rarely follow-up with their decision making and (b) identifying misinformation is difficult because identification of incorrect information, at least in this context, requires a human-in-the-loop.

While the computational approach gives a high-level idea of the r/STD community,

qualitative analysis allows for a more nuanced view of the behaviors of the community. Future research should explore whether identification of health misinformation is computationally possible. One potential path forward could be to use semantic textual similarity to compare layperson text with text from an authoritative source (such as the CDC). A limitation to this approach is the contrast between layperson and expert language.

4.6.5 RQ3: Psycholinguistic Properties

Overall, OPs primarily focused on cognitive processes (M = 13.17%, 95% CI [12.83, 13.52]) using present focus (M = 11.59%, 95% CI [11.30, 11.90] and describing biological processes (M = 8.98%, 95% CI [8.55, 9.40]. We suspect this reflects OPs' focus on presenting symptomology and crowdsourcing information about transmission and diagnosis. OPs commonly made references to a specific or relative time (M = 6.34%, 95% CI [6.12, 6.56]). There was less emphasis on past (M = 5.34%, 95% CI [5.15, 5.52]) and future focus (M = 0.87%, 95% CI [0.78, 0.96]) than present focus in the OPs. The affect, or emotional tone, was surprisingly low for the OPs (M = 4.31, 95% CI [4.09, 4.53]) despite one of the community's stated objectives being to "calm the anxiety" associated with a potential STD. However, LIWC is a word count software and can only measure directly observable emotional content by counting word frequency using a lexicon and is unable to decipher implied emotional tone. For example, the following would receive a score of 4.17% for affect although it is obvious to the human reader that the person is experiencing severe distress.

"... I then got obsessed into the world of STDs and have studied (literally spent probably days at this point) researching everything I could. I could talk about why the fuck no one (particularly school) has any ever told me about window periods, incubation periods, chances of transmission, etc. ..."

4.7 Discussion

r/STD is primarily a transient community, where most OPs make a one-time post to elicit information. The most common type of information OPs seek is a crowdsourced diagnosis. OPs describe their symptoms and often provide an image to aid commenters in diagnosing if they potentially have a STD. By and far, the OPs are mostly concerned about skin-to-skin infections such as HPV, HSV, and molluscum contagiosum, the most prevalent STDs in the U.S. OPs range in their interactions with health care coming to r/STD as either a first or second opinion.

In comparison to crowdsourcing a diagnosis, inquiring about other types of information is limited. Most OPs seek information about the risk, transmission, and presentation of HPV and HSV. Another common pattern is the uncertainty associated with the risks of oral sex (especially for chlamydia, gonorrhea, and HIV) and test window periods (especially for HIV). Comments focus on providing information in response to the OP and occasionally share online resources. The most common online resource is from a trusted information source followed by an advocacy or support organization (at the surface-level these organizations do not appear to be providing misinformation). Only a handful of re-

search articles are shared. We hypothesize this is because (a) most are not open access and (b) the language used is beyond layperson terminology. Most OPs and comments rarely mention preventative efforts. In general, misinformation was limited in the content reviewed during qualitative analysis and online resources that were shared. However, as mentioned previously, there is confusion about transmission of STDs from oral sex, test window periods, and general information about HSV and HPV.

Contrary to the stated mission to help "calm anxieties" most OPs do not explicitly ask for emotional support, however, it is clear from the qualitative analysis that many OPs experience anxiety related to concerns about risk and transmission from high risk encounters and symptoms that the OP views as potentially STD-related. OPs are also anxious regarding interpretation of test results ranging from interpreting raw test results to trusting negative results are not false negatives. Finally, several OPs in the qualitative analysis were seeking information at a level that interfered with their daily activities.

4.7.1 Health Information Seeking Models

This study used an atheoretical, bottom-up approach to examine the recent activity and content on r/STD. Rather than base our analysis on pre-conceived beliefs about the content, we used a data-driven approach to identify major themes discussed on r/STD. We compare our results of the data-driven approach to the health information seeking models presented in Chapter 2, which were primarily built to analyze individual-level survey or interview data with structured questions allowing for testing of formal hypotheses or elici-

tation of evidence to support working hypotheses. More specifically, below we discuss (1) the limitations of the RISP (see Figure 2.3), HIAM (see Figure 2.1), ECMISB (see Figure 2.2) for the passively collected data from r/STD, (2) how the CMIS (see Figure 2.4) could potentially be operationalized for future studies with r/STD or similar data from online communities, and (3) the suitability of current health information seeking models for r/STD or similar data from online communities.

Key components (individual characteristics, individual beliefs, and perceived information gathering capacity) related to an individual's health information search according to the RISP would be difficult to operationalize using passively collected data from r/STD. Similarly, the HIAM focuses on a sequential search process, which is difficult to elicit from the passively collected r/STD data. The sequential process could potentially be operationalized by examining an OP's posting history. However, there are a number of serious limitations associated with that including usage of accounts created specifically for anonymity and a low follow-up rate of response according to our qualitative analysis. Operationalizing the HIAM model would be more suitable for other types of passively collected social data (e.g., individual-level, temporal search engine query data). Finally, the ECMISB integrates contextual and personal demographics into the model, which are difficult to elicit from passively collected Reddit data. OPs are often one-time posters and we can't infer the contextual variables that affected their information search unless they explicitly share them. Additionally, unlike other forms of social media, Reddit user names are often not distinguishable for gender references and "profiles" are only a legacy page of all posts (i.e., they do not necessarily contain information about the user). Operationlizing the ECMISB may

be better suited for survey or interview data, where questions could elicit the contextual and personal demographics.

According to the CMIS, information seeking actions are the active, purposeful search for information. Although we cannot determine the process an OP takes in their information search, we can operationalize the purpose of seeking information leveraging the computational and qualitative approaches in this study. For example, topic modeling combined with regular expressions of STD mentions can identify the types of information sought by the OP. As discussed earlier, the gold-standard for the passively collected r/STD is human annotation of the types of information the OP seeks because it allows for a richer level of detail. In the subsample for qualitative analysis, 97% of OPs were seeking information, which is in accordance with the qualitative review done during the labeling of topics that the topic model uncovered.

Antecedents, as identified by the CMIS, could potentially be determined using qualitative analysis, but are difficult to capture using a computational approach because of the nuance of language. Here, we focus on how these could potentially be identified or operationalized during a qualitative analysis. Demographics could be captured if an OP explicitly mentions their identifying information. Occasionally, OPs mention their gender, however, most OPs do not. Gender was often implied by the image or description of their symptoms (e.g., bumps on penis) that the OP provided. Experience is the proximity of the OP to disease. We found evidence during our qualitative analysis that most OPs are in the third and fourth stage of proximity of disease, where they are either confronted with the possibility of being infected with a STD (third) or have received a confirmatory diagnosis of a STD

(fourth). Evidence of salience or perceived risk is often obvious because the OP states the relevance of their concern(s). OPs occasionally identify their beliefs when discussing their perceived severity, social risk, or level of control. The fact that these characteristics are difficult to capture computationally is logical given that these models were designed to primarily be used with survey or interview data with structured questions to elicit an individual's characteristics and behavior.

Another input into information seeking actions, as identified by the CMIS, is the characteristics of the information carrier. Unfortunately, there is not a way to identify this for each user based on the text data itself. However, r/STD offers a number of attractive features including accessibility, simulating face-to-face interactions with others without social risk, and near real-time information.

By characterizing r/STD, we find evidence of Kuhlthau's "uncertainty principle", which posits that uncertainty increases anxiety and reduces confidence [25]. It is clear and implied in the OPs that uncertainty results in increased anxiety and contributes to their search for information as a problem- and emotion-focused coping strategy. We also observed that the commenters engage with OPs by providing information, diagnoses, support, and sharing online resources similar to online support communities. Ultimately, acquiring new information from r/STD may modify an individual's beliefs, attitudes, and decisions to seek treatment, inform about risk, and provide resources to support behavioral changes [24]. Occasionally, the OP provides a follow-up where they discuss the behavioral implications of the comments and, in these cases, the CMIS could be thought of as an input into the HBM.

Computational analysis can identify a high-level view of the content of information seeking actions. However, current health information seeking models were primarily developed for analysis of individual-level data obtained from surveys, in-depth interviews, or focus groups. In comparison to health information seeking field, infodemiology is a relatively new field. Health information seeking models that account for passively collected data have yet to be developed, but development of such models, specific to the today's rapidly changing information environment, would be beneficial for comparison across platforms.

4.7.2 Research and Practical Implications

As the Institute of Medicine stated and the CDC reiterated in their most recent surveillance report, STDs remain "hidden epidemics of tremendous health and economic consequence in the United States [177, 175]." Despite this, resources for STDs are dwindling requiring innovative strategies for public outreach [175]. A key contribution of this study is characterizing recent online health information seeking and needs of the r/STD community, a previously unstudied online community. Although the number of individuals in our study is limited, the affected audience is likely much larger because of passive viewing of content. Our findings indicate the prevalence of sexual health information crowdsourcing on the r/STD social media platform and highlights the potential to aid in targeted health communication.

STDs carry differing experiences in symptomology, fear, uncertainty, stigma, and per-

ceptions of risk [212]. These various risk perceptions likely contribute to intention of accessing the health care systems for subsequent testing and treatment [213]. For those who do not have routine access, the uncertainty is likely heightened [213]. The ability to access 'lay health expertise' through unofficial online sources may contribute to the decision to seek care and is an area of needed future study, particularly as it relates to credibility of these sources.

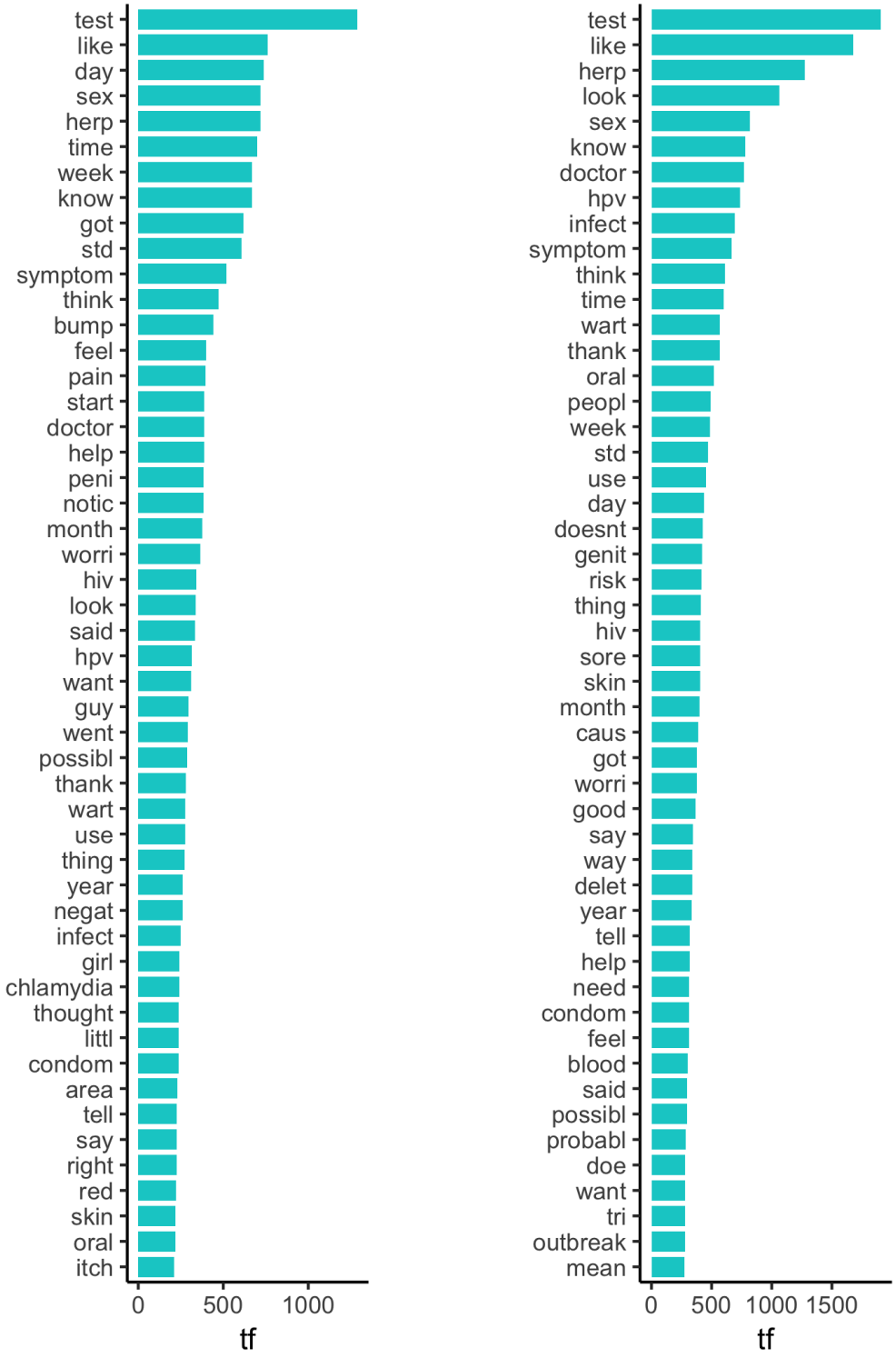
It is imperative for researchers and public health officials to understand the cultural norms and information needs of an audience in order to create effective, tailored health communication specific to an online community. Beyond tailoring health communication, monitoring of social media may be an effective, innovative strategy to complement and modernize traditional surveillance methods that rely on individuals accessing health care or participating in studies [175, 56]. This information can be valuable for health communication efforts to correct misunderstandings about transmission of STDs, combat stigmatization, encourage appropriate testing and disclosure, and supply credible resources. For example, r/STD would most likely be interested in information about risk and transmission during oral sex, risk and transmission of HPV and HSV, and test window periods for various STDs. This information could be used to create engaging infographics, since they are focused on defining many of these concerns numerical, that would inform this community about some of their top concerns [214]. Furthermore, monitoring may inform public health officials of trending concerns and information about emerging STDs such as mycoplasma genitalium.

Future research needs to take into consideration the immense power of online commu-

nities in influencing health seeking behavior and wellness, especially in vulnerable populations, and temporal changes in the public's information needs. According to Flanders et al. (2017) [188], "...it is important to understand the different needs of different groups of people in relation to online sexual health information to better tailor online resources."

4.7.3 Limitations

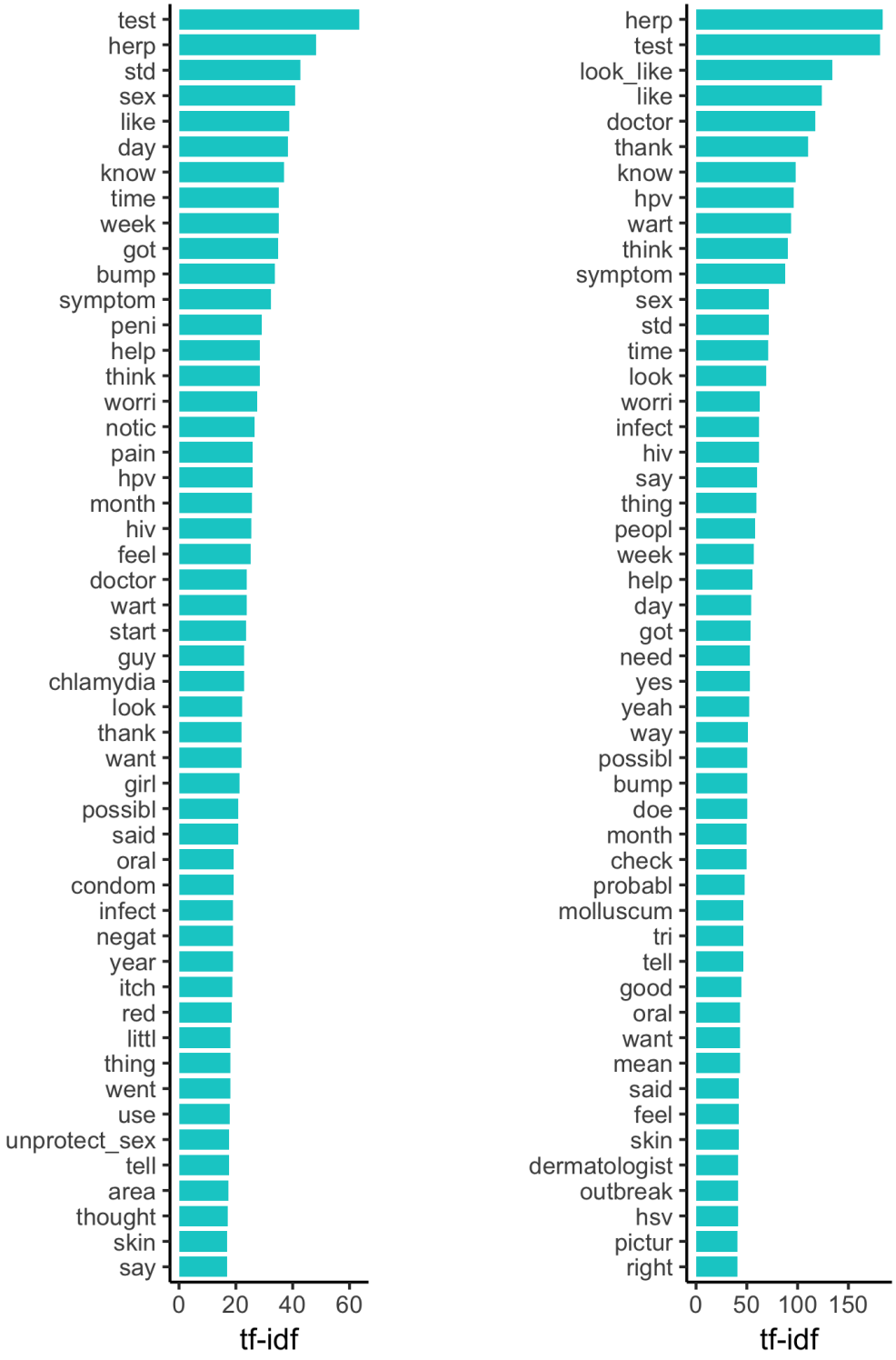
Since users of r/STD do not need to identify demographic information, the true demographic is unknown and we caution in making inferences beyond this community. We acknowledge that there may be other subreddits where sexual health is discussed. We did not account for phonetic spellings or slang terms. This could impact our analysis, however, this was not prevalent in the aforementioned analyses, so we suspect this has a negligible effect on the overall interpretation. Finally, qualitative labeling of the topics from the topic model was conducted by one person, therefore a measure of inter-rater reliability is not available.



(a) Original Posts

(b) Comments

Figure 4.2: Term frequency (tf) of original posts and comments.



(a) Original Posts

(b) Comments

Figure 4.3: Term frequency-inverse document frequency (tf-idf) of original posts and comments.

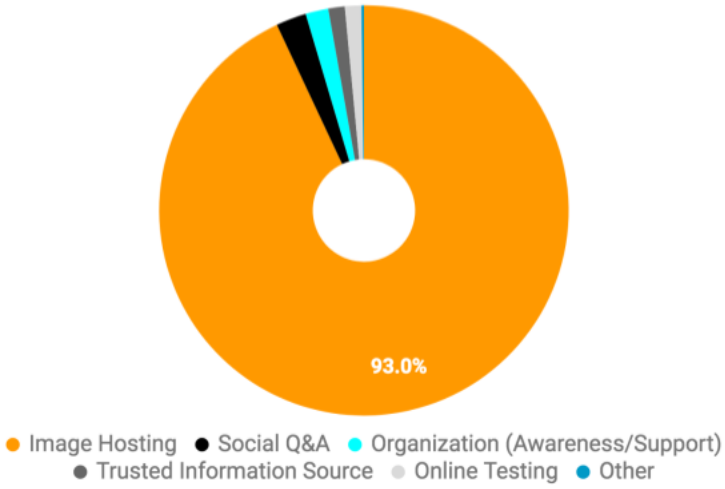


Figure 4.4: Types of online resources shared in posts.

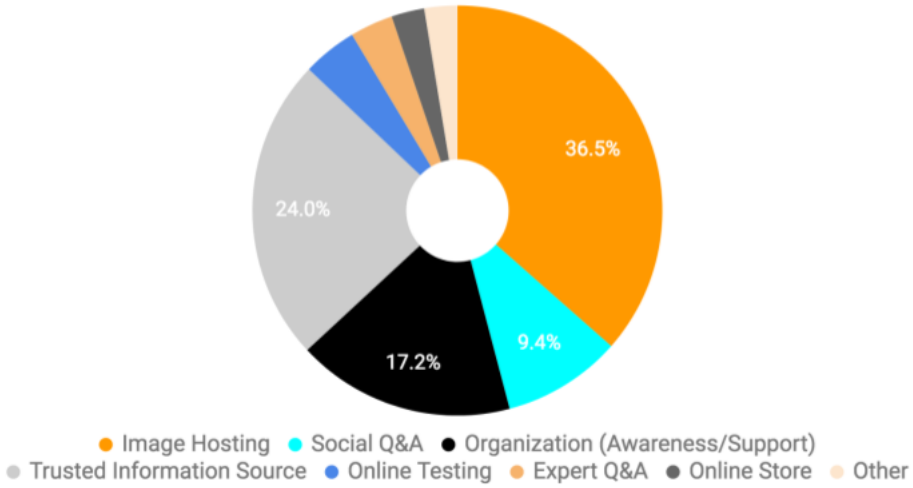


Figure 4.5: Types of online resources shared in comments.

CHAPTER 5

GENERAL DISCUSSION AND CONCLUSION

In the final chapter of this dissertation, I summarize previous chapters and contributions, discuss relevant implications, and acknowledge the limitations of the approaches used in this dissertation. Additionally, I discuss opportunities for future work to build on this dissertation and outline several fertile areas for future research.

5.1 Summary of Chapters and Contributions

In **Chapter 1**, I set the stage for this dissertation. Motivated by the need for data-driven health communication, I explained how infodemiology could be a potential strategy for supporting tailored health communication. Traditionally, health communication used a top-down approach to determine what information is most beneficial to the consumer. By leveraging the digital footprint left behind from online information seeking, the public can be passively involved in the development and dissemination of health communication. An additional benefit of such an approach is the reduced resources needed to examine the abundance of data generated by online interactions in near real-time, which can supplement more traditional methods of tailoring health communication that are often resource-intensive and lag in time.

In **Chapter 2**, I first provide a general backdrop for online information seeking followed by identifying existing models of health information seeking from a review of health behavior, health information seeking, and human information interaction literature. It is clear from the review that the existing health information seeking models were developed primarily based on historical information carriers, such as providers or newsprint, using individual-level data. I reflect upon how these models could be operationalized for the case studies and the need for development or expansion of models that capture the contemporary information field in the discussion of the two case studies in Chapters 3 and 4. Finally, I describe the state of the field of infodemiology and it's application to infectious diseases, chronic diseases, behavioral health issues, and mental health issues. In doing so, I identify reproductive and sexual health as an understudied application.

In **Chapter 3**, I focus on the first case study, which examines online information seeking for contraceptives during a period of heightened uncertainty regarding healthcare reform in the U.S. To motivate this case study, I explain the burden of unintended pregnancies in modern-day society and the cost-effectiveness of contraceptives revealing that the IUD is one of the most effective (for both cost and prevention of untended pregnancies) contraceptives available. I then provide an overview of the contraceptive mandate included in the ACA and the ongoing political battle over healthcare reform in the U.S., which creates uncertainty surrounding the future of the contraceptive mandate. I contribute empirical findings showing that Americans were searching for information about IUDs at an unprecedented volume. The etiology of the shift is not clear from the aggregate data, but I find supporting evidence that IUDs may be seen as a way to confer protection during

the uncertain climate of U.S. healthcare reform. I situate these findings by comparing the results to an analysis performed by an EHR management company that concluded IUD-related visits were elevated, which is consistent with my findings. Finally, I discuss how existing health information seeking models could be operationalized and the limitations of aggregate search engine query data in the framework of existing models.

In **Chapter 4**, I focus on the second case study, which examines online information seeking for STDs. To motivate this case study, I describe the recent increase in STD rates and burden on the healthcare system in the U.S. I then discuss the attractiveness of the online community, Reddit, for health information seeking and recent interest from medical librarians to utilize Reddit as a platform for health communication. A key contribution of this case study is characterizing recent online information seeking and health needs of the r/STD community, a previously unstudied online community. From empirical findings, I contribute insights about the community's most salient information needs and types of online resources that are shared on the platform. Additionally, I compare results of a computational approach to identify prevailing latent themes in the data, specifically topic modeling, to those uncovered by qualitative analysis performed on a subsample of the data to examine agreement between the two methods. Similar to Chapter 3, I discuss how existing health information seeking models could be operationalized and the limitations of anonymous online text data in the framework of existing models.

This brings us to the current chapter, *Chapter 5*, where I reflect on the research presented in this dissertation along with opportunities for future work.

5.2 Discussion

Although many people prefer to obtain reproductive and sexual health information from a healthcare provider, they often turn to the Internet as their first source of information given the sensitivity of these concerns. This makes online information seeking ripe to examine the salient information needs of the public.

From the case studies presented in this dissertation, we identified temporal trends and salient content of information seeking for reproductive and sexual health. As described earlier, using a counterfactual scenario simulated by an ARIMA model, I found evidence of unprecedented search engine queries related to IUDs originating from the U.S. in response to heightened uncertainty of U.S. healthcare policy [215]. Using multiple text mining approaches, I identified salient content for information seeking of STD-specific Reddit platform, r/STD. In particular, individuals most commonly seek information about HPV, HSV, test window periods, and risk and transmission of STDs during oral sex [216].

Although both studies utilized an atheoretical approach to examine online information seeking, I found supporting evidence of experience (either confrontation of out-of-pocket costs or potential symptoms of an STI) and salience (either external uncertainty of healthcare reform or internal uncertainty of risk from an encounter) as a unifying motivators for active information seeking. This is consistent with other studies that indicate higher rates of information seeking are related to heightened uncertainty [158], which supports Kuhlthau's "uncertainty principle" that uncertainty increases anxiety and reduces confidence [25]. Online information seeking may be seen as a way to gain control during periods of uncertainty

(for both external and internal contributors to uncertainty) [149].

Ultimately, new information acquired from online information seeking may inform an individual about their level of risk and provide resources to support medical decision-making or behavioral changes [24]. However, due to the limitations of passively collected aggregate data, we cannot ascertain if their search did indeed result in changes to health behavior or medical decision making. That said, searches have been found to predict behavior such as medical testing [104], and at minimum show engagement during periods of uncertainty. Active information seekers are most influential to health behavior changes, so it would be advantageous to not only supply health communication using online resources but also to tailor it to real-time, salient information needs. I further reflect on the potential promise of these methods for tailoring health communication in the future works section.

5.2.1 Limitations

Both online platforms used in this study - Google Trends and Reddit - only track a segment of the population that use these particular services [26], therefore we do not claim to model information needs for all individuals in the U.S. and caution against making inferences about the general public. However, studies such as these do give insight into the segment of the population that are likely to use the Internet to seek online information making this population opportune for online health communication. A limitation for both case studies is the lack of detailed information on the individuals, so we can not track individual-level characteristics that may contribute to active online information seeking. Finally, it is

unclear if the active information seeking resulted in a change in health behavior or medical decision making.

5.3 Opportunities for Future Work

Beyond identification of temporal trends and content analysis, I identify several areas ripe for future exploration in the following subsections.

5.3.1 Identification of Misinformation in Online Content

One area worthwhile for further exploration is automated identification of inaccurate information that is presented as factual. Although this particular type of misinformation was rarely identified in the qualitative coding of the subsample of r/STD data, it may be more readily apparent in a larger sample or for a different health issue. In the future, I believe this could be explored using semantic textual similarity (STS), which measures the degree to which two sentences are semantically equivalent (capturing gradations of similarity) using sentence-level comparison between layperson text and expert text from authoritative resources. Unfortunately, STS is not trivial and is a significant, ongoing research task, especially for non-English languages [217]. An additional layer of complexity for this particular proposed task is the comparison between the presentation of information from a layperson versus an expert.

The benefit of such a task is automated identification of the public's misunderstandings. If we can better understand the misconceptions and misunderstandings, then health com-

munication could be tailored to combat this particular type of misinformation. Similarly, this methodology could be applied to compare information available in the shared online resources to authoritative text to identify websites containing misinformation. Users often have trouble evaluating the quality and credibility of information encountered and incorporating the information into their decision-making. For example, studies have shown that health blogs are perceived as containing credible content [218]. Based on this, we can better understand deficiencies in digital health literacy to tailor health communication and education.

5.3.2 Identification of Misdiagnosis in Online Content

Another area for future exploration is identification of misdiagnosis. OPs posting on r/STD frequently shared a visual image of their symptoms when inquiring about a crowdsourced diagnosis. Many commenters focused on providing a diagnosis in return. This dissertation did not look at identifying misdiagnosis. Future work could focus on identifying commonly "misdiagnosed" STDs to identify deficiencies in the public's knowledge of STD symptomology. To do this, an expert could label the visual symptoms of an infection, which in turn could be compared to the commenter's crowdsourced diagnoses. If comments and expert diagnosis differ, especially if they differ for a specific STD, this could identify a gap in consumer's knowledge of symptomology; in other words, a need for further health communication. Using this knowledge, health communication could be tailored to provide information to fill the identified knowledge gaps regarding symptomology of STDs.

5.3.3 Leveraging Digital Knowledge to Tailor Health Communication

A vast amount of digital data is left behind from online health information seeking. These digital footprints present the opportunity to identify and address trending, salient information needs. Health communication has traditionally taken a top-down approach, where health communicators assume the audience is a passive recipient of information and decide the content the audience should receive, which is in turn is delivered by generic messages to large segments of the population [219]. Such approaches often have low success rates for changes in health behavior [220]. These approaches result in a persuasive, one-size-fits-all message, however, a pitfall of this approach is that it does not explicitly consider active information seekers that are looking for specific information to meet specific needs [20, 27]. Assuming any audience, not just active information seekers, will utilize available authoritative information or apply a one-size-fits-all message in decision making and health behavior change is a fallacy [221, 222]. Active information seekers are the most likely to have a need for information [223] and are more likely to alter health behavior or medical decision making based on the information they find during their search [20, 224, 225].

In contrast to campaigns that are developed for an entire segment of the population, tailored communication focuses on approaches that consider unique characteristics of an individual and their situation [226, 20, 227, 228] to "...meet the user on the user's terms ...[229]." This increases the perception that the message is relevant to the individual, which consequently increases the chance it is read, enhances persuasion, and creates an environment suitable for behavior change [221, 219]. Despite tailoring being a more effective

form of health communication, it seldom occurs in practice [219, 221].

Tailored health communication has been through multiple iterations of refinement. The first generation of tailored communication consisted of tailored print communication [219]. The second generation consisted of web-delivered or computer-tailored interventions [230, 221]. These interventions provide participants with tailored feedback after an online assessment or tailored content to achieve the participant's specific health goals and range from self-guided to health professional-guided modules mostly provided online [221]. Although evidence suggests that computer-based tailoring is cost-effective [228], these campaigns are time- and resource-intensive to collect information about participants' needs. The particular emphasis on the cost of resources to design such programs is not a trivial consideration, especially given fiscal cuts to public health programs [175]. Creating engaging, relevant, and timely health communication is difficult and, as documented by years of research, often ineffective [231]. It is critical that we consider innovative approaches to health communication that overcome these limitations, especially when the budgets of public health programs are becoming evermore squeezed.

Recently there have been calls in the health communication community to involve the audience in tailoring the messages. Typically, identifying information needs has been done using in-depth interviews, surveys, or focus groups. However, the infodemiology community has been working on analyzing trends in search engine queries and public discourse in social media using a bottom-up, data-driven approach for a number of years. For example, computational approaches have been used to examine public discourse on a variety of issues related to infectious disease, chronic disease, behavioral health, and mental health.

Infodemiology combined with a qualitative review of a sample of data could provide a robust picture of online health information seeking behavior and needs to develop timely, engaging, and relevant health communication. Furthermore, deriving information needs using a bottom-up, data-driven approach allows discovery of layperson language. Using language consistent with a consumers' level of communication could potentially ensure that the consumers are not cognitively overloaded [231]. One of the highlights from the 2018 Agents of Change Summit, which focuses on using marketing and technology to change peoples health behaviors for social good, is the conclusion that it is necessary to "understand the best day of the week/time to post content . . . , cultivat[e] an understanding of who your audience is, what is important to them, and what type of information they need from you [to] transform your content and the way you use your digital tools [232]."

Online health information seeking is ubiquitous and its time to rethink and create innovative ways to disseminate health information to the public. The Internet offers an opportunity for health communication outside of the traditional health care setting [233]. For example, social media has a number of attractive features for delivery of health communication including large audiences, high engagement and retention rates, and users that are actively engaged [234, 235, 236]. The use of social media as health communication tool is a current research interest [237] and social media has already been adopted many local public health departments across the U.S. [238]. As a matter of fact, some organizations such as the Atlanta Community Food Bank, are already using data-driven approaches to examine public discourse and adjust their social media strategies [239]. In an era of limited resources, it's critical to examine all methods that may be used to monitor and improve

public health.

5.4 Concluding Remarks

In summary, the main contribution of this dissertation is extracting recent information needs for reproductive and sexual health using novel online data sources thereby filling a gap in existing infodemiology literature. More specifically, this dissertation makes three contributions to the infodemiology community. First, I identified and addressed a literature gap related to reproductive and sexual health in the field of infodemiology. Second, I identified online information seeking behavior in response to changing healthcare policy. More specifically, I identified unprecedented information seeking for IUDs in response to increased uncertainty in U.S. healthcare policy reform. Finally, third, I identified salient information needs related of a previously unstudied online community. More specifically, I identified salient information needs related to sexual health and STDs on the subreddit r/STD. In doing so, I also identified online resources that are being shared by consumers, which has been understudied in public health. Both case studies contribute knowledge about reproductive and sexual health information seeking to the infodemiology community.

This dissertation only scratches the surface of infodemiology for reproductive and sexual health. Due to the sensitivity of these issues, many people choose to turn to the Internet for information. The opportunities to examine and learn about human behavior by leveraging digital footprints left behind from online information seeking are vast. My long-term

research agenda is focused on leveraging the digital footprints left behind when searching for information related to health or emergency events, such as natural disasters, to monitor and learn about human behavior. I believe that data-driven approaches to examine online information seeking can result into insights into human behavior that were previously only attainable through time- and resource-intensive focus groups, in-depth interviews, or surveys. I also believe infodemiology holds promise to complement more traditional methods of data collection, in near real-time with minimal resources, to reveal and meet information needs of the public outside of the traditional health care setting. Finally, data-driven approaches to mine patterns of information seeking across platforms can be useful in constructing hypotheses, which can then be evaluated and established into verified theories.

Appendices

APPENDIX A

STATE-LEVEL TEMPORAL VISUALIZATIONS

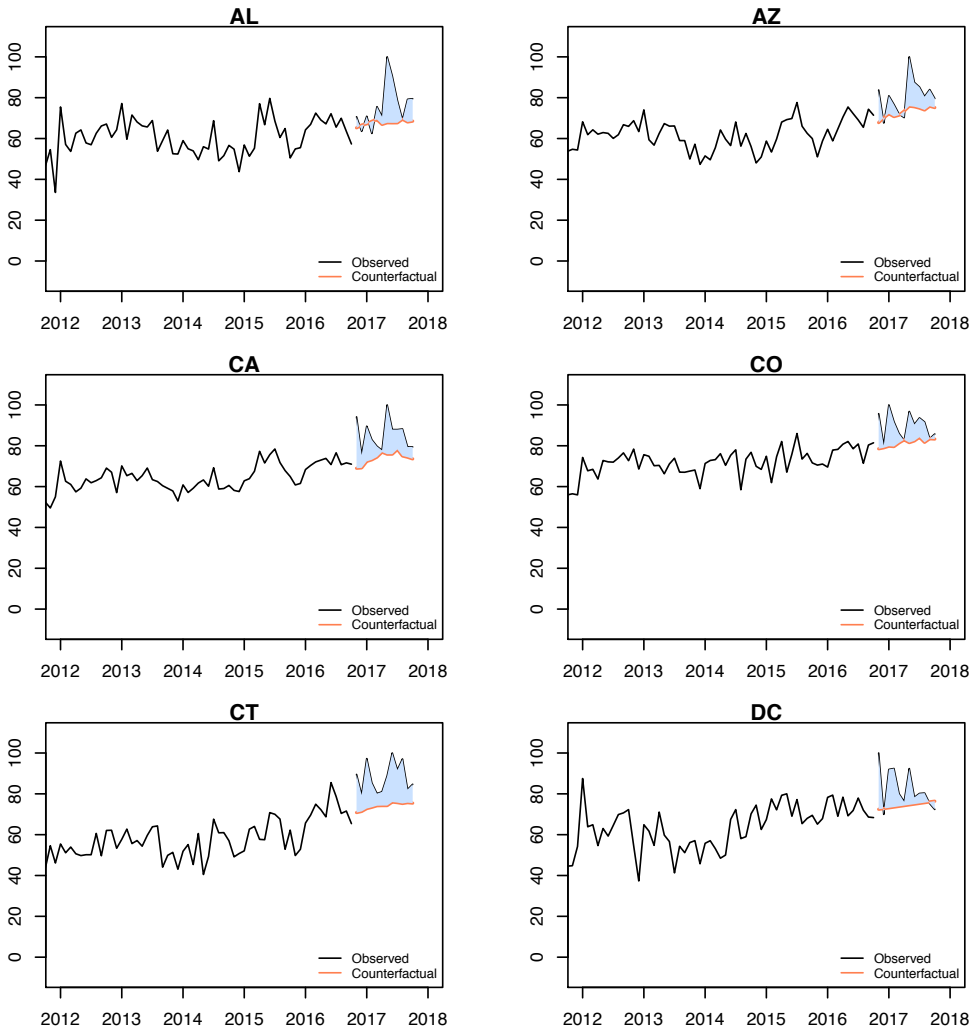


Figure A.1: Temporal Visualizations of States

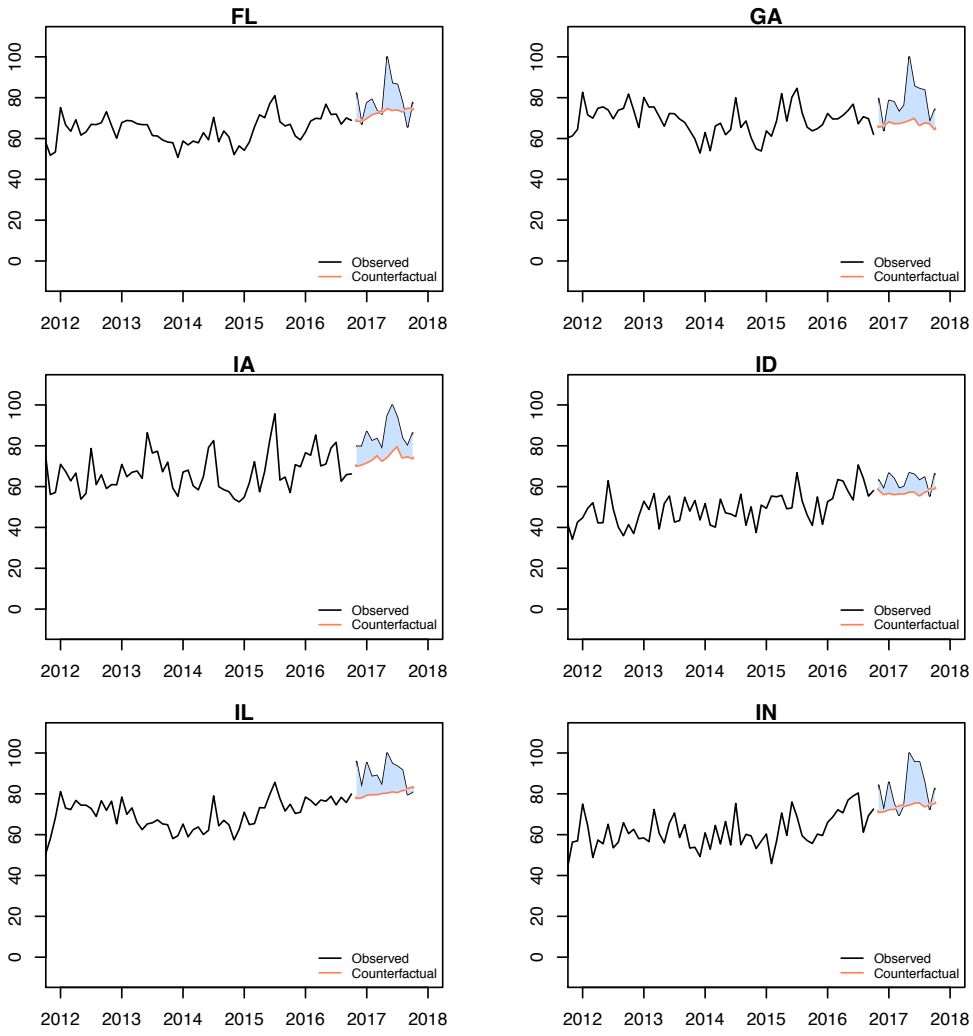


Figure A.1: Temporal Visualizations of States (cont.)

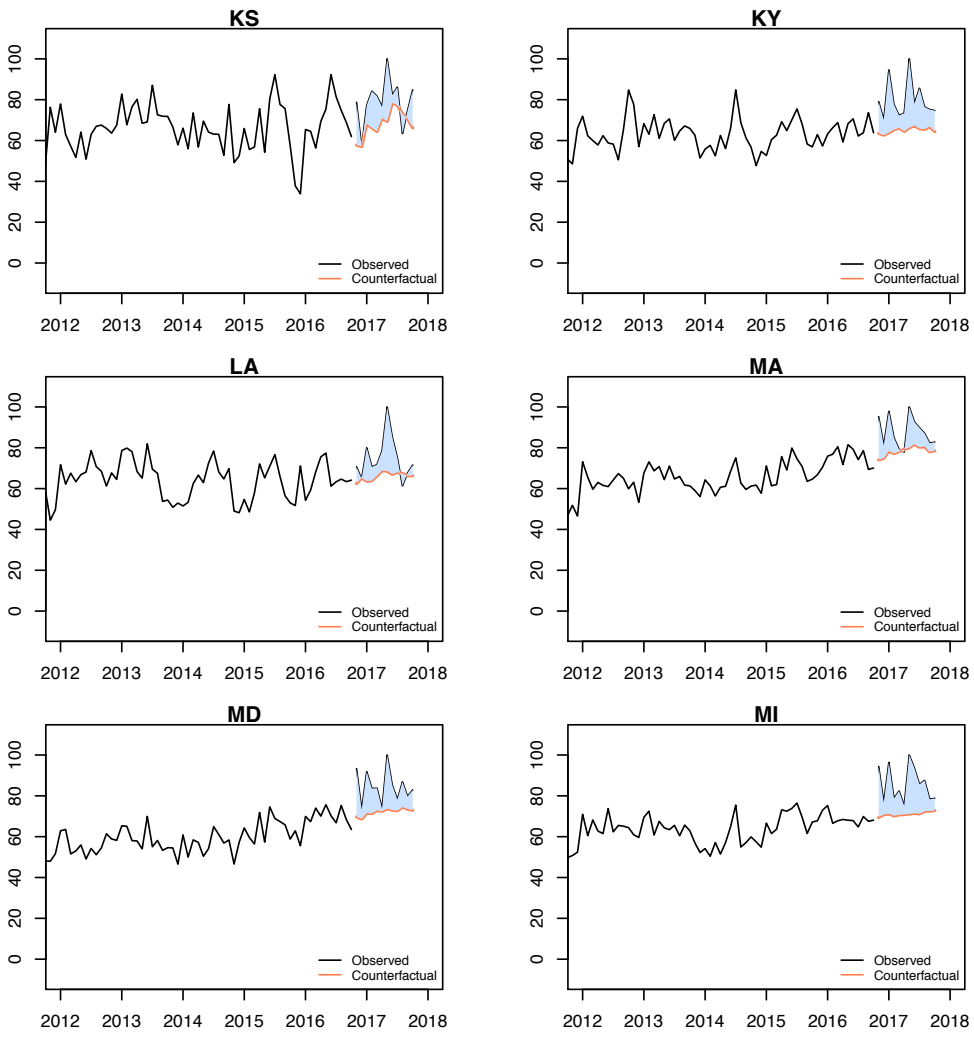


Figure A.1: Temporal Visualizations of States (cont.)

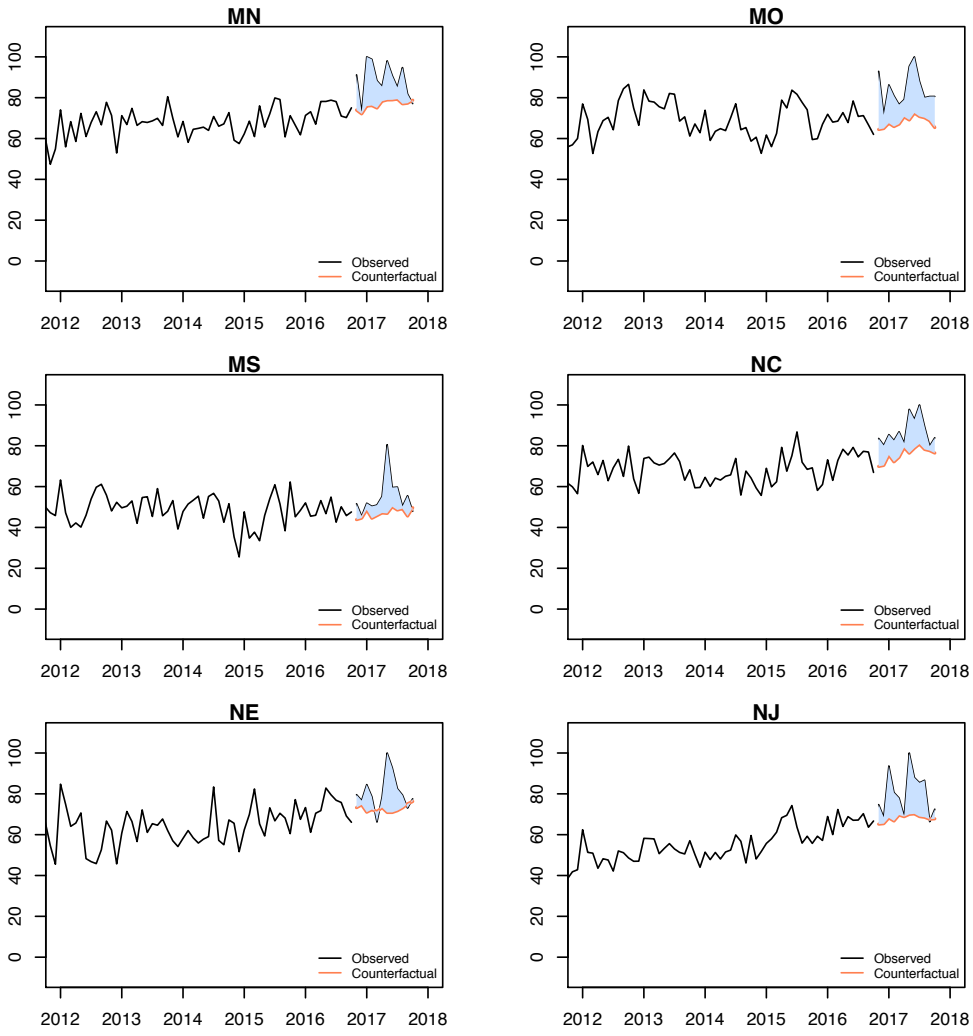


Figure A.1: Temporal Visualizations of States (cont.)

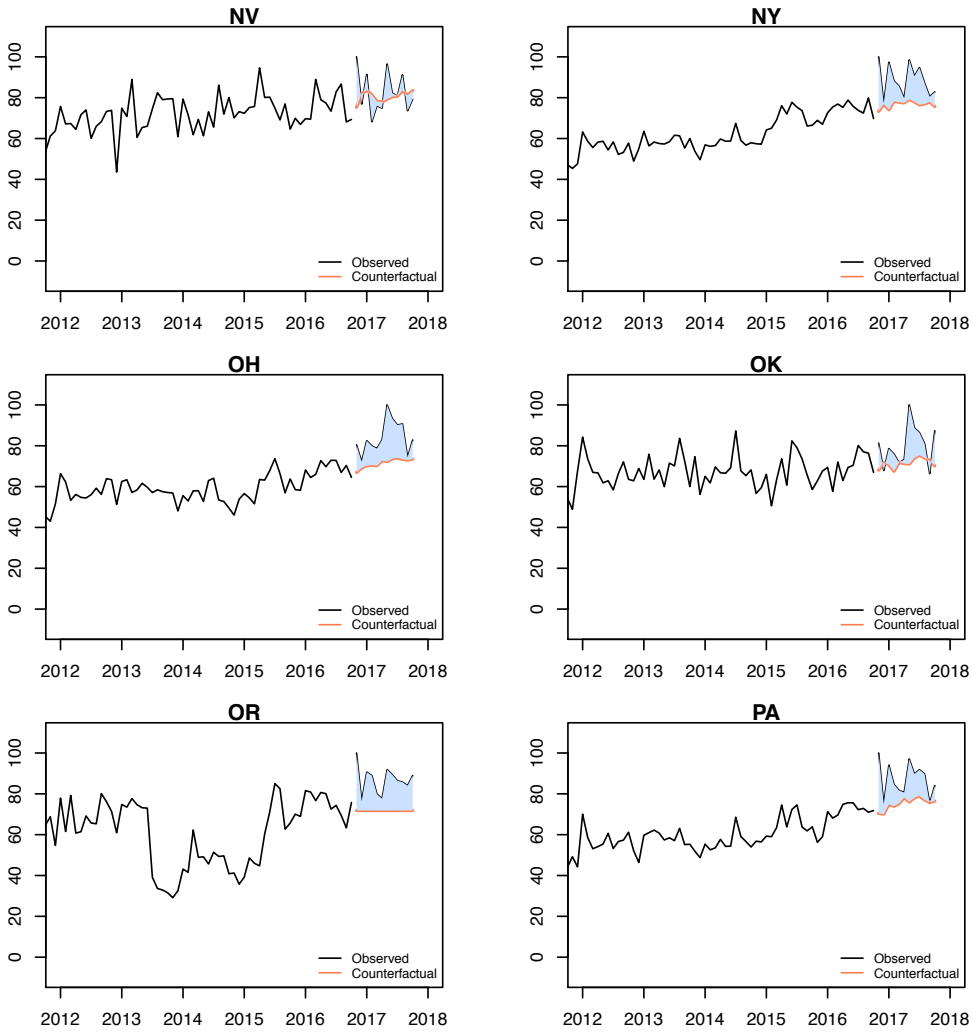


Figure A.1: Temporal Visualizations of States (cont.)

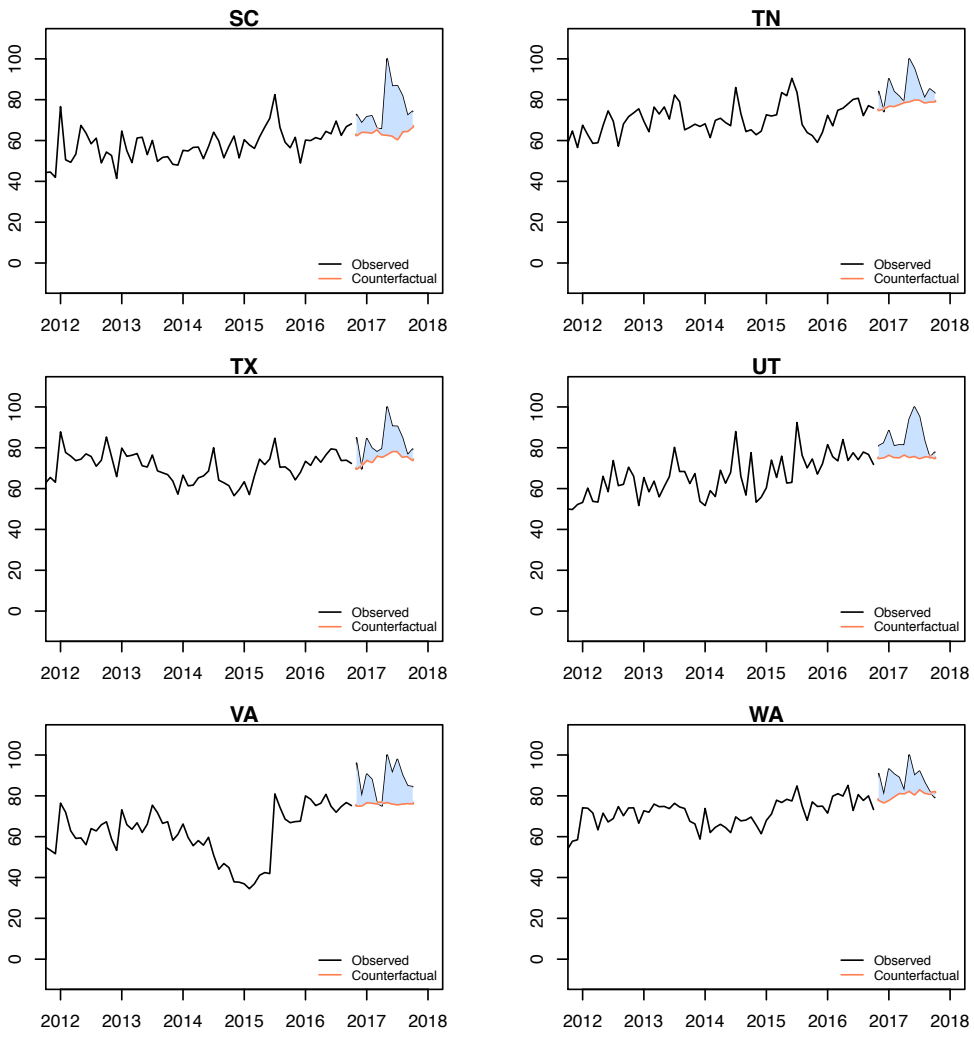


Figure A.1: Temporal Visualizations of States (cont.)

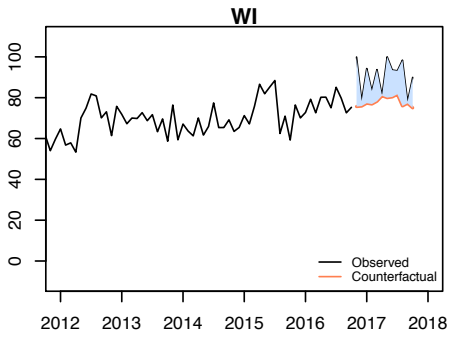


Figure A.1: Temporal Visualizations of States (cont.)

APPENDIX B

ONLINE RESOURCES SHARED ON R/STD

Domain	Category	No. of Posts
advancefornp.com	professional education/organization	1
aidsmap.com	informational site (std)	1
amazon.com	online store	1
anonimag.es	image hosting	3
arhp.org	professional education/organization	1
ashampoo.com	image hosting	1
ashastd.org	informational site (std)	1
askthedoctor.com	expert Q & A	1
atoute.org	unknown	1
baymoon.com	unknown	1
californiaherpes.com	testing center	1
cdc.gov	government resource	4
cityofchicago.org	government resource	1
com.s3.amazonaws.com	image hosting	1
cotbox.moe	image hosting	1
curezone.org	alternative medicine site	1
eroshare.com	image hosting	1
greenjournal.org	informational site (reproductive health)	1
gyazo.com	image hosting	2
hercampus.com	lifestyle site	1
herpesdatingsupport.com	herpes dating and support	3
ibb.co	image hosting	7
imgsafe.org	image hosting	1
imgur.com	image hosting	393
justanswer.com	expert Q & A	1
liveherpesvaccine.com	herpes vaccination site	1
medhelp.org	social Q & A	1
meetup.com	support group	1
mocha.us	herpes vaccination site	1
mylabbox.com	online testing	1
nih.gov	research article	2
OHIV.org	informational site (std)	1
patient.info	social Q & A	1
plannedparenthood.org	informational site (reproductive/sexual health)	1
postimg.org	image hosting	6
prntscr.com	image hosting	1
puu.sh	image hosting	1
quora.com	social Q & A	1
reddit.com	social Q & A	8

Continued on next page

Domain	Category	No. of Posts
researchgate.com	research article	1
sli.mg	image hosting	2
smh.com.au	news	1
stdaz.com	online testing	1
stdcheck.com	online testing	3
texas.gov	government resource	1
thebody.com	informational site (std)	1
thebodypro.com	informational site (std)	1
tinypic.com	image hosting	3
tumblr.com	image hosting	1
vgy.me	image hosting	1
wikimedia.org	wiki	1
womenshealthchannel.com	informational site (women's health)	1
yourstdhelp.com	online testing	1

Table B.1: Top-level domains of online resources shared in original posts.

Domain	Category	No. of Comments
aafp.org	professional education/organization	1
about.com	informational site (general health)	1
acog.org	professional education/organization	1
advocatesaz.org	advocacy organization	2
aidsvancouver.org	informational site (AIDS)	1
anal-fissure.org	support forum	1
anxietycentre.com	informational site (mental health)	1
ashasexualhealth.org	informational site (sexual health)	1
askexpertsnow.com	expert Q&A	6
i-base.info	informational site (STDs)/advocacy	1
brown.edu	student health services	1
cdc.gov	government resource	15
checkhimout.ca	informational site (men's health)	1
compareshack.com	online testing	1
craveonline.com	lifestyle website	1
cupidslibrary.com	dating site	1
cure4hpv.com	unknown	2
diureticspill.com	unknown	3
drugs.com	informational site (medication)	1
ehealthforum.com	expert Q&A	2
everydayhealth.com	informational site (general health)	1
familyeducation.com	informational site (family health)	1
gyazo.com	image hosting	1
healthline.com	informational site (general health)	2
herpes.org.uk	informational site (STDs)	1
herpesite.org	informational site (STDs)/support	1
herpesopportunity.com	informational site (STDs)/support	3
hivlawandpolicy.org	informational site (HIV)/advocacy	1
hollywoodclinic.net	urgent care center	1
hopkinsmedicine.org	university medical center	1
hpsc.ie	government resource	1
hpvandme.org	informational site (STDs)/support	1
hsvfish.com	dating site	1
hsvsingles.biz	dating site	1
hubstatic.com	image hosting	1
ibb.co	image hosting	2
imgur.com	image hosting	73
inspire.com	support site	1
iocdf.org	informational site (mental health)	1
jamanetwork.com	research article	1
juntoscontraloshongos.com	unknown	1
labtestsonline.org	online testing	1
liveherpesvaccine.com	herpes vaccination site	1
livescience.com	informational site (general)	2
lmgtfy.com	unknown	2
lww.com	research article	1
medhelp.org	social Q&A	4
medicalnewstoday.com	news (health)	1
medicinenet.com	informational site (general health)	1
medlineplus.gov	government resource	2

Continued on next page

Domain	Category	No. of Comments
medscape.com	news (health)	4
mens-health.sg	health clinic (men's)	1
mylastoutbreak.com	support site	1
naturalthrifty.com	lifestyle website	1
nccc-online.org	informational site (cervical cancer)	1
neatclub.org	dating site	1
netdoctor.co.uk	informational site (general health)	1
nhs.uk	government resource	2
nih.gov	research article	2
oxfordjournals.org	research article	1
pastebin.com	unknown	17
paythebestprice.com	coupon site	1
pinimg.com	image hosting	1
plannedparenthood.org	informational site (reprodutive/sexual health)	2
postimg.org	image hosting	1
poz.com	informational site (HIV/AIDS)	2
prepfacts.org	informational site (PrEP)	1
prostatitisradicalcure.com	alternative medicine	1
reddit.com	social Q&A	18
slidesharecdn.com	image hosting	1
smartsexresource.com	informational site (sexual health)	1
std-gov.org	informational site (STDs)	1
stdcheck.com	online testing	8
teed.io	link sharing	1
tinypic.com	image hosting	1
topherpesdatingsites.com	dating site	1
toppcock.com	online store (personal care)	3
treatcurefast.com	informational site (general health)	1
treatnheal.com	informational site (general health)	1
trojanbrands.com	condom brand's site	1
tumblr.com	image hosting	1
tyginta.com	online store (genital wart removal)	1
unsee.cc	image hosting	3
visual-science.com	informational site (general)	1
washington.edu	university informational site	2
westoverheights.com	informational site (STDs)	4
who.int	health organization	1
wikipedia.org	wiki	3
womenshealth.gov	government resource	1
youonlywetter.co.uk	online store (adult toys)	1
youtube.com	video hosting	4

Table B.2: Top-level domains of online resources shared in comments.

APPENDIX C

SUB-CODES FROM QUALITATIVE ANALYSIS

Theme	Sub-Code
Decision Making	Does not intend to see a provider
	Sees a provider or is tested
	Intends to see a provider
	Decides that it is not a STD
Emotional Content	Anxiety
	Depression
Health Care Interaction	Seeks first opinion prior to seeing a provider
	Seeks second opinion after seeing a provider
	Mentions testing or diagnosis status
	Waiting on a scheduled appointment
Information Seeking	Seeks information about dating with an STD
	Seeks a crowdsourced diagnosis
	Seeks information about dating and disclosing STD status
	Nonspecific
	Asks others to share their experience
	Seeks information about prevention
	Seeks information about providers
	Seeks information about symptoms
	Seeks information about testing
	Seeks information about transmission or risk
	Seeks information about treatment
Misinformation	Corrects misinformation
	Presents misinformation
Provides Information	Provides information to reduce stigma
	Provides a diagnosis
	Provides information about disclosing to a partner
	Provides emotional support
	Provides information about prevention
	Shares personal experience
	Suggests posting on different subreddit
	Suggests follow-up with a provider, testing, or treatment
	Suggests original poster may need mental health support
	Provides information about symptoms
	Provides information about testing
	Provides information about transmission or risk

Continued on next page

Theme	Sub-Code
	Provides information about treatment
	Provides information about what kind of provider to see
STDs	Chlamydia
	Gonorrhea
	Hepatitis
	Human immunodeficiency virus (HIV)
	Human papillomavirus (HPV)
	Herpes simplex virus (HSV)
	Molluscum contagiosum
	Mycoplasma
	Nongonococcal urethritis
	Scabies
	Syphillis
	Trichomoniasis
	Ureaplasma

Table C.1: Themes and sub-codes of original posts and comments.

REFERENCES

[1] G. Eysenbach, "Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet," *Journal of Medical Internet Research*, vol. 11, no. 1, e11, 2009.

[2] E. Yom-Tov and L. Fernandez-Luque, "Information is in the eye of the beholder: Seeking information on the mmr vaccine through an internet search engine," in *AMIA Annual Symposium Proceedings*, 2014, 12381247.

[3] M. De Choudhury and S. De, "Mental health discourse on reddit: Self-disclosure, social support, and anonymity," in *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media (ICWSM)*, 2014, pp. 71–80.

[4] J. Paparrizos, R. W. White, and E. Horvitz, "Screening for pancreatic adenocarcinoma using signals from web search logs: Feasibility study and results," *Journal of Oncology Practice*, vol. 12, 737744, 8 2016.

[5] A. L. Friedman, R. E. Kachur, S. M. Noar, and M. McFarlane, "Health communication and social marketing campaigns for sexually transmitted disease prevention and control: What is the evidence of their effectiveness?" *Sexually Transmitted Diseases*, vol. 43, no. 2S, S83S101, 2016.

[6] U.S. Department of Health & Human Services, *Save the date: New social media data to drive your planning (facebook live)*, Retrieved July 9, 2018 from `https://www.hiv.gov/blog/save-date-new-social-media-data-drive-your-planning-facebook-live`, 2018.

[7] C. Appleby, "Net gain or net loss? health care consumers become internet savvy," *Trustee*, vol. 52, no. 2, pp. 20–23, 1999.

[8] S. Fox, "Health topics: 80% of internet users look for health information online," Pew Research Center, Washington, D.C., Tech. Rep., 2011.

[9] J. Manganello, G. Gerstner, K. Pergolino, Y. Graham, A. Falisi, and D. Strogatz, "The Relationship of Health Literacy With Use of Digital Technology for Health Information: Implications for Public Health Practice," *Journal of Public Health Management and Practice*, vol. 23, no. 4, pp. 380–387, 2017.

[10] D. J. Amante, T. P. Hogan, S. L. Pagoto, T. M. English, and K. L. Lapane, "Access to care and use of the Internet to search for health information: results from the US National Health Interview Survey," *Journal of Medical Internet Research*, vol. 17, no. 4, e106, 2015.

[11] S. Fox and M. Duggan, "Mobile health 2012," Pew Research Center, Washington, D.C., Tech. Rep., 2012.

[12] M. L. Antheunis, K. Tates, and T. E. Nieboer, "Patients' and health professionals' use of social media in health care: Motives, barriers and expectations," *Patient Education and Counseling*, vol. 92, no. 3, pp. 426–431, 2013.

[13] Y. Zhang, "Beyond quality and accessibility: Source selection in consumer health information searching," *Journal of the Association for Information Science and Technology*, vol. 65, no. 5, pp. 911–927, 2014.

[14] R. J. W. Cline and K. M. Haynes, "Consumer health information seeking on the internet: The state of the art," *Health Education Research*, vol. 16, no. 6, pp. 671–692, 2001.

[15] L. Sbaffi and J. Rowley, "Trust and credibility in web-based health information: A review and agenda for future research," *Journal of Medical Internet Research*, vol. 19, no. 6, e218, 2017.

[16] F. Mapp, K. Wellings, F. Hickson, and C. H. Mercer, "Understanding sexual healthcare seeking behaviour: Why a broader research perspective is needed," *BMC Health Services Research*, vol. 17, no. 1, p. 462, 2017.

[17] K. Wood and P. Aggleton, "Promoting Young Peoples Sexual and Reproductive Health: Stigma, Discrimination and Human Rights. Guide to Good Practice, Safe Passages to Adulthood," University of Southampton, UK, Tech. Rep., 2004.

[18] J. S. Rumsfeld, S. C. Brooks, T. P. Aufderheide, M. Leary, S. M. Bradley, C. Nkonde-Price, L. H. Schwamm, M. Jessup, J. M. E. Ferrer, and R. M. Merchant, "Use of mobile devices, social media, and crowdsourcing as digital strategies to improve emergency cardiovascular care: A scientific statement from the american heart association.," *Circulation*, vol. 134, no. 8, e87 –e108, 2016.

[19] S. Fox, "The social life of health information, 2011," Pew Research Center, Washington, D.C., Tech. Rep., 2011.

[20] J. D. Johnson and D. O. Case, *Health Information Seeking*. Peter Lang Inc., International Academic Publishers, 2012.

[21] R. Fidel, *Human Information Interaction: An Ecological Approach to Information Behavior*. MIT Press, 2012.

[22] K. Glanz, B. K. Rimer, and K. Viswanath, *Health Behavior and Health Education: Theory, Research, and Practice*, Fourth. San Francisco, CA: Jossey-Bass, 2008.

[23] K. Glanz, B. K. Rimer, and K. Viswanath, *Health Behavior: Theory, Research, and Practice*, Fifth. San Francisco, CA: Jossey-Bass, 2015.

[24] S. D. Lambert and C. G. Loiselle, "Health information-seeking behavior," *Qualitative Health Research*, vol. 17, no. 8, pp. 1006–1019, 2007.

[25] C. C. Kuhlthau, *Seeking meaning: A process approach to library and information services*. Westport, CT: Libraries Unlimited, 2004.

[26] K. Zeraatkar and M. Ahmadi, "Trends of infodemiology studies: A scoping review," *Health Information and Libraries Journal*, 2018.

[27] V. S. Freimuth, J. A. Stein, and T. J. Kean, *Searching for health information: the Cancer Information Service model*. Philadelphia: University of Pennsylvania Press, 1989.

[28] S. Criss, J. A. Woo Baidal, R. E. Goldman, M. Perkins, C. Cunningham, and E. M. Taveras, "The role of health information sources in decision-making among hispanic mothers during their children's first 1000-days of life," *Maternal and Child Health Journal*, vol. 19, no. 11, pp. 2536–2543, 2015.

[29] T. Ntlotlang and B. Grand, "The role of libraries in the dissemination of health information in botswana: A study of mochudi and molepolole public libraries," *Library Review*, vol. 65, pp. 320–349, 4/5 2016.

[30] D. R. Longo, "Understanding health information, communication, and information seeking of patients and consumers: A comprehensive and integrated model," *Health Expectations*, vol. 8, no. 3, pp. 189–194, 2005.

[31] D. R. Longo, B. Ge, M. E. Radina, A. Greiner, C. D. Williams, G. S. Longo, D. M. Mouzon, A. Natale-Pereira, and D. Salas-Lopez, "Understanding breast-cancer patients' perceptions: Health information-seeking behaviour and passive information receipt," *Journal of Communication in Healthcare*, vol. 2, no. 2, pp. 184–206, 2009.

[32] D. R. Longo, S. L. Schubert, B. A. Wright, J. LeMaster, C. D. Williams, and J. N. Clore, "Health information seeking, receipt, and use in diabetes self-management.," *Annals of Family Medicine*, vol. 8, no. 4, pp. 334 –340, 2010.

[33] A. H. Eagly and S. Chaiken, *The Psychology of Attitudes*. Fort Worth: Harcourt Brace Jovanovich College Publishers, 1993.

[34] I. Ajzen, "The theory of planned behavior," *Organizational Behavior and Human Decision Processes*, vol. 50, no. 2, pp. 179 –211, 1991, Theories of Cognitive Self-Regulation.

[35] E. F. J. ter Huurne, R. J. Griffin, and J. M. Gutteling, "Risk information seeking among u.s. and dutch residents: An application of the model of risk information seeking and processing.," *Science Communication*, vol. 31, no. 2, pp. 215 –237, 2009.

[36] R. J. Griffin, K. Neuwirth, S. Dunwoody, and J. Giese, "Information sufficiency and risk communication," *Media Psychology*, vol. 6, pp. 23–61, 2004.

[37] R. J. Griffin, J. Yang, E. Huurne, F. ter Boerner, S. Ortiz, and S. Dunwoody, "After the flood: Anger, attribution, and the seeking of information," *Science Communication*, vol. 29, pp. 285–315, 2008.

[38] B. B. Johnson, "Testing and expanding a model of cognitive processing of risk information," *Risk Analysis*, vol. 25, pp. 631–650, 3 2005.

[39] L. A. Kahlor, S. Dunwoody, R. J. Griffin, and K. Neuwirth, "Seeking and processing information about impersonal risk," *Science Communication*, vol. 28, pp. 163–194, 1998.

[40] L. A. Kahlor, "An augmented risk information seeking model: The case of global warming," *Media Psychology*, vol. 10, pp. 414–435, 2007.

[41] C. E. Clarke and K. McComas, "Seeking and processing influenza vaccine information: A study of health care workers at a large urban hospital," *Health Communication*, vol. 27, no. 3, pp. 244–256, 2012.

[42] J. R. Allen Catellier and Z. J. Yang, "Trust and affect: How do they impact risk information seeking in a health context?" *Journal of Risk Research*, vol. 15, no. 8, pp. 897–911, 2012.

[43] M. Kuttschreuter, P. Rutsaert, F. Hilverda, A. Regan, J. Barnett, and W. Verbeke, "Seeking information about food-related risks: The contribution of social media.," *Food Quality and Preference*, vol. 37, pp. 10–18, 2014.

[44] R. J. Griffin, S. Dunwoody, and K. Neuwirth, "Proposed model of the relationship of information seeking and processing to the development of preventive behaviours," *Environmental Research*, vol. 80, S230–S245, 1999, Theories of Cognitive Self-Regulation.

[45] J. D. Johnson and H. Meischke, "A comprehensive model of cancer-related information seeking applied to magazines," *Human Communication Research*, vol. 19, pp. 343–367, 1993.

[46] W. A. Afifi, "Uncertainty management theories," in *Encyclopedia of Communication Theory*, S. W. Littlejohn and K. A. Foss, Eds., First Ed, SAGE Publications, Inc, 2009, pp. 973–976.

[47] D. O. Case and L. M. Given, *Looking for Information: A Survey of Research on Information Seeking, Needs, and Behavior*, Fourth Ed. Bingley, UK: Emerald, 2016.

[48] C. R. Eheman, Z. Berkowitz, J. Lee, S. Mohile, J. Purnell, E. Marie Rodriguez, J. Roscoe, D. Johnson, J. Kirshner, and G. Morrow, "Information-seeking styles among cancer patients before and after treatment by demographics and use of information sources," *Journal of Health Communication*, vol. 14, no. 5, pp. 487 –502, 2009.

[49] J. D. Johnson, "A test of a model of magazine exposure and appraisal in India," *Communication Monographs*, vol. 50, pp. 148–157, 1983.

[50] J. D. Johnson, *Cancer-related information seeking*. Cresskill, NJ: Hampton Press, 1997.

[51] N. Hartoonian, S. R. Ormseth, E. R. Hanson, E. O. Bantum, and J. E. Owen, "Information-Seeking in Cancer Survivors: Application of the Comprehensive Model of Information Seeking to HINTS 2007 Data," *Journal of Health Communication*, vol. 19, pp. 1308–1325, 11 2014.

[52] E. K. Ruppel, "Scanning health information sources: Applying and extending the comprehensive model of information seeking," *Journal of Health Communication*, vol. 21, no. 2, pp. 208–216, 2016.

[53] K. M. Oh, G. L. Kreps, J. Jun, E. Chong, and L. Ramsey, "Examining the health informationseeking behaviors of korean americans," *Journal of Health Communication*, vol. 17, no. 7, pp. 779–801, 2012.

[54] S. N. Sweet, M.-J. Perrier, C. Podzyhun, and A. E. Latimer-Cheung, "Identifying physical activity information needs and preferred methods of delivery of people with multiple sclerosis," *Disability and Rehabilitation*, vol. 35, no. 24, pp. 2056–2063, 2013.

[55] U.S. Centers for Disease Control and Prevention, *Principles of epidemiology in public health practice, third edition, an introduction to applied epidemiology and biostatistics*, Retrieved May 18, 2018 from `https://www.cdc.gov/ophss/csels/dsepd/ss1978/`, 2011.

[56] M. J. Paul and M. Dredze, "Social Monitoring for Public Health," *Synthesis Lectures on Information Concepts, Retrieval, and Services*, vol. 9, no. 5, G. Marchionini, Ed., pp. 1–183, 2017.

[57] T. M. Bernardo, A. Rajic, I. Young, K. Robiadek, M. T. Pham, and J. A. Funk, "Scoping review on search queries and social media for disease surveillance: A chronology of innovation," *Journal of Medical Internet Research*, vol. 15, e147, 2013.

[58] M. Santillana, D. Zhang, B. Althouse, and J. Ayers, "What can digital disease detection learn from (an external revision to) Google Flu Trends?" *American Journal of Preventive Medicine*, vol. 47, pp. 341–347, 3 2014.

[59] X. Sun, J. Ye, and F. Ren, "Detecting influenza states based on hybrid model with personal emotional factors from social networks," *Neurocomputing*, vol. 210, pp. 257–268, 2016.

[60] R. Priedhorsky, D. Osthus, A. Daughton, K. Moran, N. Generous, G. Fairchild, A. Deshpande, and S. Del Valle, "Measuring global disease with wikipedia: Success, failure, and a research agenda," in *Proceedings of Computer Supported Cooperative Work and Social Computing (CSCW)*, 2017.

[61] E. Nsoesie, D. Buckeridge, and J. Brownstein, "Guess whos not coming to dinner? evaluating online restaurant reservations for disease surveillance," *Journal of Medical Internet Research*, vol. 16, e22, 1 2014.

[62] B. M. Althouse, S. V. Scarpino, L. A. Meyers, J. W. Ayers, M. Bargsten, J. Baumbach, J. S. Brownstein, L. Castro, H. Clapham, D. A. Cummings, S. Del Valle, S. Eubank, G. Fairchild, L. Finelli, N. Generous, D. George, D. R. Harper, L. Hébert-Dufresne, M. A. Johansson, K. Konty, M. Lipsitch, G. Milinovich, J. D. Miller, E. O. Nsoesie, D. R. Olson, M. Paul, P. M. Polgreen, R. Priedhorsky, J. M. Read, I. Rodríguez-Barraquer, D. J. Smith, C. Stefansen, D. L. Swerdlow, D. Thompson, A. Vespignani, and A. Wesolowski, "Enhancing disease surveillance with novel data streams: Challenges and opportunities," *EPJ Data Science*, vol. 4, no. 1, p. 17, 2015.

[63] J. Gomide, A. Veloso, J. W. Meira, V. Almeida, F. Benevenuto, F. Ferraz, and M. Teixeira, "Dengue surveillance based on a computational model of spatio-temporal locality of Twitter," in *Web Science Conference*, 2011.

[64] A. Seifter, A. Schwarzwalder, K. Geis, and J. Aucott, "The utility of "Google Trend" for epidemiological research: Lyme disease as an example," *Geospatial Health*, vol. 4, pp. 135–137, 2 2010.

[65] R. Desai, A. J. Hall, B. A. Lopman, Y. Shimshoni, M. Rennick, N. Efron, Y. Matias, M. M. Patel, and U. D. Parashar, "Norovirus disease surveillance using Google Internet query share data," *Clinical Infectious Diseases*, vol. 55, e7578, 8 2012.

[66] E. Diaz-Aviles and A. Stewart, "Tracking Twitter for epidemic intelligence: Case study: EHECHUS outbreak in Germany, 2011," in *Web Science Conference*, 2012.

[67] A. Ocampo, R. Chunara, and J. Brownstein, "Using search queries for malaria surveillance, thailand," *Malaria Journal*, vol. 12, p. 390, 1 2013.

[68] O. B. Da'ar, F. Yunus, N. Md Hossain, and M. Househ, "Impact of Twitter intensity, time, and location on message lapse of bluebirds pursuit of fleas in Madagascar," *Journal of Infection Public Health*, vol. 10, pp. 396–402, 4 2017.

[69] B. Zou, V. Lampos, R. Gorton, and I. Cox, "On infectious intestinal disease surveillance using social media content," in *International Conference on Digital Health Conference*, 2016, 157161.

[70] S. Towers, S. Afzal, G. Bernal, B. S. Bliss N., B. Espinoza, J. Jackson, J. Judson-Garcia, M. Khan, M. Lin, R. Mamada, V. M. Moreno, F. Nazari, K. Okuneye, M. L. Ross, C. Rodriguez, J. Medlock, D. Ebert, and C. Castillo-Chavez, "Mass media and the contagion of fear: E case of ebola in america," *PLoS ONE*, vol. 10, e0129179, 6 2015.

[71] A. Stefanidis, E. Vraga, G. Lamprianidis, J. Radzikowski, L. P. Delamater, H. K. Jacobsen, D. Pfoser, A. Croitoru, and A. Crooks, "Zika in Twitter: Temporal variations of locations, actors, and concepts," *JMIR Public Health Surveillance*, vol. 3, e22, 2 2017.

[72] M. Deiner, T. Lietman, S. McLeod, J. Chodosh, and T. Porco, "Surveillance tools emerging from search engines and social media data for determining eye disease patterns," *Journal of the American Medical Association*, vol. 134, pp. 1024–1030, 9 2016.

[73] S. D. Young, C. Rivers, and B. Lewis, "Methods of using real-time social media technologies for detection and remote monitoring of HIV outcomes," *Preventive Medicine*, vol. 63, pp. 112–115, 2014.

[74] S. D. Young, W. Yu, and W. Wang, "Toward automating HIV identification: Machine learning for rapid identification of HIV-related social media data," *Journal of Acquired Immune Deficiency Syndromes*, vol. 74, S128S131, 2017.

[75] S. D. Young, N. Mercer, R. E. Weiss, E. A. Torrone, and S. O. Aral, "Using social media as a tool to predict syphilis," *Preventive Medicine*, vol. 109, pp. 58–61, 2018.

[76] M. E. Ireland, H. A. Schwartz, Q. Chen, L. H. Ungar, and D. Albarracn, "Future-oriented tweets predict lower county-level HIV prevalence in the United States," *Health Psychology*, vol. 34, no. Suppl, pp. 1252 –1260, 2015.

[77] S. Oh and M. S. Park, "Text mining as a method of analyzing health questions in social Q&A," *Proceedings of the American Society for Information Science and Technology*, vol. 50, no. 1, pp. 1–4,

[78] A. Domnich, E. K. Arbuzova, A. Signori, D. Amicizia, D. Panatto, and R. Gasparini, "Demand-based web surveillance of sexually transmitted infections in russia," *International Journal of Public Health*, vol. 59, pp. 841–849, 2014.

[79] P. Breen, J. Kelly, T. Heckman, and S. Quinn, "Mining pre-exposure prophylaxis trends in social media," in *2016 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, 2016, pp. 214–221.

[80] P. J. Tighe, R. C. Goldsmith, M. Gravenstein, H. R. Bernard, and R. B. Fillingim, "The painful tweet: Text, sentiment, and community structure analyses of tweets pertaining to pain," *Journal of Medical Internet Research*, vol. 17, e84, 4 2015.

[81] H. Dai, B. R. Lee, and J. Hao, "Predicting asthma prevalence by linking social media data and traditional surveys," *ANNALS of the American Academy of Political and Social Science*, vol. 669, pp. 75–92, 1 2017.

[82] Y. Liu, Q. Mei, A. D. Hanauer, K. Zheng, and M. J. Lee, "Use of social media in the diabetes community: An exploratory analysis of diabetes-related tweets," *Journal of Medical Internet Research: Diabetes*, vol. 1, e4, 2 2016.

[83] D. T. Nascimento, F. M. DosSantos, T. Danciu, M. DeBoer, H. van Holsbeeck, R. S. Lucas, C. Aiello, L. Khatib, A. M. Bender, J.-K. Zubieta, and F. A. DaSilva, "Real-Time Sharing and Expression of Migraine Headache Suffering on Twitter: A Cross-Sectional Infodemiology Study," *Journal of Medical Internet Research*, vol. 16, no. 4, e96, 2014.

[84] F. Brigo and R. Erro, "Why do people google movement disorders? an infodemiological study of information seeking behaviors," *Neurological Sciences*, vol. 37, no. 5, pp. 781–787, 2016.

[85] F Brigo, W. Otte, S. Igwe, H Ausserer, R Nardone, F Tezzon, and E Trinka, "Information-seeking behavior for epilepsy: An infodemiological study of searches for wikipedia articles," *Epileptic Disorders*, vol. 17, pp. 460–466, 2015.

[86] F. Brigo, P. Lochner, F. Tezzon, and R. Nardone, "Web search behavior for multiple sclerosis: An infodemiological study," *Multiple Sclerosis and Related Disorders*, vol. 3, no. 4, pp. 440–443, 2014.

[87] H. W. Wang, D. R. Chen, H. W. Yu, and Y. M. Chen, "Forecasting the incidence of dementia and dementia-related outpatient visits with Google Trends: Evidence from Taiwan," *Journal of Medical Internet Research*, vol. 17, e264, 2015.

[88] S. Abbar, Y. Mejova, and I. Weber, "You tweet what you eat: Studying food consumption through Twitter," in *Proceedings of the Conference on Human Factors in Computing Systems (CHI)*, 2015.

[89] S. Sharma and M. De Choudhury, "Detecting and characterizing nutritional information of food and ingestion content in Instagram," in *Proceedings of the International Conference on World Wide Web (WWW)*, 2015.

[90] N. Zhang, S. Campo, K. F. Janz, P. Eckler, J. Yang, L. G. Snetselaar, and A. Signorini, "Electronic word of mouth on Twitter about physical activity in the U.S.: exploratory infodemiology study," *Journal of Medical Internet Research*, vol. 15, e261, 11 2013.

[91] T. Chang, V. Chopra, C. Zhang, and S. J. Woolford, "The role of social media in online weight management: Systematic review," *Journal of Medical Internet Research*, vol. 15, e262, 11 2013.

[92] C. Leggatt-Cook and K. Chamberlain, "Blogging for weight loss: Personal accountability, writing selves, and the weight-loss blogosphere," *Social Health Illness*, vol. 34, pp. 963–977, 7 2012.

[93] S. Chancellor, A. Hu, and M. De Choudhury, "Norms matter: Contrasting social support around behavior change in online weight loss communities," in *ACM Conference on Human Factors in Computing Systems*, ser. CHI, 2018.

[94] J. W. Ayers, B. M. Althouse, J. P. Allem, E. C. Leas, M. Dredze, and R. S. Williams, "Revisiting the rise of electronic nicotine delivery systems using search query surveillance," *American Journal of Preventive Medicine*, vol. 50, e173–e181, 6 2016.

[95] J. H. West, P. C. Hall, K. Prier, C. L. Hanson, C. Giraud-Carrier, E. S. Neeley, and M. D. Barnes, "Temporal variability of problem drinking on Twitter," *Open Journal of Preventative Medicine*, vol. 2, 1 2012.

[96] T. Caputi, E. Leas, M Dredze, and J. W. Ayers, "Online sales of marijuana: An unrecognized public health dilemma," *American Journal of Preventative Medicine*, vol. 54, pp. 719–721, 5 2018.

[97] B. Curtis, K. Alanis-Hirsch, O. Kaynak, A. T. McLellan, and K. Meyers, "Using web searches to track interest in synthetic cannabinoids (aka herbal incense)," *Drug & Alcohol Review*, vol. 34, pp. 105–108, 2015.

[98] A. Zheluk, C. Quinn, and P. Meylakhs, "Internet search and krokodil in the russian federation: An infoveillance study," *Journal of Medical Internet Research*, vol. 18, e212, 2014.

[99] I. Seaman and C. Giraud-Carrier, "Prevalence and attitudes about illicit and prescription drugs on Twitter," in *IEEE International Conference on Healthcare Informatics (ICHI)*, 2016, pp. 14–17.

[100] T. Katsuki, T. K. Mackey, and R. Cuomo, "Establishing a link between prescription drug abuse and illicit online pharmacies: Analysis of Twitter data," *Journal of Medical Internet Research*, vol. 17, e280, 12 2015.

[101] M Dredze, D. Broniatowski, M Smith, and K. Hilyard, "Understanding vaccine refusal: Why we need social media now," *American Journal of Preventative Medicine*, vol. 50, pp. 550–552, 4 2016.

[102] D. Surian, Q. D. Nguyen, G. Kennedy, M. Johnson, E. Coiera, and G. A. Dunn, "Characterizing Twitter discussions about HPV vaccines using topic modeling and community detection," *Journal of Medical Internet Research*, vol. 18, e232, 8 2016.

[103] N. L. Bragazzi, I. Barberis, R. Rosselli, V. Gianfredi, D. Nucci, M. Moretti, T. Salvatori, G. Martucci, and M. Martini, "How often people google for vaccination: Qualitative and quantitative insights from a systematic search of the web-based activities using Google Trends," *Human Vaccines & Immunotherapeutics*, vol. 13, no. 2, pp. 464–469, 2017.

[104] J.-P. Allem, E. C. Leas, T. L. Caputi, M. Dredze, B. M. Althouse, S. M. Noar, and J. W. Ayers, "The charlie sheen effect on rapid in-home human immunodeficiency virus test sales," *Prevention Science*, vol. 18, pp. 541–544, 5 2017.

[105] M. Kumar, M. Dredze, G. Coppersmith, and M. De Choudhury, "Detecting changes in suicide content manifested in social media following celebrity suicides," in *Proceedings of the 26th ACM Conference on Hypertext and Social Media*, ser. HT '15, San Jose, California, USA: ACM, 2015, pp. 85–94, ISBN: 978-1-4503-3395-5.

[106] S. M. Noar, B. M. Althouse, J. W. Ayers, D. B. Francis, and K. M. Ribisl, "Cancer information seeking in the digital age: Effects of angelina jolies prophylactic mastectomy announcement," *Medical Decision Making*, vol. 35, pp. 16–21, 1 2015.

[107] C. Priest, A. Knopf, D. Groves, S. J. Carpenter, C. Furrey, A. Krishnan, R. W. Miller, L. J. Otte, M. Palakal, S. Wiehe, and J. Wilson, "Finding the patient's voice using big data: Analysis of users' health-related concerns in the chacha question-and-answer service (2009–2012)," *Journal of Medical Internet Research*, vol. 18, no. 3, e44, 2016.

[108] A. X. Zhang and S. Counts, "Gender and ideology in the spread of anti-abortion policy," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI)*, San Jose, California, USA: ACM, 2016, pp. 3378–3389.

[109] G. L. Pappa, T. O. Cunha, P. V. Bicalho, A. Ribeiro, A. P. Couto Silva, W. Meira, and A. M. R. Beleigoli, "Factors associated with weight change in online weight management communities: A case study in the loseit reddit community," *Journal of Medical Internet Research*, vol. 19, no. 1, e17, 2017.

[110] M. De Choudhury and M. Gamon, "Predicting depression via social media," in *Proceedings of International Conference on Weblogs and Social Media (ICWSM)*, 2014.

[111] M. A. Moreno, L. A. Jelenchick, K. G. Egan, E. Cox, H. Young, K. E. Gannon, and T. Becker, "Feeling bad on facebook: Depression disclosures by college students on a social networking site," *Depression and Anxiety*, vol. 28, 447455, 6 2011.

[112] A. C. Yang, N. E. Huang, C.-K. Peng, and S.-J. Tsai, "Do seasons have an influence on the incidence of depression? the use of an internet search engine query data as a proxy of human affect," *PLoS ONE*, vol. 5, e13728, 10 2010.

[113] J. Ayers, B. Althouse, J. Allem, J. Rosequist, and D. Ford, "Seasonality in seeking mental health information on Google," *American Journal of Preventive Medicine*, vol. 44, no. 5, pp. 520–5, 2013.

[114] M. De Choudhury, E. Kiciman, M. Dredze, G. Coppersmith, and M. Kumar, "Discovering shifts to suicidal ideation from mental health content in social media," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, ser. CHI '16, San Jose, California, USA: ACM, 2016, pp. 2098–2110, ISBN: 978-1-4503-3362-7.

[115] S. R. Braithwaite, C. Giraud-Carrier, J. West, M. D. Barnes, and C. L. Hanson, "Validating Machine Learning Algorithms for Twitter Data Against Established Measures of Suicidality," *Journal of Medical Internet Research Mental Health*, vol. 3, no. 2, e21, 2016.

[116] S. Chancellor, Y. Kalantidis, J. Pater, M. De Choudhury, and D. A. Shamma, "Multimodal classification of moderated online pro-eating disorder content," in *ACM Conference on Human Factors in Computing Systems (CHI)*, 2017.

[117] B. Althouse, J.-P. Allem, M. Childers, M. Dredze, and J. W. Ayers, "Population health concerns during the united states great recession," *American Journal of Preventive Medicine*, vol. 46, pp. 166–170, 2 2014.

[118] M. De Choudhury, S. Counts, and E. Horvitz, "Predicting postpartum changes in emotion and behavior via social media," in *Conference on Human Factors in Computing Systems*, ser. CHI, 2013.

[119] M. De Choudhury, S. Counts, E. J. Horvitz, and A. Hoff, "Characterizing and predicting postpartum depression from shared facebook data," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work; Social Computing*, ser. CSCW '14, Baltimore, Maryland, USA: ACM, 2014, pp. 626–638, ISBN: 978-1-4503-2540-0.

[120] K. Saha and M. De Choudhury, "Modeling stress with social media around incidents of gun violence on college campuses," *Proc. ACM Hum.-Comput. Interact.*, vol. 1, no. CSCW, 92:1–92:27, Dec. 2017.

[121] L. B. Finer and M. R. Zolna, "Shifts in intended and unintended pregnancies in the united states, 20012008," *American Journal of Public Health*, vol. 104, no. S1, S43–S48, 2014.

[122] A Sonfield and K Kost, "Public Costs from Unintended Pregnancies and the Role of Public Insurance Programs in Paying for Pregnancy-Related Care: National and State Estimates for 2010," New York: Guttmacher Institute, New York, Tech. Rep., 2015, `http://www.guttmacher.org/pubs/public-costs-of-UP-2010.pdf`.

[123] Institute of Medicine (US) Committee on Unintended Pregnancy, *The Best Intentions: Unintended Pregnancy and the Well-Being of Children and Families*. Washington, DC: The National Academies Press, 1995.

[124] N. V. Becker and D. Polsky, "Women saw large decrease in out-of-pocket spending for contraceptives after aca mandate removed cost sharing," *Health Affairs*, vol. 34, pp. 1204–1211, 7 2015.

[125] Bailey, MJ and Hershbein, B and Miller, AR, "The opt-in revolution? Contraception and the gender gap in wages," National Bureau of Economic Research, Cambridge, MA, Tech. Rep., 2012, `http://www.nber.org/papers/w17922.pdf`.

[126] C Goldin and L. Katz, "The power of the pill: Oral contraceptives and womens career and marriage decisions," *J Polit Econ*, vol. 110, no. 4, pp. 730–770, 2002.

[127] B. SP and L. S., "The effects of contraception on female poverty," *American Journal of Public Health*, vol. 33, no. 3, pp. 602–22, 2014.

[128] Daniels, K and Mosher, WD and Jones, J, "Contraceptive methods women have ever used: United States, 19822010," National Health Statistics Reports, No. 62,

Atlanta, GA, Tech. Rep., 2013, http://www.cdc.gov/nchs/data/nhsr/nhsr062.pdf.

[129] J. M. Bearak, L. B. Finer, J. Jerman, and M. L. Kavanaugh, "Changes in out-of-pocket costs for hormonal iuds after implementation of the affordable care act: An analysis of insurance benefit inquiries," *Contraception*, vol. 93, no. 2, 139144, 2016.

[130] K. Daniels, J. Daugherty, and J. Jones, "Current Contraceptive Status Among Women Aged 15-44: United States, 2011-2013," NCHS Data Brief; 20141-8, Cambridge, MA, Tech. Rep., 2014.

[131] Guttmatcher Institute, *Contraceptive Use in the United States*, Retrieved July 11, 2018 from https://www.guttmacher.org/fact-sheet/contraceptive-use-united-states, 2016.

[132] M. L. Kavanaugh, J. Jerman, and L. B. Finer, "Changes in use of long-acting reversible contraceptive methods among united states women," *Obstetrics & Gynecology*, vol. 83, pp. 2009–12, 2015.

[133] U. C. for Disease Control and Prevention, *Effectiveness of family planning methods*, Retrieved July 11, 2018 from https://www.cdc.gov/reproductivehealth/unintendedpregnancy/pdf/contraceptive_methods_508.pdf.

[134] A. M. Gariepy, E. J. Simon, D. A. Patel, M. D. Creinin, and E. B. Schwarz, "The impact of out-of-pocket expense on iud utilization among women with private," *Contraception*, vol. 84, no. 6, e39–e42, 2016.

[135] A. C. of Obstetricians and Gynecologists, "Increasing use of contraceptive implants and intrauterine devices to reduce unintended pregnancy. acog committee opinion no. 450.," *Obstetrics & Gynecology*, vol. 114, pp. 1434–8, 2009.

[136] A. Bergin, S. Tristan, M. Terplan, M. L. Gilliam, and A. K. Whitaker, "A missed opportunity for care: Two-visit iud insertion protocols inhibit placement," *Contraception*, vol. 86, 694697, 6 2012.

[137] J. C. Durante and E. J. Woodhams, "Patient education about the affordable care act contraceptive coverage requirement increases interest in using long-acting reversible contraception," *Women's Health Issues*, vol. 27, no. 2, pp. 152–157, 2017.

[138] C. H. Chuang, J. L. Mitchell, D. L. Velott, R. S. Legro, E. B. Lehman, L. Confer, and C. S. Weisman, "Womens awareness of their contraceptive benefits under the patient protection and affordable care act," *American Journal of Public Health*, vol. 105, no. S5, S713–S715, 2015.

[139] C. Rice, *The Trump effect? IUD visits rise after election*, Retrieved July 13, 2018 from https://www.athenahealth.com/insight/trump-effect-iud-visits-rise-after-election, 2017.

[140] C. McNicholas, M. Tessa, G. Secura, and J. F. Peipert, "The contraceptive choice project round up: What we did and what we learned," *Clinical obstetrics and gynecology*, vol. 57, pp. 635–643, 4 2014.

[141] C. Richards, "Protecting and expanding access to birth control," *New England Journal of Medicine*, vol. 374, no. 9, pp. 801–803, 2016, PMID: 26962898.

[142] *FAQs about Affordable Care Act implementation (part XXVI)*, Retrieved July 13, 2018 from http://www.cms.gov/CCIIO/Resources/Fact-Sheets-and-FAQs/Downloads/aca_implementation_faqs26.pdf.

[143] J. E. McDonough, "Health system reform in the united states," *International Journal of Health Policy and Management*, vol. 2, no. 1, pp. 5–8, 2014.

[144] J. E. McDonough, "The United States Health System in Transition," *Health Systems & Reform*, vol. 1, no. 1, pp. 39–51, 2015.

[145] J. E. McDonough, "The Road Ahead for the Affordable Care Act," *New England Journal of Medicine*, vol. 367, no. 3, pp. 199–201, 2012.

[146] T. Berenson, "Reminder: The house voted to repeal obamacare more than 50 times," *TIME Magazine*, Mar. 2017.

[147] L. Gostin, D. Hyman, and P. Jacobson, "The Affordable Care Act: Moving forward in the coming years," *Journal of the American Medical Association*, vol. 317, no. 1, pp. 19–20, 2017.

[148] M. P. Garretson, A. Gomez, H. Grabenstein, and M. Lee, *Trump's vow to repeal obamacare spurs women's rush to get birth control*, Retrieved July 13, 2018 from https://www.pbs.org/newshour/politics/trumps-vow-repeal-obamacare-spurs-womens-rush-get-birth-control, 2016.

[149] F. W. Song, J. E. West, L. Lundy, and N. S. Dahmen, "Women, pregnancy, and health information online: The making of informed patients and ideal mothers," *Gender & Society*, vol. 26, no. 5, pp. 773–798, 2012.

[150] T. H. Laz and A. B. Berenson, "Racial and ethnic disparities in internet use for seeking health information among young women," *Journal of Health Communication*, vol. 18, pp. 250–260, 2 2013.

[151] E. R. Buhi, E. M. Daley, H. J. Fuhrmann, and S. A. Smith, "An observational study of how young people search for online sexual health information," *Journal of American College Health*, vol. 48, pp. 101–111, 2 2009.

[152] J. W. Ayers, B. M. Althouse, and M. Dredze, "Could behavioral medicine lead the web data revolution?" *Journal of the American Medical Association*, vol. 311, no. 14, pp. 1399–1400, 2014.

[153] R. Hyndman and Y Khandakar, "Automatic time series forecasting: the forecast package for R," *Journal of Statistical Software*, vol. 27, no. 3, pp. 1–22, 2008.

[154] C. Rice, *As ACA debate continues, IUD visits keep climbing*, Retrieved July 13, 2018 from `https://www.athenahealth.com/insight/aca-debate-continues-iud-visits-keep-climbing`, 2017.

[155] J. Weiss, *IUD anxiety has decreased but for how long?* Retrieved July 13, 2018 from `https://www.athenahealth.com/insight/iud-anxiety-has-decreased-how-long`, 2017.

[156] D. E. Brashers, "Communication and uncertainty management," *Journal of Communication*, vol. 51, no. 3, pp. 477–497, 2006.

[157] N. M. Thompson, J. L. Bevan, and L. Sparks, "Healthcare reform information-seeking: Relationships with uncertainty, uncertainty discrepancy, and health self-efficacy," *Journal of Communication in Healthcare*, vol. 5, no. 1, pp. 56–66, 2012.

[158] N. Rosen and B. Knäuper, "A little uncertainty goes a long way: State and trait differences in uncertainty interact to increase information seeking but also increase worry," *Health Communication*, vol. 24, no. 3, pp. 228–38, 2009.

[159] T. Hogan and D. Brashers, "The theory of communication and uncertainty management: Implications from the wider realm of information behavior," in *Uncertainty, information management, and disclosure decisions: theories and application*. New York: Routledge, 2008, pp. 45–66.

[160] A. Bandura, *Self-efficacy: the exercise of control*. New York: Freeman, 1997.

[161] S. Lee, H Hwang, R Hawkins, and S Pingree, "Interplay of negative emotion and health self-efficacy on the use of health information and its outcomes," *Communication Research*, vol. 35, pp. 358–81, 2008.

[162] K. Lee, K. Hoti, J. D. Hughes, and L. Emmerton, "Dr Google and the Consumer: A Qualitative Study Exploring the Navigational Needs and Online Health Information-Seeking Behaviors of Consumers With Chronic Health Conditions," *Journal of Medical Internet Research*, vol. 16, e262, 12 2014.

[163] M. Biggs, C. Rocca, C. Brindis, H Hirsch, and D Grossman, "Did increasing use of highly effective contraception contribute to declining abortions in Iowa?" *Contraception*, vol. 91, 16773, 2015.

[164] S Ricketts, G Klingler, and S. R, "Game change in Colorado: widespread use of long-acting reversible contraceptives and rapid decline in births among young, low-income women," *Perspectives on Sexual and Reproductive Health*, vol. 46, pp. 125–32, 2014.

[165] J. Peipert, T Madden, J. Allsworth, and G. Secura, "Preventing unintended pregnancies by providing no-cost contraception," *Obstetrics & Gynecology*, vol. 120, pp. 1291–7, 2012.

[166] E. Heisel, G. E. Kolenic, M. M. Moniz, E. K. Kobernik, L. Minadeo, N. S. Kamdar, and V. K. Dalton, "Intrauterine device insertion before and after mandated health care coverage: The importance of baseline costs," *Obstetrics & Gynecology*, vol. 131, no. 5, pp. 843–849, 2018.

[167] A. M. Gomez and B. Freihart, "Motivations for interest, disinterest and uncertainty in intrauterine device use among young women," *Maternal and Child Health Journal*, vol. 21, no. 9, pp. 1753–62, 2017.

[168] N. Anderson, J. Steinauer, T. Valente, J. Koblentz, and C. Dehlendorf, "Motivations for interest, disinterest and uncertainty in intrauterine device use among young women," *Perspectives on Sexual and Reproductive Health*, vol. 46, no. 3, pp. 141–148, 2014.

[169] S. E. Rubin and I. Winrob, "Urban female family medicine patients perceptions about intrauterine contraception," *Journal of Women's Health*, vol. 19, no. 4, pp. 735–740, 2010.

[170] K. Harris, K. Byrd, M. Engel, K. Weeks, and C. R. Ahlers-Schmidt, "Internet-based information on long-acting reversible contraception for adolescents," *Journal of Primary Care & Community Health*, vol. 7, no. 2, pp. 76–80, 2016.

[171] D. Hubacher and D. Cheng, "Intrauterine devices and reproductive health: American women in feast and famine," *Contraception*, vol. 69, no. 6, pp. 437–446, 2004.

[172] CDC, *National survey of family growth*, Retrieved January 15, 2017 from `https://www.cdc.gov/nchs/nsfg/index.htm`, 2017.

[173] Health Care Cost Institute, *Accessing HCCI Data*, Retrieved January 4, 2017 from `http://www.healthcostinstitute.org/access-data/accessing-hcci-data-2018`, 2018.

[174] PublicHealthTrends, *Mental health trends provides the world's first real-time estimates of population mental wellbeing*, Retrieved January 15, 2017 from http://www.mentalhealthtrends.org, 2017.

[175] CDC, "Sexually Transmitted Disease Surveillance 2016," U.S. Department of Health and Human Services, Atlanta, GA, Tech. Rep., 2017.

[176] CDC, *HIV/AIDS & STDs*, Retrieved January 5, 2017 from https://www.cdc.gov/std/hiv/default.htm, 2017.

[177] Institute of Medicine, *The Hidden Epidemic: Confronting Sexually Transmitted Diseases*. Washington, DC: The National Academies Press, 1997.

[178] Reddit Metrics, *Fastest growing: /r/announcements*, Retrieved December 12, 2017 from http://redditmetrics.com, 2017.

[179] M. Duggan and A. Smith, "6% of Online Adults are reddit Users," Pew Research Center, Washington, D.C., Tech. Rep., 2013.

[180] Alexa Internet, Inc., *Top Sites in the U.S.* Retrieved June 6, 2018 from https://www.alexa.com/topsites/countries/US, 2018.

[181] Reddit, *r/STD*, Retrieved November 11, 2017 from https://www.reddit.com/r/STD, 2017.

[182] D. Abbey, *Reddit community finds health answers from reference collaborative*, Retrieved August 4, 2017 from https://dspace.library.colostate.edu/bitstream/handle/10968/1791/CUHSLMCM_M451.pdf?sequence=1, 2016.

[183] A. Litras, S. Latreille, and M. Temple-Smith, "Dr Google, porn and friend-of-a-friend: where are young men really getting their sexual health information?" *Sexual Health*, vol. 12, pp. 488–494, 2015.

[184] R. C. Aicken, S. C. Estcourt, M. A. Johnson, P. Sonnenberg, K. Wellings, and H. C. Mercer, "Use of the internet for sexual health among sexually experienced persons aged 16 to 44 years: Evidence from a nationally representative survey of the british population," *Journal of Medical Internet Research*, vol. 18, no. 1, e14, 2016.

[185] K. Mitchell, M. Ybarra, J. Korchmaros, and J. Kosciw, "Accessing sexual health information online: Use, motivations and consequences for youth with different sexual orientations," *Health Education Research*, vol. 29, pp. 147–157, 1 2014.

[186] M. S. Lim, A. Vella, R. Sacks-Davis, and M. E. Hellard, "Young peoples comfort receiving sexual health information via social media and other sources," *International Journal of STD & AIDS*, vol. 25, no. 14, pp. 1003–1008, 2014.

[187] J. B. Colditz, M. S. Woods, and B. A. Primack, "Adolescents seeking online health information: Topics, approaches, and challenges," in *Technology and Adolescent Mental Health*, M. A. Moreno and A. Radovic, Eds. Springer International Publishing, 2018.

[188] C. E. Flanders, L. Pragg, C. Dobinson, and C. Logie, "Young sexual minority women's use of the internet and other digital technologies for sexual health information seeking," *The Canadian Journal of Human Sexuality*, vol. 26, pp. 17–25, 1 2017.

[189] J. C. Magee, L. Bigelow, S. DeHaan, and B. S. Mustanski, "Sexual health information seeking online: A mixed-methods study among lesbian, gay, bisexual, and transgender young people," *Health Education & Behavior*, vol. 39, no. 3, pp. 276–289, 2012.

[190] A. Lyons, G. Mikolajczak, W. Heywood, B. Fileborn, V. Minichiello, S. Hinchliff, S. Malta, B. Dow, C. Barrett, and G. Brown, "Sources of information-seeking on sexually transmitted infections and safer sex by older heterosexual australian men and women," *Educational Gerontology*, vol. 44, no. 2-3, pp. 186–195, 2018.

[191] D. J. Johnson, W. A. Donohue, C. K. Atkin, and S. Johnson, "A comprehensive model of information seeking: Tests focusing on a technical organization," *Sci Communication*, vol. 16, no. 3, pp. 274–303, 1995.

[192] A. Park, M. Conway, and A. Chen, "Examining thematic similarity, difference, and membership in three online mental health communities from reddit: A text mining and visualization approach," *Computers in Human Behavior*, vol. 78, pp. 98–112, 2018.

[193] R. Kavuluru, M. Ramos-Morales, T. Holaday, A. G. Williams, L. Haye, and J. Cerel, "Classification of helpful comments on online suicide watch forums," in *Proceedings of the ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics (BCB)*, Seattle, WA, USA: ACM, 2016, pp. 32–40.

[194] K. Saha and M. De Choudhury, "Modeling stress with social media around incidents of gun violence on college campuses," *Proceedings of the ACM on Human-Computer Interaction*, vol. 1, no. CSCW, 92:1–92:27, 2017.

[195] T. Nguyen, M. E. Larsen, B. O'Dea, D. Phung, S. Venkatesh, and H. Christensen, "Estimation of the prevalence of adverse drug reactions from social media," *International Journal of Medical Informatics*, vol. 102, pp. 130–137, 2017.

[196] Reddit, *Who in the World is reddit? Results are in.* Retrieved December 15, 2017 from https://redditblog.com/2011/09/12/who-in-the-world-is-reddit-results-are-in, 2017.

[197] B. Boe, *PRAW: The Python Reddit API Wrapper*, Retrieved August 15, 2017 from https://praw.readthedocs.io, 2017.

[198] Reddit, *Reddit API Documentation*, Retrieved August 15, 2017 from https://www.reddit.com/dev/api, 2017.

[199] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[200] W. McKinney, "Data structures for statistical computing in python," in *Proceedings of the Python in Science Conference*, ser. SCIPY 2010, 2010, pp. 51–56.

[201] S. Bird, E. Klein, and E. Loper, *Natural language processing with Python*. O'Reilly Media, Inc., 2009.

[202] R. Řehůřek and P. Sojka, "Software Framework for Topic Modelling with Large Corpora," in *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 2010, pp. 45–50.

[203] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Advances in Neural Information Processing Systems 26*, Curran Associates, Inc., 2013, pp. 3111–3119.

[204] D. O'Callaghan, D. Greene, J. Carthy, and P. Cunningham, "An analysis of the coherence of descriptors in topic modeling," *Expert Systems with Applications*, vol. 42, no. 13, pp. 5645–5657, 2015.

[205] K. Stevens, P. Kegelmeyer, D. Andrzejewski, and D. Buttler, "Exploring topic coherence over many models and many topics," in *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, ser. EMNLP-CoNLL '12, Jeju Island, Korea: Association for Computational Linguistics, 2012, pp. 952–961.

[206] D. O'Callaghan, D. Greene, M. Conway, J. Carthy, and P. Cunningham, "An analysis of interactions within and between extreme right communities in social media," in *Ubiquitous Social Media Analysis: Third International Workshops, MUSE 2012, Bristol, UK, September 24, 2012, and MSM 2012, Milwaukee, WI, USA, June 25, 2012, Revised Selected Papers*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 88–107.

[207] C. Klein, P. Clutton, and V. Polito, "Topic modeling reveals distinct interests within an online conspiracy forum," *Frontiers in Psychology*, vol. 9, p. 189, 2018.

[208] M. Luo, N. Feiping, X. Change, Y. Yang, A. Hauptmann, and Q. Zheng, "Probabilistic non-negative matrix factorization and its robust extensions for topic modeling," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, ser. AAAI-17, 2017, pp. 2308–2314.

[209] D Da Kuang, J. Choo, and H. Park, "Nonnegative matrix factorization for interactive topic modeling and document clustering," in *Partitional Clustering Algorithms*, Boston, MA: Springer, 2015, pp. 215–243.

[210] J. Lipovský, *Urlextract*, Retrieved April 2, 2018 from `https://github.com/lipoja/URLExtract`, 2018.

[211] J. Pennebaker, R. Booth, R. Boyd, and M. Francis, *Linguistic Inquiry and Word Count: LIWC2015. Austin, TX: Pennebaker Conglomerates*, `www.LIWC.net`, 2015.

[212] E. Meites, C. L. Satterwhite, E. Torrone, E. Meites, E. F. Dunne, R. Mahajan, M. C. B. Ocfemia, J. Su, F. Xu, and H. Weinstock, "Sexually transmitted infections among us women and men: Prevalence and incidence estimates, 2008.," *Sexually Transmitted Diseases*, vol. 40, no. 3, pp. 187 –193, 2013.

[213] P Du, F Camacho, J Zurlo, E. Lengerich, P. Du, F. Camacho, J. Zurlo, and E. J. Lengerich, "Human immunodeficiency virus testing behaviors among us adults: The roles of individual factors, legislative status, and public health resources.," *Sexually Transmitted Diseases*, vol. 38, no. 9, pp. 858 –864, 2011.

[214] U.S. Department of Health & Human Services, *The Top 17 Digital Tools Blog Posts of 2017*, Retrieved July 9, 2018 from `https://www.hiv.gov/blog/top-17-digital-tools-blog-posts-2017`, 2017.

[215] A. L. Nobles, M. Dredze, and J. W. Ayers, ""Repeal and Replace": Increased Demand for Intrauterine Devices Following the 2016 Presidential Election," under review.

[216] A. L. Nobles, C. Dreisbach, J. Keim-Malpass, and L. E. Barnes, ""Is this a STD? Please help!": Online Information Seeking for Sexually Transmitted Diseases on Reddit," in *Proceedings of the AAAI International Conference of Web and Social Media*, ser. ICWSM-18, Stanford, CA, 2018.

[217] D. Cer, M. Diabb, E. Agirrec, I. Lopez-Gazpio, and L. Speciad, "Semeval-2017 task 1: Semantic textual similarity multilingual and cross-lingual focused evalua-

tion," in *Proceedings of the 11th International Workshop on Semantic Evaluations (SemEval-2017)*, Vancouver, Canada, 2017, pp. 1–14.

[218] S. S. Sundar, R. E. Rice, K. Hyang-Sook, and C. N. Sciamanna, "Online health information," in *The Routledge Handbook of Health Communication*, T. L. Thompason, R. Parrott, and J. F. Nussbaum, Eds., 2nd ed, New Yowk, NY: Taylor and Francis Group, 2011, pp. 181–202.

[219] S. M. Noar, C. N. Benac, and M. S. Harris, "Does tailoring matter? meta-analytic review of tailored print health behavior change interventions," *Psychological Bulletin*, vol. 133, no. 4, 67393, 2007.

[220] S. M. Noar, "A 10-year retrospective of research in health mass media campaigns: Where do we go from here?" *Journal of Health Communication*, vol. 11, no. 1, pp. 21–42, 2006.

[221] M. L. A. Lustria, J. Cortese, S. M. Noar, and R. L. Glueckauf, "Computer-tailored health interventions delivered over the web: Review and analysis of key components," *Patient Education and Counseling*, vol. 74, no. 2, pp. 156–173, 2009.

[222] S. M. Sutton, G. I. Balch, and R. C. Lefebvre, "Strategic questions for consumer-based health communications," *Public Health Reports*, vol. 110, pp. 725–733, 1995.

[223] I. Lichter, *Communication in Cancer Care*. New York: Churchill Livingstone, 1987.

[224] J. W. Swinehart, "Voluntary exposure to health communications," *American Journal of Public Health*, vol. 58, pp. 1265–1275, 1968.

[225] R. B. Cialdini, *Influence: Science and Practice*, 4th Edition. Boston, MA: Allyn and Bacon, 1968.

[226] R. N. Rimal and A. D. Adkins, "Using computers to narrowcast health messages: The role of audience segmentation, targeting, and tailoring in health promotion," in *Handbook of health communication*, I. L. Thompson, A. M. Dorsey, K. I. Miller, and R. Parrott, Eds., First Ed, Mahwah, NJ: Lawrence Erlbaum, 2003, pp. 497–513.

[227] M. Kreuter, C. Caburnay, J. Chen, and M. Donlin, "Effectiveness of individually-tailored calendars in promoting childhood immunization in urban public health centers," *American Public Health Association*, vol. 94, pp. 122–127, 2004.

[228] M. W. Kreuter, D. Farrell, L. Olevitch, and L. Brennan, *Tailoring health messages: Customizing communication with computer technology*. Mahwah, NJ: Erlbaum, 2000.

[229] A. West, "Coming soon to a location near you," *Government Information Quarterly*, vol. 25, no. 1, pp. 61–65, 2008.

[230] M. A. Wakefield, B. Loken, and R. C. Hornik, "Use of mass media campaigns to change health behaviour," *Lancet*, vol. 376, pp. 1261–71, 9748 2010.

[231] G. Kreps and L. Neuhauser, "Designing health information programs to promote the health and well-being of vulnerable populations: The benefits of evidence-based strategic health communication," in *Meeting Health Information Needs Outside Of Healthcare*, C. A. Smith and A. Keselman, Eds., Chandos Publishing, 2015.

[232] HIV.gov, *Highlights from 2018 Agents of Change Summit*, Retrieved July 29, 2018 from `https://www.hiv.gov/blog/highlights-2018-agents-change-summit`, 2018.

[233] C. T. Lence and K. Capozza, "Health information in bits and bytes," in *Meeting Health Information Needs Outside Of Healthcare*, C. A. Smith and A. Keselman, Eds., Chandos Publishing, 2015.

[234] A. C. Maher, K. L. Lewis, K. Ferrar, S. Marshall, I. De Bourdeaudhuij, and C. Vandelanotte, "Are health behavior change interventions that use online social networks effective? a systematic review," *Journal of Medical Internet Research*, vol. 16, no. 2, e40, 2014.

[235] J. T. Constine, *Facebooks Q2: Monthly users up 21% YOY to 1.15B, dailies up 27% to 699M, mobile monthlies up 51% to 819M*, Retrieved April 20, 2018 from `http://techcrunch.com/2013/07/24/facebook-growth-2/`, 2013.

[236] R. Thackeray, B. Neiger, C. Hanson, and J. McKenzie, "Enhancing promotional strategies within social marketing programs: Use of web 2.0 social media," *Health Promotion Practice*, vol. 9, no. 4, pp. 338–343, 2008.

[237] A. Jha, L. Lin, and E. Savoia, "The use of social media by state health departments in the us: Analyzing health communication through facebook," *Journal of Community Health*, vol. 41, no. 1, pp. 174–179, 2016.

[238] J. K. Harris, N. L. Mueller, and D. Snider, "Social media adoption in local health departments nationwide," *American Journal of Public Health*, vol. 103, no. 9, pp. 1700–1707, 2013, PMID: 23865660. eprint: `https://doi.org/10.2105/AJPH.2012.301166`.

[239] M. Chappelka, J. Oh, D. Scott, and M. Walker-Holmes, "Food for thought: Analyzing public opinion on the supplemental nutrition assistance program," *Bloomberg*

Data for Good Exchange Conference, 2017. eprint: https://arxiv.org/ftp/arxiv/papers/1710/1710.02443.pdf.